2-20-2007

Supplementary Data for “Sequence and annotation of the 369-kb NY-2A and the 345-kb AR158 viruses that infect *Chlorella* NC64A”: Appendix D: Gene Names C006R – C815L

Lisa A. Fitzgerald
University of Nebraska-Lincoln, lisa.fitzgerald@nrl.navy.mil

Michael V. Graves
University of Massachusetts–Lowell, Michael_Graves@uml.edu

Xiao Li
University of Massachusetts–Lowell

Tamara Feldblyum
The Institute for Genomic Research, Rockville, MD

Willaim C. Nierman
The Institute for Genomic Research, Rockville, MD, wnierman@tigr.org

See next page for additional authors

Fitzgerald, Lisa A.; Graves, Michael V.; Li, Xiao; Feldblyum, Tamara; Nierman, Willaim C.; and Van Etten, James L., "Supplementary Data for “Sequence and annotation of the 369-kb NY-2A and the 345-kb AR158 viruses that infect *Chlorella* NC64A”: Appendix D: Gene Names C006R – C815L" (2007). *Virology Papers*. 10.
https://digitalcommons.unl.edu/virologypub/10

Authors

Lisa A. Fitzgerald, Michael V. Graves, Xiao Li, Tamara Feldblyum, Willaim C. Nierman, and James L. Van Etten

Main article published in *Virology* 358:2 (February 20, 2007), pp. 472-484. doi:10.1016/j.virol.2006.08.033 Submitted July 27, 2006; revised August 17, 2006; accepted August 23, 2006; published online October 4, 2006.

SUPPLEMENTARY DATA FOR

Sequence and annotation of the 369-kb NY-2A and the 345-kb AR158 viruses that infect *Chlorella* NC64A

Lisa A. Fitzgerald[a], Michael V. Graves[b], Xiao Li[b], Tamara Feldblyum[c], William C. Nierman[c, d], and James L. Van Etten[e, *]

[a] Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588-0304
[b] Department of Biological Sciences, University of Massachusetts-Lowell, Lowell, MA 01854
[c] The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850
[d] The George Washington University School of Medicine, Department of Biochemistry and Molecular Biology, Washington, DC 20037
[e] Department of Plant Pathology, University of Nebraska-Lincoln, Lincoln, NE 68583-0722 and Nebraska Center for Virology, University of Nebraska-Lincoln, Lincoln, NE 68588-0666
* Corresponding author. Email: jvanetten@unlnotes.unl.edu

Abstract: Viruses NY-2A and AR158, members of the family *Phycodnaviridae*, genus *Chlorovirus*, infect the fresh water, unicellular, eukaryotic, chlorella-like green alga, *Chlorella* NC64A. The 368,683-bp genome of NY-2A and the 344,690-bp genome of AR158 are the two largest chlorella virus genomes sequenced to date; NY-2A contains 404 putative protein-encoding and 7 tRNA-encoding genes and AR158 contains 360 putative protein-encoding and 6 tRNA-encoding genes. The protein-encoding genes are almost evenly distributed on both strands, and intergenic space is minimal. Two of the NY-2A genes encode inteins, the large subunit of ribonucleotide reductase and a superfamily II helicase. These are the first inteins to be detected in the chlorella viruses. Approximately 40% of the viral gene products resemble entries in the public databases, including some that are unexpected for a virus. These include GDP-d-mannose dehydratase, fucose synthase, aspartate transcarbamylase, Ca^{++} transporting ATPase and ubiquitin. Comparison of NY-2A and AR158 protein-encoding genes with the prototype chlorella virus PBCV-1 indicates that 85% of the genes are present in all three viruses.

Keywords: Chlorella viruses, Phycodnaviridae, Virus NY-2A, Virus AR158, Genome sequence

Supplementary data associated with this article is archived in this repository as 4 separate files: Appendices A–D. Each document, in spreadsheet format, shows Gene Name, Genome Position, A.A. length, Peptide Mw, pI, CDD Hit Number, COGs, COG Definition, Bit Score, E-value, % Identity, % Positive, Query from-to, Hit from-to, BLASTp Hit Number, Hit Accession, BLASTp Definition, Bit Score, E-value, % Identity, % Positive, Query from-to, and Hit from-to.

Appendix A: Gene Names b002R – b797R
Appendix B: Gene Names B001L – B886R
Appendix C: Gene Names c001R – c814L
Appendix D: Gene Names C006R – C815L

Appendix D: Gene Names C006R – C815L

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
C006R	2557--3555	333	37,474	7.08	1	cd00315	Cyt_C5_DNA_methylase, Cytosine-C5 specific DNA methylases; Methyl transfer reactions play an important role in many aspects of biology. Cytosine-specific DNA methylases are found both in prokaryotes and eukaryotes. DNA methylation, or the covalent addition of a methyl group to cytosine within the context of the CpG dinucleotide, has profound effects on the mammalian genome. These effects include transcriptional repression via inhibition of transcription factor binding or the recruitment of methyl-binding proteins and their associated chromatin remodeling factors, X chromosome inactivation, imprinting and the suppression of parasitic DNA sequences. DNA methylation is also essential for proper embryonic development and is an important player in both DNA repair and genome stability	201.30	1.08E-52	32%	51%	4--328	1--314	1	CAD33712	putative DNA methylase	146.75	9.36E-34	31%	49%	4--333	1--312
					2	pfam00145	DNA methylase, C-5 cytosine-specific DNA methylase..	188.98	5.76E-49	31%	47%	4--328	1--323	2	CAD33713	putative DNA methylase	145.21	2.72E-33	32%	50%	4--317	1--293
					3	COG0270	Dcm, Site-specific DNA methylase [DNA replication, recombination, and repair].	177.58	1.53E-45	29%	49%	1--330	1--324	3	ZP_00504119	C-5 cytosine-specific DNA methylase	133.27	1.07E-29	32%	46%	1--322	9--309
														4	CAA59690	site-specific DNA-methyltransferase (cytosine-specific)	133.27	1.07E-29	30%	48%	3--326	44--365
														5	BAB07227	cytosine-specific methyltransferase	130.18	9.07E-29	29%	47%	3--330	5--309
														6	AAP07928	Type II restriction-modification system methylation subunit	129.80	1.18E-28	32%	50%	4--320	18--341
														7	CAA68505	DdeI methylase	128.64	2.64E-28	29%	44%	4--326	1--366
														8	CAE52348	putative cytosine-specific methyltransferase	125.18	2.92E-27	29%	43%	1--316	1--342
														9	AAC97192	modification methylase M.NspHI	124.79	3.81E-27	31%	44%	7--320	63--363
														10	YP_208922	putative 5-methylcytosine methyltransferase	123.64	8.48E-27	31%	47%	4--326	5--326
C008L	4643--3618	342	40,112	8.33		No Hit Found								1	AAX46848	hypothetical protein PSSM4 047	59.69	1.56E-07	33%	49%	1--164	1--155
														2	CAC11914	hypothetical protein	56.61	1.32E-06	25%	39%	67--336	94--423
C011L	5826--4708	373	42,903	6.06	1	COG4123	COG4123, Predicted O-methyltransferase [General function prediction only].	64.50	1.52E-11	24%	40%	41--168	39--176	1	AAC03124	DNA adenine methyltransferase	676.40	0.00E+00	89%	92%	6--373	1--368
					2	COG2890	HemK, Methylase of polypeptide chain release factors [Translation, ribosomal structure and biogenesis].	56.93	3.25E-09	25%	40%	47--195	111--264	2	CAA29835	unnamed protein product	560.07	4.17E-158	71%	83%	5--373	3--376
					3	COG2813	RsmC, 16S RNA G1207 methylase RsmC [Translation, ribosomal structure and biogenesis].	50.65	2.27E-07	23%	43%	44--162	156--266	3	AAC57945	DNA adenine methyltransferase	358.22	2.41E-97	49%	68%	9--373	3--357
					4	COG0286	HsdM, Type I restriction-modification system methyltransferase subunit [Defense mechanisms].	48.49	1.06E-06	21%	39%	11--181	149--346	4	P52284	Modification methylase CviRI (Adenine-specific methyltransferase CviRI) (M.CviRI)	211.85	2.78E-53	36%	53%	14--373	10--377
					5	smart00650	rADc, Ribosomal RNA adenine dimethylases; .	43.62	3.06E-05	28%	50%	44--119	11--87	5	AAC03125	DNA adenine methyltransferase	206.45	1.17E-51	33%	53%	14--373	10--380
					6	COG0030	KsgA, Dimethyladenosine transferase (rRNA methylation) [Translation, ribosomal structure and biogenesis].	41.00	2.11E-04	27%	51%	30--119	12--105	6	AAC57943	DNA adenine methyltransferase	204.91	3.40E-51	32%	55%	6--373	1--367
					7	COG1041	COG1041, Predicted DNA modification methylase [DNA replication, recombination, and repair].	39.94	3.85E-04	26%	43%	1--163	151--311	7	ZP_00518729	hypothetical protein CwatDRAFT_0727	81.65	4.35E-14	30%	50%	28--210	19--217
					8	pfam03602	Cons hypoth95, Conserved hypothetical protein 95..	36.76	3.64E-03	24%	40%	32--119	24--123	8	AAC60387	methyltransferase; M-AccI	80.11	1.26E-13	28%	47%	9--220	11--238
					9	COG4106	Tam, Trans-aconitate methyltransferase [General function prediction only].	36.43	4.31E-03	20%	38%	44--119	28--102	9	AAA50500	AccI methylase	80.11	1.26E-13	28%	47%	9--220	11--238
					10	COG0421	SpeE, Spermidine synthase [Amino acid transport and metabolism].	36.49	4.96E-03	29%	42%	46--160	76--188	10	JU0470	site-specific DNA-methyltransferase (adenine-specific) (EC 2.1.1.72) type II - Acinetobacter calcoaceticus	80.11	1.26E-13	28%	47%	9--220	11--238
C015L	7324--5864	487	54,050	5.46	1	cd00204	ANK, ankyrin repeats; ankyrin repeats mediate protein-protein interactions in very diverse families of proteins. The number of ANK repeats in a protein can range from 2 to over 20 (ankyrins, for example). ANK repeats may occur in combinations with other types of domains. The structural repeat unit contains two antiparallel helices and a beta-hairpin, repeats are stacked in a superhelical arrangement; this alignment contains 4 consecutive repeats	115.95	5.90E-27	41%	68%	80--205	1--126	1	XP_782809	PREDICTED: similar to ankyrin repeat domain 28	190.27	1.26E-46	30%	51%	24--396	420--786
					2	COG0666	Arp, FOG: Ankyrin repeat [General function prediction only].	76.85	3.28E-15	31%	47%	238--409	59--234	2	XP_784414	PREDICTED: similar to Ankyrin-2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid)	188.35	4.78E-46	32%	52%	24--396	111--477
					3	pfam00023	Ank, Ankyrin repeat. There's no clear separation between noise and signal on the HMM search Ankyrin repeats generally consist of a beta, alpha, alpha, beta order of secondary structures. The repeats associate to form a higher order structure..	45.82	6.45E-06	50%	69%	317--349	1--33	3	XP_796846	PREDICTED: similar to Ankyrin-1 (Erythrocyte ankyrin) (Ankyrin R), partial	188.35	4.78E-46	31%	49%	24--396	22--388
					4	smart00248	ANK, ankyrin repeats; Ankyrin repeats are about 33 amino acids long and occur in at least four consecutive copies. They are involved in protein protein interactions. The core of the repeat seems to be an helix-loop-helix structure. .	38.79	8.88E-04	55%	69%	317--346	1--30	4	XP_783930	PREDICTED: similar to ankyrin 3, epithelial isoform b	187.58	8.15E-46	31%	51%	22--387	338--698
														5	XP_786227	PREDICTED: similar to ankyrin 3, epithelial isoform b	186.42	1.82E-45	31%	52%	22--396	120--488
														6	XP_792296	PREDICTED: similar to Ankyrin-1 (Erythrocyte ankyrin) (Ankyrin R)	185.65	3.10E-45	33%	52%	22--396	138--506
														7	XP_798405	PREDICTED: similar to Ankyrin-1 (Erythrocyte ankyrin) (Ankyrin R), partial	185.27	4.04E-45	30%	52%	20--396	738--1108
														8	XP_783273	PREDICTED: similar to ankyrin 3, epithelial isoform b	184.50	6.90E-45	30%	52%	22--387	327--686
														9	EAL87814	NACHT domain protein, putative	183.34	1.54E-44	32%	51%	24--387	123--1481
														10	XP_787863	PREDICTED: similar to ankyrin 3, epithelial isoform b	183.34	1.54E-44	31%	49%	22--396	259--614
C018R	7386--8585	400	47,051	6.53		No Hit Found								1	NP_048711	A354R	145.59	2.72E-33	38%	54%	155--398	2--235
C021L	10161--8566	532	58,475	7.41	1	cd00204	ANK, ankyrin repeats; ankyrin repeats mediate protein-protein interactions in very diverse families of proteins. The number of ANK repeats in a protein can range from 2 to over 20 (ankyrins, for example). ANK repeats may occur in combinations with other types of domains. The structural repeat unit contains two antiparallel helices and a beta-hairpin, repeats are stacked in a superhelical arrangement; this alignment contains 4 consecutive repeats	114.79	1.22E-26	40%	58%	255--381	1--126	1	EAL87814	NACHT domain protein, putative	211.85	4.51E-53	33%	52%	26--437	115--1519
					2	COG0666	Arp, FOG: Ankyrin repeat [General function prediction only].	71.84	1.04E-13	31%	54%	226--360	71--212	2	XP_788092	PREDICTED: similar to Ankyrin-2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid), partial	204.14	9.41E-51	33%	51%	29--443	402--1809
					3	pfam00023	Ank, Ankyrin repeat. There's no clear separation between noise and signal on the HMM search Ankyrin repeats generally consist of a beta, alpha, alpha, beta order of secondary structures. The repeats associate to form a higher order structure..	40.82	2.15E-04	48%	61%	360--391	1--32	3	XP_787863	PREDICTED: similar to ankyrin 3, epithelial isoform b	203.37	1.61E-50	33%	49%	28--440	35--429
														4	XP_786227	PREDICTED: similar to ankyrin 3, epithelial isoform b	202.22	3.58E-50	33%	51%	28--440	50--455
														5	XP_782887	PREDICTED: similar to Ankyrin-1 (Erythrocyte ankyrin) (Ankyrin R)	198.36	5.17E-49	34%	53%	29--399	209--578
														6	XP_783273	PREDICTED: similar to ankyrin 3, epithelial isoform b	197.98	6.75E-49	32%	48%	29--443	4--434
														7	XP_784360	PREDICTED: similar to ankyrin 1, erythroid	197.59	8.81E-49	29%	52%	32--442	205--608
														8	AAH51456	Ankyrin repeat domain 28	196.44	1.96E-48	30%	50%	32--454	43--459
														9	XP_534254	PREDICTED: similar to ankyrin repeat domain 28	196.44	1.96E-48	30%	50%	32--454	43--459
														10	XP_224620	PREDICTED: similar to Ankyrin repeat domain protein 28	196.05	2.56E-48	30%	49%	32--454	43--459
C024L	11328--10249	360	38,832	4	1	COG3889	COG3889, Predicted solute binding protein [General function prediction only].	51.96	9.77E-08	24%	43%	176--313	736--870		No Hit Found	No Hit Found						
C029R	11423--12115	231	25,842	10.65		No Hit Found									No Hit Found	No Hit Found						
C031L	16597--12143	1485	154,837	5.55	1	COG2911	COG2911, Uncharacterized protein conserved in bacteria [Function unknown].	40.68	2.66E-04	20%	39%	490--820	743--1048	1	BAB83467	Vp260 like protein	1820.82	0.00E+00	80%	85%	1--1158	1--1162
					2	COG1664	CcmA, Integral membrane protein CcmA involved in cell shape determination [Cell envelope biogenesis, outer membrane].	35.67	8.02E-03	18%	36%	677--796	10--112	2	BAB83469	Vp260 like protein	696.04	0.00E+00	33%	49%	11--1477	1--1458
														3	BAB83468	Vp260 like protein	673.32	0.00E+00	33%	49%	11--1469	1--1448
														4	BAB83470	Vp260 like protein	654.44	0.00E+00	34%	50%	11--1334	1--1306
														5	BAB83471	Vp260 like protein	593.96	1.37E-167	38%	54%	3--953	2--955
														6	NP_048470	PBCV-1 Vp260 protein	231.88	1.37E-58	27%	42%	26--824	44--871
														7	AAA86307	glycoprotein Vp260	205.30	1.37E-50	28%	40%	26--794	44--778
														8	NP_048366	Asn/Thr/Ser/Ile rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	134.42	2.97E-29	22%	37%	34--1217	54--1293
														9	NP_048362	Asn/Thr/Ser/Val rich protein	130.95	3.29E-28	21%	36%	18--1092	180--1305
														10	NP_048377	Asn/Thr/Ser/Val rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	119.40	9.90E-25	22%	38%	62--823	19--805

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
C036L	17881--16628	418	48,765	7.27	1	pfam01541	GIY-YIG, GIY-YIG catalytic domain. This domain called GIY-YIG is found in the amino terminal region of excinuclease abc subunit c (uvrC), bacteriophage T4 endonucleases segA, segB, segC, segD and segE; it is also found in putative endonucleases encoded by group I introns of fungi and phage. The structure of I-TevI a GIY-YIG endonuclease, reveals a novel alpha/beta-fold with a central three-stranded antiparallel beta-sheet flanked by three helices. The most conserved and putative catalytic residues are located on a shallow, concave surface and include a metal coordination site	36.29	5.12E-03	32%	52%	34--104	13--79		No Hit Found	No Hit Found						
C037L	22512--18112	1467	151,992	5.05		No Hit Found								1	BAB83469	Vp260 like protein	1878.99	0.00E+00	65%	76%	10--1467	1--1464
														2	BAB83468	Vp260 like protein	1847.40	0.00E+00	64%	75%	10--1467	1--1462
														3	BAB83470	Vp260 like protein	1699.10	0.00E+00	65%	76%	10--1322	1--1313
														4	BAB83467	Vp260 like protein	634.80	6.91E-180	37%	53%	4--1114	5--1118
														5	BAB83471	Vp260 like protein	524.24	1.32E-146	35%	50%	1--939	1--949
														6	NP_048470	PBCV-1 Vp260 protein	217.24	3.44E-54	27%	39%	6--912	17--1025
														7	AAA86307	glycoprotein Vp260	199.52	7.42E-49	29%	40%	54--819	21--815
														8	NP_048377	Asn/Thr/Ser/Val rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	150.60	3.95E-34	25%	39%	57--808	13--759
														9	NP_048362	Asn/Thr/Ser/Val rich protein	144.05	3.70E-32	22%	36%	20--970	183--1238
														10	NP_048366	Asn/Thr/Ser/Ile rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	130.18	5.53E-28	22%	37%	109--1466	21--1287
C044L	25412--22554	953	96,111	8.08		No Hit Found								1	BAB83471	Vp260 like protein	1523.45	0.00E+00	81%	87%	1--953	1--955
														2	BAB83467	Vp260 like protein	615.54	2.71E-174	38%	56%	2--953	3--954
														3	BAB83468	Vp260 like protein	609.37	1.94E-172	39%	56%	10--947	1--930
														4	BAB83469	Vp260 like protein	594.73	4.94E-168	38%	55%	10--947	1--930
														5	BAB83470	Vp260 like protein	591.65	4.18E-167	38%	55%	10--947	1--931
														6	NP_048470	PBCV-1 Vp260 protein	210.31	2.62E-52	28%	40%	51--948	15--905
														7	AAA86307	glycoprotein Vp260	191.43	1.26E-46	28%	40%	51--912	15--815
														8	NP_048362	Asn/Thr/Ser/Val rich protein	159.84	4.06E-37	24%	40%	25--847	18--930
														9	NP_048377	Asn/Thr/Ser/Val rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	152.14	8.47E-35	23%	40%	22--840	18--800
														10	AAC35172	cell surface antigen rOmpA	149.44	5.49E-34	24%	36%	11--824	216--1026
C048R	25510--26742	411	46,503	5.93	1	pfam04451	Capsid_Iridovir, Iridovirus major capsid protein. This family includes the major capsid protein of iridoviruses, chlorella virus and Spodoptera ascovirus, which are all dsDNA viruses with no RNA stage. This is the most abundant structural protein and can account for up to 45% of virion protein. In Chlorella virus AR158 the major capsid protein is a glycoprotein	281.04	1.20E-76	37%	53%	6--407	3--443	1	AAC27494	putative capsid protein	429.10	1.27E-118	84%	94%	22--251	1--230
														2	NP_048787	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	283.49	8.64E-75	39%	54%	6--411	3--437
														3	BAA76601	major capsid protein MCP1	282.34	1.92E-74	39%	54%	6--411	3--437
														4	BAA76600	major capsid protein	276.56	1.06E-72	39%	54%	6--411	3--436
														5	AAC27492	major capsid protein Vp49	275.79	1.80E-72	40%	54%	6--411	3--432
														6	BAA22198	major capsid protein Vp54	275.79	1.80E-72	38%	53%	6--411	3--437
														7	1M3Y_D	Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid Containing, Dna Virus	258.07	3.89E-67	37%	53%	28--411	1--413
														8	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	257.68	5.08E-67	37%	53%	28--411	1--413
														9	NP_048914	similar to Chlorella virus PBCV-1 ORF A11L, corresponds to GenBank Accession Number U42580	246.90	8.96E-64	36%	54%	6--411	2--400
														10	NP_048359	contains aminoacyl-tRNA synthetase class-II signature	232.26	2.28E-59	35%	53%	6--411	3--403
C051R	26946--27905	320	36,755	10.71	1	cd00180	S_TKc, Serine/Threonine protein kinases, catalytic domain. Phosphotransferases of the serine or threonine-specific kinase subfamily. The enzymatic activity of these protein kinases is controlled by phosphorylation of specific residues in the activation segment of the catalytic domain, sometimes combined with reversible conformational changes in the C-terminal autoregulatory tail	49.43	5.92E-07	25%	41%	84--249	8--157	1	NP_048382	similar to E. coli LPS core biosynthesis protein, corresponds to Swiss-Prot Accession Number P27240	580.10	3.10E-164	92%	95%	13--320	1--308
C052L	30582--27913	890	101,073	8.11	1	pfam05887	Trypan_PARP, Procyclic acidic repetitive protein (PARP). This family consists of several Trypanosoma brucei procyclic acidic repetitive protein (PARP) like sequences. The procyclic acidic repetitive protein (parp) genes of Trypanosoma brucei encode a small family of abundant surface proteins whose expression is restricted to the procyclic form of the parasite. They are found at two unlinked loci, parpA and parpB; transcription of both loci is developmentally regulated	62.69	5.98E-11	32%	54%	762--830	60--128	1	BAA11344	DNA binding protein	340.50	1.56E-91	42%	53%	319--757	54--471
					2	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	57.08	2.90E-09	30%	36%	752--855	50--158	2	NP_048383	contains Pro-rich Px motif, PAPK (19X); similar to Arabidopsis anter-specific Pro-rich protein, corresponds to Swiss-Prot Accession Number P40602	246.13	4.00E-63	34%	46%	1--403	1--375
C055L	30928--30614	105	12,896	10.34		No Hit Found								1	NP_048385	A37L	183.34	1.78E-45	81%	90%	1--105	1--105
C056L	31404--30946	153	17,702	4.42	1	COG5201	SKP1, SCF ubiquitin ligase, SKP1 component [Posttranslational modification, protein turnover, chaperones].	107.45	1.78E-24	38%	57%	3--145	4--155	1	NP_048387	contains ATP/GTP-binding motif A; similar to Dictyostelium FP21 glycoprotein, corresponds to Swiss-Prot Accession Number P52285	162.16	4.14E-39	52%	69%	1--153	1--151
					2	pfam01466	Skp1, Skp1 family, dimerisation domain..	94.91	1.07E-20	58%	83%	69--146	1--77	2	AAT99735	SKP1	90.12	2.00E-17	41%	58%	43--146	46--154
					3	smart00512	Skp1, Found in Skp1 protein family; Family of Skp1 (kinetochore protein required for cell cycle progression) and elongin C (subunit of RNA polymerase II transcription factor SIII) homologues. .	72.60	6.50E-14	33%	55%	1--97	2--107	3	AAD34458	Skp1	88.20	7.60E-17	34%	51%	3--146	5--152
					4	pfam03931	Skp1_POZ, Skp1 family, tetramerisation domain..	55.68	8.75E-09	35%	57%	1--62	1--66	4	AAC63273	SKP1-like protein	88.20	7.60E-17	34%	51%	7--146	8--152
														5	AAX83944	Skp1	88.20	7.60E-17	34%	51%	7--146	11--155
														6	EAL48742	Skp1 protein, putative	86.66	2.21E-16	36%	55%	2--125	7--138
														7	AAO85510	SKP1	86.27	2.89E-16	33%	50%	7--146	8--152
														8	EAL91988	sulfur metabolism regulator SkpA, putative	85.50	4.93E-16	35%	57%	2--124	3--135
														9	BAE55074	unnamed protein product	85.11	6.43E-16	35%	57%	2--124	5--137
														10	XP_499961	hypothetical protein	84.34	1.10E-15	29%	47%	3--145	2--159
C057R	31475--32719	415	44,464	10.96	1	pfam05887	Trypan_PARP, Procyclic acidic repetitive protein (PARP). This family consists of several Trypanosoma brucei procyclic acidic repetitive protein (PARP) like sequences. The procyclic acidic repetitive protein (parp) genes of Trypanosoma brucei encode a small family of abundant surface proteins whose expression is restricted to the procyclic form of the parasite. They are found at two unlinked loci, parpA and parpB; transcription of both loci is developmentally regulated	54.98	1.19E-08	30%	52%	30--99	53--122	1	NP_048389	contains Pro-rich Px motif, PAPK (8X); similar to Thermoproteus virus protein TPX, corresponds to Swiss-Prot Accession Number P19275	575.09	1.46E-162	92%	93%	104--415	101--412
					2	pfam05616	Neisseria_TspB, Neisseria meningitidis TspB protein. This family consists of several Neisseria meningitidis TspB virulence factor proteins..	52.44	7.79E-08	31%	37%	29--112	317--406	2	NP_048519	similar to PBCV-1 ORF A41R, corresponds to GenBank Accession Number U17055	497.28	3.86E-139	79%	85%	104--415	76--387
					3	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	50.91	2.10E-07	35%	41%	16--104	58--146	3	BAD86968	hypothetical protein	73.94	1.05E-11	26%	44%	110--342	95--329
					4	pfam04625	DEC-1_N, DEC-1 protein, N terminal region. The defective chorion-1 gene (dec-1) in Drosophila encodes follicle cell proteins necessary for proper eggshell assembly. Multiple products of the dec-1 gene are formed by alternative RNA splicing and proteolytic processing. Cleavage products include S80 (80 kDa) which is incorporated into the eggshell, and further proteolysis of S80 gives S60 (60 kDa)	47.18	2.77E-06	36%	40%	4--92	61--145	4	NP_916095	P0481E12.18	73.94	1.05E-11	26%	44%	110--342	72--306
					5	COG3147	DedD, Uncharacterized protein conserved in bacteria [Function unknown].	46.94	2.91E-06	42%	47%	28--90	88--150	5	AAM63817	unknown	67.01	1.29E-09	24%	44%	103--346	55--299
					6	pfam05518	Totivirus_coat, Totivirus coat protein..	44.97	1.33E-05	19%	24%	10--103	659--752	6	NP_191439	unknown protein	67.01	1.29E-09	24%	44%	103--346	55--299
					7	pfam02318	RPH3A_effector, Rabphilin-3A effector domain. This is a family of proteins involved in protein transport in synaptic vesicles. Rabphilin-3A has been shown to contact Rab3A, a small G protein important in neurotransmitter release, in two distinct areas..	43.14	4.81E-05	20%	33%	25--109	170--254	7	XP_464783	unknown protein	52.76	2.52E-05	24%	42%	119--354	64--311
C061L	34530--32722	603	68,716	6.77	1	COG0464	SpoVK, ATPases of the AAA+ class [Posttranslational modification, protein turnover, chaperones].	58.26	1.34E-09	36%	59%	234--293	252--311	1	NP_048392	contains ATP/GTP-binding site motif A; contains DDC/GAD/HDC/TyrDC pyridoxal phosphate attachment site; similar to yeast BCS1 protein, corresponds to Swiss-Prot Accession Number P32839	1081.24	0.00E+00	89%	95%	1--603	1--599
					2	pfam00004	AAA, ATPase family associated with various cellular activities (AAA). AAA family proteins often perform chaperone-like functions that assist in the assembly, operation, or disassembly of protein complexes..	54.25	1.96E-08	32%	50%	492--558	82--148	2	EAN88268	hypothetical protein, conserved	102.83	3.46E-20	25%	45%	212--542	267--562
					3	COG1222	RPT1, ATP-dependent 26S proteasome regulatory subunit [Posttranslational modification, protein turnover, chaperones].	51.38	1.70E-07	45%	67%	231--280	158--207	3	EAN97512	hypothetical protein, conserved	101.68	7.70E-20	24%	43%	212--542	267--562

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					4	cd00009	AAA, AAA-superfamily of ATPases associated with a wide variety of cellular activities, including membrane fusion, proteolysis, and DNA replication..	48.05	1.36E-06	21%	45%	261--349	2--94	4	AAZ12359	hypothetical protein, conserved	100.52	1.72E-19	23%	41%	136--542	191--535
					5	COG0465	HflB, ATP-dependent Zn proteases [Posttranslational modification, protein turnover, chaperones].	43.39	4.04E-05	55%	70%	496--536	272--312	5	CAJ04397	hypothetical protein, conserved	99.75	2.93E-19	24%	43%	212--552	292--598
					6	COG5271	MDN1, AAA ATPase containing von Willebrand factor type A (vWA) domain [General function prediction only].	42.01	9.95E-05	55%	64%	260--307	545--1589	6	CAJ04393	hypothetical protein, conserved	99.75	2.93E-19	24%	43%	212--552	279--585
					7	COG0593	DnaA, ATPase involved in DNA replication initiation [DNA replication, recombination, and repair].	40.64	2.34E-04	28%	50%	247--335	103--185	7	CAC28656	related to human BCS1 protein	94.74	9.41E-18	26%	43%	248--558	371--710
					8	COG0714	COG0714, MoxR-like ATPases [General function prediction only].	39.72	5.60E-04	26%	44%	261--335	46--119	8	YP_142926	unknown	94.36	1.23E-17	31%	52%	429--591	18--188
					9	COG1484	DnaC, DNA replication protein [DNA replication, recombination, and repair].	39.22	7.17E-04	20%	39%	206--399	58--237	9	NP_803852	ORF286	91.66	7.97E-17	25%	42%	226--549	198--460
					10	COG1223	COG1223, Predicted ATPase (AAA+ superfamily) [General function prediction only].	38.80	9.55E-04	39%	63%	261--304	154--203	10	CAJ06263	hypothetical protein, conserved	74.33	1.32E-11	24%	40%	220--551	287--573
C062R	34593--34961	123	14,000	6.66		No Hit Found								1	NP_048396	A48R	223.79	1.19E-57	90%	91%	1--123	1--123
C063L	35641--34958	228	26,158	8.2	1	pfam03009	GDPD, Glycerophosphoryl diester phosphodiesterase family. E. coli has two sequence related isozymes of glycerophosphoryl diester phosphodiesterase (GDPD) - periplasmic and cytosolic. This family also includes agrocinopine synthase, the similarity to GDPD has been noted. This family appears to have weak but not significant matches to mammalian phospholipase C pfam00388, which suggests that this family may adopt a TIM barrel fold.	153.26	3.21E-38	30%	48%	14--224	1--238	1	NP_048397	similar to Escherichia coli glycerophosphoryl diester phosphodiesterase, corresponds to Swiss-Prot Accession Number P10908	417.93	1.12E-115	94%	96%	10--228	1--219
					2	COG0584	UgpQ, Glycerophosphoryl diester phosphodiesterase [Energy production and conversion].	121.47	1.15E-28	27%	44%	10--224	7--247	2	BAB65669	226aa long hypothetical glycerophosphoryl diester phosphodiesterase	115.93	9.10E-25	34%	54%	12--226	4--215
														3	ZP_00816323	glycerophosphoryl diester phosphodiesterase, putative	111.69	1.72E-23	30%	51%	10--224	2--230
														4	BAD64029	glycerophosphoryl diester phosphodiesterase	102.06	1.36E-20	32%	48%	10--224	4--232
														5	BAB80127	probable glycerophosphodiester phosphodiesterase	101.29	2.32E-20	31%	48%	12--226	7--235
														6	CAG35064	probable glcerophosphoryl diester phosphodiesterase	100.91	3.03E-20	30%	47%	11--225	5--225
														7	AAK78410	Glycerophosphoryl diester phosphodiesterase	100.91	3.03E-20	29%	50%	12--225	8--236
														8	CAB12801	yhdW	100.52	3.96E-20	30%	49%	11--228	3--235
														9	BAB81583	probable glycerophosphodiester phosphodiesterase	100.14	5.17E-20	29%	48%	11--228	3--234
														10	YP_433908	Glycerophosphoryl diester phosphodiesterase	98.98	1.15E-19	29%	48%	10--224	2--230
C064L	35981--35634	116	13,656	9.92	1	pfam03013	Pyr_excise, Pyrimidine dimer DNA glycosylase. Pyrimidine dimer DNA glycosylases excise pyrimidine dimers by hydrolysis of the glycosylic bond of the 5' pyrimidine, followed by the intra-pyrimidine phosphodiester bond. Pyrimidine dimers are the major UV-lesions of DNA.	157.51	1.82E-39	53%	68%	1--112	26--135	1	AAD33377	pyrimidine dimer-specific glycosylase	239.58	2.08E-62	100%	100%	1--116	26--141
														2	AAD33382	pyrimidine dimer-specific glycosylase	237.65	7.92E-62	99%	99%	1--116	26--141
														3	NP_048398	PBCV-1 pyrimidine dimer-specific glycosylase	208.38	5.15E-53	88%	92%	1--116	26--141
														4	AAD33379	pyrimidine dimer-specific glycosylase	207.22	1.15E-52	87%	92%	1--116	26--141
														5	AAD33375	pyrimidine dimer-specific glycosylase	207.22	1.15E-52	87%	92%	1--116	26--141
														6	AAD33381	pyrimidine dimer-specific glycosylase	206.45	1.96E-52	87%	91%	1--116	26--141
														7	AAD33352	pyrimidine dimer-specific glycosylase	206.45	1.96E-52	87%	91%	1--116	26--141
														8	AAD33353	pyrimidine dimer-specific glycosylase	206.07	2.55E-52	87%	92%	1--116	26--141
														9	AAD33367	pyrimidine dimer-specific glycosylase	206.07	2.55E-52	87%	91%	1--116	26--141
														10	AAD33374	pyrimidine dimer-specific glycosylase	205.68	3.34E-52	87%	92%	1--116	26--141
C065R	36068--36358	97	11,095	10.05		No Hit Found									No Hit Found	No Hit Found						
C066L	36664--36386	93	10,785	6.23		No Hit Found								1	NP_048399	contains type I hydrophobic transmembrane region and ATP/GTP binding motif	179.87	1.92E-44	89%	95%	1--93	111--203
														2	ZP_00279033	hypothetical protein Bcep02006353	83.96	1.43E-15	36%	64%	1--92	113--210
														3	AAQ58722	conserved hypothetical protein	66.63	2.37E-10	37%	58%	1--92	102--198
														4	AAQ60907	conserved hypothetical protein	61.23	9.95E-09	33%	52%	1--92	98--197
														5	NP_792595	hypothetical protein PSPTO2790	58.15	8.42E-08	33%	57%	1--93	171--270
														6	YP_235595	hypothetical protein Psyr_2518	55.07	7.13E-07	29%	55%	1--92	127--225
														7	ABA75491	conserved hypothetical protein	54.68	9.31E-07	30%	53%	1--92	125--223
														8	AAZ37859	conserved hypothetical protein	54.68	9.31E-07	29%	55%	1--92	127--225
														9	ZP_00902424	conserved hypothetical protein	54.68	9.31E-07	31%	54%	1--92	121--219
														10	NP_746428	hypothetical protein PP4312	51.22	1.03E-05	30%	52%	1--92	121--219
C068R	36818--38461	548	61,710	10.9	1	COG0675	COG0675, Transposase and inactivated derivatives [DNA replication, recombination, and repair].	84.75	1.58E-17	21%	36%	182--546	1--347	1	ABA24789	Transposase, IS891/IS1136/IS1341	147.13	1.42E-33	27%	44%	58--546	20--533
					2	pfam07282	Transposase_35, Putative transposase DNA-binding domain. This putative domain is found at the C-terminus of a large number of transposase proteins. This domain contains four conserved cysteines suggestive of a zinc binding domain. Given the need for transposases to bind DNA as well as the large number of DNA-binding zinc fingers we hypothesise this domain is DNA-binding.	72.96	5.02E-14	41%	51%	476--544	1--69	2	BAB78230	transposase	145.98	3.16E-33	28%	45%	102--546	20--489
					3	pfam01385	Transposase_2, Probable transposase. This family includes IS891, IS1136 and IS1341..	48.37	1.33E-06	21%	38%	187--453	1--277	3	AAU06281	putative transposase	124.02	1.29E-26	29%	44%	180--544	43--419
					4	pfam05471	Podocalyxin, Podocalyxin. This family consists of several eukaryotic podocalyxin proteins. Podocalyxin is a major membrane protein of the glomerular epithelium and is thought to be involved in maintenance of the architecture of the foot processes and filtration slits characteristic of this unique epithelium by virtue of its high negative charge. Podocalyxin functions as an anti-adhesin that maintains an open filtration pathway between neighbouring foot processes in the glomerular epithelium by charge repulsion.	37.81	1.82E-03	18%	35%	15--165	65--218	4	ZP_00158267	COG0675: Transposase and inactivated derivatives	120.94	1.09E-25	31%	48%	240--546	39--359
					5	pfam05109	Herpes_BLLF1, Herpes virus major outer envelope glycoprotein (BLLF1). This family consists of the BLLF1 viral late glycoprotein, also termed gp350/220. It is the most abundantly expressed glycoprotein in the viral envelope of the Herpesviruses and is the major antigen responsible for stimulating the production of neutralising antibodies in vivo.	35.99	6.29E-03	20%	33%	8--206	573--792	5	NP_048981	similar to Synechocystis transposase, corresponds to GenBank Accession Number D90909	119.01	4.14E-25	28%	44%	180--544	37--413
														6	YP_142458	putative transposase	104.76	8.08E-21	30%	48%	311--545	289--536
C069R	36823--37380	186	21,414	9.83	1	COG2452	COG2452, Predicted site-specific integrase-resolvase [DNA replication, recombination, and repair].	111.93	8.87E-26	38%	60%	4--179	3--182	1	YP_142457	putative resolvase	139.81	3.78E-32	42%	59%	9--174	15--183
					2	pfam00239	Resolvase, Resolvase, N terminal domain. The N-terminal domain of the resolvase family (this family) contains the active site and the dimer interface. The extended arm at the C-terminus of this domain connects to the C-terminal helix-turn-helix domain of resolvase - see pfam02796..	62.96	5.18E-11	30%	49%	55--174	4--126	2	YP_143125	putative resolvase	133.27	3.54E-30	37%	60%	6--179	8--190
														3	YP_142434	putative resolvase	132.49	6.04E-30	37%	59%	6--179	8--190
														4	BAD84844	predicted site-specific integrase/resolvase	120.94	1.82E-26	39%	59%	4--179	9--193
														5	AAL82147	DNA invertase	115.16	9.97E-25	38%	57%	4--185	15--207
														6	AAM23803	predicted site-specific integrase-resolvase	114.78	1.30E-24	35%	55%	1--181	1--194
														7	AAM23946	predicted site-specific integrase-resolvase	114.01	2.22E-24	35%	55%	1--181	1--193
														8	BAE47874	putative IS transposase (OrfA)	114.01	2.22E-24	34%	55%	17--186	18--203
														9	ZP_00682177	Resolvase, N-terminal	113.62	2.90E-24	34%	57%	10--174	16--189
														10	AAM24002	predicted site-specific integrase-resolvase	113.62	2.90E-24	35%	55%	1--181	1--192
														7	AAS54227	AGL264Wp	104.38	1.06E-20	26%	43%	186--544	71--451
C072L	38922--38650	91	9,864	10.83		No Hit Found								1	NP_048399	contains type I hydrophobic transmembrane region and ATP/GTP binding motif	115.55	4.48E-25	96%	98%	21--77	23--79
														2	ZP_00279033	hypothetical protein Bcep02006353	66.63	2.38E-10	65%	82%	32--77	36--81
														3	AAQ58722	conserved hypothetical protein	63.93	1.55E-09	63%	78%	32--77	25--70
														4	AAQ60907	conserved hypothetical protein	58.15	8.48E-08	59%	73%	32--80	21--69
														5	NP_792595	hypothetical protein PSPTO2790	55.84	4.21E-07	52%	77%	32--79	92--139
														6	NP_746428	hypothetical protein PP4312	55.45	5.50E-07	52%	75%	32--79	42--89
														7	YP_235595	hypothetical protein Psyr_2518	55.07	7.18E-07	50%	77%	32--79	48--95
														8	AAZ37859	conserved hypothetical protein	55.07	7.18E-07	50%	77%	32--79	48--95
														9	ZP_00902424	conserved hypothetical protein	55.07	7.18E-07	52%	75%	32--79	42--89
														10	AAY91559	conserved hypothetical protein	52.37	4.65E-06	47%	72%	32--79	52--99
C073L	40057--39056	334	38,714	6.24		No Hit Found									No Hit Found	No Hit Found						
C074L	40530--40105	142	15,855	4.27		No Hit Found									No Hit Found	No Hit Found						

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
C075L	41615--40587	343	39,184	8.24	1	pfam00145	DNA methylase, C-5 cytosine-specific DNA methylase..	143.91	2.17E-35	27%	42%	5--288	1--289	1	NP_048873	M.CviAII cytosine DNA methyltransferase	323.17	7.58E-87	50%	65%	4--339	1--341
					2	cd00315	Cyt_C5_DNA_methylase, Cytosine-C5 specific DNA methylases; Methyl transfer reactions play an important role in many aspects of biology. Cytosine-specific DNA methylases are found both in prokaryotes and eukaryotes. DNA methylation, or the covalent addition of a methyl group to cytosine within the context of the CpG dinucleotide, has profound effects on the mammalian genome. These effects include transcriptional repression via inhibition of transcription factor binding or the recruitment of methyl-binding proteins and their associated chromatin remodeling factors, X chromosome inactivation, imprinting and the suppression of parasitic DNA sequences. DNA methylation is also essential for proper embryonic development and is an important player in both DNA repair and genome stability	136.59	3.36E-33	36%	52%	5--165	1--166	2	NP_048886	M.CviAIV cytosine DNA methyltransferase	286.57	7.87E-76	46%	61%	4--328	2--332
					3	COG0270	Dcm, Site-specific DNA methylase [DNA replication, recombination, and repair].	106.31	4.27E-24	27%	45%	4--181	3--185	3	AAC64006	cytosine methyltransferase	281.18	3.30E-74	43%	58%	5--339	3--355
														4	AAC55063	cytosine methyltransferase	269.24	1.30E-70	40%	57%	6--339	4--359
														5	NP_049039	nonfunctional M.CviAV cytosine DNA methyltransferase	267.70	3.78E-70	40%	58%	6--339	4--359
														6	AAV84097	CviPII m5C DNA methyltransferase	255.76	1.49E-66	41%	60%	3--339	11--356
														7	NP_818425	gp127	101.68	3.59E-20	39%	52%	5--157	3--165
														8	YP_358798	putative methylase	98.98	2.33E-19	31%	46%	5--227	1--261
														9	BAD65383	site-specific DNA-methyltransferase	98.21	3.97E-19	28%	46%	5--222	1--280
														10	AAV83360	DNA cytosine methylase	95.90	1.97E-18	35%	53%	5--160	99--258
C078L	42275--41646	210	23,885	4.58		No Hit Found									No Hit Found	No Hit Found						
C081R	42351--43109	253	29,734	5.72		No Hit Found									No Hit Found	No Hit Found						
C086R	43131--44066	312	34,055	4.04	1	pfam05887	Trypan_PARP, Procyclic acidic repetitive protein (PARP). This family consists of several Trypanosoma brucei procyclic acidic repetitive protein (PARP) like sequences. The procyclic acidic repetitive protein (parp) genes of Trypanosoma brucei encode a small family of abundant surface proteins whose expression is restricted to the procyclic form of the parasite. They are found at two unlinked loci, parpA and parpB; transcription of both loci is developmentally regulated	43.81	3.21E-05	40%	40%	164--199	80--115	1	NP_048415	contains Pro-rich Px motif EPSPEPxP (5X), and PEST sequence; similar to trypanosome procyclin precursor, corresponds to Swiss-Prot Accession Number P08469	211.08	3.63E-53	85%	94%	200--312	197--309
C089R	44100--45158	353	42,019	6.93		No Hit Found								1	NP_048419	similar to Mycoplasma hypothetical protein MG366, corresponds to Swiss-Prot Accession Number P47606	613.22	3.84E-174	83%	91%	2--353	3--354
														2	NP_048411	A63L	149.83	1.20E-34	36%	52%	126--353	5--232
														3	AAZ58444	conserved hypothetical protein	90.12	1.13E-16	40%	63%	5--117	14--130
														4	ZP_00680654	conserved hypothetical protein	86.66	1.25E-15	38%	56%	4--117	6--126
														5	AAO29065	conserved hypothetical protein	85.11	3.64E-15	38%	57%	4--117	6--126
														6	AAM40292	conserved hypothetical protein	83.19	1.38E-14	38%	55%	4--117	6--126
														7	BAE67710	conserved hypothetical protein	83.19	1.38E-14	38%	55%	4--117	6--126
														8	CAJ22732	conserved hypothetical protein	83.19	1.38E-14	38%	55%	4--117	6--126
														9	BAE53285	hypothetical protein	83.19	1.38E-14	36%	59%	4--117	5--125
														10	ZP_00683615	conserved hypothetical protein	82.03	3.08E-14	37%	55%	6--117	1--119
C090R	45224--46165	314	36,108	8.14		No Hit Found								1	NP_048502	A154L	481.87	1.11E-134	70%	82%	3--314	37--347
														2	NP_048920	similar to Chlorella virus PBCV-1 ORF A154L, corresponds to GenBank Accession Number U42580	480.33	3.24E-134	72%	82%	3--314	41--350
														3	NP_048477	similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903	454.14	2.49E-126	61%	73%	3--314	5--356
														4	NP_077492	EsV-1-7	67.78	5.01E-10	38%	52%	40--144	4--116
														5	AAG37861	variant-specific surface protein VSP136-4	50.83	6.34E-05	26%	37%	23--170	579--744
C093L	47028--46165	288	34,217	6.02		No Hit Found								1	NP_048423	A75L	516.15	4.63E-145	86%	95%	9--288	1--280
C097L	47570--47040	177	20,853	4.9		No Hit Found									No Hit Found	No Hit Found						
C100R	47598--47942	115	13,367	8.91		No Hit Found									No Hit Found	No Hit Found						
C102L	48118--47864	85	9,668	6.52		No Hit Found								1	NP_048425	contains type 1 hydrophobic transmembrane region	84.34	1.13E-15	68%	83%	10--69	14--73
C104R	48197--49093	299	33,469	5.48	1	COG0388	COG0388, Predicted amidohydrolase [General function prediction only].	189.60	3.71E-49	35%	50%	3--296	1--270	1	NP_048426	contains ATP/GTP-binding site motif A; similar to rat beta-alanine synthetase, corresponds to Swiss-Prot Accession Number Q03248	503.83	2.54E-141	80%	87%	1--297	1--297
					2	pfam00795	CN_hydrolase, Carbon-nitrogen hydrolase. This family contains hydrolases that break carbon-nitrogen bonds. The family includes: Nitrilase EC:3.5.5.1, Aliphatic amidase EC:3.5.1.4, Biotidinase EC:3.5.1.12, Beta-ureidopropionase EC:3.5.1.6..	143.94	2.28E-35	34%	53%	6--179	1--174	2	ZP_00403767	COG0388: Predicted amidohydrolase	286.96	4.88E-76	48%	65%	3--296	2--289
					3	COG0815	Lnt, Apolipoprotein N-acyltransferase [Cell envelope biogenesis, outer membrane].	56.57	4.81E-09	20%	33%	18--217	233--451	3	AAK99627	Beta-alanine synthase or beta-ureidopropionase	285.03	1.85E-75	48%	65%	3--296	2--289
														4	ABA40443	beta-alanine synthase-like protein	281.95	1.57E-74	48%	63%	3--293	6--287
														5	CAB45873	beta-alanine synthase	281.18	2.68E-74	48%	63%	3--293	6--287
														6	NP_565650	NLP1 (NITRILASE-LIKE PROTEIN 1); hydrolase, acting on carbon-nitrogen (but not peptide) bonds	278.49	1.74E-73	46%	63%	3--293	8--289
														7	CAG77171	putative carbon-nitrogen hydrolase	278.49	1.74E-73	47%	62%	3--296	2--285
														8	ZP_00140725	COG0388: Predicted amidohydrolase	278.10	2.27E-73	47%	63%	1--296	1--283
														9	ZP_00417184	Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase	276.17	8.61E-73	46%	63%	1--296	1--283
														10	ZP_00910860	probable hydratase	273.09	7.29E-72	47%	64%	3--293	2--279
C110R	49457--50296	280	32,423	7.21		No Hit Found								1	NP_049003	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	185.65	1.38E-45	45%	62%	70--280	1--189
														2	NP_049005	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	87.81	3.91E-16	67%	83%	64--118	1--55
														3	NP_048807	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	85.50	1.94E-15	70%	93%	70--117	1--48
														4	NP_048525	A177R	69.71	1.10E-10	54%	70%	67--121	1--55
														5	NP_048427	A79R	66.24	1.22E-09	50%	76%	70--119	1--50
														6	NP_048629	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	63.54	7.89E-09	49%	72%	67--117	1--51
C112R	50572--51303	244	27,530	9.61		No Hit Found								1	NP_048629	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	412.54	5.36E-114	78%	88%	1--244	1--252
														2	NP_048807	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	364.39	1.68E-99	70%	83%	4--244	1--249
														3	NP_049005	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	364.00	2.19E-99	69%	81%	1--243	4--255
														4	NP_048427	A79R	351.67	1.12E-95	73%	86%	4--228	1--226
														5	AAU06304	hypothetical protein A275R	334.34	1.86E-90	94%	98%	74--244	1--171
														6	AAU06301	hypothetical protein A275R	333.18	4.14E-90	94%	98%	74--244	1--171
														7	NP_048525	A177R	317.78	1.80E-85	62%	77%	1--242	1--243
														8	AAU06302	hypothetical protein A275R	272.71	6.63E-72	98%	99%	110--244	1--135
														9	AAU06303	hypothetical protein A275R	115.55	1.36E-24	93%	96%	186--244	1--59
														10	AAU06299	hypothetical protein A275R	112.85	8.78E-24	91%	94%	186--244	5--63
C114L	51896--51321	192	22,245	6.96		No Hit Found								1	NP_048429	A81L	287.73	1.20E-76	73%	88%	4--190	1--187
C116L	52522--51959	188	22,359	7.42		No Hit Found								1	NP_048432	A84L	328.18	7.58E-89	82%	92%	1--188	1--188
C118R	52637--53326	230	26,661	4.8	1	smart00702	P4Hc, Prolyl 4-hydroxylase alpha subunit homologues. Mammalian enzymes catalyse hydroxylation of collagen, for example. Prokaryotic enzymes might catalyse hydroxylation of antibiotic peptides. These are 2-oxoglutarate-dependent dioxygenases, requiring 2-oxoglutarate and dioxygen as cosubstrates and ferrous iron as a cofactor	116.32	3.97E-27	33%	45%	46--228	6--178	1	NP_048433	PBCV-1 prolyl 4-hydroxylase	347.44	1.89E-94	76%	81%	8--229	15--240

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					2	pfam03171	2OG-FeII_Oxy, 2OG-Fe(II) oxygenase superfamily. This family contains members of the 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily. This family includes the C-terminal of prolyl 4-hydroxylase alpha subunit. The holoenzyme has the activity EC:1.14.11.2 catalysing the reaction: Procollagen L-proline + 2-oxoglutarate + O2 <=> procollagen trans- 4-hydroxy-L-proline + succinate + CO2. The full enzyme consists of a alpha2 beta2 complex with the alpha subunit contributing most of the parts of the active site. The family also includes lysyl hydrolases, isopenicillin synthases and AlkB	48.19	1.59E-06	28%	40%	126--229	2--96	2	AAZ62310	Procollagen-proline,2-oxoglutarate-4-dioxygenase	84.34	2.99E-15	33%	46%	50--228	99--274
														3	ZP_00984285	hypothetical protein BdolA 01003928	82.80	8.70E-15	29%	45%	38--228	70--257
														4	ZP_00238502	prolyl 4-hydroxylase alpha subunit	80.88	3.31E-14	29%	50%	46--228	44--211
														5	AAT63151	prolyl 4-hydroxylase, alpha subunit	80.49	4.32E-14	29%	50%	46--228	60--227
														6	YP_142947	prolyl 4-hydroxylase	78.95	1.26E-13	26%	39%	1--228	1--237
														7	AAU16279	prolyl 4-hydroxylase, alpha subunit	78.95	1.26E-13	29%	49%	46--228	60--227
														8	ZP_00740932	Prolyl 4-hydroxylase alpha subunit	78.95	1.26E-13	29%	49%	50--228	80--243
														9	AAS43215	prolyl 4-hydroxylase, alpha subunit domain protein	77.80	2.80E-13	29%	49%	46--228	44--211
														10	AAP11148	Prolyl 4-hydroxylase alpha subunit	77.41	3.66E-13	29%	49%	50--228	64--227
C121R	53408--53890	161	18,151	4.15		No Hit Found								1	NP_048486	A138R	50.06	2.64E-05	24%	51%	46--151	151--255
C123R	53950--54411	154	18,470	5		No Hit Found								1	NP_048438	contains phenyl group binding site (CAAX box)	303.91	8.77E-82	89%	96%	1--154	1--154
C124L	55696--54428	423	48,826	10.89		No Hit Found								1	NP_048441	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563	535.03	1.72E-150	90%	93%	1--284	1--296
														2	NP_048439	a91L	228.02	4.49E-58	86%	92%	296--423	1--126
														3	NP_048636	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055	72.79	2.42E-11	37%	53%	3--104	417--543
														4	NP_048632	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662	70.09	1.57E-10	40%	58%	3--81	516--610
														5	NP_049032	similar to Chlorella virus PBCV-1 ORF A282L, corresponds to GenBank Accession Number U42580	65.08	5.04E-09	40%	54%	41--139	1--102
C126L	56709--55750	320	36,869	4.18	1	cd02180	GH16_laminarinase, Laminarinase, also known as glucan endo-1,3-beta-D-glucosidase, is a glycosyl hydrolase family 16 member that hydrolyzes 1,3-beta-D-glucosidic linkages in 1,3-beta-D-glucans such as laminarins, curdlans, paramylons, and pachymans, with very limited action on mixed-link (1,3-1,4-)-beta-D-glucans	174.30	1.40E-44	38%	54%	68--318	1--237	1	AAX16367	1,3(4)-beta-glucanase	158.69	2.23E-37	37%	54%	68--318	68--307
					2	cd02182	GH16_laminarinase_like, A beta-1,3-glucanase (laminarinase)-like protein exists in the bacterial genus Streptomyces as well as the fungal class Sordariomycetes. The laminarinases belong to glycosyl hydrolase family 16 all of which have a conserved jelly roll fold with an active site channel. The bacterial members contain an additional C-terminal carbohydrate-binding module (CBM)	109.61	5.12E-25	30%	46%	67--318	3--257	2	P23903	Glucan endo-1,3-beta-glucosidase A1 precursor ((1->3)-beta-glucan endohydrolase) ((1->3)-beta-glucanase A1)	153.68	7.17E-36	35%	50%	63--318	420--679
					3	cd00413	Glyco_hydrolase_16, The O-Glycosyl hydrolases are a widespread group of enzymes that hydrolyse the glycosidic bond between two or more carbohydrates, or between a carbohydrate and a non-carbohydrate moiety. A glycosyl hydrolase classification system based on sequence similarity has led to the definition of more than 95 different families inlcuding glycoside hydrolase family 16. Family 16 includes lichenase, xyloglucan endotransglycosylase (XET), beta-agarase, kappa-carrageenase, endo-beta-1,3-glucanase, endo-beta-1,3-1,4-glucanase, and endo-beta-galactosidase, all of which have a conserved jelly roll fold with a deep active site channel harboring the catalytic residues..	101.61	1.11E-22	32%	46%	72--319	1--218	3	BAD63242	endo-beta-1,3-glucanase	152.53	1.60E-35	37%	51%	68--319	36--279
					4	cd02179	GH16_beta_GRP, Beta-GRP (beta-1,3-glucan recognition protein) is one of several pattern recognition receptors (PRRs), also referred to as biosensor proteins, that complexes with pathogen-associated beta-1,3-glucans and then transduces signals necessary for activation of an appropriate immune response. Their structures adopt a jelly roll fold with a deep active site channel harboring the catalytic residues, like those of other glycosyl hydrolase family 16 members	82.66	6.37E-17	28%	44%	153--294	117--275	4	ZP_00504674	Glycoside hydrolase, family 16:S-layer protein (SLH domain):Carbohydrate-binding, CenC-like	150.21	7.93E-35	35%	52%	68--318	427--668
					5	COG2273	SKN1, Beta-glucanase/Beta-glucan synthetase [Carbohydrate transport and metabolism].	65.85	7.28E-12	23%	37%	63--319	38--264	5	CAA61884	endo-1,3(4)-beta-glucanase	150.21	7.93E-35	35%	52%	68--318	427--668
					6	pfam00722	Glyco_hydro_16, Glycosyl hydrolases family 16..	52.53	6.14E-08	26%	39%	154--316	47--182	6	ZP_00767179	Glycoside hydrolase, family 16	143.67	7.42E-33	32%	51%	67--318	39--268
					7	cd02177	GH16_kappa_carrageenase, Kappa-carrageenase degrades kappa-carrageenans which are the gel-forming, sulfated 1,3-alpha-1,4-beta-galactans that make up the cell walls of marine red algae such as Rhodophyceaea. Kappa-carrageenases exist in bacteria belonging to at least three phylogenetically distant branches, including pseudoalteromonas, planctomycetes, and baceroidetes. This domain adopts a curved beta-sandwich conformation, with a tunnel-shaped active site cavity, referred to as a jellyroll fold	44.33	2.22E-05	26%	40%	67--318	9--268	7	EAN71367	Glycoside hydrolase, family 16	143.28	9.69E-33	34%	48%	68--318	51--326
					8	pfam03935	SKN1, Beta-glucan synthesis-associated protein (SKN1). This family consists of the beta-glucan synthesis-associated proteins KRE6 and SKN1. Beta1,6-Glucan is a key component of the yeast cell wall, interconnecting cell wall proteins, beta1,3-glucan, and chitin. It has been postulated that the synthesis of beta1,6-glucan begins in the endoplasmic reticulum with the formation of protein-bound primer structures and that these primer structures are extended in the Golgi complex by two putative glucosyltransferases that are functionally redundant, Kre6 and Skn1. This is followed by maturation steps at the cell surface and by coupling to other cell wall macromolecules	40.84	2.61E-04	31%	51%	245--318	622--689	8	AAC60453	beta-1,3-glucanase	142.90	1.27E-32	34%	48%	52--318	408--682
														9	ZP_00908236	Carbohydrate-binding family V/XII:Fibronectin, type III	142.90	1.27E-32	34%	50%	68--318	35--263
														10	YP_435911	Beta-glucanase/Beta-glucan synthetase	140.20	8.20E-32	35%	51%	67--318	332--572
C128R	56788--58311	508	58,792	8.88	1	pfam03142	Chitin_synth_2, Chitin synthase. Members of this family are fungal chitin synthase EC:2.4.1.16 enzymes. They catalyse chitin synthesis as follows: UDP-N-acetyl-D-glucosamine + {(1,4)-(N-acetyl-beta-D-glucosaminyl)}(N) <=> UDP + {(1,4)-(N-acetyl-beta-D-glucosaminyl)}(N+1)	82.27	6.97E-17	22%	38%	88--506	30--496	1	BAE48153	chitin synthase	984.17	0.00E+00	94%	97%	1--507	1--507
					2	COG1215	COG1215, Glycosyltransferases, probably involved in cell wall biogenesis [Cell envelope biogenesis, outer membrane].	65.34	9.34E-12	20%	39%	40--497	10--397	2	BAB83509	chitin synthase	294.66	4.98E-78	37%	55%	26--505	11--501
														3	EAA78335	hypothetical protein FG06550.1	174.10	9.80E-42	26%	45%	7--502	130--649
														4	EAA68628	hypothetical protein FG10619.1	172.17	3.72E-41	26%	46%	38--502	202--684
														5	BAE60326	unnamed protein product	154.84	6.15E-36	27%	44%	8--502	107--580
														6	BAE48158	chitin synthase	122.48	3.38E-26	37%	56%	8--215	4--209
														7	EAA72910	hypothetical protein FG03170.1	105.92	3.28E-21	27%	44%	87--389	257--563
														8	XP_503779	hypothetical protein	104.38	9.54E-21	24%	40%	33--504	737--1228
														9	NP_077569	ESV-1-84	101.29	8.07E-20	24%	42%	44--502	33--484
														10	BAA74449	Csm1	96.67	1.99E-18	23%	42%	47--504	192--1675
C132R	58472--60256	595	65,399	5.89	1	COG0449	GlmS, Glucosamine 6-phosphate synthetase, contains amidotransferase and phosphosugar isomerase domains [Cell envelope biogenesis, outer membrane].	618.74	2.34E-178	44%	63%	1--593	1--595	1	NP_048448	PBCV-1 glucosamine synthetase	1084.32	0.00E+00	91%	94%	1--595	1--595
					2	cd00714	GFAT, Glutamine amidotransferases class-II (Gn-AT)_GFAT-type. This domain is found at the N-terminus of glucosamine 6-phosphate (GlcN-6-P) synthase (GLMS or GFAT).The glutaminase domain catalyzes amide nitrogen transfer from glutamine to the appropriate substrate. In this process, glutamine is hydrolyzed to glutamic acid and ammonia. GFAT catalyzes the formation of glucosamine 6-phosphate from fructose 6-phosphate and glutamine, the initiating step in the biosynthesis of UDP-GlcN-6-P	262.76	3.50E-71	47%	68%	2--216	1--215	2	BAD15299	glutamine:fructose-6-phosphate amidotransferase GFAT	1083.17	0.00E+00	91%	95%	1--595	1--596
					3	COG2222	AgaS, Predicted phosphosugar isomerases [Cell envelope biogenesis, outer membrane].	177.78	1.45E-45	29%	46%	254--594	5--338	3	CAE39493	glucosamine--fructose-6-phosphate aminotransferase	528.09	3.27E-148	46%	65%	1--593	1--608

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					4	cd00715	GPATase_N, Glutamine amidotransferases class-II (GN-AT)_GPAT-type. This domain is found at the N-terminus of glutamine phosphoribosylpyrophosphate (Prpp) amidotransferase (GPATase) . The glutaminase domain catalyzes amide nitrogen transfer from glutamine to the appropriate substrate. In this process, glutamine is hydrolyzed to glutamic acid and ammonia. GPATase catalyzes the first step in purine biosynthesis, an amide transfer from glutamine to PRPP, resulting in phosphoribosylamine, pyrophosphate and glutamate. GPATase crystalizes as a homotetramer, but can also exist as a homdimer..	120.62	2.47E-28	29%	48%	2--252	1--244	4	CAE44992	glucosamine--fructose-6-phosphate aminotransferase	526.17	1.24E-147	46%	65%	1--593	1--608
					5	COG0034	PurF, Glutamine phosphoribosylpyrophosphate amidotransferase [Nucleotide transport and metabolism].	118.00	1.26E-27	28%	52%	1--232	4--233	5	ZP_00244599	COG0449: Glucosamine 6-phosphate synthetase, contains amidotransferase and phosphosugar isomerase domains	523.09	1.05E-146	45%	63%	1--593	1--617
					6	pfam00310	GATase 2, Glutamine amidotransferases class-II..	114.63	1.42E-26	41%	61%	2--136	1--136	6	ZP_00509192	Glucosamine-fructose-6-phosphate aminotransferase, isomerising	515.38	2.20E-144	44%	63%	1--593	1--614
					7	cd00352	Gn_AT_II, Glutamine amidotransferases class-II (GATase). The glutaminase domain catalyzes an amide nitrogen transfer from glutamine to the appropriate substrate. In this process, glutamine is hydrolyzed to glutamic acid and ammonia. This domain belongs t	109.41	5.79E-25	28%	47%	2--221	1--220	7	ZP_00942990	Glucosamine--fructose-6-phosphate aminotransferase (isomerizing)	509.99	9.22E-143	44%	63%	1--593	1--610
					8	pfam01380	SIS, SIS domain. SIS (Sugar ISomerase) domains are found in many phosphosugar isomerases and phosphosugar binding proteins. SIS domains are also found in proteins that regulate the expression of genes involved in synthesis of phosphosugars. Presumably the SIS domains bind to the end-product of the pathway.	99.98	3.50E-22	37%	53%	285--417	2--135	8	YP_438548	glucosamine--fructose-6-phosphate aminotransferase, isomerizing	508.06	3.51E-142	45%	63%	1--593	1--603
					9	COG0367	AsnB, Asparagine synthase (glutamine-hydrolyzing) [Amino acid transport and metabolism].	83.95	2.24E-17	30%	46%	1--246	1--209	9	CAD13706	PROBABLE GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE PROTEIN	506.52	1.02E-141	44%	63%	1--593	1--610
					10	cd00712	AsnB, Glutamine amidotransferases class-II (GATase) asparagine synthase_B type. Asparagine synthetase B catalyses the ATP-dependent conversion of aspartate to asparagine. This enzyme is a homodimer, with each monomer composed of a glutaminase domain and a synthetase domain. The N-terminal glutaminase domain hydrolyzes glutamine to glutamic acid and ammonia	77.56	2.21E-15	33%	50%	2--200	1--164	10	ZP_00983220	COG0449: Glucosamine 6-phosphate synthetase, contains amidotransferase and phosphosugar isomerase domains	504.21	5.06E-141	45%	64%	1--593	1--603
C134L	61817--60372	482	54,659	4.59	1	COG3534	AbfA, Alpha-L-arabinofuranosidase [Carbohydrate transport and metabolism].	57.61	1.97E-09	28%	40%	50--295	41--285	1	AAC45377	endoglucanase F precursor	62.39	3.90E-08	22%	37%	42--351	642--915
					2	pfam04495	GRASP55_65, GRASP55/65 family. GRASP55 (Golgi reassembly stacking protein of 55 kDa) and GRASP65 (a 65 kDa) protein are highly homologous. GRASP55 is a component of the Golgi stacking machinery. GRASP65, an N-ethylmaleimide- sensitive membrane protein required for the stacking of Golgi cisternae in a cell-free system..	36.26	5.30E-03	31%	55%	263--326	64--129	2	BAA10965	cellulose-binding protein	62.00	5.09E-08	22%	37%	42--351	642--915
														3	CAD86595	cellulase precursor	60.08	1.93E-07	24%	36%	69--310	280--517
														4	CAF22222	putative cellulase	55.84	3.65E-06	24%	40%	64--255	263--416
														5	ZP_00526510	Alpha-L-arabinofuranosidase	51.22	8.98E-05	24%	38%	64--254	70--262
														6	CAA61937	putative alpha-L-arabinofuranosidase	51.22	8.98E-05	25%	41%	25--254	17--244
C137R	61928--62908	327	38,067	9.09	1	pfam01331	mRNA_cap_enzyme, mRNA capping enzyme, catalytic domain. This family represents the ATP binding catalytic domain of the mRNA capping enzyme..	145.43	7.36E-36	27%	49%	57--233	1--192	1	NP_048451	PBCV-1 mRNA guanylyltransferase	518.46	1.15E-145	73%	90%	8--327	11--330
					2	COG5226	CEG1, mRNA capping enzyme, guanylyltransferase (alpha) subunit [RNA processing and modification].	102.42	6.92E-23	25%	40%	52--324	42--362	2	1CKN_B	Chain B, Structure Of Guanylylated Mrna Capping Enzyme Complexed With Gtp	516.15	5.70E-145	73%	90%	8--327	11--330
					3	pfam03919	mRNA_cap_C, mRNA capping enzyme, C-terminal domain..	57.68	2.13E-09	26%	43%	236--322	1--110	3	NP_974263	mRNA guanylyltransferase/ phosphoprotein phosphatase/ protein tyrosine/serine/threonine phosphatase	94.74	4.10E-18	25%	46%	40--325	337--650
														4	AAT68133	mRNA capping enzyme	92.43	2.03E-17	26%	46%	54--308	273--547
														5	CAG09212	unnamed protein product	91.28	4.53E-17	25%	46%	34--308	249--541
														6	EAL46815	mRNA capping enzyme, putative	89.35	1.72E-16	27%	45%	47--321	358--662
														7	ABA92070	mRNA capping enzyme - like protein	88.20	3.83E-16	24%	47%	52--321	370--668
														8	P78587	mRNA capping enzyme alpha subunit (mRNA guanylyltransferase) (GTP--RNA guanylyltransferase) (GTase)	84.73	4.24E-15	26%	41%	54--321	41--371
														9	CAB10156	ceg1	83.57	9.44E-15	26%	41%	54--321	42--357
														10	1P16_B	Chain B, Structure Of An Mrna Capping Enzyme Bound To The Phosphorylated Carboxyl-Terminal Domain Of Rna Polymerase Ii	83.19	1.23E-14	26%	40%	54--321	41--371
C140L	63853--63002	284	33,296	6.83	1	pfam00443	UCH, Ubiquitin carboxyl-terminal hydrolase..	128.20	1.24E-30	25%	41%	3--279	5--312	1	NP_048453	contains ubquitin carboxy-terminal hydrolase active sites; similar to human ubiquitin carboxy-terminal hydrolase, coresponds to Swiss-Prot Accession Number Q09879	454.91	1.23E-126	73%	88%	1--283	1--282
					2	cd02257	Peptidase_C19, Peptidase C19 contains ubiquitinyl hydrolases. They are intracellular peptidases that remove ubiquitin molecules from polyubiquinated peptides by cleavage of isopeptide bonds. They hydrolyse bonds involving the carboxyl group of the C-terminal Gly residue of ubiquitin The purpose of the de-ubiquitination is thought to be editing of the ubiquitin conjugates, which could rescue them from degradation, as well as recycling of the ubiquitin. The ubiquitin/proteasome system is responsible for most protein turnover in the mammalian cell, and with over 50 members, family C19 is one of the largest families of peptidases in the human genome.	80.74	2.45E-16	19%	33%	4--280	2--320	2	EAA08027	ENSANGP00000018711	70.09	8.55E-11	21%	38%	1--279	557--886
					3	cd02661	Peptidase_C19E, A subfamily of Peptidase C19. Peptidase C19 contains ubiquitinyl hydrolases. They are intracellular peptidases that remove ubiquitin molecules from polyubiquinated peptides by cleavage of isopeptide bonds. They hydrolyze bonds involving the carboxyl group of the C-terminal Gly residue of ubiquitin. The purpose of the de-ubiquitination is thought to be editing of the ubiquitin conjugates, which could rescue them from degradation, as well as recycling of the ubiquitin. The ubiquitin/proteasome system is responsible for most protein turnover in the mammalian cell, and with over 50 members, family C19 is one of the largest families of peptidases in the human genome..	73.37	3.32E-14	22%	44%	5--279	5--303	3	AAN09567	CG14619-PB, isoform B	63.93	6.13E-09	23%	40%	1--281	11--336
					4	cd02674	Peptidase_C19R, A subfamily of peptidase C19. Peptidase C19 contains ubiquitinyl hydrolases. They are intracellular peptidases that remove ubiquitin molecules from polyubiquinated peptides by cleavage of isopeptide bonds. They hydrolyze bonds involving the carboxyl group of the C-terminal Gly residue of ubiquitin. The purpose of the de-ubiquitination is thought to be editing of the ubiquitin conjugates, which could rescue them from degradation, as well as recycling of the ubiquitin. The ubiquitin/proteasome system is responsible for most protein turnover in the mammalian cell, and with over 50 members, family C19 is one of the largest families of peptidases in the human genome..	73.45	3.58E-14	19%	39%	5--280	3--335	4	AAN09566	CG14619-PC, isoform C	63.93	6.13E-09	23%	40%	1--281	612--937
					5	cd02659	peptidase_C19C, A subfamily of Peptidase C19. Peptidase C19 contains ubiquitinyl hydrolases. They are intracellular peptidases that remove ubiquitin molecules from polyubiquinated peptides by cleavage of isopeptide bonds. They hydrolyze bonds involving the carboxyl group of the C-terminal Gly residue of ubiquitin. The purpose of the de-ubiquitination is thought to be editing of the ubiquitin conjugates, which could rescue them from degradation, as well as recycling of the ubiquitin. The ubiquitin/proteasome system is responsible for most protein turnover in the mammalian cell, and with over 50 members, family C19 is one of the largest families of peptidases in the human genome..	51.02	1.83E-07	35%	59%	225--273	252--301	5	AAN09565	CG14619-PE, isoform E	63.93	6.13E-09	23%	40%	1--281	530--855
					6	cd02660	Peptidase_C19D, A subfamily of Peptidase C19. Peptidase C19 contains ubiquitinyl hydrolases. They are intracellular peptidases that remove ubiquitin molecules from polyubiquinated peptides by cleavage of isopeptide bonds. They hydrolyze bonds involving the carboxyl group of the C-terminal Gly residue of ubiquitin. The purpose of the de-ubiquitination is thought to be editing of the ubiquitin conjugates, which could rescue them from degradation, as well as recycling of the ubiquitin. The ubiquitin/proteasome system is responsible for most protein turnover in the mammalian cell, and with over 50 members, family C19 is one of the largest families of peptidases in the human genome..	47.24	2.45E-06	29%	43%	6--105	5--102	6	EAL32347	GA13118-PA	61.62	3.04E-08	23%	40%	1--281	512--837
					7	COG5560	UBP12, Ubiquitin C-terminal hydrolase [Posttranslational modification, protein turnover, chaperones].	45.39	1.07E-05	23%	43%	6--169	270--446	7	CAA20678	SPCC1682.12c	60.08	8.85E-08	23%	40%	6--281	129--428

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					8	cd02657	Peptidase_C19A, A subfamily of Peptidase C19. Peptidase C19 contains ubiquitinyl hydrolases. They are intracellular peptidases that remove ubiquitin molecules from polyubiquinated peptides by cleavage of isopeptide bonds. They hydrolyse bonds involving the carboxyl group of the C-terminal Gly residue of ubiquitin. The purpose of the de-ubiquitination is thought to be editing of the ubiquitin conjugates, which could rescue them from degradation, as well as recycling of the ubiquitin. The ubiquitin/proteasome system is responsible for most protein turnover in the mammalian cell, and with over 50 members, family C19 is one of the largest families of peptidases in the human genome..	41.46	1.57E-04	19%	34%	4--203	2--204	8	CAG61157	unnamed protein product	56.23	1.28E-06	23%	40%	6--281	389--720
					9	COG5077	COG5077, Ubiquitin carboxyl-terminal hydrolase [Posttranslational modification, protein turnover, chaperones].	40.37	2.85E-04	32%	58%	221--273	427--480	9	CAA89492	UBP12	50.45	7.01E-05	21%	44%	6--169	368--554
					10	cd02663	Peptidase_C19G, A subfamily of Peptidase C19. Peptidase C19 contains ubiquitinyl hydrolases. They are intracellular peptidases that remove ubiquitin molecules from polyubiquinated peptides by cleavage of isopeptide bonds. They hydrolyze bonds involving the carboxyl group of the C-terminal Gly residue of ubiquitin. The purpose of the de-ubiquitination is thought to be editing of the ubiquitin conjugates, which could rescue them from degradation, as well as recycling of the ubiquitin. The ubiquitin/proteasome system is responsible for most protein turnover in the mammalian cell, and with over 50 members, family C19 is one of the largest families of peptidases in the human genome..	40.30	2.93E-04	40%	58%	225--264	291--331	10	EAL29601	GA18919-PA	50.06	9.16E-05	48%	66%	225--267	727--771
C142L	64830--63889	314	35,651	7.08	1	COG1405	SUA7, Transcription initiation factor TFIIIB, Brf1 subunit/Transcription initiation factor TFIIB [Transcription].	83.04	5.03E-17	23%	41%	35--294	2--263	1	NP_048455	similar to Pyrococcus woesei factor TFIIB homolog, corresponds to GenBank Accession Number X70668	421.01	2.33E-116	69%	86%	25--314	1--290
														2	EAM94502	Transcription factor TFIIB	62.39	2.11E-08	21%	40%	39--312	18--307
														3	YP_142604	putative transcription initiation factor IIB	61.62	3.59E-08	22%	43%	44--301	154--432
														4	AAT43923	transcription initiation factor IIB	61.62	3.59E-08	21%	39%	39--312	17--306
														5	BAD85469	transcription initiation factor IIB	60.85	6.13E-08	23%	40%	39--312	10--295
														6	XP_626996	transcription initiation factor TFIIB Sua7p; ZnR+2cyclins	57.77	5.19E-07	24%	46%	36--291	154--410
														7	AAL81501	transcription initiation factor IIB chain b; (TFIIB)	57.77	5.19E-07	23%	41%	39--312	10--295
														8	S34116	transcription factor IIB - Pyrococcus woesei	57.77	5.19E-07	23%	41%	39--312	10--295
														9	EAM94531	Transcription factor TFIIB	57.00	8.84E-07	21%	41%	39--313	12--302
														10	AAL64842	transcription initiation factor IIB (TFIIB)	56.61	1.16E-06	24%	43%	31--293	8--275
C143R	63923--64162	80	9,142	8.7		No Hit Found									No Hit Found	No Hit Found						
C146L	65311--64823	163	19,405	7.62		No Hit Found								1	NP_048457	A109L	184.11	1.19E-45	84%	90%	61--162	1--102
C150R	65432--68011	860	98,810	6.08	1	pfam00535	Glycos_transf_2, Glycosyl transferase. Diverse family, transferring sugar from UDP-glucose, UDP-N-acetyl- galactosamine, GDP-mannose or CDP-abequose, to a range of substrates including cellulose, dolichol phosphate and teichoic acids..	44.83	1.57E-05	22%	39%	262--429	2--167	1	NP_048462	A114R	889.80	0.00E+00	86%	94%	380--860	5--485
					2	COG0463	WcaA, Glycosyltransferases involved in cell wall biogenesis [Cell envelope biogenesis, outer membrane].	42.43	7.53E-05	27%	48%	258--364	3--104	2	NP_048459	A111R	776.93	0.00E+00	96%	98%	1--379	1--379
					3	COG1216	COG1216, Predicted glycosyltransferases [General function prediction only].	39.34	6.81E-04	18%	35%	258--529	3--279	3	ZP_00202013	COG0463: Glycosyltransferases involved in cell wall biogenesis	99.37	5.81E-19	31%	51%	260--472	48--264
					4	COG1215	COG1215, Glycosyltransferases, probably involved in cell wall biogenesis [Cell envelope biogenesis, outer membrane].	37.61	2.26E-03	25%	45%	254--366	50--160	4	CAG34747	hypothetical protein	98.98	7.59E-19	30%	48%	23--235	23--234
					5	pfam05598	DUF772, Sulfolobus solfataricus protein of unknown function (DUF772). This family consists of several proteins from Sulfolobus solfataricus described as first ORF in transposon ISC1212..	36.59	3.93E-03	36%	55%	741--833	23--100	5	AAU37201	unknown	97.44	2.21E-18	26%	44%	260--486	6--241
														6	AAP95345	possible glycosyltransferase	96.29	4.92E-18	31%	46%	255--475	1--225
														7	ZP_00154867	COG1216: Predicted glycosyltransferases	89.35	6.01E-16	29%	42%	260--486	29--261
														8	AAY61346	unknown	63.16	4.61E-08	24%	45%	3--206	53--254
														9	ZP_00340129	hypothetical protein RakaH01000503	62.39	7.87E-08	24%	47%	3--206	53--254
														10	AAU03775	conserved hypothetical protein	60.08	3.91E-07	23%	46%	3--206	49--250
C155R	68043--69071	343	38,606	6.71	1	COG1089	Gmd, GDP-D-mannose dehydratase [Cell envelope biogenesis, outer membrane].	488.25	4.43E-139	59%	74%	2--343	1--341	1	NP_048466	PBCV-1 GDP-D-mannose dehydratase	624.78	1.22E-177	88%	95%	1--340	1--340
					2	pfam01370	Epimerase, NAD dependent epimerase/dehydratase family. This family of proteins utilise NAD as a cofactor. The proteins in this family use nucleotide-sugar substrates for a variety of chemical reactions..	146.12	4.75E-36	31%	47%	6--334	1--300	2	CAB63300	GDP-mannose-4,6-dehydratase	399.44	8.33E-110	57%	73%	3--343	2--353
					3	COG0451	WcaG, Nucleoside-diphosphate-sugar epimerases [Cell envelope biogenesis, outer membrane / Carbohydrate transport and metabolism].	134.32	1.63E-32	30%	46%	4--343	1--311	3	BAC93113	GDP-mannose-4,6-dehydratase	394.82	2.05E-108	56%	74%	1--343	1--354
					4	COG1088	RfbB, dTDP-D-glucose 4,6-dehydratase [Cell envelope biogenesis, outer membrane].	117.24	2.18E-27	27%	48%	6--337	3--313	4	YP_113616	GDP-mannose 4,6-dehydratase	393.66	4.57E-108	57%	73%	1--340	1--350
					5	COG1087	GalE, UDP-glucose 4-epimerase [Cell envelope biogenesis, outer membrane].	105.64	7.91E-24	26%	45%	6--334	3--315	5	ZP_00826187	COG1089: GDP-D-mannose dehydratase	393.66	4.57E-108	56%	73%	4--343	2--352
					6	COG1091	RfbD, dTDP-4-dehydrorhamnose reductase [Cell envelope biogenesis, outer membrane].	64.90	1.36E-11	25%	42%	6--335	3--276	6	BAB03208	putative GDP-mannose dehydratase	392.50	1.02E-107	56%	74%	4--343	2--337
					7	COG0702	COG0702, Predicted nucleoside-diphosphate-sugar epimerases [Cell envelope biogenesis, outer membrane / Carbohydrate transport and metabolism].	52.61	7.31E-08	35%	51%	6--101	3--91	7	AAR38453	GDP-mannose 4,6-dehydratase	390.58	3.87E-107	55%	71%	4--342	2--351
					8	pfam02716	Isoflavone_redu, Isoflavone reductase. This is a family of isoflavone reductases from plants. Isoflavone reductase enzymes EC:1.3.1.45 catalyse the penultimate step in the synthesis of the phytoalexin medicarpin..	50.63	2.50E-07	30%	51%	6--72	6--76	8	ZP_00533296	GDP-mannose 4,6-dehydratase	389.81	6.60E-107	57%	73%	3--342	2--351
					9	pfam02719	Polysacc_synt_2, Polysaccharide biosynthesis protein. This is a family of diverse bacterial polysaccharide biosynthesis proteins including the CapD protein, WalL protein, mannosyl-transferase, and several putative epimerases (e.g. WbiI)..	45.64	7.34E-06	29%	45%	1--116	194--318	9	ZP_00826190	COG1089: GDP-D-mannose dehydratase	389.04	1.13E-106	57%	72%	4--343	2--352
					10	COG1086	COG1086, Predicted nucleoside-diphosphate sugar epimerases [Cell envelope biogenesis, outer membrane / Carbohydrate transport and metabolism].	41.42	1.51E-04	23%	40%	2--168	249--405	10	ABA58008	GDP-mannose 4,6-dehydratase	387.88	2.51E-106	54%	73%	4--342	2--353
C157R	69151--70401	417	48,993	7.25	1	pfam01541	GIY-YIG, GIY-YIG catalytic domain. This domain called GIY-YIG is found in the amino terminal region of excinuclease abc subunit c (uvrC), bacteriophage T4 endonucleases segA, segB, segC, segD and segE; it is also found in putative endonucleases encoded by group I introns of fungi and phage. The structure of I-TevI a GIY-YIG endonuclease, reveals a novel alpha/beta-fold with a central three-stranded antiparallel beta-sheet flanked by three helices. The most conserved and putative catalytic residues are located on a shallow, concave surface and include a metal coordination site	36.68	3.73E-03	31%	53%	25--104	4--85	1	ZP_00738535	hypothetical protein RBTH_06729	50.83	9.63E-05	23%	41%	65--318	16--260
C160R	70425--70736	104	12,564	11.18		No Hit Found								1	NP_048469	A121R	168.70	4.41E-41	76%	91%	8--104	1--97
														2	AAL73467	Tlr 6Fp protein	85.11	6.41E-16	41%	66%	13--98	14--97
														3	AAL73477	Tlr 6Fp protein	83.57	1.86E-15	40%	63%	13--98	14--97
														4	AAL73473	Tlr 6Fp protein	83.57	1.86E-15	40%	66%	13--98	14--97
C162R	70794--74930	1379	143,202	4.96		No Hit Found								1	NP_048470	PBCV-1 Vp260 protein	1132.47	0.00E+00	57%	69%	1--1064	1--1028
														2	AAA86307	glycoprotein Vp260	1034.63	0.00E+00	67%	77%	1--789	1--788
														3	NP_048471	contains a leucine zipper motif; similar to Bacillus subtilis phage P2A preneck appendage protein, corresponds to Swiss-Prot Accession Number P07537	555.44	4.98E-156	84%	93%	1069--1379	1--311
														4	BAB83467	Vp260 like protein	265.77	7.87E-69	28%	43%	12--874	2--816
														5	BAB83469	Vp260 like protein	258.07	1.64E-66	31%	46%	18--799	39--811
														6	BAB83468	Vp260 like protein	255.37	1.06E-65	31%	45%	32--799	46--811
														7	BAB83470	Vp260 like protein	241.12	2.08E-61	30%	44%	32--799	46--811
														8	BAB83471	Vp260 like protein	232.26	9.65E-59	28%	44%	44--866	32--834
														9	NP_048362	Asn/Thr/Ser/Val rich protein	177.56	2.82E-42	25%	39%	15--896	250--1161
														10	NP_048366	Asn/Thr/Ser/Ile rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	169.86	5.88E-40	24%	40%	14--910	124--1117
C167L	75477--74938	180	21,061	9.78	1	pfam01096	TFIIS, Transcription factor S-II (TFIIS)..	66.92	3.34E-12	68%	74%	141--179	1--39	1	NP_048472	contains a zinc ribbon domain; similar to Chlorella virus CVU1 TFIIS-like transcription factor, corresponds to GenBank Accession Number D29631	362.46	3.26E-99	91%	97%	1--180	1--180

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					2	smart00440	ZnF_C2C2, C2C2 Zinc finger; Nucleic-acid-binding motif in transcriptional elongation factor TFIIS and RNA polymerases. .	61.45	1.31E-10	56%	77%	141--180	1--40	2	BAA04187	transcription elongation factor SII	362.46	3.26E-99	91%	97%	1--180	1--180
					3	COG1594	RPB9, DNA-directed RNA polymerase, subunit M/Transcription elongation factor TFIIS [Transcription].	58.12	1.52E-09	34%	53%	101--179	34--111	3	S47662	transcription elongation factor TFIIS homolog - Chlorella virus CV-U1	355.14	5.20E-97	89%	96%	1--180	1--180
														4	BAA04186	transcription elongation factor SII	320.86	1.09E-86	89%	96%	1--164	1--164
														5	AAS54500	AGR011Wp	90.51	2.39E-17	46%	60%	82--179	207--302
														6	CAG98214	unnamed protein product	89.74	4.07E-17	35%	52%	19--179	130--290
														7	CAG86394	unnamed protein product	87.43	2.02E-16	33%	52%	41--179	149--288
														8	CAG58585	unnamed protein product	86.27	4.50E-16	39%	53%	59--179	184--304
														9	AAF71710	transcription elongation factor TFIIS	78.57	9.39E-14	33%	52%	55--179	202--322
														10	EAL34421	GA17632-PA	78.18	1.23E-13	33%	52%	55--179	191--311
C169R	75507--76223	239	26,551	10.3		No Hit Found								1	NP_048475	A127R	426.79	2.67E-118	84%	89%	1--238	1--244
C171L	76630--76226	135	15,967	5.02		No Hit Found								1	NP_048479	A131L	218.39	4.93E-56	76%	86%	1--134	1--135
C174R	77335--78078	248	28,371	10.22		No Hit Found								1	NP_048481	similar to Synechoccystis 3-pyrophosphohydrolase, corresponds to GenBank Accession Number D90911	383.65	2.76E-105	95%	97%	42--239	1--198
														2	AAM64943	unknown	107.84	2.92E-22	31%	55%	42--239	57--264
														3	AAU82110	chloroplast inositol phosphatase-like protein	107.46	3.82E-22	29%	57%	49--239	57--257
														4	NP_565491	THF1	107.46	3.82E-22	31%	55%	42--239	57--264
														5	AAR24582	chloroplast Ptr ToxA-binding protein	106.69	6.51E-22	29%	57%	49--239	57--257
														6	AAQ19850	light-regulated chloroplast-localized protein	105.92	1.11E-21	30%	56%	50--239	66--265
														7	XP_478693	inositol phosphatase-like protein	105.92	1.11E-21	30%	56%	49--239	58--258
														8	ZP_00675630	hypothetical protein TeryDRAFT_0334	93.97	4.37E-18	32%	53%	51--235	6--208
														9	ABA24174	conserved hypothetical protein	83.96	4.52E-15	30%	52%	51--205	6--171
														10	ZP_00514928	conserved hypothetical protein	83.19	7.71E-15	27%	50%	51--238	6--212
C175L	78553--78059	165	19,504	10.22	1	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	45.45	8.83E-06	29%	45%	9--89	3--78	1	NP_048482	similar to bacteriophage T4 intron-associated endonuclease, corresponds to Swiss-Prot Accession Number P13299	273.48	1.56E-72	80%	85%	1--165	1--165
														2	NP_048671	A315L	57.00	2.30E-07	34%	55%	9--95	2--87
														3	YP_293795	putative endonuclease	52.76	4.34E-06	32%	54%	8--92	2--88
														4	NP_048641	PBCV-1 33kd peptide	51.22	1.26E-05	35%	57%	20--95	17--92
														5	NP_048895	A539R	50.83	1.65E-05	47%	71%	11--56	34--79
														6	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	49.29	4.80E-05	34%	52%	9--93	2--87
C177R	78630--79064	145	16,394	11.55		No Hit Found								1	NP_048484	a136R	233.03	1.92E-60	81%	90%	4--145	5--146
C179R	79125--79322	66	8,127	11.16		No Hit Found								1	NP_048485	A137R	78.57	6.12E-14	66%	78%	6--61	13--69
C180L	79614--79306	103	12,005	10.4		No Hit Found								1	NP_048487	A139L	143.67	1.52E-33	72%	84%	19--103	19--103
C182R	79670--82966	1099	118,630	11.07	1	pfam05887	Trypan_PARP, Procyclic acidic repetitive protein (PARP). This family consists of several Trypanosoma brucei procyclic acidic repetitive protein (PARP) like sequences. The procyclic acidic repetitive protein (parp) genes of Trypanosoma brucei encode a small family of abundant surface proteins whose expression is restricted to the procyclic form of the parasite. They are found at two unlinked loci, parpA and parpB; transcription of both loci is developmentally regulated	57.30	2.68E-09	26%	48%	1012--1092	43--123	1	NP_048488	PBCV-1 surface protein	1320.45	0.00E+00	68%	72%	1--1022	1--1027
					2	pfam05616	Neisseria_TspB, Neisseria meningitidis TspB protein. This family consists of several Neisseria meningitidis TspB virulence factor proteins..	48.97	8.32E-07	37%	41%	1033--1087	323--377	2	BAD22850	surface protein	1307.74	0.00E+00	67%	72%	1--1022	1--1027
					3	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	43.98	2.44E-05	40%	42%	1024--1098	59--132	3	BAD12236	surface protein	1283.09	0.00E+00	66%	71%	1--1022	1--1027
					4	pfam04625	DEC-1_N, DEC-1 protein, N terminal region. The defective chorion-1 gene (dec-1) in Drosophila encodes follicle cell proteins necessary for proper eggshell assembly. Multiple products of the dec-1 gene are formed by alternative RNA splicing and proteolytic processing. Cleavage products include S80 (80 kDa) which is incorporated into the eggshell, and further proteolysis of S80 gives S60 (60 kDa)	43.72	2.89E-05	39%	39%	1033--1084	94--145	4	BAE02830	surface protein	632.10	3.28E-179	40%	52%	1--901	1--882
					5	COG3147	DedD, Uncharacterized protein conserved in bacteria [Function unknown].	43.47	3.82E-05	39%	41%	1030--1091	86--147	5	T17636	proline-rich protein A145R - Chlorella virus PBCV-1	286.19	4.41E-75	65%	69%	792--1022	1--236
					6	COG5373	COG5373, Predicted membrane protein [Function unknown].	42.27	8.33E-05	31%	37%	1026--1090	59--124	6	NP_048762	Pro-, Lys-rich, PAPK (30x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	80.11	4.80E-13	58%	74%	961--1022	3--64
					7	pfam05518	Totivirus_coat, Totivirus coat protein..	41.51	1.39E-04	22%	30%	966--1095	619--750	7	NP_048519	similar to PBCV-1 ORF A41R, corresponds to GenBank Accession Number U17055	58.92	1.14E-06	60%	71%	901--945	3--47
C187L	83371--82979	131	14,999	4.67		No Hit Found								1	NP_048498	A150L	172.94	2.34E-42	79%	89%	1--106	1--107
C189R	83484--83849	122	13,562	8.88		No Hit Found								1	NP_048499	A151R	237.27	1.01E-61	95%	96%	3--122	16--135
														2	NP_048496	A148R	61.62	7.60E-09	36%	57%	3--97	8--107
C191R	83927--85303	459	51,942	9.16	1	COG1061	SSL2, DNA or RNA helicases of superfamily II [Transcription / DNA replication, recombination, and repair].	132.14	7.94E-32	24%	39%	85--448	32--432	1	NP_048501	similar to phage T5 helicase, corresponds to Swiss-Prot Accession Number P11107	726.47	0.00E+00	78%	87%	1--459	1--459
					2	smart00487	DEXDc, DEAD-like helicases superfamily; .	60.63	2.58E-10	23%	38%	96--246	13--184	2	YP_142750	VV A18 helicase	192.59	2.33E-47	30%	51%	7--425	9--445
					3	cd00269	DEXHc, DEXH-box helicases. A diverse family of proteins involved in ATP-dependent DNA or RNA unwinding, needed in a variety of cellular processes. The name derives from the sequence of the Walker B motif (motif II). This domain contains the ATP-binding region..	52.39	7.84E-08	22%	39%	112--232	2--143	3	NP_077551	EsV-1-66	191.82	3.98E-47	31%	50%	24--428	17--443
					4	cd00046	DEXDc, DEAD-like helicases superfamily. A diverse family of proteins involved in ATP-dependent RNA or DNA unwinding. This domain contains the ATP-binding region..	52.01	1.02E-07	23%	41%	112--232	2--144	4	ZP_00754006	COG1061: DNA or RNA helicases of superfamily II	153.68	1.20E-35	31%	45%	79--432	413--776
					5	COG4096	HsdR, Type I site-specific restriction-modification system, R (restriction) subunit and related helicases [Defense mechanisms].	44.58	1.60E-05	27%	43%	114--233	189--321	5	YP_437404	DNA or RNA helicase of superfamily II	147.13	1.12E-33	31%	46%	89--432	421--776
					6	pfam00270	DEAD, DEAD/DEAH box helicase. Members of this family include the DEAD and DEAH box helicases. Helicases are involved in unwinding nucleic acids. The DEAD box helicases are involved in various aspects of RNA metabolism, including nuclear transcription, pre mRNA splicing, ribosome biogenesis, nucleocytoplasmic transport, translation, RNA decay and organellar gene expression.	42.33	7.69E-05	18%	35%	96--268	15--206	6	BAB34728	hypothetical protein	145.59	3.27E-33	31%	48%	83--435	417--780
					7	COG4889	COG4889, Predicted helicase [General function prediction only].	41.50	1.58E-04	25%	42%	350--451	523--633	7	NP_287072	putative helicase	145.59	3.27E-33	31%	48%	83--435	425--788
					8	cd00079	HELICc, Helicase superfamily c-terminal domain; associated with DEXDc-, DEAD-, and DEAH-box proteins, yeast initiation factor 4A, Ski2p, and Hepatitis C virus NS3 helicases; this domain is found in a wide variety of helicases and helicase related proteins; may not be an autonomously folding unit, but an integral part of the helicase; 4 helicase superfamilies at present according to the organization of their signature motifs; all helicases share the ability to unwind nucleic acid duplexes with a distinct directional polarity; they utilize the free energy from nucleoside triphosphate hydrolysis to fuel their translocation along DNA, unwinding the duplex in the process	39.52	5.86E-04	17%	36%	304--398	17--124	8	EAM63429	Type III restriction enzyme, res subunit:DEAD/DEAH box helicase, N-terminal	128.64	4.14E-28	31%	48%	82--432	429--790
														9	ZP_00665630	Type III restriction enzyme, res subunit:DEAD/DEAH box helicase, N-terminal	128.64	4.14E-28	28%	44%	70--432	469--849
														10	ZP_00798957	Helicase, C-terminal:Type III restriction enzyme, res subunit:DEAD/DEAH box helicase, N-terminal	127.10	1.20E-27	28%	45%	82--458	437--847
C193L	85643--85311	111	12,206	3.97		No Hit Found								1	NP_048505	A157L	163.70	1.43E-39	78%	92%	1--101	1--101
C194L	85967--85683	95	11,354	3.34		No Hit Found								1	NP_048506	A158L	85.89	3.88E-16	53%	75%	10--89	18--97
C195R	86156--86464	103	11,747	11.19		No Hit Found								1	NP_048509	A161R	67.40	1.39E-10	45%	52%	4--103	14--123
C197L	86992--86657	112	12,780	6.78		No Hit Found								1	NP_048513	A165L	173.33	1.80E-42	74%	85%	1--112	168--279
C198L	87447--87004	148	17,418	9.99		No Hit Found								1	NP_048513	A165L	153.30	1.97E-36	65%	75%	8--120	22--134
C200R	87527--88330	268	31,261	4.58	1	COG5377	COG5377, Phage-related protein, predicted endonuclease [DNA replication, recombination, and repair].	42.72	6.40E-05	21%	38%	34--214	17--190	1	NP_048514	PBCV-1 exonuclease	477.25	2.11E-133	80%	91%	1--268	1--268
														2	AAG28903	F12A21.19	91.66	2.49E-17	31%	50%	19--211	84--276
														3	NP_176934	unknown protein	91.66	2.49E-17	31%	50%	19--211	105--297

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
														4	AAR26918	FirrV-1-B43 precursor	75.10	2.41E-12	30%	49%	14--212	4--196
														5	XP_480865	unknown protein	74.33	4.12E-12	30%	44%	30--209	135--317
														6	XP_472944	OSJNBa0081L15.20	73.56	7.02E-12	33%	47%	30--169	142--296
														7	CAF27185	Exonuclease	72.79	1.20E-11	25%	43%	30--219	3--201
														8	CAF27488	Exonuclease	72.40	1.56E-11	25%	43%	30--212	3--195
														9	NP_077549	EsV-1-64	68.17	2.95E-10	27%	49%	15--202	1--179
														10	YP_142708	Lambda-type exonuclease	68.17	2.95E-10	31%	46%	27--174	178--339
C202R	88372--88872	167	18,380	4.71		No Hit Found								1	NP_048516	A168R	247.67	9.59E-65	73%	78%	2--167	1--166
C203R	88932--90146	405	47,296	6.76		No Hit Found								1	NP_048711	A354R	140.58	8.92E-32	35%	54%	156--395	2--233
														2	NP_048779	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	72.40	2.98E-11	26%	45%	68--367	25--319
														3	AAL73456	endonuclease	52.76	2.45E-05	28%	46%	2--170	21--191
C204R	90172--91134	321	36,544	6.3	1	COG0540	PyrB, Aspartate carbamoyltransferase, catalytic chain [Nucleotide transport and metabolism].	305.61	4.83E-84	43%	60%	24--318	8--310	1	NP_048517	PBCV-1 aspartate transcarbamylase	552.75	5.33E-156	84%	92%	1--321	1--321
					2	COG0078	ArgF, Ornithine carbamoyltransferase [Amino acid transport and metabolism].	161.88	7.35E-41	31%	49%	24--320	7--310	2	CAC85728	aspartate carbamoyltransferase	285.03	2.07E-75	49%	69%	26--315	72--370
					3	pfam02729	OTCace_N, Aspartate/ornithine carbamoyltransferase, carbamoyl-P binding domain..	145.37	7.73E-36	46%	62%	24--163	1--143	3	Q43064	Aspartate carbamoyltransferase 3, chloroplast precursor (Aspartate transcarbamylase 3) (ATCase 3)	282.72	1.03E-74	48%	66%	13--315	75--384
					4	pfam00185	OTCace, Aspartate/ornithine carbamoyltransferase, Asp/Orn binding domain..	114.21	1.82E-26	33%	52%	166--315	1--156	4	Q43087	Aspartate carbamoyltransferase 2, chloroplast precursor (Aspartate transcarbamylase 2) (ATCase 2)	280.41	5.10E-74	48%	66%	13--315	69--378
														5	CAA50687	aspartate carbamoyltransferase	279.64	8.71E-74	50%	68%	23--315	84--383
														6	NP_188668	amino acid binding / aspartate carbamoyltransferase/ carboxyl- and carbamoyltransferase/ ornithine carbamoyltransferase	278.49	1.94E-73	49%	68%	23--315	84--383
														7	Q43086	Aspartate carbamoyltransferase 1, chloroplast precursor (Aspartate transcarbamylase 1) (ATCase 1)	275.02	2.14E-72	48%	69%	26--315	83--379
														8	AAL90999	AT3g20330/MQC12_8	274.25	3.66E-72	49%	67%	23--315	84--383
														9	XP_480986	aspartate carbamoyltransferase	265.77	1.30E-69	46%	65%	26--315	55--356
														10	CAA52201	aspartate carbamoyltransferase	264.62	2.90E-69	52%	70%	57--315	1--261
C206R	91147--92250	368	40,447	10.2	1	COG3509	LpqC, Poly(3-hydroxybutyrate) depolymerase [Secondary metabolites biosynthesis, transport, and catabolism].	36.11	6.13E-03	21%	37%	71--204	46--178	1	NP_048519	similar to PBCV-1 ORF A41R, corresponds to GenBank Accession Number U17055	607.83	1.72E-172	77%	84%	4--368	1--387
														2	NP_048389	contains Pro-rich Px motif, PAPK (8X); similar to Thermoproteus virus protein TPX, corresponds to Swiss-Prot Accession Number P19275	508.06	1.86E-142	80%	87%	56--368	101--412
														3	NP_048488	PBCV-1 surface protein	65.86	2.44E-09	60%	72%	6--55	905--954
														4	BAD12236	surface protein	65.86	2.44E-09	60%	72%	6--55	905--954
														5	BAD22850	surface protein	65.86	2.44E-09	60%	72%	6--55	905--954
														6	T17636	proline-rich protein A145R - Chlorella virus PBCV-1	65.86	2.44E-09	60%	72%	6--55	114--163
														7	AAM63817	unknown	59.31	2.28E-07	25%	45%	79--288	79--289
														8	NP_191439	unknown protein	59.31	2.28E-07	25%	45%	79--288	79--289
														9	BAD86968	hypothetical protein	57.38	8.67E-07	23%	39%	62--314	95--340
														10	NP_916095	P0481E12.18	57.38	8.67E-07	23%	39%	62--314	72--317
C208L	93093--92257	279	30,651	7.75	1	pfam01734	Patatin, Patatin-like phospholipase. This family consists of various patatin glycoproteins from plants. The patatin protein accounts for up to 40% of the total soluble protein in potato tubers. Patatin is a storage protein but it also has the enzymatic activity of lipid acyl hydrolase, catalysing the cleavage of fatty acids from membrane lipids. Members of this family have been found also in vertebrates.	123.14	3.95E-29	33%	49%	19--192	1--179	1	NP_048521	similar to E. coli hypothetical protein, corresponds to Swiss-Prot Accession Number P39407	513.07	3.74E-144	90%	96%	4--279	13--288
					2	COG1752	RssA, Predicted esterase of the alpha-beta hydrolase superfamily [General function prediction only].	87.45	2.42E-18	28%	48%	18--191	13--185	2	ZP_00240206	Patatin-like phospholipase family	92.82	1.21E-17	30%	49%	19--193	8--195
					3	COG4667	COG4667, Predicted esterase of the alpha-beta hydrolase superfamily [General function prediction only].	56.04	5.76E-09	25%	41%	18--213	13--202	3	AAS94389	phospholipase, patatin family	90.51	5.99E-17	30%	50%	19--193	5--198
														4	ABB37620	esterase of the alpha-beta hydrolase superfamily-like	89.35	1.33E-16	31%	48%	19--193	5--198
														5	XP_800775	PREDICTED: hypothetical protein XP_795682, partial	84.34	4.29E-15	32%	48%	13--192	71--268
														6	NP_149926	463L	83.19	9.56E-15	28%	51%	2--199	13--214
														7	AAQ66865	conserved hypothetical protein	81.65	2.78E-14	30%	49%	19--193	8--199
														8	CAG23338	hypothetical protein	79.72	1.06E-13	26%	47%	19--238	8--236
														9	XP_789091	PREDICTED: hypothetical protein XP_783998	74.33	4.44E-12	29%	46%	13--192	81--278
														10	YP_142800	patatin-like phospholipase (463L)	73.56	7.57E-12	24%	44%	5--276	48--326
C212L	94287--93142	382	43,488	6.26	1	COG4123	COG4123, Predicted O-methyltransferase [General function prediction only].	59.87	4.11E-10	28%	40%	48--162	45--171	1	AAC03125	DNA adenine methyltransferase	769.62	0.00E+00	98%	99%	1--382	1--382
					2	COG0286	HsdM, Type I restriction-modification system methyltransferase subunit [Defense mechanisms].	55.04	1.18E-08	21%	35%	3--220	145--386	2	P52284	Modification methylase CviRI (Adenine-specific methyltransferase CviRI) (M.CviRI)	620.54	2.70E-176	79%	88%	1--381	1--378
					3	COG2813	RsmC, 16S RNA G1207 methylase RsmC [Translation, ribosomal structure and biogenesis].	49.49	6.04E-07	27%	43%	40--118	151--234	3	CAA29835	unnamed protein product	218.78	2.36E-55	33%	54%	10--381	12--377
					4	COG2263	COG2263, Predicted RNA methylase [Translation, ribosomal structure and biogenesis].	46.38	4.98E-06	29%	48%	28--124	26--124	4	AAC03124	DNA adenine methyltransferase	210.31	8.40E-53	34%	53%	10--380	9--368
					5	COG2890	HemK, Methylase of polypeptide chain release factors [Translation, ribosomal structure and biogenesis].	44.98	1.42E-05	23%	41%	33--129	93--203	5	AAC57945	DNA adenine methyltransferase	203.76	7.86E-51	32%	54%	8--380	6--357
					6	COG0421	SpeE, Spermidine synthase [Amino acid transport and metabolism].	43.03	4.93E-05	35%	47%	47--113	76--154	6	AAC57943	DNA adenine methyltransferase	200.29	8.69E-50	32%	55%	4--381	3--368
					7	pfam01170	UPF0020, Putative RNA methylase family UPF0020. This domain is probably a methylase. It is associated with the THUMP domain that also occurs with RNA modification domains..	40.25	3.26E-04	20%	46%	25--118	5--107	7	YP_063409	cpp14	67.40	8.80E-10	30%	47%	14--169	273--429
					8	COG0116	COG0116, Predicted N6-adenine-specific DNA methylase [DNA replication, recombination, and repair].	39.13	8.42E-04	29%	43%	29--118	218--309	8	AAW34165	unknown	67.40	8.80E-10	30%	47%	14--169	273--429
					9	COG4106	Tam, Trans-aconitate methyltransferase [General function prediction only].	35.66	8.08E-03	25%	41%	38--162	21--130	9	AAR29548	cpp14	67.40	8.80E-10	30%	47%	14--169	273--429
														10	ZP_00371038	helicase, Snf2 family	64.31	7.45E-09	29%	46%	14--169	195--351
C216L	94654--94346	103	12,059	9.39		No Hit Found								1	NP_049043	A687R	113.62	1.69E-24	71%	86%	26--101	1--75
C217L	95632--94661	324	37,294	7.7	1	COG3392	COG3392, Adenine-specific DNA methylase [DNA replication, recombination, and repair].	200.98	1.36E-52	38%	58%	4--303	1--311	1	AAC03127	DNA adenine methyltransferase	606.68	3.16E-172	93%	93%	1--324	1--324
					2	pfam02086	MethyltransfD12, D12 class N6 adenine-specific DNA methyltransferase..	150.10	3.03E-37	33%	47%	6--284	1--253	2	NP_048600	M.CviAII adenine DNA methyltransferase	494.58	1.75E-138	76%	84%	1--321	1--324
														3	S27901	site-specific DNA-methyltransferase (adenine-specific) (EC 2.1.1.72) CviAII - Chlorella virus PBCV-1	492.66	6.65E-138	76%	84%	1--321	1--324
														4	AAC57944	DNA adenine methyltransferase	491.89	1.13E-137	75%	84%	1--321	1--324
														5	YP_392646	Site-specific DNA-methyltransferase (adenine-specific)	177.56	4.73E-43	34%	52%	4--320	1--332
														6	ZP_00371322	ulcer associated adenine specific DNA methyltransferase	177.56	4.73E-43	36%	55%	4--305	1--316
														7	CAA38356	NlaIII methyltransferase	175.64	1.80E-42	35%	51%	4--316	1--327
														8	AAP77647	conserved hypothetical protein	171.79	2.60E-41	37%	49%	4--305	1--319
														9	AAC45814	methylase HpyI	171.40	3.39E-41	37%	54%	4--305	1--311
														10	AAG49536	CATG-specific methyltransferase	169.47	1.29E-40	36%	53%	4--305	25--335
C220R	95750--98170	807	88,535	8.39	1	smart00637	CBD_II, CBD_II domain..	71.13	1.60E-13	27%	38%	11--105	3--101	1	BAA78554	vChti-1	1313.13	0.00E+00	78%	84%	1--806	1--835
					2	pfam00704	Glyco_hydro_18, Glycosyl hydrolases family 18..	62.84	5.19E-11	25%	37%	540--716	4--195	2	NP_048529	PBCV-1 chitinase	1309.66	0.00E+00	79%	85%	1--806	1--829
														3	BAC72964	putative sugar hydrolase	216.47	3.02E-54	32%	46%	3--408	35--483
														4	ZP_00570566	Cellulose-binding, bacterial type	214.16	1.50E-53	37%	53%	79--410	197--535
														5	CAC10108	putative secreted sugar hydrolase	212.62	4.36E-53	31%	45%	2--414	33--489
														6	CAH35762	putative exported chitinase	177.56	1.56E-42	36%	50%	124--410	35--339
														7	ZP_00479579	COG3979: Uncharacterized protein contain chitin-binding domain type 3	177.18	2.03E-42	37%	51%	124--410	31--335
														8	ZP_00501511	COG3979: Uncharacterized protein contain chitin-binding domain type 3	177.18	2.03E-42	37%	51%	124--410	31--335
														9	ZP_00488193	COG3979: Uncharacterized protein contain chitin-binding domain type 3	177.18	2.03E-42	37%	51%	124--410	31--335
														10	YP_442923	glycosyl hydrolase, family 18	176.79	2.65E-42	36%	50%	124--410	82--386
C227L	98997--98389	203	23,838	4.34		No Hit Found									No Hit Found	No Hit Found						
C228L	99329--99105	75	9,483	11.36		No Hit Found									No Hit Found	No Hit Found						

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
C230-34R	99389--102216	943	104,868	6.82	1	pfam00136	DNA_pol_B, DNA polymerase family B. This region of DNA polymerase B appears to consist of more than one structural domain, possibly including elongation, DNA-binding and dNTP binding activities..	350.06	1.82E-97	38%	56%	432--863	1--442	1	P30320	DNA polymerase	1771.52	0.00E+00	95%	96%	1--913	1--913
					2	COG0417	PolB, DNA polymerase elongation subunit (family B) [DNA replication, recombination, and repair].	340.87	1.08E-94	30%	49%	27--863	11--775	2	NP_048532	PBVC-1 DNA polymerase	1633.62	0.00E+00	86%	92%	1--913	1--913
					3	smart00486	POLBc, DNA polymerase type-B family; DNA polymerase alpha, delta, epsilon and zeta chain (eukaryota), DNA polymerases in archaea, DNA polymerase II in e. coli, mitochondrial DNA polymerases and and virus DNA polymerases .	309.07	4.59E-85	33%	52%	181--649	1--475	3	BAA35142	DNA polymerase	1629.38	0.00E+00	86%	92%	1--913	1--913
					4	cd00145	POLBc, DNA polymerase type-B family; DNA directed DNA polymerase. Posseses DNA binding, polymerase and 3'-5' exonuclease activity..	291.44	7.56E-80	34%	50%	181--685	1--511	4	AAK28923	DNA polymerase	442.58	2.98E-122	99%	100%	473--692	1--220
					5	pfam03104	DNA_pol_B_exo, DNA polymerase family B, exonuclease domain. This domain has 3' to 5' exonuclease activity and adopts a ribonuclease H type fold..	222.67	4.44E-59	26%	42%	32--359	1--334	5	AAK28956	DNA polymerase	440.27	1.48E-121	97%	100%	473--692	1--220
														6	AAK28951	DNA polymerase	439.50	2.53E-121	98%	99%	473--692	1--220
														7	AAK28952	DNA polymerase	437.57	9.60E-121	97%	99%	473--692	1--220
														8	AAK28963	DNA polymerase	422.17	4.17E-116	92%	97%	473--692	1--220
														9	AAK28965	DNA polymerase	421.78	5.45E-116	92%	97%	473--692	1--220
														10	AAK28930	DNA polymerase	420.62	1.21E-115	91%	97%	473--692	1--220
C236R	102276--103385	370	42,862	5.23	1	pfam01541	GIY-YIG, GIY-YIG catalytic domain. This domain called GIY-YIG is found in the amino terminal region of excinuclease abc subunit c (uvrC), bacteriophage T4 endonucleases segA, segB, segC, segD and segE; it is also found in putative endonucleases encoded by group I introns of fungi and phage. The structure of I-TevI a GIY-YIG endonuclease, reveals a novel alpha/beta-fold with a central three-stranded antiparallel beta-sheet flanked by three helices. The most conserved and putative catalytic residues are located on a shallow, concave surface and include a metal coordination site	39.76	4.44E-04	30%	44%	31--105	13--83	1	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	51.60	4.80E-05	28%	42%	187--339	52--218
C239R	103403--107509	1369	152,191	11.6	1	COG1196	Smc, Chromosome segregation ATPases [Cell division and chromosome partitioning].	50.80	2.14E-07	13%	40%	1087--1352	169--439	1	NP_048536	similar to SWI/SNF chromatin remodeling complex subunit OSA2	1108.21	0.00E+00	61%	66%	2--994	12--986
					2	COG3064	TolA, Membrane protein involved in colicin uptake [Cell envelope biogenesis, outer membrane].	47.86	1.83E-06	28%	43%	1079--1254	127--288	2	T17682	hypothetical protein A192R - Chlorella virus PBCV-1	261.92	1.13E-67	48%	56%	1028--1368	8--360
					3	COG5022	COG5022, Myosin heavy chain [Cytoskeleton].	44.97	1.22E-05	16%	41%	1091--1366	761--1024	3	NP_704635	hypothetical protein	73.56	5.69E-11	22%	36%	143--451	129--483
					4	COG0419	SbcC, ATPase involved in DNA repair [DNA replication, recombination, and repair].	43.12	4.53E-05	14%	45%	1119--1353	481--704	4	XP_731097	hypothetical protein PY03130	58.92	1.45E-06	23%	36%	162--398	85--339
					5	COG4942	COG4942, Membrane-bound metallopeptidase [Cell division and chromosome partitioning].	40.33	3.21E-04	20%	39%	1099--1256	143--299	5	T17681	hypothetical protein a191R - Chlorella virus PBCV-1	57.77	3.23E-06	56%	70%	811--869	3--60
C241L	108302--107517	262	29,685	4.82	1	pfam00705	PCNA_N, Proliferating cell nuclear antigen, N-terminal domain. N-terminal and C-terminal domains of PCNA are topologically identical. Three PCNA molecules are tightly associated to form a closed ring encircling duplex DNA..	83.01	4.18E-17	28%	62%	7--129	1--122	1	NP_048540	similar to human PCNA, corresponds to Swiss-Prot Accession Number P12004	514.61	1.14E-144	97%	99%	1--262	1--262
					2	pfam02747	PCNA_C, Proliferating cell nuclear antigen, C-terminal domain. N-terminal and C-terminal domains of PCNA are topologically identical. Three PCNA molecules are tightly associated to form a closed ring encircling duplex DNA..	77.67	1.83E-15	31%	54%	137--259	4--127	2	XP_395519	PREDICTED: similar to ENSANGP00000012272	150.21	5.65E-35	29%	54%	7--259	533--785
					3	COG0592	DnaN, DNA polymerase sliding clamp subunit (PCNA homolog) [DNA replication, recombination, and repair].	61.08	2.07E-10	21%	41%	19--261	72--323	3	XP_514499	PREDICTED: proliferating cell nuclear antigen	149.06	1.26E-34	29%	57%	7--259	1--253
														4	AAX43349	proliferating cell nuclear antigen	149.06	1.26E-34	29%	57%	7--259	1--253
														5	AAX36355	proliferating cell nuclear antigen	149.06	1.26E-34	29%	57%	7--259	1--253
														6	XP_534355	PREDICTED: similar to proliferating cell nuclear antigen	148.67	1.65E-34	29%	56%	3--259	202--458
														7	AAB27811	PCNA	148.29	2.15E-34	30%	55%	7--259	1--253
														8	1AXC_E	Chain E, Human Pcna	148.29	2.15E-34	30%	56%	8--259	2--253
														9	BAE47145	proliferating cell nuclear antigen	147.52	3.66E-34	30%	56%	7--259	1--252
														10	CAG46598	PCNA	147.13	4.79E-34	29%	57%	7--259	1--253
C242L	108785--108330	152	17,364	7.89		No Hit Found								1	NP_048543	A196L	285.80	2.49E-76	86%	95%	1--152	1--152
C244R	108836--109135	100	11,047	10.62		No Hit Found								1	NP_048546	A199R	154.07	1.14E-36	75%	91%	1--100	1--101
C245L	109711--109142	190	22,074	8.36	1	pfam01753	zf-MYND, MYND finger..	40.80	2.47E-04	50%	65%	116--150	4--38	1	NP_849969	SDG37	56.61	4.38E-07	46%	63%	114--157	57--105
														2	AAD03568	putative SET-domain transcriptional regulator	56.61	4.38E-07	46%	63%	114--157	57--105
														3	XP_598182	PREDICTED: similar to SET and MYND domain-containing protein 3 (Zinc finger MYND domain-containing protein 1), partial	54.68	1.67E-06	39%	56%	98--150	14--66
														4	CAG07000	unnamed protein product	54.30	2.18E-06	36%	56%	73--150	407--482
														5	CAE59608	Hypothetical protein CBG03016	53.53	3.71E-06	44%	55%	95--150	4--64
														6	CAA86783	Hypothetical protein R06F6.4	53.14	4.85E-06	40%	57%	95--150	4--64
														7	CAJ02073	MYND finger domain-like protein	52.37	8.27E-06	58%	70%	119--152	133--166
														8	AAL75997	putative SET-domain transcriptional regulator	51.99	1.08E-05	34%	56%	84--150	38--97
														9	CAD43192	egg laying nine 1 protein	51.60	1.41E-05	44%	61%	112--158	13--59
														10	AAM97151	unknown protein,3'-partial	51.22	1.84E-05	30%	53%	73--150	30--97
C246R	109841--110197	119	13,381	9.85	1	pfam00383	dCMP_cyt_deam, Cytidine and deoxycytidylate deaminase zinc-binding region..	48.44	1.09E-06	28%	47%	4--108	3--103	1	NP_048547	contains cytidine and deoxycytidine deaminase Zn-binding region signature	191.05	8.40E-48	75%	87%	1--118	1--118
					2	COG0590	CumB, Cytosine/adenosine deaminases [Nucleotide transport and metabolism / Translation, ribosomal structure and biogenesis].	48.02	1.59E-06	27%	49%	1--108	5--107	2	AAR26853	FirrV-1-A29	50.45	1.77E-05	28%	47%	22--111	24--105
C250L	110901--110563	113	12,262	5.6		No Hit Found								1	NP_048549	A202L	200.29	1.37E-50	84%	88%	1--113	1--113
C252R	110964--111611	216	23,910	6.02		No Hit Found								1	NP_048550	A203R	348.59	7.48E-95	87%	89%	19--216	19--216
C255R	111623--112234	204	22,312	12		No Hit Found								1	NP_048552	A205R	215.31	8.75E-55	60%	67%	17--204	10--206
C256R	112300--113415	372	42,192	4.59	1	pfam02784	Orn_Arg_deC_N, Pyridoxal-dependent decarboxylase, pyridoxal binding domain. These pyridoxal-dependent decarboxylases acting on ornithine, lysine, arginine and related substrates This domain has a TIM barrel fold..	201.64	8.84E-53	39%	57%	25--260	4--246	1	NP_048554	PBCV-1 arginine decarboxylase	661.76	0.00E+00	83%	92%	1--372	1--372
					2	COG0019	LysA, Diaminopimelate decarboxylase [Amino acid transport and metabolism].	199.35	4.62E-52	27%	46%	8--372	18--394	2	CAE02644	ornithine decarboxylase	270.78	5.03E-71	39%	59%	1--364	43--407
					3	pfam00278	Orn_DAP_Arg_deC, Pyridoxal-dependent decarboxylase, C-terminal sheet domain. These pyridoxal-dependent decarboxylases act on ornithine, lysine, arginine and related substrates. .	89.49	5.65E-19	33%	47%	263--365	1--103	3	EAA00421	ENSANGP00000020224	268.47	2.50E-70	38%	58%	19--370	34--389
					4	COG1166	SpeA, Arginine decarboxylase (spermidine biosynthesis) [Amino acid transport and metabolism].	50.25	3.37E-07	28%	46%	148--270	238--374	4	AAV88093	ornithine decarboxylase 1	268.47	2.50E-70	38%	57%	18--372	39--408
														5	AAX36104	ornithine decarboxylase 1	268.47	2.50E-70	38%	57%	18--372	39--408
														6	XP_856678	PREDICTED: similar to Ornithine decarboxylase (ODC) isoform 8	267.70	4.26E-70	38%	58%	18--372	39--395
														7	XP_856512	PREDICTED: similar to Ornithine decarboxylase (ODC) isoform 4	267.31	5.56E-70	38%	57%	18--372	39--404
														8	NP_776555	ornithine decarboxylase 1	266.54	9.49E-70	38%	56%	18--372	39--408
														9	AAD02222	ornithine decarboxylase	266.16	1.24E-69	38%	57%	19--372	60--428
														10	P07805	Ornithine decarboxylase (ODC)	266.16	1.24E-69	38%	57%	19--372	38--406
C259R	113520--114221	234	27,145	4.65	1	pfam03154	Atrophin-1, Atrophin-1 family. Atrophin-1 is the protein product of the dentatorubral-pallidoluysian atrophy (DRPLA) gene. DRPLA OMIM:125370 is a progressive neurodegenerative disorder. It is caused by the expansion of a CAG repeat in the DRPLA gene on c	42.21	9.20E-05	47%	62%	7--101	645--711	1	NP_048555	KAEKA (6X), SDDD (7X)	73.56	5.48E-12	33%	40%	1--152	30--177
C260L	114667--114224	148	16,630	4.69		No Hit Found								1	NP_048560	A213L	256.91	1.26E-67	83%	95%	1--146	1--146
C261L	115095--114697	133	15,240	7.93		No Hit Found								1	NP_048561	A214L	179.49	2.56E-44	81%	90%	1--110	1--110
C264L	117332--116184	383	43,996	10.19	1	COG1231	COG1231, Monoamine oxidase [Amino acid transport and metabolism].	50.37	2.79E-07	17%	32%	6--383	9--441	1	NP_048564	similar to bovine monoamine oxidase, corresponds to Swiss-Prot Accession Number P21398	692.19	0.00E+00	89%	92%	1--383	12--394
					2	COG1233	COG1233, Phytoene dehydrogenase and related proteins [Secondary metabolites biosynthesis, transport, and catabolism].	46.62	4.49E-06	33%	46%	6--74	5--74	2	BAA83788	orf1	691.42	0.00E+00	89%	92%	1--383	15--397
C263L	116179--115262	306	34,117	9.73		No Hit Found								1	NP_048562	PBCV-1 alginate lyase	403.29	4.86E-111	62%	71%	7--304	1--320
														2	BAB19127	vAL-1	399.05	9.17E-110	57%	66%	7--304	1--348

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
														3	BAA83789	alginate lyase	398.67	1.20E-109	59%	69%	7--304	1--332
														4	BAE48156	hypothetical methionyl-tRNA synthetase	89.74	1.19E-16	31%	44%	86--299	24--247
														5	NP_048917	similar to Chlorella virus CVK2 DNA binding protein, corresponds to GenBank Accession Number D78305	61.62	3.47E-08	25%	40%	107--305	431--648
														6	BAA11342	DNA binding protein	61.23	4.53E-08	26%	40%	107--284	430--622
														7	NP_048519	similar to PBCV-1 ORF A41R, corresponds to GenBank Accession Number U17055	51.22	4.69E-05	42%	68%	8--57	3--52
C265R	117460--119361	634	72,852	6.35	1	COG1215	COG1215, Glycosyltransferases, probably involved in cell wall biogenesis [Cell envelope biogenesis, outer membrane].	107.71	1.73E-24	17%	34%	68--489	1--423	1	NP_048569	similar to cellulose synthase catalytic subunit (UDP-forming)	1179.08	0.00E+00	90%	92%	1--632	44--675
					2	pfam00535	Glycos_transf_2, Glycosyl transferase. Diverse family, transferring sugar from UDP-glucose, UDP-N-acetyl- galactosamine, GDP-mannose or CDP-abequose, to a range of substrates including cellulose, dolichol phosphate and teichoic acids..	36.74	3.58E-03	20%	35%	134--302	10--168	2	T17709	hypothetical protein A219R - Chlorella virus PBCV-1	303.52	1.41E-80	82%	85%	1--184	44--227
														3	CAC48842	putative cellulose synthase protein	291.20	7.26E-77	38%	53%	122--559	101--531
														4	ZP_00913218	similar to cellulose synthase	286.96	1.37E-75	39%	56%	122--527	167--566
														5	ZP_00919442	similar to cellulose synthase	282.72	2.58E-74	36%	54%	122--589	111--568
														6	ABA79331	cellulose synthase-like protein	281.57	5.75E-74	35%	54%	122--589	111--568
														7	AAL44127	cellulose synthase	277.72	8.30E-73	38%	54%	122--525	103--500
														8	BAB54246	mlr7873	276.17	2.42E-72	40%	55%	122--510	113--496
														9	ZP_00958843	cellulose synthase-like protein	275.40	4.12E-72	36%	53%	122--581	103--555
														10	AAQ87082	Cellulose synthase catalytic subunit	274.63	7.03E-72	36%	52%	122--595	103--570
C270L	119784--119371	138	16,053	10.17		No Hit Found								1	NP_048575	A227L	219.55	2.19E-56	85%	95%	18--138	17--137
C271L	120028--119804	75	8,356	7.94		No Hit Found								1	NP_048577	A229L	111.31	8.59E-24	67%	81%	1--75	1--77
C272R	120002--120634	211	24,052	9.99		No Hit Found								1	NP_048578	A230R	305.06	8.97E-82	78%	91%	19--210	3--195
C275L	121760--120642	373	42,982	10.01		No Hit Found								1	NP_048579	contains ATP/GTP-binding motif A	611.68	1.21E-173	78%	86%	5--373	9--383
														2	NP_048983	similar to Chlorella virus PBCV-1 ORF A231L, corresponds to GenBank Accession Number U42580	309.69	9.81E-83	44%	64%	6--329	103--427
C280R	121869--122201	111	13,444	10.24		No Hit Found								1	NP_048581	A233R	174.10	1.06E-42	74%	83%	6--111	7--112
C281R	122282--123385	368	42,617	5.14	1	COG0751	GlyS, Glycyl-tRNA synthetase, beta subunit [Translation, ribosomal structure and biogenesis].	38.64	9.34E-04	21%	41%	40--166	25--148		No Hit Found	No Hit Found						
C285L	123792--123469	108	12,669	10.77		No Hit Found								1	NP_048582	A234L	154.84	6.72E-37	65%	85%	1--107	1--107
C286R	123855--125375	507	57,077	9.68	1	pfam06408	Homo_sperm_syn, Homospermidine synthase. This family consists of several homospermidine synthase proteins (EC:2.5.1.44). Homospermidine synthase (HSS) catalyses the synthesis of the polyamine homospermidine from 2 mol putrescine in an NAD(+)-dependent reaction.	635.84	0.00E+00	42%	57%	31--502	1--471	1	NP_048585	PBCV-1 homospermidine synthase	908.67	0.00E+00	89%	92%	22--507	33--518
					2	COG1748	LYS9, Saccharopine dehydrogenase and related proteins [Amino acid transport and metabolism].	49.90	4.63E-07	22%	42%	32--463	2--369	2	ZP_00589757	Homospermidine synthase	312.00	3.00E-83	38%	56%	31--479	5--452
					3	pfam03435	Saccharop_dh, Saccharopine dehydrogenase. This family comprised of three structural domains that can not be separated in the linear sequence. In some organisms this enzyme is found as a bifunctional polypeptide with lysine ketoglutarate reductase. The saccharopine dehydrogenase can also function as a saccharopine reductase.	48.71	1.09E-06	26%	50%	34--190	1--138	3	AAM05046	homospermidine synthase	285.80	2.31E-75	35%	54%	33--479	15--459
					4	cd01483	E1_enzyme_family, Superfamily of activating enzymes (E1) of the ubiquitin-like proteins. This family includes classical ubiquitin-activating enzymes E1, ubiquitin-like (ubl) activating enzymes and other mechanistic homologes, like MoeB, Thif1 and others. The common reaction mechanism catalyzed by MoeB, ThiF and the E1 enzymes begins with a nucleophilic attack of the C-terminal carboxylate of MoaD, ThiS and ubiquitin, respectively, on the alpha-phosphate of an ATP molecule bound at the active site of the activating enzymes, leading to the formation of a high-energy acyladenylate intermediate and subsequently to the formation of a thiocarboxylate at the C termini of MoaD and ThiS.	36.57	4.52E-03	18%	35%	33--134	1--121	4	AAM29862	homospermidine synthase	281.95	3.33E-74	34%	54%	33--479	8--452
					5	cd00757	ThiF_MoeB_HesA_family, ThiF_MoeB_HesA. Family of E1-like enzymes involved in molybdopterin and thiamine biosynthesis family. The common reaction mechanism catalyzed by MoeB and ThiF, like other E1 enzymes, begins with a nucleophilic attack of the C-terminal carboxylate of MoaD and ThiS, respectively, on the alpha-phosphate of an ATP molecule bound at the active site of the activating enzymes, leading to the formation of a high-energy acyladenylate intermediate and subsequently to the formation of a thiocarboxylate at the C termini of MoaD and ThiS. MoeB, as the MPT synthase (MoaE/MoaD complex) sulfurase, is involved in the biosynthesis of the molybdenum cofactor, a derivative of the tricyclic pterin, molybdopterin (MPT). ThiF catalyzes the adenylation of ThiS, as part of the biosynthesis pathway of thiamin pyrophosphate (vitamin B1).	36.62	4.63E-03	19%	41%	29--140	19--149	5	ZP_00543002	Homospermidine synthase	274.63	5.31E-72	34%	53%	33--479	8--452
														6	ZP_00584810	Homospermidine synthase	242.66	2.24E-62	33%	52%	33--479	13--446
														7	AAP94695	putative homospermidine synthase	232.65	2.32E-59	33%	51%	23--483	1--449
														8	AAU28555	homospermidine synthase	232.26	3.02E-59	33%	50%	37--494	19--460
														9	CAH16656	hypothetical protein	231.49	5.16E-59	33%	50%	37--494	19--460
														10	BAE51070	Homospermidine synthase	230.72	8.80E-59	32%	50%	33--479	30--462
C290L	125816--125382	145	16,932	6.95		No Hit Found								1	NP_048587	A239L	126.33	2.53E-28	54%	75%	37--144	1--109
C292R	125974--128148	725	82,788	6.59	1	COG4581	COG4581, Superfamily II RNA helicase [DNA replication, recombination, and repair].	354.66	8.30E-99	37%	57%	10--429	119--586	1	NP_048589	contains ATP-GTP binding motif; similar to Saccharomyces cervisiae antiviral protein SKI2, corresponds to Swiss-Prot Accession Number P35207	1380.54	0.00E+00	92%	96%	1--725	1--725
					2	COG1204	COG1204, Superfamily II helicase [General function prediction only].	177.89	1.43E-45	30%	50%	25--410	47--436	2	XP_667944	ATP-dependent RNA helicase; ATP-dependent RNA helicase	317.78	8.51E-85	35%	54%	11--518	100--640
					3	COG1202	COG1202, Superfamily II helicase, archaea-specific [General function prediction only].	128.54	1.02E-30	28%	51%	29--398	236--577	3	XP_627152	Mtr4p like SKI family SFII helicase	317.78	8.51E-85	35%	54%	11--518	100--640
					4	pfam00270	DEAD, DEAD/DEAH box helicase. Members of this family include the DEAD and DEAH box helicases. Helicases are involved in unwinding nucleic acids. The DEAD box helicases are involved in various aspects of RNA metabolism, including nuclear transcription, pre mRNA splicing, ribosome biogenesis, nucleocytoplasmic transport, translation, RNA decay and organellar gene expression.	110.51	2.55E-25	24%	40%	9--192	9--206	4	NP_703683	ATP dependent RNA helicase, putative	317.39	1.11E-84	32%	52%	6--587	250--831
					5	smart00487	DEXDc, DEAD-like helicases superfamily; .	96.06	5.35E-21	24%	40%	3--184	1--195	5	CAD25317	ATP-DEPENDENT RNA HELICASE (SKI2 FAMILY)	316.24	2.47E-84	28%	47%	12--725	66--881
					6	COG1201	Lhr, Lhr-like helicases [General function prediction only].	80.68	2.33E-16	24%	46%	15--424	27--411	6	XP_624031	PREDICTED: similar to ENSANGP00000020973	313.54	1.60E-83	39%	60%	8--432	120--578
					7	cd00046	DEXDc, DEAD-like helicases superfamily. A diverse family of proteins involved in ATP-dependent RNA or DNA unwinding. This domain contains the ATP-binding region..	71.27	1.46E-13	20%	41%	26--156	1--143	7	XP_680351	ATP dependent RNA helicase	313.15	2.10E-83	38%	59%	6--429	208--667
					8	COG1205	COG1205, Distinct helicase family with a unique C-terminal domain including a metal-binding cysteine cluster [General function prediction only].	71.20	1.71E-13	21%	41%	10--389	70--429	8	XP_729688	Homo sapiens KIAA0052 protein	312.77	2.74E-83	37%	58%	6--429	208--667
					9	cd00269	DEXHc, DEXH-box helicases. A diverse family of proteins involved in ATP-dependent DNA or RNA unwinding, needed in a variety of cellular processes. The name derives from the sequence of the Walker B motif (motif II). This domain contains the ATP-binding region..	64.33	2.16E-11	23%	43%	26--156	1--142	9	XP_662016	hypothetical protein AN4412.2	310.07	1.77E-82	38%	60%	11--416	153--599
					10	COG0513	SrmB, Superfamily II DNA and RNA helicases [DNA replication, recombination, and repair / Transcription / Translation, ribosomal structure and biogenesis].	57.86	1.86E-09	20%	41%	15--417	56--414	10	XP_688106	PREDICTED: similar to superkiller viralicidic activity 2-like	309.69	2.32E-82	36%	56%	11--460	146--628
C298R	128175--129083	303	32,359	10.85	1	pfam00967	Barwin, Barwin family..	156.34	4.00E-39	43%	54%	175--296	1--119	1	NP_048594	Pro-rich, PAPK (20X); similar to Arabidopsis anter-specific Pro-rich protein, corresponds to Swiss-Prot Accession Number P40602	332.03	1.35E-89	97%	98%	145--303	129--288
					2	pfam05616	Neisseria_TspB, Neisseria meningitidis TspB protein. This family consists of several Neisseria meningitidis TspB virulence factor proteins..	49.74	4.72E-07	28%	34%	64--144	322--402	2	NP_187123	PR4 (PATHOGENESIS-RELATED 4)	70.09	9.59E-11	35%	49%	155--297	60--192
					3	pfam02993	MCPVI, Minor capsid protein VI. This minor capsid protein may act as a link between the external capsid and the internal DNA-protein core. The C-terminal 11 residues may function as a protease cofactor leading to enzyme activation..	46.94	3.39E-06	26%	36%	41--145	111--214	3	AAC33732	PR-4 type protein	68.17	3.65E-10	37%	46%	175--297	23--142
					4	COG5373	COG5373, Predicted membrane protein [Function unknown].	45.73	7.30E-06	27%	35%	30--120	35--124	4	BAC16357	hevein-like protein	66.63	1.06E-09	34%	44%	173--297	71--212

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					5	pfam05518	Totivirus coat, Totivirus coat protein..	43.82	2.85E-05	20%	26%	25--145	628--750	5	CAA42820	PR-4a protein	62.39	2.00E-08	34%	46%	175--296	27--145
					6	pfam03276	Gag spuma, Spumavirus gag protein..	42.38	8.20E-05	31%	36%	11--123	167--277	6	CAA41437	pathogenesis-related protein 4A	61.62	3.41E-08	34%	46%	175--296	27--145
					7	pfam05887	Trypan_PARP, Procyclic acidic repetitive protein (PARP). This family consists of several Trypanosoma brucei procyclic acidic repetitive protein (PARP) like sequences. The procyclic acidic repetitive protein (parp) genes of Trypanosoma brucei encode a small family of abundant surface proteins whose expression is restricted to the procyclic form of the parasite. They are found at two unlinked loci, parpA and parpB; transcription of both loci is developmentally regulated	42.27	8.40E-05	29%	58%	81--154	60--133	7	CAA41438	pathogenesis-related protein 4B	61.23	4.46E-08	34%	46%	175--296	19--137
					8	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	42.05	9.01E-05	31%	40%	37--167	28--159	8	CAA42821	PR-4b protein	61.23	4.46E-08	34%	46%	175--296	27--145
					9	COG3147	DedD, Uncharacterized protein conserved in bacteria [Function unknown].	41.93	9.40E-05	25%	32%	11--122	39--150	9	BAD11073	pathogenesis-related protein 4b	60.85	5.82E-08	33%	46%	175--297	70--189
					10	COG3087	FtsN, Cell division protein [Cell division and chromosome partitioning].	41.95	9.83E-05	24%	36%	16--152	54--190	10	AAF61434	pathogenesis-related protein 4A	60.85	5.82E-08	34%	48%	168--296	21--145
C299R	129177--130316	380	43,593	6.08	1	cd00204	ANK, ankyrin repeats; ankyrin repeats mediate protein-protein interactions in very diverse families of proteins. The number of ANK repeats in a protein can range from 2 to over 20 (ankyrins, for example). ANK repeats may occur in combinations with other types of domains. The structural repeat unit contains two antiparallel helices and a beta-hairpin, repeats are stacked in a superhelical arrangement; this alignment contains 4 consecutive repeats.	53.16	4.18E-08	42%	65%	293--359	11--76	1	NP_048596	4 ankyrin repeats; similar to Drosophila melanogaster ankyrin, corresponds to GenBank Accession Number L35601	438.34	1.88E-121	59%	74%	1--376	1--375
					2	pfam00023	Ank, Ankyrin repeat. There's no clear separation between noise and signal on the HMM search Ankyrin repeats generally consist of a beta, alpha, alpha, beta order of secondary structures. The repeats associate to form a higher order structure..	38.89	8.05E-04	50%	69%	322--354	1--33	2	AAY61232	Guanosine polyphosphate pyrophosphohydrolases/synthetases homolog	116.70	1.25E-24	28%	48%	26--356	683--997
														3	EAL29245	GA14074-PA	95.13	3.91E-18	26%	45%	33--355	187--496
														4	XP_797753	PREDICTED: similar to ankyrin repeat domain protein 17 isoform b, partial	94.36	6.67E-18	25%	44%	17--359	736--1098
														5	XP_782887	PREDICTED: similar to Ankyrin-1 (Erythrocyte ankyrin) (Ankyrin R)	92.82	1.94E-17	23%	46%	51--359	214--543
														6	ZP_00373467	ankyrin repeat domain protein	92.05	3.31E-17	27%	44%	38--379	95--481
														7	XP_788897	PREDICTED: similar to ankyrin 3, epithelial isoform d	91.66	4.32E-17	25%	47%	51--356	41--367
														8	AAV85825	ankyrin domain protein	90.51	9.63E-17	28%	43%	38--379	81--467
														9	AAM11327	GH01626p	89.35	2.14E-16	26%	44%	29--353	36--351
														10	AAN12046	CG7462-PC, isoform C	89.35	2.14E-16	26%	44%	29--353	186--501
C300R	130486--131349	288	33,837	6.81	1	smart00220	S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-specific kinase subfamily. .	177.71	1.59E-45	32%	51%	27--284	1--256	1	AAU06280	protein kinase A248R	559.68	3.65E-158	95%	95%	1--288	22--309
					2	cd00180	S_TKc, Serine/Threonine protein kinases, catalytic domain. Phosphotransferases of the serine or threonine-specific kinase subfamily. The enzymatic activity of these protein kinases is controlled by phosphorylation of specific residues in the activation segment of the catalytic domain, sometimes combined with reversible conformational changes in the C-terminal autoregulatory tail.	166.92	2.68E-42	32%	53%	26--284	1--256	2	AAU06275	protein kinase A248R	559.68	3.65E-158	95%	95%	1--288	1--288
					3	pfam00069	Pkinase, Protein kinase domain..	155.83	5.38E-39	30%	50%	27--284	1--258	3	AAU06282	protein kinase A248R	551.21	1.30E-155	94%	94%	1--288	22--309
					4	COG0515	SPS1, Serine/threonine protein kinase [General function prediction only / Signal transduction mechanisms / Transcription / DNA replication, recombination, and repair].	99.85	4.12E-22	26%	41%	26--284	1--278	4	AAU06274	protein kinase A248R	424.48	1.83E-117	93%	93%	68--288	22--242
					5	smart00219	TyrKc, Tyrosine kinase, catalytic domain; Phosphotransferases. Tyrosine-specific kinase subfamily. .	69.05	6.80E-13	23%	42%	28--279	2--255	5	NP_048597	PBCV-1 protein kinase	416.77	3.82E-115	71%	81%	4--288	22--308
					6	cd00192	TyrKc, Tyrosine kinase, catalytic domain. Phosphotransferases; tyrosine-specific kinase subfamily. Enzymes with TyrKc domains belong to an extensive family of proteins which share a conserved catalytic core common to both serine/threonine and tyrosine protein kinases. Enzymatic activity of tyrosine protein kinases is controlled by phosphorylation of specific tyrosine residues in the activation segment of the catalytic domain or a C-terminal tyrosine (tail) residue with reversible conformational changes.	66.37	4.58E-12	23%	42%	25--277	7--261	6	AAA87065	serine/threonine protein kinase	412.15	9.41E-114	70%	81%	4--288	18--304
					7	pfam06293	Kdo, Lipopolysaccharide kinase (Kdo/WaaP) family. These lipopolysaccharide kinases are related to protein kinases pfam00069. This family includes waaP (rfaP) gene product is required for the addition of phosphate to O-4 of the first heptose residue of the lipopolysaccharide (LPS) inner core region. It has previously been shown that WaaP is necessary for resistance to hydrophobic and polycationic antimicrobials in E. coli and that it is required for virulence in invasive strains of S. enterica.	38.71	9.12E-04	24%	42%	104--154	103--153	7	AAU06285	protein kinase A248R	399.82	4.83E-110	89%	90%	69--288	1--221
														8	AAU06286	protein kinase A248R	223.40	6.20E-57	68%	78%	134--288	1--156
														9	AAU06270	protein kinase A248R	207.61	3.52E-52	66%	77%	139--288	1--151
														10	AAU06273	protein kinase A248R	197.21	4.76E-49	66%	77%	147--288	1--143
C303R	131539--132834	432	50,042	10.77	1	COG0675	COG0675, Transposase and inactivated derivatives [DNA replication, recombination, and repair].	81.67	1.09E-16	23%	38%	37--423	1--356	1	AAU06281	putative transposase	822.00	0.00E+00	98%	98%	22--432	30--439
					2	pfam07282	Transposase_35, Putative transposase DNA-binding domain. This putative domain is found at the C-terminus of a large number of transposase proteins. This domain contains four conserved cysteines suggestive of a zinc binding domain. Given the need for transposases to bind DNA as well as the large number of DNA-binding zinc fingers we hypothesise this domain is DNA-binding.	69.88	3.83E-13	41%	53%	345--412	1--69	2	NP_048981	similar to Synechocystis transposase, corresponds to GenBank Accession Number D90909	714.92	0.00E+00	84%	91%	24--432	26--433
					3	pfam01385	Transposase_2, Probable transposase. This family includes IS891, IS1136 and IS1341. .	48.76	1.03E-06	23%	42%	42--314	1--261	3	YP_142458	putative transposase	108.61	4.10E-22	26%	43%	8--412	95--535
														4	AAS54227	AGL264Wp	105.53	3.47E-21	26%	43%	41--414	71--453
														5	ABA24789	Transposase, IS891/IS1136/IS1341	96.29	2.10E-18	26%	42%	6--414	105--533
														6	BAB78230	transposase	96.29	2.10E-18	26%	42%	6--414	61--489
														7	ZP_00158267	COG0675: Transposase and inactivated derivatives	82.42	3.14E-14	27%	43%	114--414	43--359
														8	YP_238637	ORF021	80.11	1.56E-13	26%	44%	40--409	3--364
														9	BAE47830	putative IS transposase (OrfB)	76.26	2.25E-12	23%	44%	42--409	6--384
														10	BAB80584	probable transposase	74.71	6.55E-12	23%	43%	34--409	1--381
C306L	133916--133179	246	27,399	9.63		No Hit Found								1	NP_048629	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	391.35	1.30E-107	74%	86%	4--246	1--252
														2	NP_048427	A79R	375.17	9.65E-103	78%	88%	7--230	1--226
														3	NP_048807	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	372.86	4.79E-102	73%	83%	7--246	1--249
														4	NP_049005	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	364.77	1.30E-99	69%	80%	1--245	1--255
														5	AAU06304	hypothetical protein A275R	337.04	2.91E-91	95%	98%	76--246	1--171
														6	AAU06301	hypothetical protein A275R	335.88	6.49E-91	95%	98%	76--246	1--171
														7	NP_048525	A177R	283.11	4.99E-75	56%	70%	4--244	1--243
														8	AAU06302	hypothetical protein A275R	274.25	2.32E-72	99%	100%	112--246	1--135
														9	AAU06303	hypothetical protein A275R	116.70	6.19E-25	94%	96%	188--246	1--59
														10	AAU06299	hypothetical protein A275R	114.01	4.01E-24	93%	94%	188--246	5--63
C309R	135004--135228	75	9,326	8.76		No Hit Found								1	NP_048615	A261R	57.38	1.47E-07	92%	96%	35--61	23--49
C311L	135921--135277	215	24,312	9.08		No Hit Found								1	NP_048616	A262L	189.50	5.75E-47	84%	92%	100--190	1--91
														2	NP_048617	A263L	91.66	1.63E-17	55%	64%	1--81	43--123
C313L	136882--136136	249	28,131	9.43	1	cd00529	RuvC_resolvase, Holliday junction resolvases (HJRs) are endonucleases that specifically resolve Holliday junction DNA intermediates during homologous recombination. HJR's occur in archaea, bacteria, and in the mitochondria of certain fungi, however this CD includes only the bacterial and mitochondrial HJR's. These are referred to as the RuvC family of Holliday junction resolvases, RuvC being the E.coli HJR. RuvC and its orthologs are homodimers and are structurely similar to RNase H and Hsp70.	51.38	1.59E-07	29%	44%	5--162	2--148	1	NP_048619	A265L	256.91	3.93E-67	72%	80%	80--249	1--170
C315L	137704--136913	264	30,214	6.44	1	COG2267	PldB, Lysophospholipase [Lipid metabolism].	87.79	1.53E-18	21%	39%	1--264	15--295	1	NP_048625	A271L	249.60	6.95E-65	74%	85%	31--187	1--157
					2	COG1647	COG1647, Esterase/lipase [General function prediction only].	42.94	5.67E-05	24%	40%	24--252	22--233	2	AAD36421	lipase, putative	70.86	4.42E-11	24%	44%	19--243	15--234
					3	COG0657	Aes, Esterase/lipase [Lipid metabolism].	42.22	9.41E-05	20%	35%	6--155	68--221	3	NP_568327	catalytic/ hydrolase	66.63	8.34E-10	23%	43%	18--260	82--333
					4	COG3571	COG3571, Predicted hydrolase of the alpha/beta-hydrolase fold [General function prediction only].	39.61	5.58E-04	34%	45%	8--108	5--106	4	ABB46702	lipase, putative	65.47	1.86E-09	24%	43%	8--262	93--356

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					5	COG0596	MhpC, Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily) [General function prediction only].	39.68	5.67E-04	22%	36%	1--142	5--139	5	AAB89497	lysophospholipase	63.93	5.41E-09	24%	43%	18--264	23--266
														6	AAM60954	lysophospholipase isolog, putative	63.16	9.22E-09	24%	42%	14--260	117--371
														7	NP_177867	catalytic/ hydrolase	63.16	9.22E-09	24%	42%	14--260	117--371
														8	ZP_00486825	COG2267: Lysophospholipase	62.00	2.06E-08	23%	40%	18--264	26--276
														9	NP_191845	catalytic/ hydrolase	61.23	3.51E-08	25%	41%	10--243	26--268
														10	EAL46579	conserved hypothetical protein	60.85	4.58E-08	25%	45%	15--192	24--201
C318L	138160--137744	139	15,771	10.35	1	pfam03713	DUF305, Domain of unknown function (DUF305). Domain found in small family of bacterial secreted proteins with no known function. Also found in Paramecium bursaria chlorella virus 1. This domain is short and found in one or two copies. The domain has a conserved HH motif that may be functionally important.	50.35	3.16E-07	47%	59%	87--139	1--52	1	NP_048627	A273L	246.90	1.27E-64	88%	94%	4--139	3--138
					2	COG3544	COG3544, Uncharacterized protein conserved in bacteria [Function unknown].	43.12	4.60E-05	27%	40%	50--137	6--92	2	YP_142507	unknown	100.52	1.47E-20	41%	61%	9--139	7--141
														3	EAN09897	Protein of unknown function DUF305	53.14	2.70E-06	29%	52%	9--139	8--142
														4	AAZ98258	hypothetical protein Tbd_2305	51.60	7.85E-06	34%	60%	43--139	82--175
														5	ZP_00333569	COG3544: Uncharacterized protein conserved in bacteria	51.60	7.85E-06	34%	60%	43--139	57--150
														6	EAN06281	Protein of unknown function DUF305	51.22	1.03E-05	38%	58%	43--139	57--150
														7	EAN05440	Protein of unknown function DUF305	49.29	3.90E-05	29%	49%	10--139	22--159
														8	ZP_00376873	hypothetical protein ELI2114	48.91	5.09E-05	28%	50%	33--139	48--154
C319R	138495--139187	231	27,629	4.81		No Hit Found								1	NP_048628	encodes Asp/Lys rich sequence	251.91	1.09E-65	64%	73%	28--229	61--261
C321L	141006--139195	604	68,839	11	1	cd00180	S_TKc, Serine/Threonine protein kinases, catalytic domain. Phosphotransferases of the serine or threonine-specific kinase subfamily. The enzymatic activity of these protein kinases is controlled by phosphorylation of specific residues in the activation segment of the catalytic domain, sometimes combined with reversible conformational changes in the C-terminal autoregulatory tail.	57.14	2.54E-09	28%	52%	58--254	3--152	1	NP_048632	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662	1106.66	0.00E+00	90%	92%	1--603	1--609
					2	pfam00069	Pkinase, Protein kinase domain..	56.45	4.22E-09	32%	54%	58--254	2--151	2	NP_048636	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055	622.85	9.96E-177	55%	65%	1--603	15--568
					3	smart00220	S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-specific kinase subfamily. .	55.99	6.63E-09	28%	52%	58--254	2--151	3	NP_048970	RPQT-like (9x)	331.26	5.99E-89	49%	66%	10--341	6--337
					4	COG0661	AarF, Predicted unusual protein kinase [General function prediction only].	46.09	6.02E-06	32%	54%	221--278	285--341	4	NP_048441	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563	84.34	1.27E-14	47%	66%	509--602	1--85
					5	cd00142	PI3Kc_family, Phosphoinositide 3-kinase, catalytic domain; Phosphoinositide 3-kinase isoforms participate in a variety of processes, including cell motility, the Ras pathway, vesicle trafficking and secretion, and apoptosis. These homologues may be either lipid kinases and/or protein kinases: the former phosphorylate the 3-position in the inositol ring of inositol phospholipids. The ataxia telangiectesia-mutated gene product, the targets of rapamycin (TOR) and the DNA-dependent kinase have not been found to possess lipid kinase activity. Some of this family possess PI-4 kinase activities.	41.70	1.28E-04	39%	59%	219--287	139--195	5	NP_049032	similar to Chlorella virus PBCV-1 ORF A282L, corresponds to GenBank Accession Number U42580	71.63	8.55E-11	82%	92%	564--604	1--41
					6	cd00891	PI3Kc, Phosphoinositide 3-kinase (PI3K), catalytic domain; PI3Ks phosphorylate the 3-position in the inositol ring of inositol phospholipids. PI3Ks play an important role in a variety of fundamental cellular processes, including cell motility, the Ras pathway, vesicle trafficking and secretion, and apoptosis. They can be divided into 3 main classes, defined by their substrate specificity and domain structure..	40.96	2.35E-04	38%	52%	192--255	174--240	6	XP_757661	protein kinase Fuz7	54.30	1.41E-05	25%	44%	69--270	80--275
					7	COG0515	SPS1, Serine/threonine protein kinase [General function prediction only / Signal transduction mechanisms / Transcription / DNA replication, recombination, and repair].	40.14	3.49E-04	14%	30%	177--481	76--370	7	Q99078	Dual specificity protein kinase FUZ7	54.30	1.41E-05	25%	44%	69--270	80--275
					8	cd00895	PI3Kc_II, Phosphoinositide 3-kinase (PI3K) class II, catalytic domain; Phosphoinositide 3-kinase isoforms participate in a variety of processes, including cell motility, the Ras pathway, vesicle trafficking and secretion, and apoptosis. They phosphorylate the 3-position in the inositol ring of inositol phospholipids. PI3K class II phosphorylate phosphoinositol (PtdIns), PtdIns(4)-phosphate, but not PtdIns(4,5)-bisphosphate. They are larger, having a C2 domain at the C-terminus.	40.31	3.75E-04	33%	43%	174--246	146--226	8	EAA70129	hypothetical protein FG09903.1	51.60	9.16E-05	24%	40%	72--270	38--232
C324L	142580--141063	506	56,477	10.95	1	pfam00069	Pkinase, Protein kinase domain..	65.31	9.59E-12	29%	47%	85--283	6--183	1	NP_048636	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055	890.18	0.00E+00	88%	90%	1--504	15--509
					2	cd00180	S_TKc, Serine/Threonine protein kinases, catalytic domain. Phosphotransferases of the serine or threonine-specific kinase subfamily. The enzymatic activity of these protein kinases is controlled by phosphorylation of specific residues in the activation segment of the catalytic domain, sometimes combined with reversible conformational changes in the C-terminal autoregulatory tail.	63.30	4.36E-11	31%	51%	85--257	7--151	2	NP_048632	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662	607.45	3.45E-172	55%	63%	1--506	1--610
					3	smart00220	S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-specific kinase subfamily. .	61.77	1.07E-10	32%	49%	85--257	6--150	3	NP_048970	RPQT-like (9x)	294.66	4.95E-78	41%	59%	10--388	6--386
					4	smart00219	TyrKc, Tyrosine kinase, catalytic domain; Phosphotransferases. Tyrosine-specific kinase subfamily. .	45.17	1.23E-05	26%	47%	58--266	2--161	4	NP_048441	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563	87.04	1.57E-15	49%	66%	411--504	1--85
					5	cd00192	TyrKc, Tyrosine kinase, catalytic domain. Phosphotransferases; tyrosine-specific kinase subfamily. Enzymes with TyrKc domains belong to an extensive family of proteins which share a conserved catalytic core common to both serine/threonine and tyrosine protein kinases. Enzymatic activity of tyrosine protein kinases is controlled by phosphorylation of specific tyrosine residues in the activation segment of the catalytic domain or a C-terminal tyrosine (tail) residue with reversible conformational changes.	43.64	3.51E-05	28%	45%	58--262	10--169	5	NP_049032	similar to Chlorella virus PBCV-1 ORF A282L, corresponds to GenBank Accession Number U42580	72.40	4.00E-11	84%	94%	466--504	1--39
C327L	143478--142642	279	30,690	9.41	1	cd00542	PVA, Penicillin V acylase (PVA), also known as conjugated bile salt acid hydrolase (CBAH), catalyzes the hydrolysis of penicillin V to yield 6-amino penicillanic acid (6-APA), an important key intermediate of semisynthetic penicillins. PVA has an N-terminal nucleophilic cysteine as do other Ntn Hydrolases which is exposed by prossessing of the PVA precursor. PVA forms a homotetramer.	202.75	4.32E-53	33%	47%	2--276	1--297	1	NP_048638	PBCV-1 amidase	493.04	4.00E-138	86%	93%	1--278	1--278
					2	COG3049	COG3049, Penicillin V acylase and related amidases [Cell envelope biogenesis, outer membrane].	172.44	5.27E-44	28%	43%	1--279	22--323	2	AAU25651	Choloylglycine hydrolase	148.29	2.42E-34	31%	50%	1--278	1--300
					3	pfam02275	CBAH, Linear amide C-N hydrolases, choloylglycine hydrolase family. This family includes several hydrolases which cleave carbon-nitrogen bonds, other than peptide bonds, in linear amides. These include choloylglycine hydrolase (conjugated bile acid hydrolase, CBAH) EC:3.5.1.24, penicillin acylase EC:3.5.1.11 and acid ceramidase EC:3.5.1.23.	159.01	5.38E-40	32%	49%	2--278	1--300	3	ABB11030	Penicillin amidase	142.90	1.02E-32	31%	49%	1--273	1--309
					4	cd01935	CGH_like, Choloylglycine hydrolase (CGH)_like. This family of choloylglycine hydrolases-like proteins includes conjugated bile acid hydrolase (CBAH), penicillin acylase and acid ceramidase which cleave carbon-nitrogen bonds, other than peptide bonds, in linear amides..	114.52	1.47E-26	32%	46%	2--269	1--270	4	AAL51724	CHOLOYLGLYCINE HYDROLASE	136.73	7.28E-31	31%	47%	2--279	37--339
					5	cd01902	CGH, CGH Choloylglycine hydrolase (also known as bile salt hydrolase) is an intestinal bacterial enzyme responsible for the deconjugation and subsequent dehydroxylation of conjugated cholic acid (CA) to form deoxycholic acid (DCA). CGH has a conserved Ntn hydrolase fold similar to those of penicillin v acylase (PVA) and acid ceramidase (AC)..	76.20	5.79E-15	27%	45%	2--237	1--242	5	AAN30379	choloylglycine hydrolase family protein	136.73	7.28E-31	31%	47%	2--279	3--305
														6	YP_414856	Choloylglycine hydrolase	136.73	7.28E-31	31%	47%	2--279	3--305
														7	AAP08002	Choloylglycine hydrolase	134.04	4.72E-30	28%	49%	1--271	1--293
														8	ZP_00238968	choloylglycine hydrolase family protein	133.65	6.17E-30	29%	50%	1--271	1--293
														9	AAT60567	choloylglycine hydrolase	132.88	1.05E-29	29%	50%	1--271	1--293
														10	AAT33011	choloylglycine hydrolase family protein	130.18	6.82E-29	28%	50%	1--271	1--293
C329L	144805--143792	338	39,711	7.38	1	cd00204	ANK, ankyrin repeats; ankyrin repeats mediate protein-protein interactions in very diverse families of proteins. The number of ANK repeats in a protein can range from 2 to over 20 (ankyrins, for example). ANK repeats may occur in combinations with other types of domains. The structural repeat unit contains two antiparallel helices and a beta-hairpin, repeats are stacked in a superhelical arrangement; this alignment contains 4 consecutive repeats.	82.05	8.98E-17	38%	59%	208--317	3--112	1	EAL87814	NACHT domain protein, putative	79.72	1.43E-13	34%	52%	186--316	057--1190
					2	COG0666	Arp, FOG: Ankyrin repeat [General function prediction only].	56.05	6.34E-09	25%	44%	159--323	18--192	2	1N0R_A	Chain A, 4ank: A Designed Ankyrin Repeat Protein With Four Identical Consensus Repeats	77.03	9.24E-13	40%	62%	198--302	18--125

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
														3	AAO25691	ankyrin repeat protein E4_2	76.64	1.21E-12	33%	55%	185--317	17--152
														4	AAN41651	tankyrase 1	75.49	2.69E-12	33%	50%	185--319	124--262
														5	AAS14030	ankyrin repeat domain protein	74.33	5.99E-12	37%	55%	197--311	169--286
														6	XP_849388	PREDICTED: similar to Tankyrase 1 (TANK1) (Tankyrase I) (TNKS-1) (TRF1-interacting ankyrin-related ADP-ribose polymerase) isoform 3	73.17	1.33E-11	32%	48%	185--326	185--330
														7	XP_696745	PREDICTED: similar to Tnks protein	73.17	1.33E-11	32%	47%	185--326	127--272
														8	XP_695922	PREDICTED: similar to Tnks protein	73.17	1.33E-11	32%	47%	185--326	127--272
														9	XP_693089	PREDICTED: similar to Tankyrase 1 (TANK1) (Tankyrase I) (TNKS-1) (TRF1-interacting ankyrin-related ADP-ribose polymerase)	73.17	1.33E-11	32%	47%	185--326	110--255
														10	AAH57370	Tnks protein	72.79	1.74E-11	32%	48%	185--326	178--323
C334R	145085--146200	372	41,903	9.53		No Hit Found								1	NP_048640	A286R	583.95	2.68E-165	73%	86%	1--372	7--378
C337L	147006--146194	271	30,562	7.15	1	smart00220	S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-specific kinase subfamily. .	160.77	1.62E-40	31%	51%	18--268	1--256	1	NP_048643	similar to PBCV-1 serine/threonine protein kinase, corresponds to GenBank Accession Number U14660	389.42	5.91E-107	67%	81%	1--268	1--279
					2	cd00180	S_TKc, Serine/Threonine protein kinases, catalytic domain. Phosphotransferases of the serine or threonine-specific kinase subfamily. The enzymatic activity of these protein kinases is controlled by phosphorylation of specific residues in the activation segment of the catalytic domain, sometimes combined with reversible conformational changes in the C-terminal autoregulatory tail	160.76	1.94E-40	30%	50%	18--268	2--256	2	AAU06282	protein kinase A248R	135.96	1.18E-30	30%	54%	9--268	39--305
					3	pfam00069	Pkinase, Protein kinase domain..	152.36	6.39E-38	29%	47%	18--268	1--258	3	AAU06280	protein kinase A248R	134.04	4.47E-30	30%	54%	9--268	39--305
					4	COG0515	SPS1, Serine/threonine protein kinase [General function prediction only / Signal transduction mechanisms / Transcription / DNA replication, recombination, and repair].	90.22	2.74E-19	26%	46%	17--269	1--279	4	AAU06275	protein kinase A248R	134.04	4.47E-30	30%	54%	9--268	18--284
					5	smart00219	TyrKc, Tyrosine kinase, catalytic domain; Phosphotransferases. Tyrosine-specific kinase subfamily. .	74.83	1.35E-14	23%	44%	22--264	5--253	5	NP_048631	similar to PBCV-1 serine/threonine protein kinase, corresponds to GenBank Accession Number U14660	132.49	1.30E-29	30%	52%	24--269	25--277
					6	cd00192	TyrKc, Tyrosine kinase, catalytic domain. Phosphotransferases; tyrosine-specific kinase subfamily. Enzymes with TyrKc domains belong to an extensive family of proteins which share a conserved catalytic core common to both serine/threonine and tyrosine protein kinases. Enzymatic activity of tyrosine protein kinases is controlled by phosphorylation of specific tyrosine residues in the activation segment of the catalytic domain or a C-terminal tyrosine (tail) residue with reversible conformational changes	70.99	1.83E-13	23%	45%	22--262	13--262	6	AAU06285	protein kinase A248R	121.71	2.30E-26	30%	55%	61--268	2--217
														7	AAU06274	protein kinase A248R	117.86	3.32E-25	31%	54%	59--268	22--238
														8	NP_048597	PBCV-1 protein kinase	116.70	7.39E-25	26%	51%	18--268	45--304
														9	XP_629689	putative CAM kinase (CAMK)	115.55	1.65E-24	31%	52%	22--270	113--363
														10	AAA87065	serine/threonine protein kinase	114.39	3.67E-24	27%	53%	18--268	41--300
C339R	147063--147761	233	27,221	9.16	1	cd00283	GIY-YIG_Cterm, GIYX(10-11)YIG family of class I homing endonucleases C-terminus (GIY-YIG_Cterm). Homing endonucleases promote the mobility of intron or intein by recognizing and cleaving a homologous allele that lacks the sequence. They catalyze a double	55.01	1.24E-08	47%	60%	124--209	23--106	1	NP_048641	PBCV-1 33kd peptide	236.88	3.71E-61	49%	61%	24--220	1--250
					2	pfam07453	NUMOD1, NUMOD1 domain..	47.35	2.34E-06	52%	64%	167--200	1--34	2	NP_048671	A315L	136.35	6.82E-31	35%	51%	31--218	3--241
					3	smart00497	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished). .	46.28	4.97E-06	42%	56%	167--217	1--51	3	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	128.26	1.86E-28	38%	52%	31--217	3--223
					4	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	43.91	2.65E-05	32%	46%	27--117	1--83	4	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	96.67	5.99E-19	46%	63%	41--156	16--128
					5	pfam01541	GIY-YIG, GIY-YIG catalytic domain. This domain called GIY-YIG is found in the amino terminal region of excinuclease abc subunit c (uvrC), bacteriophage T4 endonucleases segA, segB, segC, segD and segE; it is also found in putative endonucleases encoded by group I introns of fungi and phage. The structure of I-TevI a GIY-YIG endonuclease, reveals a novel alpha/beta-fold with a central three-stranded antiparallel beta-sheet flanked by three helices. The most conserved and putative catalytic residues are located on a shallow, concave surface and include a metal coordination site	38.99	7.99E-04	26%	44%	27--114	1--89	5	NP_899393	SegD	57.38	4.03E-07	28%	48%	28--191	2--169
														6	CAA38804	GIY COII i1 grp IB protein	53.53	5.82E-06	23%	37%	8--205	56--278
														7	AAC49244	ORF301	49.68	8.40E-05	28%	48%	42--164	88--220
C342L	148734--147757	326	37,215	7.86	1	pfam01374	Glyco_hydro_46, Glycosyl hydrolase family 46. This family are chitosanase enzymes..	259.20	3.87E-70	39%	53%	98--309	1--216	1	NP_048646	PBCV-1 chitosanase	544.66	1.49E-153	78%	87%	3--326	1--328
					2	cd00978	chitosanase_glyco_hydro_46, Glycosyl hydrolase family 46 chitosanase domain. This family are composed of the chitosanase enzymes which hydrolyzes chitosan, a biopolymer of beta (1,4)-linked -D-glucosamine (GlcN) residues produced by partial or full deacetylation of chitin. Chitosanases play a role in defense against pathogens such as fungi and are found in microorganisms, fungi, viruses, and plants. Microbial chitosanases who members are the most prevalent can be divided into 3 subclasses based on the specificity of the cleavage positions for partial acetylated chitosan. Subclass I chitosanases such as N174 can split GlcN-GlcN and GlcNAc-GlcN linkages, whereas subclass II chitosanases such as Bacillus sp. no. 7-M can cleave only GlcN-GlcN linkages. Subclass III chitosanases such as MH-K1 chitosanase are the most versatile and can split both GlcN-GlcN and GlcN-GlcNAc linkages..	216.74	2.91E-57	29%	45%	88--309	1--221	2	BAA20342	vChta-1	533.87	2.63E-150	76%	87%	3--326	1--328
														3	ABC17783	secreted chitosanase precursor	85.89	1.89E-15	28%	45%	71--309	34--275
														4	P48846	Chitosanase precursor	79.34	1.77E-13	27%	46%	95--313	53--266
														5	BAA94840	chitosanase	76.26	1.50E-12	26%	46%	95--313	60--274
														6	CAB14630	chitosanase	76.26	1.50E-12	28%	43%	100--325	49--276
														7	BAB19276	chitosanase	74.71	4.36E-12	27%	42%	100--325	50--277
														8	BAA01474	chitosanase	73.94	7.45E-12	25%	42%	65--309	39--286
														9	1QGI_A	Chain A, Chitosanase From Bacillus Circulans	73.56	9.72E-12	25%	42%	71--309	3--244
														10	ZP_00111326	COG1652: Uncharacterized protein containing LysM domain	73.17	1.27E-11	25%	39%	65--302	63--282
C344L	149665--148706	320	36,183	6.9	1	COG0451	WcaG, Nucleoside-diphosphate-sugar epimerases [Cell envelope biogenesis, outer membrane / Carbohydrate transport and metabolism].	160.51	2.29E-40	29%	50%	6--310	2--314	1	AAO67556	GDP-4-keto-6-deoxy-D-mannose epimerase/reductase	556.98	2.81E-157	84%	93%	1--313	3--315
					2	pfam01370	Epimerase, NAD dependent epimerase/dehydratase family. This family of proteins utilise NAD as a cofactor. The proteins in this family use nucleotide-sugar substrates for a variety of chemical reactions..	115.31	7.85E-27	26%	45%	7--308	1--310	2	NP_048649	PBCV-1 fucose synthase	553.90	2.38E-156	84%	92%	1--313	3--315
					3	COG1087	GalE, UDP-glucose 4-epimerase [Cell envelope biogenesis, outer membrane].	90.61	2.61E-19	25%	42%	6--297	2--313	3	ZP_00544248	NAD-dependent epimerase/dehydratase	375.56	1.16E-102	62%	76%	1--298	1--301
					4	COG1088	RfbB, dTDP-D-glucose 4,6-dehydratase [Cell envelope biogenesis, outer membrane].	84.88	1.36E-17	24%	45%	6--307	2--319	4	AAM30354	GDP-fucose synthetase	371.70	1.67E-101	61%	75%	1--298	1--301
					5	COG1091	RfbD, dTDP-4-dehydrorhamnose reductase [Cell envelope biogenesis, outer membrane].	59.12	6.79E-10	27%	44%	6--289	2--266	5	ZP_00308009	COG0451: Nucleoside-diphosphate-sugar epimerases	367.47	3.16E-100	61%	78%	15--297	1--283
					6	COG1089	Gmd, GDP-D-mannose dehydratase [Cell envelope biogenesis, outer membrane].	52.98	4.78E-08	24%	39%	9--310	7--344	6	ABA57735	NAD-dependent epimerase/dehydratase	365.16	1.57E-99	55%	76%	1--305	1--305
														7	AAS70777	gdp-l-fucose synthetase	358.22	1.92E-97	55%	71%	1--304	1--310
														8	ABB39695	GDP-fucose synthetase NAD dependent epimerase/dehydratase	348.59	1.52E-94	54%	71%	1--304	3--306
														9	AAR33958	GDP-fucose synthetase	348.59	1.52E-94	55%	73%	1--298	1--298
														10	ZP_00662732	NAD-dependent epimerase/dehydratase	348.21	1.98E-94	55%	73%	1--307	1--307
C346R	149780--150559	260	30,909	6.72	1	pfam02086	MethyltransfD12, D12 class N6 adenine-specific DNA methyltransferase..	148.56	9.69E-37	33%	47%	7--237	1--251	1	AAC03126	DNA adenine methyltransferase	519.62	3.54E-146	94%	96%	1--260	1--261
					2	COG0338	Dam, Site-specific DNA methylase [DNA replication, recombination, and repair].	142.35	5.62E-35	30%	51%	1--260	2--265	2	Q01511	Modification methylase CviBI (Adenine-specific methyltransferase CviBI) (M.CviBI)	487.65	1.49E-136	85%	95%	1--260	1--260
														3	AAK08495	N.BstNBI methyltransferase	191.43	2.21E-47	38%	55%	1--260	1--283
														4	CAD92748	putative adenine methyltransferase	179.10	1.13E-43	38%	55%	1--260	1--283
														5	AAK81290	Site-specific DNA methylase dam	169.47	8.98E-41	35%	55%	4--260	1--280
														6	AAK27214	adenine methyltransferase M.PleI	167.93	2.61E-40	34%	53%	1--260	1--283
														7	ZP_00886307	DNA adenine methylase	134.42	3.20E-30	33%	50%	3--260	14--290
														8	ZP_00575937	N6 adenine-specific DNA methyltransferase, D12 class	132.49	1.22E-29	31%	51%	1--260	7--270
														9	AAB98590	modification methylase, type II R/M system 2	130.18	6.04E-29	32%	50%	1--260	3--289
														10	BAD74251	type IIs modification methyltransferase	129.03	1.35E-28	33%	50%	2--253	16--276
C348L	151114--150623	164	19,448	10.18		No Hit Found									No Hit Found	No Hit Found						
C349L	152009--151197	271	32,151	6.81		No Hit Found									No Hit Found	No Hit Found						

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
C351R	152159--152554	132	15,146	12.05		No Hit Found								1	NP_048650	A296R	135.19	5.56E-31	75%	87%	47--132	72--156
C354R	152565--153704	380	42,895	8.79		No Hit Found								1	NP_048920	similar to Chlorella virus PBCV-1 ORF A154L, corresponds to GenBank Accession Number U42580	540.04	4.59E-152	72%	81%	22--380	1--350
														2	NP_048502	A154L	502.29	1.06E-140	67%	79%	25--380	3--347
														3	NP_048477	similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903	497.66	2.61E-139	64%	79%	24--380	4--356
														4	NP_077492	EsV-1-7	61.62	4.79E-08	28%	50%	32--205	50--222
C358L	154816--153707	370	42,027	9.04	1	pfam00145	DNA_methylase, C-5 cytosine-specific DNA methylase..	138.52	8.51E-34	27%	41%	2--234	1--225	1	NP_048873	M.CviAII cytosine DNA methyltransferase	313.15	8.84E-84	44%	61%	1--367	1--342
					2	cd00315	Cyt_C5_DNA_methylase, Cytosine-C5 specific DNA methylases; Methyl transfer reactions play an important role in many aspects of biology. Cytosine-specific DNA methylases are found both in prokaryotes and eukaryotes. DNA methylation, or the covalent addition of a methyl group to cytosine within the context of the CpG dinucleotide, has profound effects on the mammalian genome. These effects include transcriptional repression via inhibition of transcription factor binding or the recruitment of methyl-binding proteins and their associated chromatin remodeling factors, X chromosome inactivation, imprinting and the suppression of parasitic DNA sequences. DNA methylation is also essential for proper embryonic development and is an important player in both DNA repair and genome stability	136.20	4.74E-33	31%	47%	2--173	1--178	2	AAC64006	cytosine methyltransferase	306.61	8.28E-82	45%	60%	2--367	3--356
					3	COG0270	Dcm, Site-specific DNA methylase [DNA replication, recombination, and repair].	109.78	4.06E-25	23%	41%	1--198	3--211	3	AAC55063	cytosine methyltransferase	300.44	5.93E-80	42%	60%	4--367	5--360
														4	AAV84097	CviPII m5C DNA methyltransferase	299.67	1.01E-79	43%	62%	2--367	16--357
														5	NP_049039	nonfunctional M.CviAV cytosine DNA methyltransferase	298.13	2.94E-79	42%	60%	4--368	5--361
														6	NP_048886	M.CviAIV cytosine DNA methyltransferase	282.34	1.67E-74	44%	58%	1--355	2--332
														7	CAD80133	gp9.1	96.67	1.30E-18	36%	52%	5--157	4--159
														8	ZP_00783160	C-5 cytosine-specific DNA methylase	90.89	7.14E-17	33%	47%	5--161	4--171
														9	ZP_00874816	C-5 cytosine-specific DNA methylase	90.12	1.22E-16	34%	47%	5--161	4--171
														10	AAK75434	type II DNA modification methyltransferase Spn5252IP	88.20	4.63E-16	33%	45%	1--159	1--170
C363R	154884--155999	372	42,557	8.53	1	COG4106	Tam, Trans-aconitate methyltransferase [General function prediction only].	60.70	2.64E-10	31%	49%	35--156	22--130	1	AAC57943	DNA adenine methyltransferase	327.02	5.91E-88	44%	65%	1--370	1--368
					2	COG4123	COG4123, Predicted O-methyltransferase [General function prediction only].	57.56	2.04E-09	19%	35%	30--206	30--216	2	AAC03124	DNA adenine methyltransferase	199.13	1.86E-49	31%	54%	1--369	1--368
					3	COG0286	HsdM, Type I restriction-modification system methyltransferase subunit [Defense mechanisms].	52.72	6.37E-08	22%	35%	21--239	164--409	3	AAC57945	DNA adenine methyltransferase	177.95	4.44E-43	28%	54%	7--369	6--357
					4	pfam01170	UPF0020, Putative RNA methylase family UPF0020. This domain is probably a methylase. It is associated with the THUMP domain that also occurs with RNA modification domains..	45.64	8.88E-06	28%	43%	46--121	32--114	4	CAA29835	unnamed protein product	177.18	7.57E-43	27%	53%	2--370	5--377
					5	COG2890	HemK, Methylase of polypeptide chain release factors [Translation, ribosomal structure and biogenesis].	44.21	2.00E-05	30%	49%	43--116	111--188	5	P52284	Modification methylase CviRI (Adenine-specific methyltransferase CviRI) (M.CviRI)	173.33	1.09E-41	31%	50%	3--370	4--378
					6	COG2813	RsmC, 16S RNA G1207 methylase RsmC [Translation, ribosomal structure and biogenesis].	39.48	5.78E-04	29%	40%	34--154	149--265	6	AAC03125	DNA adenine methyltransferase	164.47	5.08E-39	28%	49%	3--371	4--382
					7	COG1041	COG1041, Predicted DNA modification methylase [DNA replication, recombination, and repair].	36.09	6.77E-03	24%	42%	17--156	167--311	7	AAM81324	BpmI methyltransferase	63.93	9.34E-09	24%	44%	22--226	14--244
					8	COG2264	PrmA, Ribosomal protein L11 methylase [Translation, ribosomal structure and biogenesis].	35.66	8.85E-03	27%	42%	35--111	155--236	8	ZP_00371924	helicase, SNF2 family	62.00	3.55E-08	25%	42%	38--223	217--419
														9	CAI07563	N6 adenine-specific DNA methyltransferase, N12 class	60.46	1.03E-07	29%	42%	20--174	13--158
														10	AAP78031	type I restriction/modification enzyme	58.15	5.12E-07	27%	43%	14--183	440--636
C365R	156060--157001	314	36,208	8.45	1	COG2890	HemK, Methylase of polypeptide chain release factors [Translation, ribosomal structure and biogenesis].	56.16	4.97E-09	24%	44%	29--150	98--238	1	CAB92310	DNA methyltrasferase	206.84	6.91E-52	39%	57%	18--314	44--344
					2	COG1041	COG1041, Predicted DNA modification methylase [DNA replication, recombination, and repair].	54.20	2.38E-08	29%	45%	40--155	196--315	2	AAF74028	M.Hpy188I	204.53	3.43E-51	39%	56%	18--314	44--344
					3	pfam01170	UPF0020, Putative RNA methylase family UPF0020. This domain is probably a methylase. It is associated with the THUMP domain that also occurs with RNA modification domains..	51.81	1.18E-07	29%	40%	39--152	26--142	3	ZP_00783145	reticulocyte binding protein	70.09	1.01E-10	34%	53%	36--163	694--827
					4	COG2813	RsmC, 16S RNA G1207 methylase RsmC [Translation, ribosomal structure and biogenesis].	51.81	1.30E-07	34%	49%	44--144	161--260	4	ZP_00780227	SNF2 family protein	70.09	1.01E-10	34%	53%	36--163	694--827
					5	COG0286	HsdM, Type I restriction-modification system methyltransferase subunit [Defense mechanisms].	46.95	3.42E-06	20%	36%	19--249	167--409	5	AAN00153	SNF2 family protein	70.09	1.01E-10	34%	53%	36--163	694--827
					6	COG4123	COG4123, Predicted O-methyltransferase [General function prediction only].	46.39	5.24E-06	28%	41%	44--152	47--172	6	ZP_00874800	SNF2-related	69.71	1.32E-10	34%	55%	36--163	694--827
					7	COG2519	GCD14, tRNA(1-methyladenosine) methyltransferase and related methyltransferases [Translation, ribosomal structure and biogenesis].	41.76	1.31E-04	23%	35%	40--153	93--198	7	ZP_00874780	SNF2-related:Helicase, C-terminal	69.32	1.72E-10	34%	53%	36--163	694--827
					8	COG2226	UbiE, Methylase involved in ubiquinone/menaquinone biosynthesis [Coenzyme metabolism].	38.72	1.02E-03	27%	47%	40--107	50--123	8	AAS71294	site-specific modification DNA-methyltransferase	66.24	1.46E-09	31%	52%	22--156	17--179
					9	COG4122	COG4122, Predicted O-methyltransferase [General function prediction only].	36.43	4.87E-03	26%	48%	65--148	82--164	9	ZP_00413234	Helicase, C-terminal	65.86	1.90E-09	30%	52%	42--182	263--406
					10	COG2230	Cfa, Cyclopropane fatty acid synthase and related methyltransferases [Cell envelope biogenesis, outer membrane].	36.40	5.30E-03	22%	42%	40--148	71--174	10	CAD47011	Unknown	64.31	5.54E-09	34%	52%	42--163	494--618
C367L	157725--157177	183	21,592	10.09		No Hit Found								1	NP_048651	A297L	299.29	3.56E-80	81%	91%	1--170	1--170
														2	BAE64611	unnamed protein product	54.30	1.99E-06	30%	48%	3--116	6--119
C369L	158443--157763	227	25,839	6.73		No Hit Found								1	NP_048652	A298L	374.02	1.84E-102	78%	90%	3--227	1--225
														2	ZP_00463717	hypothetical protein Bcen2424DRAFT_1981	57.00	4.97E-07	24%	49%	34--201	24--205
														3	ZP_00982604	hypothetical protein BcenP_01000047	57.00	4.97E-07	24%	49%	34--201	24--205
														4	ABB08661	hypothetical protein Bcep18194_A5067	56.61	6.49E-07	25%	50%	34--201	24--205
														5	ZP_00689842	hypothetical protein BambDRAFT_1259	54.30	3.22E-06	24%	47%	34--201	24--205
C372L	159525--158473	351	41,361	10.21		No Hit Found								1	NP_048711	A354R	271.55	2.72E-71	53%	70%	112--350	2--235
														2	AAL73479	endonuclease	64.70	5.04E-09	30%	44%	17--158	15--163
														3	AAL73476	endonuclease	60.85	7.28E-08	32%	47%	6--141	9--144
														4	NP_048779	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	59.69	1.62E-07	25%	44%	41--320	31--319
														5	AAO93095	I-BasI	56.61	1.37E-06	29%	42%	18--167	14--174
														6	AAM00817	HNH endonuclease I-TwoI	54.30	6.82E-06	33%	54%	17--99	11--92
														7	YP_189077	HNH endonuclease family protein	52.76	1.98E-05	43%	53%	27--99	47--112
														8	NP_695069	putative endodeoxyribonuclease	51.60	4.42E-05	29%	47%	18--128	9--118
														9	AAA56884	endodeoxyribonuclease	51.60	4.42E-05	33%	46%	11--123	10--114
														10	NP_047162	putative HNH homing endonuclease	51.22	5.77E-05	35%	47%	17--111	13--105
C374L	160271--159642	210	24,005	5.2		No Hit Found								1	NP_048655	A301L	216.47	4.22E-55	54%	65%	1--201	1--239
C377R	160325--160564	80	9,629	9.15		No Hit Found								1	NP_048658	A304R	109.77	2.46E-23	73%	86%	4--79	1--76
C378L	161108--160563	182	20,501	10.77	1	cd00127	DSPc, Dual specificity phosphatases (DSP); Ser/Thr and Tyr protein phosphatases. Structurally similar to tyrosine-specific phosphatases but with a shallower active site cleft and a distinctive active site signature motif, HCxxGxxR. Characterized as VHR- or Cdc25-like..	128.44	1.06E-30	32%	51%	24--160	1--139	1	NP_048659	similar to human protein Tyr-phosphatase, corresponds to GenBank Accession Number U27193	322.78	2.96E-87	92%	96%	1--168	1--168
					2	smart00195	DSPc, Dual specificity phosphatase, catalytic domain; .	124.62	1.28E-29	36%	56%	24--162	1--139	2	AAB88308	Lateral-signal-induced phosphatase protein 1	76.64	3.69E-13	32%	51%	25--163	185--325
					3	pfam00782	DSPc, Dual specificity phosphatase, catalytic domain. Ser/Thr and Tyr protein phosphatases. The enzyme's tertiary fold is highly similar to that of tyrosine-specific phosphatases, except for a "recognition" region..	124.21	1.60E-29	35%	54%	24--162	1--139	3	NP_998405	dual specificity phosphatase 16	74.71	1.40E-12	33%	51%	25--163	159--298
					4	COG2453	CDC14, Predicted protein-tyrosine phosphatase [Signal transduction mechanisms].	48.55	1.15E-06	29%	52%	92--144	95--147	4	XP_543810	PREDICTED: similar to Dual specificity protein phosphatase 16 (Mitogen-activated protein kinase phosphatase 7) (MAP kinase phosphatase 7) (MKP-7)	73.94	2.39E-12	33%	52%	25--163	159--298
					5	smart00404	PTPc_motif, Protein tyrosine phosphatase, catalytic domain motif; .	36.94	3.31E-03	29%	47%	100--138	36--74	5	AAH42101	DUSP16 protein	72.79	5.33E-12	33%	52%	25--163	159--298
														6	AAH31643	Unknown (protein for IMAGE:5176724)	72.79	5.33E-12	33%	52%	25--163	110--249
														7	AAI09236	Dual specificity phosphatase 16	72.79	5.33E-12	33%	52%	25--163	159--298
														8	AAI09235	Dual specificity phosphatase 16	72.79	5.33E-12	33%	52%	25--163	159--298
														9	XP_520751	PREDICTED: similar to KIAA1700 protein	72.79	5.33E-12	33%	52%	25--163	268--407
														10	BAB21791	KIAA1700 protein	72.79	5.33E-12	33%	52%	25--163	184--323
C379L	161396--161136	87	10,613	4.08		No Hit Found								1	NP_048660	A306L	114.78	7.74E-25	87%	92%	32--87	31--86

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
C380L	161809--161450	120	13,885	4.17		No Hit Found								1	NP_048663	A308L	89.74	2.63E-17	82%	96%	53--102	20--69
														2	NP_048664	a309L	80.49	1.59E-14	67%	76%	1--56	1--56
C381L	162415--161906	170	18,518	8.42		No Hit Found								1	NP_048665	A310L	305.83	3.08E-82	85%	94%	1--170	1--170
C382L	163422--162679	248	27,538	5.07		No Hit Found								1	NP_048667	PBCV-1 33kDa translation peptide	426.79	2.84E-118	90%	93%	1--237	1--237
														2	NP_077561	EsV-1-76	73.17	7.97E-12	27%	41%	12--219	20--238
														3	AAR26966	FirrV-1-I1	61.62	2.40E-08	27%	42%	35--212	36--222
														4	AAR26885	FirrV-1-B10	55.07	2.25E-06	27%	41%	40--205	52--228
C384L	163929--163702	76	8,355	8.92		No Hit Found								1	NP_048669	A313L	95.13	6.35E-19	65%	73%	1--67	1--67
C386R	164108--164422	105	12,536	10.27		No Hit Found								1	NP_048670	A314R	120.17	1.86E-26	71%	80%	24--105	1--80
C387R	164452--164646	65	7,380	10.89		No Hit Found								1	NP_048437	A89R	124.79	7.47E-28	92%	96%	1--65	1--65
C389R	164789--166072	428	47,920	10.74	1	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	59.39	6.44E-10	37%	47%	21--115	42--135	1	NP_048674	A318R	433.34	7.16E-120	94%	97%	217--427	1--211
					2	pfam05887	Trypan_PARP, Procyclic acidic repetitive protein (PARP). This family consists of several Trypanosoma brucei procyclic acidic repetitive protein (PARP) like sequences. The procyclic acidic repetitive protein (parp) genes of Trypanosoma brucei encode a small family of abundant surface proteins whose expression is restricted to the procyclic form of the parasite. They are found at two unlinked loci, parpA and parpB; transcription of both loci is developmentally regulated	54.98	1.15E-08	33%	53%	30--96	59--125	2	NP_048672	PAPK (17X); similar to PBCV-1 ORF A41R, corresponds to Genbank Accession Number U17055	109.38	2.37E-22	75%	85%	116--180	126--194
C392R	166131--167366	412	48,580	7.72		No Hit Found								1	NP_048711	A354R	139.04	2.64E-31	36%	51%	154--409	1--235
														2	NP_048779	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	77.41	9.44E-13	27%	44%	85--376	43--323
C396R	167389--167736	116	12,925	10.38		No Hit Found								1	NP_048676	A320R	132.49	3.59E-30	59%	68%	1--116	24--139
C397R	167753--168106	118	12,792	8.8		No Hit Found								1	NP_048677	A321R	142.90	2.63E-33	60%	76%	6--118	8--119
C398L	168650--168114	179	20,825	5.88		No Hit Found								1	NP_048678	A322L	182.57	4.54E-45	55%	65%	1--179	1--176
C400L	169994--168690	435	48,594	4.72		No Hit Found								1	NP_048680	A324L	564.30	2.75E-159	69%	75%	1--435	1--453
														2	AAR26897	FirrV-1-B22	80.88	9.24E-14	27%	55%	98--269	113--278
														3	NP_077588	EsV-1-103	72.02	4.29E-11	28%	51%	128--291	232--391
														4	YP_142803	unknown	61.62	5.80E-08	28%	47%	147--290	225--361
														5	YP_294142	hypothetical protein EhV 384	55.84	3.18E-06	23%	41%	64--258	125--309
C404L	170662--170054	203	23,713	8.35		No Hit Found								1	NP_048682	A326L	348.98	5.01E-95	78%	91%	1--201	1--207
														2	XP_647817	hypothetical protein DDB0218133	50.06	4.81E-05	27%	43%	4--175	10--193
C406L	171830--170754	359	42,621	9.88		No Hit Found								1	NP_048684	A328L	589.73	4.67E-167	76%	88%	1--359	1--355
tRNAs																						
Leu	172099--172183	85 bs					anticodon TAA															
Ile	172209--172281	73 bs					anticodon TAT															
Asn	172307--172378	72 bs					anticodon GTT															
Leu	172430--172513	84 bs					anticodon CAA															
Arg	172537--172608	72 bs					anticodon TCT															
Val	172708--172780	73 bs					anticodon AAC															
C413R	173286--174422	379	41,949	5.26	1	COG0677	WecC, UDP-N-acetyl-D-mannosaminuronate dehydrogenase [Cell envelope biogenesis, outer membrane].	209.32	3.93E-55	29%	48%	3--361	9--399	1	BAE48155	UDP-glucose dehydrogenase	726.86	0.00E+00	93%	97%	1--379	1--379
					2	COG1004	Ugd, Predicted UDP-glucose 6-dehydrogenase [Cell envelope biogenesis, outer membrane].	133.78	2.53E-32	27%	44%	5--351	2--373	2	EAA78333	hypothetical protein FG06548.1	231.88	2.66E-59	37%	57%	6--368	65--437
					3	pfam03721	UDPG_MGDP_dh_N, UDP-glucose/GDP-mannose dehydrogenase family, NAD binding domain. The UDP-glucose/GDP-mannose dehydrogenaseses are a small group of enzymes which possesses the ability to catalyse the NAD-dependent 2-fold oxidation of an alcohol to an acid without the release of an aldehyde intermediate.	81.43	1.50E-16	23%	44%	5--171	2--190	3	EAA72911	hypothetical protein FG03171.1	212.62	1.67E-53	36%	53%	3--356	41--409
					4	pfam00984	UDPG_MGDP_dh, UDP-glucose/GDP-mannose dehydrogenase family, central domain. The UDP-glucose/GDP-mannose dehydrogenaseses are a small group of enzymes which possesses the ability to catalyse the NAD-dependent 2-fold oxidation of an alcohol to an acid without the release of an aldehyde intermediate.	70.99	1.73E-13	30%	42%	182--256	2--76	4	BAE60327	unnamed protein product	205.68	2.04E-51	36%	54%	6--367	61--435
					5	COG2084	MmsB, 3-hydroxyisobutyrate dehydrogenase and related beta-hydroxyacid dehydrogenases [Lipid metabolism].	53.69	3.18E-08	24%	40%	4--237	1--231	5	EAA78448	hypothetical protein FG11489.1	199.90	1.12E-49	38%	55%	3--319	41--366
					6	pfam01408	GFO_IDH_MocA, Oxidoreductase family, NAD-binding Rossmann fold. This family of enzymes utilise NADP or NAD. This family is called the GFO/IDH/MOCA family..	41.01	2.13E-04	25%	42%	5--92	2--87	6	EAA70861	hypothetical protein FG04144.1	190.66	6.81E-47	37%	52%	6--339	79--417
					7	pfam03720	UDPG_MGDP_dh_C, UDP-glucose/GDP-mannose dehydrogenase family, UDP binding domain. The UDP-glucose/GDP-mannose dehydrogenaseses are a small group of enzymes which possesses the ability to catalyse the NAD-dependent 2-fold oxidation of an alcohol to an acid without the release of an aldehyde intermediate.	39.89	4.14E-04	29%	51%	293--361	2--74	7	AAM23919	UDP-N-acetyl-D-mannosaminuronate dehydrogenase	177.18	7.79E-43	32%	52%	6--378	25--427
					8	COG0039	Mdh, Malate/lactate dehydrogenases [Energy production and conversion].	37.46	2.58E-03	29%	42%	5--76	2--80	8	AAU22397	UDP-glucose 6-dehydrogenase	176.41	1.33E-42	31%	53%	6--361	11--393
					9	cd01339	LDH-like_MDH, LDH-like structure and DMH enzymatic activity; member of the family of NAD-dependent 2-hydroxycarboxylate dehydrogenases. Tetrameric Malate dehydrogenases (MDHs), including those from phototrophic bacteria, have a higher similarity to (Lactate dehydrogenases) LDHs than to other MDHs. LDH catalyzes the last step of glycolysis in which pyruvate is converted to L-lactate. MDH is one of the key enzymes in the citric acid cycle, facilitating both the conversion of malate to oxaloacetate and replenishing levels of oxalacetate by reductive carboxylation of pyruvate. L-2-hydroxyisocaproate dehydrogenases are also members of the family.	35.44	9.53E-03	31%	51%	5--71	1--73	9	BAD63054	UDP-N-acetyl-D-mannosaminuronate dehydrogenase	174.10	6.59E-42	31%	51%	4--369	15--410
														10	NP_782810	UDP-glucose 6-dehydrogenase	172.17	2.51E-41	30%	50%	3--364	17--408
C414R	174322--175203	294	32,301	9.17		No Hit Found								1	BAE48156	hypothetical methionyl-tRNA synthetase	466.85	3.34E-130	88%	93%	44--294	1--251
														2	NP_048562	PBCV-1 alginate lyase	127.10	6.27E-28	36%	51%	52--290	87--315
														3	BAA83789	alginate lyase	124.02	5.31E-27	35%	51%	52--290	99--327
														4	BAB19127	vAL-1	121.32	3.44E-26	35%	50%	52--290	115--343
														5	BAC87758	alginate lyase	80.88	5.16E-14	32%	44%	75--293	55--263
														6	BAE45131	alginase	80.88	5.16E-14	32%	44%	75--293	55--263
														7	EAL19065	hypothetical protein CNBH1670	68.94	2.03E-10	29%	46%	98--290	315--491
														8	AAW45360	expressed protein	68.94	2.03E-10	29%	46%	98--290	315--491
														9	EAL19679	hypothetical protein CNBG3070	59.31	1.61E-07	28%	40%	81--289	155--361
														10	AAW44587	hypothetical protein CNG01710	59.31	1.61E-07	28%	40%	81--289	145--351
C415L	176330--175212	373	42,614	4.5	1	pfam01522	Polysacc_deac_1, Polysaccharide deacetylase. This domain is found in polysaccharide deacetylase. This family of polysaccharide deacetylases includes NodB (nodulation protein B from Rhizobium) which is a chitooligosaccharide deacetylase. It also includes chitin deacetylase from yeast, and endoxylanases which hydrolyses glucosidic bonds in xylan..	84.56	1.55E-17	32%	44%	44--158	5--118	1	BAE48157	hypothetical chitooligosaccharide deacetylase	704.90	0.00E+00	89%	93%	6--373	2--369
														2	EAL33604	GA14716-PA	149.06	2.22E-34	30%	48%	25--320	230--529
														3	EAA12484	ENSANGP00000011077	146.36	1.44E-33	28%	46%	25--344	141--463
														4	AAF53561	CG17905-PA	145.98	1.88E-33	29%	47%	25--334	219--533
														5	CAE68839	Hypothetical protein CBG14800	138.66	3.00E-31	29%	46%	24--320	329--635
														6	XP_624655	PREDICTED: similar to CG31973-PA, isoform A	135.19	3.31E-30	29%	46%	25--317	478--765
														7	EAA06323	ENSANGP00000021951	134.42	5.65E-30	29%	48%	24--314	104--395
														8	AAF51568	CG31973-PB, isoform B	129.41	1.82E-28	26%	46%	25--320	477--2767
														9	AAF51567	CG31973-PA, isoform A	129.41	1.82E-28	26%	46%	25--320	683--973
														10	EAL34164	GA16591-PA	127.49	6.91E-28	26%	46%	25--320	661--951

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
C418R	176510--178090	527	60,621	9.18	1	pfam03142	Chitin_synth_2, Chitin synthase. Members of this family are fungal chitin synthase EC:2.4.1.16 enzymes. They catalyse chitin synthesis as follows: UDP-N-acetyl-D-glucosamine + {(1,4)-(N-acetyl-beta-D-glucosaminyl)}(N) <=> UDP + {(1,4)-(N-acetyl-beta-D-glucosaminyl)}(N+1)	74.18	2.38E-14	26%	42%	77--435	28--424	1	BAB83509	chitin synthase	922.54	0.00E+00	91%	94%	18--507	7--496
					2	COG1215	COG1215, Glycosyltransferases, probably involved in cell wall biogenesis [Cell envelope biogenesis, outer membrane].	44.54	1.70E-05	20%	36%	23--522	4--435	2	BAE48158	chitin synthase	361.69	3.48E-98	86%	89%	1--210	1--210
														3	BAE48153	chitin synthase	280.03	1.33E-73	34%	53%	4--507	8--500
														4	EAA78335	hypothetical protein FG06550.1	165.62	3.66E-39	27%	45%	4--507	131--647
														5	EAA68628	hypothetical protein FG10619.1	153.30	1.88E-35	26%	44%	22--507	192--682
														6	BAE60326	unnamed protein product	134.42	9.04E-30	25%	41%	13--507	117--578
														7	NP_077569	ESV-1-84	105.92	3.44E-21	23%	42%	2--507	1--482
														8	XP_503779	hypothetical protein	72.79	3.23E-11	28%	44%	218--436	951--1160
														9	EAA72910	hypothetical protein FG03170.1	69.32	3.57E-10	22%	41%	76--507	255--674
														10	AAS52496	AEL189Wp	67.01	1.77E-09	23%	41%	172--507	703--1033
C423L	179281--178118	388	43,104	8.44	1	COG3147	DedD, Uncharacterized protein conserved in bacteria [Function unknown].	48.10	1.62E-06	29%	37%	180--272	46--140	1	NP_048689	PLPRNLLL (4X), SPPPSKP (3X)	417.93	2.72E-115	85%	92%	1--213	1--213
					2	pfam03346	Lipoprotein_14, Actinobacillus constitutively-expressed outer membrane lipoprotein A..	46.29	5.00E-06	32%	48%	212--273	59--121	2	NP_048688	a332L	218.78	2.42E-55	88%	93%	280--388	1--109
					3	pfam03276	Gag_spuma, Spumavirus gag protein..	46.23	6.13E-06	32%	39%	228--340	177--290	3	EAA01148	ENSANGP00000018413	78.95	2.99E-13	31%	43%	23--210	9--202
					4	COG4982	COG4982, 3-oxoacyl-[acyl-carrier protein].	41.94	1.15E-04	31%	49%	226--271	10--55	4	AAM50982	RE24790p	64.31	7.62E-09	26%	41%	1--212	36--250
					5	pfam05743	Tsg101, Tumour susceptibility gene 101 protein (TSG101). This family consists of the eukaryotic tumour susceptibility gene 101 protein (TSG101). Altered transcripts of this gene have been detected in sporadic breast cancers and many other human malignancies. However, the involvement of this gene in neoplastic transformation and tumourigenesis is still elusive. TSG101 is required for normal cell function of embryonic and adult tissues but that this gene is not a tumour suppressor for sporadic forms of breast cancer	40.09	3.45E-04	34%	42%	214--288	138--221	5	AAF46012	CG15786-PA	64.31	7.62E-09	26%	41%	1--212	17--231
					6	pfam05518	Totivirus_coat, Totivirus coat protein..	40.35	3.58E-04	24%	27%	217--272	696--751	6	EAA06469	ENSANGP00000012390	63.16	1.70E-08	27%	41%	26--212	1--197
					7	smart00494	ChtBD2, Chitin-binding domain type 2; .	39.34	6.02E-04	46%	54%	281--321	12--49	7	EAL32472	GA13958-PA	59.31	2.45E-07	25%	41%	23--212	11--210
					8	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	38.97	7.45E-04	25%	32%	213--278	50--115	8	EAL29007	GA18137-PA	54.68	6.04E-06	25%	40%	23--210	4--195
					9	pfam01607	CBM_14, Chitin binding Peritrophin-A domain. This domain is called the Peritrophin-A domain and is found in chitin binding proteins particularly peritrophic matrix proteins of insects and animal chitinases. Copies of the domain are also found in some baculoviruses. Relevant references that describe proteins with this domain include. It is an extracellular domain that contains six conserved cysteines that probably form three disulphide bridges. Chitin binding has been demonstrated for a protein containing only two of these domains	36.95	3.29E-03	38%	66%	290--321	18--47	9	EAL29006	GA18133-PA	53.53	1.35E-05	26%	40%	23--209	18--210
					10	COG5373	COG5373, Predicted membrane protein [Function unknown].	36.49	4.14E-03	25%	38%	205--272	51--120	10	EAA00829	ENSANGP00000011567	52.76	2.30E-05	26%	37%	22--209	16--215
C427L	179982--179305	226	26,408	10.12		No Hit Found								1	NP_048695	A339L	108.23	1.86E-22	41%	51%	1--135	1--156
														2	NP_048693	A337L	100.52	3.88E-20	63%	79%	150--226	1--77
C429L	180503--180087	139	15,937	4.06		No Hit Found								1	NP_048697	A341L	218.39	4.85E-56	78%	87%	8--139	4--135
C430L	182241--180586	552	60,603	8.88		No Hit Found								1	NP_048699	A342L	967.61	0.00E+00	87%	91%	3--538	23--559
C434R	182289--182657	123	15,104	10.92		No Hit Found									No Hit Found	No Hit Found						
C435R	182657--183202	182	20,806	9.65		No Hit Found								1	NP_048705	A348R	216.47	2.98E-55	71%	80%	23--179	1--156
C437L	183514--183218	99	11,367	10.16		No Hit Found								1	BAA22202	URF14.2	170.63	1.18E-41	89%	89%	4--99	16--111
														2	BAA22201	URF14.2	164.47	8.47E-40	84%	88%	4--99	16--111
														3	BAA22199	URF14.2	163.31	1.89E-39	83%	88%	4--99	16--111
														4	BAA22200	URF14.2	161.77	5.49E-39	83%	87%	4--99	16--111
														5	NP_048706	A349L	96.29	2.83E-19	93%	100%	4--50	87--133
C438L	184739--183990	250	28,312	10.37		No Hit Found								1	NP_048708	KKD (6X), mixed charge	85.50	1.58E-15	43%	64%	7--106	26--123
														2	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	61.62	2.44E-08	27%	41%	8--249	3--224
														3	NP_048671	A315L	57.00	6.01E-07	26%	39%	8--247	3--239
														4	YP_293795	putative endonuclease	55.45	1.75E-06	36%	52%	9--101	5--94
														5	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	51.60	2.53E-05	35%	51%	9--103	4--94
C439R	184925--185290	122	14,877	6.09		No Hit Found								1	NP_048707	a350R	174.48	8.04E-43	62%	76%	1--122	1--122
														2	CAG59956	unnamed protein product	48.14	8.70E-05	25%	38%	1--121	6--173
C440L	186309--185275	345	39,656	10.09	1	pfam01541	GIY-YIG, GIY-YIG catalytic domain. This domain called GIY-YIG is found in the amino terminal region of excinuclease abc subunit c (uvrC), bacteriophage T4 endonucleases segA, segB, segC, segD and segE; it is also found in putative endonucleases encoded by group I introns of fungi and phage. The structure of I-TevI a GIY-YIG endonuclease, reveals a novel alpha/beta-fold with a central three-stranded antiparallel beta-sheet flanked by three helices. The most conserved and putative catalytic residues are located on a shallow, concave surface and include a metal coordination site	40.92	1.94E-04	33%	43%	15--96	4--85	1	NP_048708	KKD (6X), mixed charge	444.89	1.74E-123	65%	77%	6--345	19--358
					2	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	36.21	5.37E-03	35%	47%	15--104	4--83	2	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	58.54	3.52E-07	38%	52%	14--116	3--103
														3	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	53.91	8.66E-06	34%	51%	15--120	4--105
C444L	187045--186425	207	23,352	3.66		No Hit Found								1	NP_048709	Asp/Glu rich; DAEDDDIYxxET (2X) negative charge cluster	359.38	3.89E-98	85%	87%	1--207	1--207
														2	YP_142843	unknown	52.76	7.79E-06	22%	41%	6--207	8--221
C447L	188051--187113	313	35,630	4.25		No Hit Found								1	NP_048714	A357L	364.00	3.36E-99	72%	84%	76--313	25--276
														2	NP_048716	a359L	54.30	5.70E-06	69%	79%	1--40	1--41
C449R	188110--191784	1225	135,910	10.97	1	pfam05110	AF-4, AF-4 proto-oncoprotein. This family consists of AF4 (Proto-oncogene AF4) and FMR2 (Fragile X E mental retardation syndrome) nuclear proteins. These proteins have been linked to human diseases such as acute lymphoblastic leukaemia and mental retardation. The family also contains a Drosophila AF4 protein homologue Lilliputian which contains an AT-hook domain. Lilliputian represents a novel pair-rule gene that acts in cytoskeleton regulation, segmentation and morphogenesis in Drosophila	40.15	3.55E-04	22%	42%	1045--1169	842--960	1	NP_048720	similar to chicken vitellogenin II, corresponds to Swiss-Prot Accession Number P02845	422.55	4.43E-116	37%	55%	411--1069	2--666
					2	pfam04484	DUF566, Family of unknown function (DUF566). Family of related proteins that is plant specific..	38.53	1.15E-03	24%	38%	1046--1194	54--194	2	NP_048717	A360R	57.77	2.87E-06	24%	37%	39--275	12--227
C459L	192186--191791	132	15,166	4.89		No Hit Found								1	NP_048723	A366L	67.40	1.42E-10	30%	47%	9--124	2--132
C462L	192950--192234	239	27,187	9.65		No Hit Found								1	NP_048723	A366L	291.58	1.34E-77	59%	76%	6--239	14--255
														2	NP_048427	A79R	57.38	4.25E-07	28%	42%	1--165	1--166
														3	NP_048525	A177R	55.45	1.62E-06	32%	47%	1--137	4--133
														4	NP_048629	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	54.68	2.76E-06	26%	42%	1--230	4--249
														5	NP_049005	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	50.83	3.98E-05	27%	39%	4--230	10--253
C466L	194279--193503	259	29,252	9.58		No Hit Found								1	NP_048735	A378L	354.37	1.95E-96	74%	76%	19--243	21--245
														2	BAA11342	DNA binding protein	83.96	4.93E-15	35%	44%	137--248	322--402
														3	NP_048921	A565R	83.57	6.43E-15	35%	44%	137--248	387--467
														4	NP_048917	similar to Chlorella virus CVK2 DNA binding protein, corresponds to GenBank Accession Number D78305	83.57	6.43E-15	35%	44%	137--248	324--404
														5	BAA11343	DNA binding protein	83.19	8.40E-15	35%	44%	137--248	379--459
														6	CAA64974	QI74 protein	76.26	1.03E-12	35%	45%	120--257	206--326

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
														7	NP_048741	Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	51.99	2.07E-05	50%	61%	206--248	85--128
C467L	194936--194310	209	24,566	10.53		No Hit Found								1	NP_048736	A379L	379.02	4.86E-104	86%	93%	1--209	1--207
C470R	195079--196533	485	55,647	6.13	1	pfam04451	Capsid_Iridovir, Iridovirus major capsid protein. This family includes the major capsid protein of iridoviruses, chlorella virus and Spodoptera ascovirus, which are all dsDNA viruses with no RNA stage. This is the most abundant structural protein and can account for up to 45% of virion protein. In Chlorella virus AR158 the major capsid protein is a glycoprotein	177.81	1.28E-45	30%	47%	95--445	3--387	1	NP_048740	similar to PBCV-1 major capsid protein, corresponds to Swiss-Prot Accession Number P30328	481.49	2.71E-134	89%	95%	162--418	1--257
														2	NP_048739	cytochrome C family heme-binding site signature	190.27	1.25E-46	63%	73%	1--147	1--147
														3	BAA22198	major capsid protein Vp54	152.53	2.89E-35	31%	44%	95--443	3--372
														4	BAA76601	major capsid protein MCP1	149.83	1.87E-34	30%	44%	95--443	3--372
														5	BAA76600	major capsid protein	148.67	4.17E-34	31%	45%	95--443	3--371
														6	BAE06835	hypothetical major capsid protein	148.67	4.17E-34	27%	46%	95--484	3--438
														7	NP_048787	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	148.29	5.45E-34	30%	44%	95--443	3--372
														8	AAC27492	major capsid protein Vp49	139.43	2.53E-31	29%	44%	95--443	3--363
														9	NP_048741	Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	125.56	3.78E-27	75%	85%	408--485	379--456
														10	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	122.87	2.45E-26	28%	42%	117--443	1--348
C475L	198676--196817	620	66,775	7.82	1	pfam05887	Trypan_PARP, Procyclic acidic repetitive protein (PARP). This family consists of several Trypanosoma brucei procyclic acidic repetitive protein (PARP) like sequences. The procyclic acidic repetitive protein (parp) genes of Trypanosoma brucei encode a small family of abundant surface proteins whose expression is restricted to the procyclic form of the parasite. They are found at two unlinked loci, parpA and parpB; transcription of both loci is developmentally regulated	46.90	3.67E-06	30%	52%	456--500	64--108	1	NP_048741	Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	451.83	3.14E-125	80%	84%	198--455	8--265
					2	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	46.68	4.46E-06	52%	54%	456--501	76--122	2	BAA11343	DNA binding protein	166.78	2.02E-39	39%	47%	159--364	244--458
					3	pfam05616	Neisseria_TspB, Neisseria meningitidis TspB protein. This family consists of several Neisseria meningitidis TspB virulence factor proteins..	45.89	6.08E-06	27%	34%	452--518	331--402	3	NP_048921	A565R	162.54	3.81E-38	37%	45%	159--364	244--466
					4	COG3147	DedD, Uncharacterized protein conserved in bacteria [Function unknown].	45.79	7.52E-06	33%	38%	442--500	87--145	4	NP_048747	Asn-rich	113.62	2.03E-23	67%	69%	56--140	52--134
					5	pfam01213	CAP, CAP protein..	44.61	1.87E-05	28%	36%	371--476	230--326	5	NP_048917	similar to Chlorella virus CVK2 DNA binding protein, corresponds to GenBank Accession Number D78305	109.38	3.83E-22	47%	57%	264--364	304--403
					6	pfam04625	DEC-1_N, DEC-1 protein, N terminal region. The defective chorion-1 gene (dec-1) in Drosophila encodes follicle cell proteins necessary for proper eggshell assembly. Multiple products of the dec-1 gene are formed by alternative RNA splicing and proteolytic processing. Cleavage products include S80 (80 kDa) which is incorporated into the eggshell, and further proteolysis of S80 gives S60 (60 kDa)	43.33	4.64E-05	39%	39%	456--500	98--142	6	BAA11342	DNA binding protein	108.61	6.53E-22	47%	57%	264--364	302--401
					7	COG0341	SecF, Preprotein translocase subunit SecF [Intracellular trafficking and secretion].	40.95	1.93E-04	24%	43%	12--110	65--156	7	NP_048688	a332L	71.25	1.15E-10	64%	78%	512--553	2--43
					8	pfam01299	Lamp, Lysosome-associated membrane glycoprotein (Lamp)..	38.03	1.50E-03	40%	44%	435--489	121--176	8	BAE02830	surface protein	56.61	2.94E-06	43%	62%	396--453	894--951
					9	COG3087	FtsN, Cell division protein [Cell division and chromosome partitioning].	38.10	1.51E-03	44%	49%	449--487	141--182	9	CAA64974	QI74 protein	55.84	5.02E-06	24%	34%	110--354	194--437
C479R	198745--199506	254	29,184	9.73	1	smart00382	AAA, ATPases associated with a variety of cellular activities; AAA - ATPases associated with a variety of cellular activities. This profile/alignment only detects a fraction of this vast family. The poorly conserved N-terminal helix is missing from the alignment. .	44.74	1.46E-05	17%	40%	16--154	2--146	1	NP_048749	contains ATP/GTP-binding site motif A; similar to frog virus 3 ATPase, corresponds to GenBank Accession Number M80551	377.10	2.71E-103	75%	78%	1--253	1--253
					2	COG4857	COG4857, Predicted kinase [General function prediction only].	36.50	4.51E-03	33%	49%	4--72	29--107	2	BAB69884	AGB-1	149.44	9.19E-35	40%	60%	33--229	39--238
														3	AAR26836	FirrV-1-A12	126.33	8.33E-28	34%	53%	33--230	37--243
														4	NP_077511	EsV-1-26	120.17	5.97E-26	32%	53%	33--230	48--254
														5	YP_293826	putative DNA-binding protein	114.39	3.28E-24	34%	54%	33--214	35--215
														6	NP_149538	075L	114.01	4.28E-24	30%	53%	33--230	46--246
														7	AAS18149	ATPase	107.46	4.00E-22	29%	54%	33--229	47--245
														8	AAV91100	ATPase-like protein	107.46	4.00E-22	29%	54%	33--229	47--245
														9	YP_003858	ATPase	106.69	6.83E-22	29%	53%	33--230	44--243
														10	AAT09674	AAA-ATPase	106.30	8.92E-22	29%	53%	33--230	44--243
C483R	199727--200104	126	14,884	3.81		No Hit Found								1	NP_048751	A394R	179.10	3.33E-44	65%	86%	5--124	2--121
C484R	200163--200408	82	9,445	10.31		No Hit Found								1	NP_048752	A395R	140.58	1.29E-32	82%	90%	1--82	1--82
C485L	200997--200548	150	18,221	6.81		No Hit Found								1	NP_048753	A396L	94.74	8.23E-19	69%	81%	83--149	1--69
C486R	201239--202132	298	34,691	10.38		No Hit Found									No Hit Found	No Hit Found						
C487L	202901--202095	269	31,138	8.32	1	smart00650	rADc, Ribosomal RNA adenine dimethylases; .	47.09	3.09E-06	18%	42%	14--115	1--115	1	AAC03123	DNA adenine methyltransferase	543.12	3.15E-153	98%	99%	1--269	1--267
					2	pfam00398	RrnaAD, Ribosomal RNA adenine dimethylase..	44.46	1.86E-05	22%	44%	14--117	18--134	2	ZP_00579458	hypothetical protein SalaDRAFT_0836	127.49	4.13E-28	31%	54%	1--256	12--258
					3	COG0030	KsgA, Dimethyladenosine transferase (rRNA methylation) [Translation, ribosomal structure and biogenesis].	41.77	1.06E-04	20%	41%	14--107	18--126	3	AAT27581	putative type II DNA methylase protein	108.61	1.98E-22	39%	54%	1--157	6--176
														4	NP_223729	putative TYPE II DNA MODIFICATION ENZYME (METHYLTRANSFERASE)	104.76	2.87E-21	36%	53%	1--168	27--198
														5	ZP_00372070	putative type II DNA modification enzyme (methyltransferase)	90.51	5.59E-17	34%	54%	2--160	66--230
														6	AAD07438	predicted coding region HP0369	70.09	7.82E-11	41%	56%	69--168	14--114
														7	AAF87840	modification methyltransferase Hpy8I	56.23	1.17E-06	34%	51%	30--124	32--146
														8	AAD07955	adenine specific DNA methyltransferase (HINDIIM)	52.76	1.29E-05	31%	52%	30--124	32--146
														9	NP_223564	TYPE II DNA MODIFICATION ENZYME (METHYLTRANSFERASE)	50.83	4.91E-05	31%	52%	30--124	36--150
														10	AAX44414	hypothetical protein PSSM2_036	50.06	8.38E-05	30%	47%	81--165	7--86
C490L	203474--203121	118	13,001	10.33		No Hit Found								1	NP_048755	A398L	229.18	2.80E-59	98%	100%	1--118	1--118
C491R	203540--204124	195	22,176	8.75	1	pfam00075	RnaseH, RNase H. RNase H digests the RNA strand of an RNA/DNA hybrid. Important enzyme in retroviral replication cycle, and often found as a domain associated with reverse transcriptases. Structure is a mixed alpha+beta fold with three a/b/a layers..	73.40	3.87E-14	30%	39%	27--158	1--125	1	NP_048756	A399R	276.56	2.89E-73	71%	80%	2--192	3--194
C492R	204131--204508	126	14,558	9.15		No Hit Found								1	NP_048757	similar to PBCV-1 ORF A214, corresponds to GenBank Accession Number U42580	189.50	2.46E-47	79%	92%	6--122	1--117
														2	NP_048561	A214L	49.68	3.04E-05	26%	54%	13--124	3--119
C493R	204533--205366	278	31,799	6.07		No Hit Found								1	NP_048758	A401R	486.88	2.85E-136	86%	91%	4--278	2--277
														2	BAC51116	bll5851	176.02	1.08E-42	39%	56%	36--276	3--249
														3	ZP_00863599	conserved hypothetical protein	166.78	6.53E-40	36%	56%	36--276	7--253
														4	YP_190685	hypothetical protein GOX0246	140.58	5.01E-32	30%	54%	40--275	11--253
														5	ZP_00577151	conserved hypothetical protein	101.29	3.37E-20	27%	50%	40--274	14--260
														6	YP_162362	hypothetical protein ZMO0627	57.77	4.27E-07	24%	42%	38--275	9--207
														7	YP_162363	hypothetical protein ZMO0628	56.23	1.24E-06	23%	42%	64--269	13--204
C494R	205515--206210	232	26,029	7.97		No Hit Found								1	NP_048759	A402R	403.29	2.96E-111	86%	94%	6--232	1--227
C496R	206248--206538	97	11,465	9.68		No Hit Found								1	NP_048760	A403R	172.56	3.13E-42	91%	96%	7--97	3--93
C497R	206574--207131	186	21,728	3.07	1	COG5271	MDN1, AAA ATPase containing von Willebrand factor type A (vWA) domain [General function prediction only].	40.47	3.20E-04	27%	51%	38--157	950--4068	1	NP_048761	Glu-, Asn-rich	117.09	2.62E-25	91%	98%	3--61	6--64
C498R	207183--208850	556	59,600	10.52	1	pfam05616	Neisseria_TspB, Neisseria meningitidis TspB protein. This family consists of several Neisseria meningitidis TspB virulence factor proteins..	50.90	2.17E-07	31%	38%	165--239	323--397	1	NP_048762	Pro-, Lys-rich, PAPK (30x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	444.51	4.39E-123	76%	87%	282--556	218--496
					2	COG5373	COG5373, Predicted membrane protein [Function unknown].	49.20	7.67E-07	31%	40%	179--264	43--124	2	BAD12236	surface protein	90.51	1.61E-16	69%	80%	2--63	506--567
					3	pfam03276	Gag_spuma, Spumavirus gag protein..	48.93	9.34E-07	32%	42%	188--289	180--278	3	NP_048488	PBCV-1 surface protein	89.35	3.58E-16	66%	79%	1--63	505--567
					4	pfam05518	Totivirus_coat, Totivirus coat protein..	47.67	1.84E-06	27%	33%	87--212	631--751	4	BAD22850	surface protein	89.35	3.58E-16	66%	79%	1--63	505--567

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					5	pfam02993	MCPVI, Minor capsid protein VI. This minor capsid protein may act as a link between the external capsid and the internal DNA-protein core. The C-terminal 11 residues may function as a protease cofactor leading to enzyme activation..	47.33	2.43E-06	31%	40%	75--171	111--206	5	T17636	proline-rich protein A145R - Chlorella virus PBCV-1	65.86	4.24E-09	52%	65%	3--63	37--97
					6	pfam06070	Herpes_UL32, Herpesvirus large structural phosphoprotein UL32. The large phosphorylated protein (UL32-like) of herpes viruses is the polypeptide most frequently reactive in immuno-blotting analyses with antisera when compared with other viral proteins..	46.29	4.72E-06	17%	27%	64--239	637--812	6	BAE02830	surface protein	52.37	4.85E-05	38%	61%	5--63	571--629
C503L	209601--208975	209	23,492	8.33		No Hit Found								1	NP_048764	A407L	331.26	1.16E-89	77%	82%	1--209	1--210
C504L	210698--209643	352	41,381	10.22		No Hit Found								1	NP_048711	A354R	298.52	2.08E-79	61%	73%	111--350	2--233
														2	NP_048779	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	68.17	4.58E-10	25%	44%	19--321	18--319
														3	AAL73476	endonuclease	63.16	1.47E-08	29%	46%	5--165	9--169
														4	AAL73479	endonuclease	60.85	7.32E-08	27%	54%	16--139	15--134
														5	NP_047162	putative HNH homing endonuclease	58.15	4.74E-07	32%	55%	14--109	11--104
														6	AAL73456	endonuclease	54.30	6.85E-06	29%	48%	14--122	84--195
														7	YP_240848	ORF027	53.53	1.17E-05	29%	44%	174--350	3--179
														8	YP_142601	HNH endonuclease	52.76	1.99E-05	23%	39%	14--350	19--361
C507L	211593--210838	252	28,231	4.67		No Hit Found								1	NP_048765	A408L	348.98	7.78E-95	82%	90%	2--214	19--233
														2	NP_077527	EsV-1-42	80.88	3.95E-14	30%	51%	21--167	6--164
														3	NP_048767	A410L	77.80	3.34E-13	37%	59%	22--120	8--107
														4	AAR26867	FirrV-1-A43	70.86	4.09E-11	26%	55%	17--155	1--144
C509L	211970--211596	125	14,949	4.61		No Hit Found								1	NP_048767	A410L	207.99	6.71E-53	90%	95%	16--125	1--110
														2	NP_048765	A408L	80.88	1.24E-14	37%	58%	21--122	39--139
														3	NP_077527	EsV-1-42	50.83	1.37E-05	29%	46%	20--121	4--110
C510R	212061--212588	176	20,083	6.78		No Hit Found								1	NP_048769	A412R	315.46	4.35E-85	83%	91%	1--176	1--179
C512R	212616--213671	352	39,640	8.32	1	cd00315	Cyt_C5_DNA_methylase, Cytosine-C5 specific DNA methylases; Methyl transfer reactions play an important role in many aspects of biology. Cytosine-specific DNA methylases are found both in prokaryotes and eukaryotes. DNA methylation, or the covalent addition of a methyl group to cytosine within the context of the CpG dinucleotide, has profound effects on the mammalian genome. These effects include transcriptional repression via inhibition of transcription factor binding or the recruitment of methyl-binding proteins and their associated chromatin remodeling factors, X chromosome inactivation, imprinting and the suppression of parasitic DNA sequences. DNA methylation is also essential for proper embryonic development and is an important player in both DNA repair and genome stability	205.54	6.66E-54	33%	52%	5--333	1--314	1	AAP07928	Type II restriction-modification system methylation subunit	151.75	3.15E-35	34%	50%	5--334	18--350
					2	COG0270	Dcm, Site-specific DNA methylase [DNA replication, recombination, and repair].	197.99	1.28E-51	33%	50%	3--336	2--328	2	CAD33713	putative DNA methylase	149.06	2.04E-34	34%	53%	5--321	1--292
					3	pfam00145	DNA_methylase, C-5 cytosine-specific DNA methylase..	195.91	5.03E-51	32%	50%	5--333	1--323	3	AAC97192	modification methylase M.NspHI	146.36	1.32E-33	34%	49%	8--321	63--359
														4	AAC97190	modification methylase M.NspI	145.21	2.95E-33	33%	50%	8--321	63--359
														5	YP_208922	putative 5-methylcytosine methyltransferase	144.44	5.03E-33	31%	49%	1--333	1--328
														6	AAX14650	BbvCI methyltransferase 1	142.90	1.46E-32	32%	47%	3--323	17--402
														7	CAA74996	Bpu10I (5m)cytosine-specific DNA modification methyltransferase (C1)	142.51	1.91E-32	29%	46%	3--337	6--397
														8	CAA68505	DdeI methylase	139.04	2.11E-31	30%	45%	5--322	1--357
														9	AAT40769	putative DNA methylase	137.12	8.03E-31	34%	50%	1--316	1--293
														10	AAF89681	cytosine-specific methyltransferase	135.58	2.34E-30	29%	49%	8--322	6--350
C513L	214715--213609	369	42,117	8.4	1	COG4123	COG4123, Predicted O-methyltransferase [General function prediction only].	59.10	7.90E-10	26%	38%	43--154	45--171	1	AAC57943	DNA adenine methyltransferase	642.88	0.00E+00	85%	93%	1--368	1--368
					2	COG2890	HemK, Methylase of polypeptide chain release factors [Translation, ribosomal structure and biogenesis].	53.84	2.65E-08	26%	45%	31--114	99--187	2	AAC57945	DNA adenine methyltransferase	209.92	1.05E-52	34%	53%	7--367	6--357
					3	COG2813	RsmC, 16S RNA G1207 methylase RsmC [Translation, ribosomal structure and biogenesis].	53.73	2.81E-08	30%	46%	45--152	161--265	3	P52284	Modification methylase CviRI (Adenine-specific methyltransferase CviRI) (M.CviRI)	209.15	1.79E-52	34%	54%	3--368	4--378
					4	COG0286	HsdM, Type I restriction-modification system methyltransferase subunit [Defense mechanisms].	51.18	1.84E-07	19%	35%	18--283	161--464	4	AAC03124	DNA adenine methyltransferase	209.15	1.79E-52	33%	55%	1--367	1--368
					5	pfam01170	UPF0020, Putative RNA methylase family UPF0020. This domain is probably a methylase. It is associated with the THUMP domain that also occurs with RNA modification domains..	43.33	4.22E-05	23%	43%	44--113	30--107	5	AAC03125	DNA adenine methyltransferase	196.44	1.20E-48	31%	54%	3--369	4--382
					6	COG2263	COG2263, Predicted RNA methylase [Translation, ribosomal structure and biogenesis].	39.83	4.03E-04	23%	46%	41--117	44--123	6	CAA29835	unnamed protein product	194.90	3.49E-48	32%	54%	11--368	14--377
					7	COG0421	SpeE, Spermidine synthase [Amino acid transport and metabolism].	39.18	6.70E-04	25%	44%	35--151	69--188	7	YP_392645	N-6 DNA methylase	71.25	5.83E-11	30%	46%	12--248	19--263
					8	COG1041	COG1041, Predicted DNA modification methylase [DNA replication, recombination, and repair].	37.63	2.15E-03	25%	38%	44--154	199--311	8	AAC60387	methyltransferase; M-AccI	70.09	1.30E-10	30%	48%	12--200	19--227
														9	AAA50500	AccI methylase	70.09	1.30E-10	30%	48%	12--200	19--227
														10	JU0470	site-specific DNA-methyltransferase (adenine-specific) (EC 2.1.1.72) type II - Acinetobacter calcoaceticus	70.09	1.30E-10	30%	48%	12--200	19--227
C515L	215523--214762	254	28,376	9.92	1	COG5028	COG5028, Vesicle coat complex COPII, subunit SEC24/subunit SFB2/subunit SFB3 [Intracellular trafficking and secretion].	38.03	1.57E-03	19%	33%	57--155	22--121	1	NP_048770	Gln-rich, QQQQM(4x); similar to human transcription factor TFIID, corresponds to Swiss-Prot Accession Number P20226	234.57	2.17E-60	53%	62%	1--254	1--244
C518R	215600--215815	72	7,973	10.63		No Hit Found								1	NP_048771	A414R	135.58	4.14E-31	90%	98%	1--72	22--93
C519L	217117--215822	432	49,842	6.77		No Hit Found								1	NP_048774	A417L	678.32	0.00E+00	78%	87%	1--423	1--424
														2	XP_729961	rhoptry protein	53.14	2.04E-05	22%	44%	68--387	326--629
														3	NP_473107	replication factor C subunit 1, putative	52.37	3.48E-05	22%	40%	55--352	412--741
														4	AAG43050	replication factor C subunit 1	52.37	3.48E-05	22%	40%	55--352	412--741
														5	XP_724804	replication factor C, 140 kDa subunit	51.22	7.76E-05	20%	40%	55--352	397--726
C521L	217375--217166	70	7,948	6.34		No Hit Found								1	NP_048777	A420L	119.78	2.36E-26	81%	94%	1--70	1--70
C522R	217421--217711	97	11,209	10.34		No Hit Found								1	NP_048778	A421R	136.73	1.90E-31	65%	78%	7--97	8--98
C523R	217951--218406	152	17,955	4.94		No Hit Found								1	NP_048780	A423R	231.11	7.29E-60	73%	83%	1--152	1--157
C527R	218762--219109	116	13,389	5.03		No Hit Found								1	NP_048783	A426R	186.42	2.09E-46	78%	88%	3--116	1--114
C528L	219468--219112	119	13,736	6.49	1	COG3118	COG3118, Thioredoxin domain-containing protein [Posttranslational modification, protein turnover, chaperones].	39.12	6.83E-04	18%	37%	6--109	22--130	1	NP_048784	contains thioredoxin active site-like sequence; similar to Synechocystis thioredoxin-like protein, corresponds to Swiss-Prot Accession Number P52232	196.05	2.61E-49	76%	87%	1--118	1--118
														2	ZP_00533498	Thioredoxin	50.45	1.77E-05	28%	53%	22--109	7--98
														3	EAL29786	GA21460-PA	50.45	1.77E-05	24%	48%	9--111	34--140
														4	AAF47638	CG8993-PA	48.14	8.80E-05	25%	51%	9--100	34--129
C529L	219825--219535	97	10,686	3.86		No Hit Found									No Hit Found	No Hit Found						
C531L	221216--219858	453	53,232	5.38	1	cd00204	ANK, ankyrin repeats; ankyrin repeats mediate protein-protein interactions in very diverse families of proteins. The number of ANK repeats in a protein can range from 2 to over 20 (ankyrins, for example). ANK repeats may occur in combinations with other types of domains. The structural repeat unit contains two antiparallel helices and a beta-hairpin, repeats are stacked in a superhelical arrangement; this alignment contains 4 consecutive repeats.	43.53	3.55E-05	28%	45%	140--250	11--118	1	NP_048786	A429L	739.95	0.00E+00	76%	87%	1--452	24--475
														2	NP_192255	protein binding	73.94	1.20E-11	24%	40%	34--364	128--472
														3	BAD29430	ankyrin-like protein	62.77	2.76E-08	23%	40%	37--353	138--470
														4	BAD28737	ankyrin repeat family protein-like	53.91	1.28E-05	31%	50%	172--290	274--392
														5	CAG31465	hypothetical protein	51.99	4.87E-05	24%	42%	16--241	772--996

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
C532L	222655--221345	437	48,357	7.43	1	pfam04451	Capsid_Iridovir, Iridovirus major capsid protein. This family includes the major capsid protein of iridoviruses, chlorella virus and Spodoptera ascovirus, which are all dsDNA viruses with no RNA stage. This is the most abundant structural protein and can account for up to 45% of virion protein. In Chlorella virus AR158 the major capsid protein is a glycoprotein	477.50	8.12E-136	52%	65%	1--433	1--443	1	BAA76601	major capsid protein MCP1	800.82	0.00E+00	90%	91%	1--437	1--437
														2	NP_048787	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	794.65	0.00E+00	89%	91%	1--437	1--437
														3	BAA22198	major capsid protein Vp54	785.41	0.00E+00	87%	90%	1--437	1--437
														4	BAA76600	major capsid protein	748.81	0.00E+00	84%	87%	1--437	1--436
														5	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	747.66	0.00E+00	88%	90%	25--437	1--413
														6	1M3Y_D	Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid Containing, Dna Virus	747.66	0.00E+00	88%	90%	25--437	1--413
														7	AAC27492	major capsid protein Vp49	636.34	0.00E+00	73%	80%	1--437	1--432
														8	BAE06835	hypothetical major capsid protein	327.41	5.70E-88	42%	55%	1--437	1--440
														9	NP_048359	contains aminoacyl-tRNA synthetase class-II signature	245.36	2.85E-63	36%	51%	1--437	1--403
														10	NP_048358	similar to PBCV-1 major capsid protein, corresponds to Swiss-Prot Accession Number P30328	231.49	4.26E-59	33%	54%	1--435	2--399
C535R	222788--223228	147	16,167	8.45		No Hit Found								1	NP_048789	Lys-rich	186.42	2.10E-46	75%	86%	1--116	1--116
C537L	223771--223460	104	11,280	11.05		No Hit Found								1	NP_048794	similar to Methanothrix chromosomal protein MC1A, corresponds to Swiss-Prot Accession Number P15251	133.27	2.05E-30	65%	72%	1--103	1--103
C538L	224034--223801	78	9,179	10.77	1	COG0695	GrxC, Glutaredoxin and related proteins [Posttranslational modification, protein turnover, chaperones].	45.34	1.11E-05	27%	45%	1--78	1--76	1	NP_048795	similar to E. coli glutaredoxin, corresponds to Swiss-Prot Accession Number P37687	134.04	1.23E-30	80%	91%	1--78	1--78
C539R	224057--224392	112	13,181	7.92		No Hit Found								1	NP_048796	A439R	173.33	1.80E-42	66%	84%	1--112	1--112
C540L	225573--224395	393	46,328	8.96		No Hit Found								1	NP_048711	A354R	189.50	1.60E-46	44%	61%	159--391	6--233
														2	NP_048779	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	63.16	1.73E-08	25%	45%	63--362	30--319
														3	NP_048435	A87R	57.77	7.27E-07	23%	43%	122--390	161--447
C542R	225618--225872	85	9,980	8.13	1	pfam04451	Capsid_Iridovir, Iridovirus major capsid protein. This family includes the major capsid protein of iridoviruses, chlorella virus and Spodoptera ascovirus, which are all dsDNA viruses with no RNA stage. This is the most abundant structural protein and can account for up to 45% of virion protein. In Chlorella virus AR158 the major capsid protein is a glycoprotein	36.06	7.08E-03	48%	58%	25--58	1--34		No Hit Found	No Hit Found						
C543L	226251--225841	137	15,550	4.42		No Hit Found								1	NP_048798	A441L	207.99	6.61E-53	73%	82%	1--137	1--137
C545R	226339--227289	317	36,407	5.8		No Hit Found								1	NP_048800	A443R	527.71	1.80E-148	85%	91%	10--317	1--308
C546L	227697--227383	105	11,928	5.02		No Hit Found								1	NP_048801	A444L	183.73	1.37E-45	91%	93%	1--105	1--104
C547L	228480--227839	214	24,184	8.96	1	cd00283	GIY-YIG_Cterm, GIYX(10-11)YIG family of class I homing endonucleases C-terminus (GIY-YIG_Cterm). Homing endonucleases promote the mobility of intron or intein by recognizing and cleaving a homologous allele that lacks the sequence. They catalyze a double	49.23	6.10E-07	38%	50%	98--210	1--113	1	NP_048671	A315L	244.59	1.49E-63	53%	66%	1--210	1--240
					2	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	47.76	1.94E-06	33%	55%	1--89	1--83	2	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	240.74	2.15E-62	53%	68%	1--212	1--225
					3	smart00497	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished). .	44.74	1.38E-05	33%	48%	159--212	1--53	3	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	159.07	8.23E-38	57%	76%	1--128	1--128
					4	pfam01541	GIY-YIG, GIY-YIG catalytic domain. This domain called GIY-YIG is found in the amino terminal region of excinuclease abc subunit c (uvrC), bacteriophage T4 endonucleases segA, segB, segC, segD and segE; it is also found in putative endonucleases encoded by group I introns of fungi and phage. The structure of I-TevI a GIY-YIG endonuclease, reveals a novel alpha/beta-fold with a central three-stranded antiparallel beta-sheet flanked by three helices. The most conserved and putative catalytic residues are located on a shallow, concave surface and include a metal coordination site	42.46	8.19E-05	23%	45%	1--86	1--89	4	NP_048641	PBCV-1 33kd peptide	145.98	7.21E-34	37%	52%	11--213	15--250
					5	pfam07453	NUMOD1, NUMOD1 domain..	35.79	8.64E-03	48%	62%	159--188	1--30	5	YP_293795	putative endonuclease	78.18	1.85E-13	31%	47%	2--182	3--176
														6	NP_899393	SegD	57.77	2.58E-07	31%	50%	1--160	1--171
														7	CAA38804	GIY COII i1 grp IB protein	52.37	1.08E-05	31%	49%	15--156	86--237
														8	NP_048482	similar to bacteriophage T4 intron-associated endonuclease, corresponds to Swiss-Prot Accession Number P13299	51.22	2.42E-05	30%	54%	2--87	9--95
														9	CAA73995	unnamed protein product	49.68	7.03E-05	26%	39%	10--186	122--326
C551L	229953--228568	462	52,742	6.71	1	COG0661	AarF, Predicted unusual protein kinase [General function prediction only].	182.07	6.72E-47	27%	43%	42--436	56--477	1	NP_048802	similar to Clostridium pasteurianum ORF, corresponds to GenBank Accession Number Z28353	810.83	0.00E+00	88%	94%	1--462	1--462
					2	pfam03109	ABC1, ABC1 family. This family includes ABC1 from yeast and AarF from E. coli. These proteins have a nuclear or mitochondrial subcellular location in eukaryotes. The exact molecular functions of these proteins is not clear, however yeast ABC1 suppresses a cytochrome b mRNA translation defect and is essential for the electron transfer in the bc 1 complex and E. coli AarF is required for ubiquinone production. It has been suggested that members of the ABC1 family are novel chaperonins. These proteins are unrelated to the ABC transporter proteins..	121.08	1.62E-28	40%	56%	94--208	6--119	2	BAB66733	488aa long conserved hypothetical protein	140.58	1.06E-31	28%	48%	54--383	50--399
														3	AAY80974	universally conserved protein	136.35	2.00E-30	27%	46%	42--383	38--399
														4	AAK42726	ABC transporter, ABC1 family, putative	134.81	5.82E-30	26%	48%	24--382	19--393
														5	ZP_00779577	2-polyprenylphenol 6-hydroxylase	131.34	6.44E-29	30%	45%	39--385	50--420
														6	XP_450284	ABC1 family protein-like	129.41	2.45E-28	26%	47%	42--383	207--575
														7	AAF13088	unknown protein	128.26	5.45E-28	25%	46%	54--402	200--573
														8	AAF21180	unknown protein	127.49	9.30E-28	26%	48%	54--383	200--554
														9	AAL10497	At3g07700/F17A17.4	127.49	9.30E-28	26%	48%	54--383	200--554
														10	AAM67100	unknown	127.49	9.30E-28	26%	48%	54--383	200--554
C554L	230362--230045	106	12,521	10.64	1	pfam00085	Thioredoxin, Thioredoxin. Thioredoxins are small enzymes that participate in redox reactions, via the reversible oxidation of an active centre disulfide bond. Some members with only the active site are not separated from the noise..	52.15	9.83E-08	30%	52%	31--105	33--106	1	NP_048805	contains cytochrome C family heme-binding site signature; similar to maize protein disulphide isomerase, correspond to Swiss-Prot Accession Number P52588	187.19	1.23E-46	82%	91%	1--106	1--106
														2	XP_532876	PREDICTED: similar to Protein disulfide-isomerase A6 precursor (Thioredoxin domain containing protein 7)	54.30	1.25E-06	40%	57%	31--99	55--124
														3	EAN99322	protein disulfide isomerase, putative	53.53	2.13E-06	30%	53%	31--105	49--123
														4	EAN94802	protein disulfide isomerase, putative	52.76	3.63E-06	30%	52%	31--105	49--123
														5	CAD99203	protein disulfide isomerase1-2	52.37	4.73E-06	31%	55%	22--102	40--122
														6	BAC86977	unnamed protein product	51.99	6.18E-06	38%	55%	31--99	107--176
														7	NP_005733	protein disulfide isomerase-associated 6	51.99	6.18E-06	38%	55%	31--99	55--124
														8	XP_515706	PREDICTED: protein disulfide isomerase-related protein	51.99	6.18E-06	38%	55%	31--99	3--72
														9	CAH92529	hypothetical protein	51.99	6.18E-06	38%	55%	31--99	55--124
														10	AAB50217	protein disulfide isomerase-related protein 5	51.99	6.18E-06	38%	55%	31--99	36--105
C556R	230524--231087	188	22,243	5.89		No Hit Found								1	NP_048806	PBCV-1 RNA triphosphatase	332.03	5.25E-90	85%	90%	2--188	7--193
C558L	232669--231359	437	48,435	7.43	1	pfam04451	Capsid_Iridovir, Iridovirus major capsid protein. This family includes the major capsid protein of iridoviruses, chlorella virus and Spodoptera ascovirus, which are all dsDNA viruses with no RNA stage. This is the most abundant structural protein and can account for up to 45% of virion protein. In Chlorella virus AR158 the major capsid protein is a glycoprotein	470.56	9.84E-134	51%	64%	1--433	1--443	1	BAA76601	major capsid protein MCP1	776.93	0.00E+00	87%	89%	1--437	1--437
														2	NP_048787	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	770.00	0.00E+00	87%	89%	1--437	1--437
														3	BAA22198	major capsid protein Vp54	761.14	0.00E+00	85%	88%	1--437	1--437
														4	BAA76600	major capsid protein	754.98	0.00E+00	85%	88%	1--437	1--436
														5	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	723.01	0.00E+00	86%	88%	25--437	1--413

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
														6	1M3Y_D	Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid Containing, Dna Virus	723.01	0.00E+00	86%	88%	25--437	1--413
														7	AAC27492	major capsid protein Vp49	634.80	1.67E-180	73%	80%	1--437	1--432
														8	BAE06835	hypothetical major capsid protein	317.39	5.90E-85	42%	54%	1--437	1--440
														9	NP_048359	contains aminoacyl-tRNA synthetase class-II signature	246.13	1.67E-63	36%	51%	1--437	1--403
														10	NP_048914	similar to Chlorella virus PBCV-1 ORF A11L, corresponds to GenBank Accession Number U42580	228.02	4.71E-58	34%	50%	3--437	2--400
C559R	232774--233583	270	32,693	5.45	1	pfam04724	Glyco_transf_17, Glycosyltransferase family 17. This family represents beta-1,4-mannosyl-glycoprotein beta-1,4-N-acetylglucosaminyltransferase (EC:2.4.1.144). This enzyme transfers the bisecting GlcNAc to the core mannose of complex N-glycans. The addition of this residue is regulated during development and has functional consequences for receptor signalling, cell adhesion, and tumour progression	90.78	2.15E-19	26%	48%	4--239	84--311	1	CAE79544	putative N-acetylglucosaminyltransferase	127.87	3.18E-28	30%	49%	6--240	2--246
														2	EAA75195	hypothetical protein FG05624.1	77.03	6.44E-13	28%	41%	4--240	89--334
														3	XP_362071	hypothetical protein MG04516.4	73.17	9.30E-12	26%	40%	4--257	79--341
														4	XP_466520	glycosyl transferase-like protein	64.31	4.32E-09	23%	44%	2--239	79--307
														5	XP_466518	glycosyl transferase-like protein	63.54	7.37E-09	22%	44%	2--239	112--340
														6	XP_472878	OSJNBa0044K18.7	60.08	8.15E-08	22%	42%	2--239	109--337
														7	NP_178963	acetylglucosaminyltransferase/ transferase, transferring glycosyl groups	60.08	8.15E-08	25%	41%	4--239	81--311
														8	XP_327741	predicted protein	56.23	1.18E-06	42%	60%	4--71	78--148
														9	NP_186811	acetylglucosaminyltransferase/ transferase, transferring glycosyl groups	54.68	3.42E-06	21%	43%	4--239	110--337
														10	AAH53040	Mannoside acetylglucosaminyltransferase 3	51.22	3.79E-05	32%	57%	4--74	210--282
C560L	234532--233672	287	31,169	4.73		No Hit Found								1	NP_048811	A454L	500.36	2.61E-140	85%	91%	1--287	1--289
C562L	236525--234564	654	75,501	5.48	1	COG3378	COG3378, Predicted ATPase [General function prediction only].	65.77	6.51E-12	25%	39%	283--533	146--390	1	NP_048813	contains ATP/GTP-binding site motif A	1243.41	0.00E+00	92%	97%	1--654	1--654
					2	pfam03288	Pox_D5, Poxvirus D5 protein-like. This family includes D5 from Poxviruses which is necessary for viral DNA replication, and is a nucleic acid independent nucleoside triphosphatase. Members of this family are also found outside of poxviruses..	65.66	7.41E-12	27%	44%	292--524	3--228	2	NP_077594	EsV-1-109	287.73	8.34E-76	31%	50%	36--618	18--575
														3	AAR26902	FirrV-1-B27	268.86	4.01E-70	33%	51%	132--619	93--574
														4	YP_294217	putative nucleic acid independent nucleoside triphosphatase	150.98	1.22E-34	25%	43%	191--646	170--675
														5	ZP_00123428	COG3378: Predicted ATPase	66.24	3.95E-09	27%	48%	340--532	310--502
														6	BAE05402	putative DNA primase-phage associated	58.15	1.08E-06	23%	43%	294--525	107--315
														7	ZP_00503756	Phage/plasmid primase P4, C-terminal	55.84	5.34E-06	23%	40%	239--501	246--492
														8	ZP_00593388	Phage/plasmid primase P4, C-terminal	55.84	5.34E-06	25%	42%	360--523	346--508
														9	AAF27348	phage phi-R73 primase-like protein	55.45	6.97E-06	23%	41%	251--532	186--465
														10	CAG34908	hypothetical protein, probably cold-shock inducible	54.68	1.19E-05	27%	40%	354--530	361--539
C568R	236869--237672	268	30,513	5.78	1	COG0571	Rnc, dsRNA-specific ribonuclease [Transcription].	167.75	1.37E-42	34%	55%	46--266	10--235	1	NP_048820	similar to Bacillus ribonuclease III, corresponds to Swiss-Prot Accession Number P51833	490.73	1.84E-137	88%	95%	1--267	8--274
					2	smart00535	RIBOc, Ribonuclease III family; .	128.05	1.27E-30	38%	58%	59--185	1--126	2	YP_445467	ribonuclease III	139.04	1.36E-31	35%	55%	35--260	9--241
					3	cd00593	RIBOc, RIBOc. Ribonuclease III C terminal domain. This group consists of eukaryotic, bacterial and archeal ribonuclease III (RNAse III) proteins. RNAse III is a double stranded RNA-specific endonuclease. Prokaryotic RNAse III is important in post-transcriptional control of mRNA stability and translational efficiency. It is involved in the processing of ribosomal RNA precursors. Prokaryotic RNAse III also plays a role in the maturation of tRNA precursors and in the processing of phage and plasmid transcripts. Eukaryotic RNase III's participate (through direct cleavage) in rRNA processing, in processing of small nucleolar RNAs (snoRNAs) and snRNA's (components of the spliceosome). In eukaryotes RNase III or RNaseIII like enzymes such as Dicer are involved in RNAi (RNA interference) and miRNA (micro-RNA) gene silencing	120.00	3.66E-28	39%	59%	72--189	16--133	3	AAM73335	ribonuclease III	132.88	9.76E-30	35%	58%	61--264	48--260
					4	pfam00636	Ribonuclease 3, RNase3 domain..	119.44	5.53E-28	50%	68%	78--168	1--91	4	ZP_00590199	Ribonuclease III	132.49	1.27E-29	33%	58%	47--257	47--265
					5	cd00048	DSRM, Double-stranded RNA binding motif. Binding is not sequence specific but is highly specific for double stranded RNA. Found in a variety of proteins including dsRNA dependent protein kinase PKR, RNA helicases, Drosophila staufen protein, E. coli RNase III, RNases H1, and dsRNA dependent adenosine deaminases	65.00	1.30E-11	36%	55%	194--260	1--68	5	ZP_00591208	Ribonuclease III	130.18	6.32E-29	31%	51%	16--257	13--265
														6	ZP_00532592	Ribonuclease III	129.41	1.08E-28	32%	54%	23--257	26--259
														7	ABB23018	Ribonuclease III	125.56	1.56E-27	34%	56%	56--257	43--252
														8	ZP_00511103	Ribonuclease III	122.09	1.72E-26	34%	57%	66--257	66--259
														9	ZP_00528534	Ribonuclease III	120.55	5.01E-26	33%	58%	61--257	76--281
														10	ZP_00661649	Ribonuclease III	120.17	6.54E-26	31%	52%	36--257	20--252
C569R	237872--238642	257	30,838	9.67	1	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	46.61	4.59E-06	41%	62%	45--124	12--83	1	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	72.02	1.91E-11	44%	54%	35--133	3--100
					2	pfam01541	GIY-YIG, GIY-YIG catalytic domain. This domain called GIY-YIG is found in the amino terminal region of excinuclease abc subunit c (uvrC), bacteriophage T4 endonucleases segA, segB, segC, segD and segE; it is also found in putative endonucleases encoded by group I introns of fungi and phage. The structure of I-TevI a GIY-YIG endonuclease, reveals a novel alpha/beta-fold with a central three-stranded antiparallel beta-sheet flanked by three helices. The most conserved and putative catalytic residues are located on a shallow, concave surface and include a metal coordination site	46.31	4.66E-06	31%	49%	39--120	4--88	2	NP_048671	A315L	67.40	4.70E-10	43%	59%	47--133	13--99
														3	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	67.40	4.70E-10	45%	56%	47--133	13--98
														4	NP_048641	PBCV-1 33kd peptide	60.46	5.74E-08	32%	51%	39--181	9--154
														5	NP_048482	similar to bacteriophage T4 intron-associated endonuclease, corresponds to Swiss-Prot Accession Number P13299	57.00	6.35E-07	38%	60%	38--120	10--93
														6	YP_293795	putative endonuclease	51.22	3.48E-05	35%	53%	38--128	4--97
C570R	238669--239022	118	13,637	10.35	1	pfam04777	Evr1_Alr, Erv1 / Alr family. Biogenesis of Fe/S clusters involves a number of essential mitochondrial proteins. Erv1p of Saccharomyces cerevisiae mitochondria is required for the maturation of Fe/S proteins in the cytosol. The ALR (augmenter of liver regeneration) represents a mammalian orthologue of yeast Erv1p. Both Erv1p and full-length ALR are located in the mitochondrial intermembrane an d it thought to operate downstream of the mitochondrial ABC transporter	74.18	2.03E-14	33%	48%	22--113	4--91	1	NP_048821	PBCV-1 thiol oxidoreductase	188.73	4.19E-47	72%	82%	1--118	1--118
					2	COG5054	ERV1, Mitochondrial sulfhydryl oxidase involved in the biogenesis of cytosolic Fe/S proteins [Posttranslational modification, protein turnover, chaperones].	47.36	2.79E-06	29%	49%	1--112	69--175	2	YP_142722	putative thiol oxidoreductase	72.79	3.35E-12	37%	54%	11--106	3--97
														3	YP_142950	thiol oxidoreductase E10R	58.15	8.53E-08	35%	51%	12--96	38--123
														4	CAG59329	unnamed protein product	57.00	1.90E-07	29%	53%	15--112	76--168
														5	XP_503294	hypothetical protein	56.61	2.48E-07	31%	52%	10--113	88--186
														6	NP_078699	Thiol oxidoreductase	52.37	4.68E-06	30%	51%	1--107	16--116
														7	NP_149810	347L	51.22	1.04E-05	29%	48%	8--99	3--97
														8	CAH02199	unnamed protein product	49.68	3.04E-05	29%	49%	4--113	74--178
														9	CAE74303	Hypothetical protein CBG22010	49.68	3.04E-05	31%	46%	15--109	59--152
														10	CAA48192	ERV1	49.68	3.04E-05	26%	51%	15--112	19--111
C571L	239975--239040	312	36,661	7.12		No Hit Found								1	NP_048823	A467L	583.95	2.06E-165	92%	94%	1--312	1--312
														2	NP_065022	Hypothetical protein	58.54	3.01E-07	28%	46%	72--223	65--218
C573R	240113--241438	442	50,724	8.68		No Hit Found								1	NP_048824	A468R	731.87	0.00E+00	79%	88%	1--442	1--443
														2	AAR26870	FirrV-1-A46	60.08	1.73E-07	24%	46%	1--239	1--226
C575R	241595--242194	200	22,715	4.51		No Hit Found								1	NP_048826	A470R	342.04	5.89E-93	82%	92%	1--199	1--203
														2	BAB69883	UKCH-2	91.66	1.39E-17	33%	56%	15--151	57--203
C577R	242243--242764	174	21,125	10.15		No Hit Found								1	NP_048827	A471R	288.89	4.21E-77	79%	87%	1--174	1--173
														2	YP_142861	unknown	115.93	4.88E-25	46%	61%	27--156	53--191
														3	AAR26829	FirrV-1-A5	83.57	2.68E-15	34%	53%	7--149	19--161
														4	NP_077626	EsV-1-141	77.41	1.92E-13	32%	60%	7--137	23--154

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
C579R	242821--243792	324	37,457	4.55	1	cd01049	RNRR2, Ribonucleotide Reductase, R2/beta subunit (RNRR2) is a member of a broad superfamily of ferritin-like diiron-carboxylate proteins. The RNR protein catalyzes the conversion of ribonucleotides to deoxyribonucleotides and is found in all eukaryotes.	312.95	3.02E-86	52%	66%	16--291	1--284	1	NP_048832	contains ribonucleotide reductase (RR) signature; similar to tobacco RR small subunit, corresponds to Swiss-Prot Accession Number P49730	657.52	0.00E+00	98%	99%	1--324	1--324
					2	pfam00268	Ribonuc red sm, Ribonucleotide reductase, small chain..	300.25	1.63E-82	46%	65%	5--285	1--281	2	AAO62422	ribonucleotide reductase small subunit	400.59	3.44E-110	58%	72%	4--324	6--333
					3	COG0208	NrdF, Ribonucleotide reductase, beta subunit [Nucleotide transport and metabolism].	245.20	6.51E-66	34%	49%	7--324	18--346	3	NP_189342	ribonucleoside-diphosphate reductase	400.59	3.44E-110	58%	73%	4--324	6--332
					4	cd00657	Ferritin_like, Ferritin-like, diiron-carboxylate proteins participate in a range of functions including iron regulation, mono-oxygenation, and reactive radical production. These proteins are characterized by the fact that they catalyze dioxygen-dependent	54.72	1.39E-08	19%	30%	66--213	1--140	4	CAF24073	probable ribonucleoside-diphosphate reductase small chain	397.51	2.91E-109	60%	74%	1--324	1--323
														5	XP_550581	putative ribonucleotide reductase R2	393.28	5.49E-108	57%	74%	4--324	20--345
														6	NP_910365	putative ribonucleoside-diphosphate reductase	393.28	5.49E-108	57%	74%	4--324	2--327
														7	BAD46317	putative ribonucleotide reductase R2	392.50	9.36E-108	57%	74%	4--324	14--339
														8	NP_701941	ribonucleotide reductase small subunit	392.12	1.22E-107	60%	76%	5--324	31--349
														9	AAF15363	ribonucleotide reductase R2 subunit	391.73	1.60E-107	61%	77%	13--324	9--320
														10	P50650	Ribonucleoside-diphosphate reductase small chain (Ribonucleotide reductase small subunit) (Ribonucleotide reductase R2 subunit)	391.73	1.60E-107	60%	76%	5--324	31--349
C580L	244724--243795	310	36,649	9.05		No Hit Found								1	NP_048834	Lys-, Arg-rich; contains eukaryotic putative RNA-binding region RNP-1 signature; similar to PBCV-1 ORF A267L, corresponds to GenBank Accession Number U42580	598.59	7.98E-170	89%	91%	1--310	1--310
														2	NP_048846	Lys-, Glu-rich	458.76	9.87E-128	68%	79%	1--310	1--310
														3	YP_142777	unknown	153.30	8.87E-36	29%	45%	1--288	135--457
														4	NP_048621	A267L	90.12	9.22E-17	28%	44%	69--290	9--246
C583L	245040--244762	93	9,813	10.52		No Hit Found								1	NP_048836	Gly-rich	120.17	1.80E-26	96%	96%	1--57	1--57
C584L	245745--245074	224	25,962	4.72	1	COG5540	COG5540, RING-finger-containing ubiquitin ligase [Posttranslational modification, protein turnover, chaperones].	37.76	1.72E-03	33%	51%	145--194	324--373	1	NP_048837	A481L	409.45	3.88E-113	83%	91%	1--224	1--224
C586R	245821--246468	216	25,092	9.75	1	COG1350	COG1350, Predicted alternative tryptophan synthase beta-subunit (paralog of TrpB) [General function prediction only].	36.03	5.98E-03	31%	46%	123--204	24--94	1	NP_048838	A482R	393.66	2.03E-108	85%	94%	1--213	1--214
														2	AAR26970	FirrV-1-I5	65.08	1.65E-09	29%	45%	30--126	32--146
														3	AAR26889	FirrV-1-B14	60.85	3.12E-08	31%	46%	33--126	32--145
														4	YP_294196	hypothetical protein EhV 438	58.92	1.18E-07	33%	48%	33--127	77--169
														5	NP_077581	EsV-1-96	58.54	1.55E-07	30%	46%	30--126	67--181
C587L	246939--246475	155	18,705	9.75		No Hit Found								1	NP_048840	A484L	274.63	5.85E-73	82%	91%	1--155	1--155
C588R	246974--247405	144	16,962	10.93		No Hit Found								1	NP_048841	A485R	243.82	1.09E-63	83%	92%	3--144	6--147
C590R	247481--248431	317	34,880	5.1		No Hit Found								1	NP_048844	Pro-,Gln-rich	451.06	2.14E-125	75%	80%	1--317	1--317
														2	YP_142771	unknown	57.38	6.88E-07	25%	43%	8--177	5--191
C593R	248472--248699	76	8,410	4.55		No Hit Found								1	NP_048847	A491R	126.72	1.97E-28	78%	92%	1--76	1--76
C595L	249249--248677	191	22,516	8.4		No Hit Found								1	NP_048848	A492L	273.86	1.77E-72	72%	81%	1--183	7--189
														2	NP_048849	A493L	193.74	2.33E-48	55%	69%	4--181	1--178
C596L	249904--249308	199	23,952	8.73		No Hit Found								1	NP_048849	A493L	352.44	4.30E-96	92%	96%	1--184	1--184
														2	NP_048848	A492L	244.59	1.26E-63	67%	81%	1--178	10--187
C597R	249964--251040	359	42,035	8.91		No Hit Found								1	NP_048850	A494R	593.58	3.23E-168	80%	87%	6--359	7--360
														2	YP_142783	PBCV1-A494R-like protein	140.97	5.74E-32	38%	54%	154--355	223--427
														3	NP_077586	EsV-1-101	131.34	4.55E-29	37%	54%	167--356	120--316
														4	AAR26895	FirrV-1-B20	114.39	5.75E-24	32%	52%	170--356	105--298
														5	YP_294161	hypothetical protein EhV 403	101.68	3.86E-20	32%	45%	154--355	156--362
														6	NP_149745	282R	61.62	4.42E-08	22%	40%	152--356	155--380
														7	YP_142784	unknown	55.84	2.43E-06	25%	43%	215--357	109--251
C600R	251093--251530	146	15,390	10.3		No Hit Found								1	NP_048853	Gly-rich, KGKDIGGG (4x)	85.11	6.42E-16	97%	100%	108--146	108--146
C602L	252685--251600	362	39,627	4.92	1	COG5373	COG5373, Predicted membrane protein [Function unknown].	57.67	2.21E-09	29%	38%	114--272	24--152	1	NP_048857	A501L	99.37	1.94E-19	100%	100%	1--50	1--50
					2	pfam05887	Trypan_PARP, Procyclic acidic repetitive protein (PARP). This family consists of several Trypanosoma brucei procyclic acidic repetitive protein (PARP) like sequences. The procyclic acidic repetitive protein (parp) genes of Trypanosoma brucei encode a small family of abundant surface proteins whose expression is restricted to the procyclic form of the parasite. They are found at two unlinked loci, parpA and parpB; transcription of both loci is developmentally regulated	56.53	4.72E-09	37%	47%	130--209	41--120	2	NP_048856	Pro-, Ser-rich	73.17	1.49E-11	61%	61%	264--325	1--62
C604L	253003--252719	95	11,130	10.12		No Hit Found								1	NP_048858	A502L	189.50	2.49E-47	93%	98%	1--95	1--95
C605L	253860--253018	281	32,432	9.72		No Hit Found								1	NP_048859	contains prokaryotic membrane lipoprotein lipid attachment site	547.74	1.39E-154	95%	97%	1--281	24--304
C608L	255371--253893	493	56,449	5.18		No Hit Found								1	NP_048861	A505L	750.36	0.00E+00	75%	82%	1--488	1--484
C613R	255555--255941	129	15,195	8.12	1	cd00162	RING, RING-finger (Really Interesting New Gene) domain, a specialized type of Zn-finger of 40 to 60 residues that binds two atoms of zinc; defined by the 'cross-brace' motif C-X2-C-X(9-39)-C-X(1-3)-H-X(2-3)-(N/C/H)-X2-C-X(4-48)C-X2-C; probably involved in mediating protein-protein interactions; identified in a proteins with a wide range of functions such as viral replication, signal transduction, and development; has two variants, the C3HC4-type and a C3H2C3-type (RING-H2 finger), which have different cysteine/histidine pattern; a subset of RINGs are associated with B-Boxes (C-X2-H-X7-C-X7-C-X2-C-H-X2-H).	40.46	3.30E-04	34%	61%	80--129	1--45	1	AAH85684	Tripartite motif protein 50	51.99	6.06E-06	37%	59%	69--129	3--60
					2	COG5175	MOT2, Transcriptional repressor [Transcription].	35.50	9.99E-03	33%	63%	76--127	12--61	2	AAL91073	tripartite motif protein 50	51.99	6.06E-06	37%	59%	69--129	3--60
														3	XP_793752	PREDICTED: similar to tripartite motif-containing 33, partial	50.83	1.35E-05	36%	56%	70--127	8--64
														4	XP_691234	PREDICTED: similar to Tripartite motif protein 32 (Zinc-finger protein HT2A) (72 kDa Tat-interacting protein)	50.06	2.30E-05	29%	47%	36--127	262--346
														5	AAN14328	CG10981-PB, isoform B	49.29	3.93E-05	43%	54%	79--129	257--302
														6	AAF51944	CG10981-PA, isoform A	49.29	3.93E-05	43%	54%	79--129	264--309
														7	XP_791398	PREDICTED: similar to Transcription intermediary factor 1-gamma (TIF1-gamma) (RET-fused gene 7 protein) (Rfg7 protein) (Tripartite motif protein 33), partial	48.91	5.13E-05	36%	57%	65--127	5--64
														8	XP_799164	PREDICTED: similar to tripartite motif-containing 33, partial	48.91	5.13E-05	34%	55%	70--127	8--64
														9	CAB39846	SPBC2A9.04c	48.52	6.70E-05	37%	45%	50--129	71--153
														10	XP_787784	PREDICTED: similar to Galectin-3 binding protein precursor (Lectin galactoside-binding soluble 3 binding protein) (Mac-2 binding protein) (Mac-2 BP) (MAC2BP) (Tumor-associated antigen 90K)	48.52	6.70E-05	38%	53%	76--127	10--59
C615L	256203--255958	82	9,200	10.22		No Hit Found								1	NP_048875	A519L	154.07	1.13E-36	91%	96%	1--81	1--81
C616L	256525--256211	105	12,459	10.87		No Hit Found								1	NP_048876	A520L	172.17	4.11E-42	86%	93%	10--105	5--100
C618L	257601--256531	357	40,255	10.74		No Hit Found								1	AAV84098	CviPII top-strand DNA nicking endonuclease	580.48	2.81E-164	94%	96%	1--310	1--310
														2	AAC55064	restriction endonuclease	96.29	1.61E-18	32%	49%	35--282	52--277

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
C619L	258703--257615	363	41,507	8.94	1	cd00315	Cyt_C5_DNA_methylase, Cytosine-C5 specific DNA methylases; Methyl transfer reactions play an important role in many aspects of biology. Cytosine-specific DNA methylases are found both in prokaryotes and eukaryotes. DNA methylation, or the covalent addition of a methyl group to cytosine within the context of the CpG dinucleotide, has profound effects on the mammalian genome. These effects include transcriptional repression via inhibition of transcription factor binding or the recruitment of methyl-binding proteins and their associated chromatin remodeling factors, X chromosome inactivation, imprinting and the suppression of parasitic DNA sequences. DNA methylation is also essential for proper embryonic development and is an important player in both DNA repair and genome stability	151.61	1.17E-37	30%	45%	16--277	1--270	1	AAV84097	CviPII m5C DNA methyltransferase	749.97	0.00E+00	99%	99%	1--363	1--363
					2	pfam00145	DNA methylase, C-5 cytosine-specific DNA methylase..	140.83	1.73E-34	37%	51%	16--192	1--178	2	AAC64006	cytosine methyltransferase	326.25	9.80E-88	46%	61%	14--363	1--362
					3	COG0270	Dcm, Site-specific DNA methylase [DNA replication, recombination, and repair].	106.31	4.36E-24	28%	51%	14--178	2--171	3	NP_048873	M.CviAII cytosine DNA methyltransferase	313.54	6.57E-84	47%	63%	16--357	2--342
														4	AAC55063	cytosine methyltransferase	288.89	1.73E-76	41%	58%	14--363	1--366
														5	NP_049039	nonfunctional M.CviAV cytosine DNA methyltransferase	285.42	1.92E-75	41%	57%	14--363	1--366
														6	NP_048886	M.CviAIV cytosine DNA methyltransferase	256.91	7.30E-67	41%	58%	14--345	1--332
														7	ZP_00783160	C-5 cytosine-specific DNA methylase	96.67	1.26E-18	32%	51%	16--182	1--176
														8	ZP_00874816	C-5 cytosine-specific DNA methylase	96.29	1.65E-18	32%	51%	16--182	1--176
														9	AAT87295	Type II restriction-modification system methylation subunit	94.36	6.27E-18	36%	51%	8--181	93--269
														10	CAD47029	unknown	92.82	1.82E-17	33%	51%	16--182	1--176
C620L	259251--258706	182	20,365	11.61		No Hit Found								1	NP_048877	contains Gln-rich, neutral zinc metallopeptidase, zinc binding region signature	265.00	7.31E-70	75%	80%	1--182	212--392
C621L	259884--259279	202	22,974	6.27		No Hit Found								1	NP_048877	contains Gln-rich, neutral zinc metallopeptidase, zinc binding region signature	298.52	7.66E-80	76%	85%	1--193	1--192
														2	YP_142679	metal-dependent hydrolase	58.92	1.02E-07	27%	48%	29--179	33--191
														3	NP_149599	136R	52.37	9.56E-06	36%	57%	98--160	98--163
C622R	259937--260440	168	18,860	9.94		No Hit Found								1	NP_048879	A523R	292.74	2.65E-78	84%	90%	4--168	7--171
C624L	261415--260453	321	36,824	8		No Hit Found								1	NP_048920	similar to Chlorella virus PBCV-1 ORF A154L, corresponds to GenBank Accession Number U42580	489.96	4.24E-137	65%	76%	5--321	4--350
														2	NP_048502	A154L	459.91	4.70E-128	67%	80%	3--321	35--347
														3	NP_048477	similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903	455.29	1.16E-126	61%	73%	1--321	1--356
														4	NP_077492	EsV-1-7	68.17	3.98E-10	34%	49%	8--151	10--153
C627R	261519--261956	146	16,278	7.3		No Hit Found								1	NP_048882	A526R	228.79	3.60E-59	84%	91%	18--146	1--129
C629R	261983--262267	95	11,300	11.07		No Hit Found								1	NP_048883	A527R	100.52	1.52E-20	85%	93%	1--60	5--63
C630R	262356--262586	77	8,346	10.06		No Hit Found									No Hit Found	No Hit Found						
C632R	262553--263596	348	39,704	8.14	1	pfam00145	DNA methylase, C-5 cytosine-specific DNA methylase..	153.16	3.42E-38	29%	45%	3--245	1--246	1	NP_048886	M.CviAIV cytosine DNA methyltransferase	599.74	4.30E-170	85%	92%	1--332	1--332
					2	cd00315	Cyt_C5_DNA_methylase, Cytosine-C5 specific DNA methylases; Methyl transfer reactions play an important role in many aspects of biology. Cytosine-specific DNA methylases are found both in prokaryotes and eukaryotes. DNA methylation, or the covalent addition of a methyl group to cytosine within the context of the CpG dinucleotide, has profound effects on the mammalian genome. These effects include transcriptional repression via inhibition of transcription factor binding or the recruitment of methyl-binding proteins and their associated chromatin remodeling factors, X chromosome inactivation, imprinting and the suppression of parasitic DNA sequences. DNA methylation is also essential for proper embryonic development and is an important player in both DNA repair and genome stability	149.68	3.70E-37	28%	42%	3--288	1--269	2	NP_048873	M.CviAII cytosine DNA methyltransferase	424.86	1.89E-117	58%	73%	2--345	1--343
					3	COG0270	Dcm, Site-specific DNA methylase [DNA replication, recombination, and repair].	111.71	1.20E-25	32%	47%	3--163	4--169	3	AAC64006	cytosine methyltransferase	300.83	4.12E-80	45%	59%	1--343	1--355
														4	AAC55063	cytosine methyltransferase	292.35	1.47E-77	43%	59%	1--343	1--359
														5	NP_049039	nonfunctional M.CviAV cytosine DNA methyltransferase	290.04	7.28E-77	43%	59%	1--343	1--359
														6	AAV84097	CviPII m5C DNA methyltransferase	283.11	8.90E-75	44%	60%	1--343	14--356
														7	AAN07998	gp80	92.43	2.23E-17	35%	50%	3--158	1--159
														8	ABA25040	C-5 cytosine-specific DNA methylase	92.43	2.23E-17	33%	51%	6--158	5--160
														9	CAD80133	gp9.1	90.12	1.11E-16	32%	50%	6--163	4--162
														10	CAC48079	putative DNA methylase	89.35	1.89E-16	36%	48%	3--158	1--164
C634L	264203--264006	66	7,644	9.33		No Hit Found								1	NP_048887	A531L	86.66	2.25E-16	85%	89%	19--66	19--67
C635L	264462--264226	79	8,738	10.09		No Hit Found								1	NP_048888	A532L	117.47	1.18E-25	72%	72%	1--79	1--79
C636L	265879--264743	379	43,680	10.54	1	COG0675	COG0675, Transposase and inactivated derivatives [DNA replication, recombination, and repair].	82.44	6.66E-17	24%	39%	53--370	59--356	1	AAU06281	putative transposase	731.48	0.00E+00	93%	97%	1--379	62--439
					2	pfam07282	Transposase_35, Putative transposase DNA-binding domain. This putative domain is found at the C-terminus of a large number of transposase proteins. This domain contains four conserved cysteines suggestive of a zinc binding domain. Given the need for transposases to bind DNA as well as the large number of DNA-binding zinc fingers we hypothesise this domain is DNA-binding	72.96	5.12E-14	41%	54%	292--359	1--69	2	NP_048981	similar to Synechocystis transposase, corresponds to GenBank Accession Number D90909	686.03	0.00E+00	87%	93%	1--379	56--433
					3	pfam01385	Transposase_2, Probable transposase. This family includes IS891, IS1136 and IS1341. .	37.20	2.75E-03	21%	40%	2--261	14--261	3	YP_142458	putative transposase	105.53	2.88E-21	31%	46%	58--359	220--535
														4	AAS54227	AGL264Wp	93.20	1.48E-17	26%	44%	1--361	84--453
														5	ABA24789	Transposase, IS891/IS1136/IS1341	92.43	2.52E-17	27%	41%	1--361	161--533
														6	BAB78230	transposase	92.43	2.52E-17	27%	41%	1--361	117--489
														7	ZP_00158267	COG0675: Transposase and inactivated derivatives	85.11	4.03E-15	28%	42%	61--361	43--359
														8	YP_238637	ORF021	67.40	8.69E-10	25%	44%	2--356	18--364
														9	BAE47830	putative IS transposase (OrfB)	65.86	2.53E-09	22%	43%	1--356	18--384
														10	YP_143208	putative transposase	65.47	3.30E-09	25%	40%	64--357	223--516
C638R	266287--267876	530	57,510	4.54		No Hit Found								1	NP_048889	A533R	698.74	0.00E+00	91%	95%	1--374	1--374
														2	AAA66400	unknown protein	696.81	0.00E+00	90%	95%	1--374	1--374
														3	NP_048890	a534R	184.50	7.68E-45	94%	97%	441--530	16--105
C641L	268677--267889	263	29,988	10.22	1	cd00283	GIY-YIG_Cterm, GIYX(10-11)YIG family of class I homing endonucleases C-terminus (GIY-YIG_Cterm). Homing endonucleases promote the mobility of intron or intein by recognizing and cleaving a homologous allele that lacks the sequence. They catalyze a double	54.62	1.42E-08	41%	53%	116--259	2--113	1	NP_048671	A315L	264.62	2.07E-69	51%	66%	1--259	1--240
					2	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	40.06	3.78E-04	30%	49%	1--89	1--83	2	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	243.43	4.94E-63	50%	65%	1--262	1--226
					3	pfam01541	GIY-YIG, GIY-YIG catalytic domain. This domain called GIY-YIG is found in the amino terminal region of excinuclease abc subunit c (uvrC), bacteriophage T4 endonucleases segA, segB, segC, segD and segE; it is also found in putative endonucleases encoded by group I introns of fungi and phage. The structure of I-TevI a GIY-YIG endonuclease, reveals a novel alpha/beta-fold with a central three-stranded antiparallel beta-sheet flanked by three helices. The most conserved and putative catalytic residues are located on a shallow, concave surface and include a metal coordination site	37.06	3.48E-03	24%	44%	1--86	1--89	3	NP_048641	PBCV-1 33kd peptide	202.60	9.67E-51	43%	63%	11--260	15--248
					4	pfam07453	NUMOD1, NUMOD1 domain..	35.40	9.95E-03	36%	48%	207--240	1--34	4	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	173.71	4.81E-42	45%	64%	1--195	1--171
														5	NP_899393	SegD	85.11	2.25E-15	32%	53%	1--204	1--202
														6	YP_293795	putative endonuclease	83.57	6.55E-15	35%	55%	2--128	3--132
														7	AAT53588	group I intron GIY-YIG endonuclease	65.47	1.85E-09	33%	51%	126--251	103--236
														8	ZP_00391680	COG0532: Translation initiation factor 2 (IF-2; GTPase)	65.47	1.85E-09	33%	51%	126--251	78--211
														9	CAA38804	GIY COII i1 grp IB protein	62.77	1.20E-08	29%	53%	40--189	114--249
														10	CAA73995	unnamed protein product	59.31	1.32E-07	30%	46%	10--157	122--276
C643L	268969--268757	71	8,274	4.48		No Hit Found								1	NP_048891	A535L	140.58	1.29E-32	90%	95%	1--71	1--71

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
C644L	269236--268994	81	9,172	11.05		No Hit Found								1	NP_048892	A536L	78.57	6.05E-14	66%	74%	23--81	19--73
C645L	270007--269264	248	28,197	9.2		No Hit Found								1	NP_048723	A366L	216.85	4.47E-55	61%	79%	73--239	86--252
C648L	270955--270179	259	30,767	6.38		No Hit Found								1	NP_048893	A537L	312.77	6.52E-84	59%	73%	2--258	9--263
C651R	271091--271609	173	20,002	10.22		No Hit Found								1	NP_048895	A539R	297.75	8.89E-80	80%	93%	1--173	1--173
														2	NP_048482	similar to bacteriophage T4 intron-associated endonuclease, corresponds to Swiss-Prot Accession Number P13299	50.83	1.89E-05	56%	75%	34--74	11--51
														3	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	48.91	7.19E-05	31%	51%	33--137	3--106
C652L	275390--271599	1264	128,452	5.89		No Hit Found								1	NP_048896	A540L	1486.86	0.00E+00	63%	75%	81--1259	1--1175
														2	AAA66404	unknown protein	389.81	3.30E-106	66%	79%	969--1259	1--291
														3	ZP_00532602	Hep Hag	175.64	9.78E-42	26%	39%	231--992	92--890
														4	ZP_00950302	outer membrane protein	175.25	1.28E-41	27%	39%	377--1102	182--866
														5	ZP_00533161	Hep Hag	172.94	6.34E-41	28%	40%	251--841	104--755
														6	AAX44675	possible T4-like proximal tail fiber	169.47	7.01E-40	30%	42%	374--910	179--717
														7	ZP_00945173	Collagen alpha 2(I) chain precursor	140.97	2.67E-31	28%	39%	319--842	211--804
														8	ZP_00458641	Outer membrane autotransporter barrel	127.49	3.05E-27	23%	35%	208--1085	137--1022
														9	CAH35630	putative outer membrane protein	126.72	5.21E-27	23%	37%	253--1098	148--991
														10	ZP_00464538	PE-PGRS family protein	124.79	1.98E-26	26%	33%	296--908	471--1167
C658R	275509--276477	323	37,214	7	1	pfam01068	DNA_ligase_A_M, ATP dependent DNA ligase domain. This domain belongs to a more diverse superfamily, including pfam01331 and pfam01653..	95.37	9.17E-21	23%	41%	31--213	1--201	1	1P8L_A	Chain A, New Crystal Structure Of Chlorella Virus Dna Ligase- Adenylate	516.15	5.58E-145	82%	90%	20--322	1--303
					2	COG1793	CDC9, ATP-dependent DNA ligase [DNA replication, recombination, and repair].	70.03	3.34E-13	19%	39%	31--318	119--417	2	NP_048900	PBCV-1 DNA ligase	514.61	1.62E-144	83%	90%	26--322	1--297
														3	1FVI_A	Chain A, Crystal Structure Of Chlorella Virus Dna Ligase-Adenylate	509.61	5.23E-143	83%	90%	28--322	2--296
														4	ABA50091	PBCV-1 DNA ligase	184.11	5.03E-45	39%	57%	30--322	9--305
														5	CAI10149	DNA ligase, ATP-dependent	80.11	1.02E-13	27%	46%	48--318	69--301
														6	AAZ12125	DNA ligase, putative	77.03	8.66E-13	26%	39%	49--322	215--509
														7	YP_189099	DNA ligase, ATP-dependent	76.64	1.13E-12	25%	43%	52--322	160--427
														8	ZP_00550374	ATP-dependent DNA ligase	76.64	1.13E-12	22%	44%	48--315	48--277
														9	ZP_00733068	ATP-dependent DNA ligase	76.64	1.13E-12	27%	44%	49--314	40--266
														10	AAZ80807	ATP-dependent DNA ligase	73.17	1.25E-11	25%	47%	48--321	53--284
C661L	277639--276452	396	44,713	6.69	1	COG0438	RfaG, Glycosyltransferase [Cell envelope biogenesis, outer membrane].	63.89	2.77E-11	20%	39%	1--394	2--381	1	NP_048902	similar to Streptomyces glucosyltransferase protein, corresponds to GenBank Accession Number AB005901	535.80	9.22E-151	81%	92%	20--328	1--309
					2	pfam00534	Glycos_transf_1, Glycosyl transferases group 1. Mutations in this domain may lead to disease (Paroxysmal Nocturnal haemoglobinuria). Members of this family transfer activated sugars to a variety of substrates, including glycogen, Fructose-6-phosphate and lipopolysaccharides. Members of this family transfer UDP, ADP, GDP or CMP linked sugars. The eukaryotic glycogen synthases may be distant members of this family..	52.25	7.81E-08	26%	45%	178--369	9--172	2	NP_835600	putative glycosyltransferases	58.92	3.30E-07	25%	40%	77--313	78--342
					3	COG0297	GlgA, Glycogen synthase [Carbohydrate transport and metabolism].	36.01	6.76E-03	25%	42%	173--307	283--402	3	ZP_00687600	hypothetical protein BambDRAFT 3460	58.15	5.63E-07	31%	46%	185--320	54--179
														4	AAZ58719	putative Glycosyltransferase	54.68	6.23E-06	25%	47%	126--311	581--756
														5	ZP_00789707	glycosyl transferase, group 1 family protein	52.37	3.09E-05	22%	44%	136--314	151--324
														6	ZP_00787178	glycosyl transferase, group 1 family protein	51.99	4.04E-05	22%	44%	136--314	151--324
C663L	279046--277673	458	53,247	8.01	1	COG0553	HepA, Superfamily II DNA/RNA helicases, SNF2 family [Transcription / DNA replication, recombination, and repair].	149.14	6.48E-37	28%	45%	5--433	337--846	1	NP_048904	similar to Caenorhabditis transcription activator, corresponds to Swiss-Prot Accession Number P41877	805.82	0.00E+00	85%	94%	1--458	1--458
					2	pfam00176	SNF2_N, SNF2 family N-terminal domain. This domain is found in proteins involved in a variety of processes including transcription regulation (e.g., SNF2, STH1, brahma, MOT1) , DNA repair (e.g., ERCC6, RAD16, RAD5), DNA recombination (e.g., RAD54), and chromatin unwinding (e.g., ISWI) as well as a variety of other proteins with little functional information (e.g. lodestar, ETL1)	120.38	2.70E-28	28%	46%	10--270	1--294	2	ZP_00404276	COG0553: Superfamily II DNA/RNA helicases, SNF2 family	173.33	1.46E-41	29%	47%	7--439	416--849
					3	smart00487	DEXDc, DEAD-like helicases superfamily; .	74.49	1.51E-14	20%	34%	2--196	4--202	3	AAK75613	Snf2 family protein	173.33	1.46E-41	29%	47%	7--439	583--1016
					4	cd00079	HELICc, Helicase superfamily c-terminal domain; associated with DEXDc-, DEAD-, and DEAH-box proteins, yeast initiation factor 4A, Ski2p, and Hepatitis C virus NS3 helicases; this domain is found in a wide variety of helicases and helicase related proteins; may not be an autonomously folding unit, but an integral part of the helicase; 4 helicase superfamilies at present according to the organization of their signature motifs; all helicases share the ability to unwind nucleic acid duplexes with a distinct directional polarity; they utilize the free energy from nucleoside triphosphate hydrolysis to fuel their translocation along DNA, unwinding the duplex in the process	73.80	2.85E-14	26%	44%	285--402	13--130	4	AAL00179	SWF/SNF family ATP-dependent RNA helicase	172.94	1.91E-41	29%	47%	7--439	583--1016
					5	COG1061	SSL2, DNA or RNA helicases of superfamily II [Transcription / DNA replication, recombination, and repair].	65.50	7.99E-12	21%	41%	1--412	31--394	5	AAX71394	phage-related DNA helicase	171.79	4.25E-41	27%	49%	10--444	589--1025
					6	cd00269	DEXHc, DEXH-box helicases. A diverse family of proteins involved in ATP-dependent DNA or RNA unwinding, needed in a variety of cellular processes. The name derives from the sequence of the Walker B motif (motif II). This domain contains the ATP-binding region..	63.56	3.00E-11	20%	42%	28--164	1--143	6	ZP_00366465	COG0553: Superfamily II DNA/RNA helicases, SNF2 family	171.79	4.25E-41	27%	49%	10--444	589--1025
					7	smart00490	HELICc, helicase superfamily c-terminal domain; .	62.56	6.01E-11	27%	46%	314--395	3--82	7	AAZ50907	SWF/SNF family helicase	171.40	5.55E-41	27%	49%	10--444	589--1025
					8	pfam00271	Helicase_C, Helicase conserved C-terminal domain. TThis domain family is found in a wide variety of helicases and helicase related proteins. It may be that this is not an autonomously folding unit, but an integral part of the helicase..	61.77	1.29E-10	29%	46%	317--395	2--78	8	NP_296465	helicase, Snf2 family	171.01	7.25E-41	30%	49%	2--437	712--1169
					9	cd00046	DEXDc, DEAD-like helicases superfamily. A diverse family of proteins involved in ATP-dependent RNA or DNA unwinding. This domain contains the ATP-binding region..	53.94	2.52E-08	17%	35%	29--164	2--144	9	AAL97089	putative SNF helicase	170.63	9.46E-41	27%	49%	10--444	588--1024
					10	COG0513	SrmB, Superfamily II DNA and RNA helicases [DNA replication, recombination, and repair / Transcription / Translation, ribosomal structure and biogenesis].	48.62	9.48E-07	19%	39%	280--446	255--420	10	AAK33394	putative SNF helicase	170.63	9.46E-41	27%	49%	10--444	589--1025
C667L	279550--279128	141	14,996	4.5	1	pfam00692	dUTPase, dUTPase. dUTPase hydrolyses dUTP to dUMP and pyrophosphate..	134.21	1.65E-32	52%	68%	13--140	3--129	1	NP_048907	similar to tomato dUTP pyrophosphatase, corresponds to GenBank Accesssion Number S40549	232.26	3.32E-60	80%	90%	1--141	1--141
					2	COG0717	Dcd, Deoxycytidine deaminase [Nucleotide transport and metabolism].	57.20	2.44E-09	29%	46%	28--118	69--158	2	AAW51452	deoxyuridine triphosphatase	226.87	1.40E-58	78%	89%	1--141	1--141
														3	AAW51453	deoxyuridine triphosphatase	224.56	6.93E-58	79%	89%	1--141	1--141
														4	EAL88001	dUTPase	171.01	9.09E-42	60%	78%	4--141	113--250
														5	XP_657875	hypothetical protein AN0271.2	166.78	1.71E-40	57%	79%	3--140	69--206
														6	NP_190278	dUTP diphosphatase/ hydrolase	163.70	1.45E-39	59%	77%	6--141	31--166
														7	BAE55800	unnamed protein product	163.70	1.45E-39	58%	77%	4--140	57--193
														8	XP_469212	putative deoxyuridine triphosphatase	162.93	2.48E-39	57%	78%	4--141	88--225
														9	AAB22611	deoxyuridine triphosphatase; dUTPase; P18	161.77	5.52E-39	58%	78%	6--141	34--169
														10	EAA67244	hypothetical protein FG00904.1	159.84	2.10E-38	59%	76%	3--140	31--168
C669R	279661--280611	317	36,435	9.79	1	pfam00352	TBP, Transcription factor TFIID (or TATA-binding protein, TBP)..	36.64	4.30E-03	22%	56%	216--283	22--85	1	NP_048908	similar to Sulfolobus TATA-binding protein, corresponds to GenBank Accession Number S55311'	459.91	4.60E-128	80%	89%	48--317	1--270
C670L	282103--280592	504	59,832	6.84	1	cd01992	PP-ATPase, N-terminal domain of predicted ATPase of the PP-loop faimly implicated in cell cycle control [Cell division and chromosome partitioning]. This is a subfamily of Adenine nucleotide alpha hydrolases superfamily.Adeninosine nucleotide alpha hydrolases superfamily includes N type ATP PPases and ATP sulphurylases. It forms a apha/beta/apha fold which binds to Adenosine group. This domain has a strongly conserved motif SGGXD at the N terminus.	114.94	1.22E-26	26%	47%	200--385	2--185	1	NP_048910	similar to MesJ cell cycle protein	996.50	0.00E+00	94%	96%	1--498	1--498
					2	pfam01171	ATP_bind_3, PP-loop family. This family of proteins belongs to the PP-loop superfamily..	107.26	2.43E-24	30%	53%	204--385	6--186	2	T18059	hypothetical protein A557L - Chlorella virus PBCV-1	213.00	1.90E-53	90%	93%	20--124	8--112
					3	COG0037	MesJ, Predicted ATPase of the PP-loop superfamily implicated in cell cycle control [Cell division and chromosome partitioning].	95.68	7.08E-21	22%	40%	184--489	6--298	3	T18058	hypothetical protein A556L - Chlorella virus PBCV-1	193.36	1.56E-47	89%	90%	121--223	4--106
					4	cd01993	Alpha_ANH_like_II, This is a subfamily of Adenine nucleotide alpha hydrolases superfamily.Adeninosine nucleotide alpha hydrolases superfamily includes N type ATP PPases and ATP sulphurylases. It forms a apha/beta/apha fold which binds to Adenosine group. This subfamily of proteins is predicted to bind ATP. This domainhas a strongly conserved motif SGGKD at the N terminus.	72.97	4.76E-14	26%	43%	200--371	2--180	4	CAD84933	conserved hypothetical protein	84.73	7.80E-15	25%	51%	195--417	24--244

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					5	cd01990	Alpha_ANH_like_I, This is a subfamily of Adenine nucleotide alpha hydrolases superfamily. Adenine nucleotide alpha hydrolases superfamily includes N type ATP PPases and ATP sulphurylases. It forms a apha/beta/apha fold which binds to Adenosine group. This subfamily of proteins probably binds ATP. This domain is about 200 amino acids long with a strongly conserved motif SGGKD at the N terminus..	38.28	1.37E-03	27%	40%	200--365	1--148	5	AAD35664	conserved hypothetical protein	81.65	6.60E-14	24%	47%	200--450	8--259
														6	YP_193201	putative cell cycle	80.11	1.92E-13	24%	44%	194--482	18--312
														7	CAG39532	conserved hypothetical protein	77.80	9.53E-13	24%	49%	200--423	14--237
														8	CAI80146	conserved hypothetical protein	77.41	1.24E-12	24%	48%	200--423	14--237
														9	ZP_00530972	PP-loop	75.49	4.73E-12	27%	45%	200--425	28--252
														10	BAE05811	unnamed protein product	75.49	4.73E-12	25%	45%	200--450	15--280
C675L	283377--282178	400	45,385	5.2	1	pfam04451	Capsid_Iridovir, Iridovirus major capsid protein. This family includes the major capsid protein of iridoviruses, chlorella virus and Spodoptera ascovirus, which are all dsDNA viruses with no RNA stage. This is the most abundant structural protein and can account for up to 45% of virion protein. In Chlorella virus AR158 the major capsid protein is a glycoprotein	427.42	1.07E-120	40%	57%	1--395	2--442	1	NP_048914	similar to Chlorella virus PBCV-1 ORF A11L, corresponds to GenBank Accession Number U42580	769.62	0.00E+00	93%	97%	1--400	1--400
														2	NP_048359	contains aminoacyl-tRNA synthetase class-II signature	419.47	9.77E-116	52%	70%	2--400	3--403
														3	NP_048358	similar to PBCV-1 major capsid protein, corresponds to Swiss-Prot Accession Number P30328	322.01	2.12E-86	41%	64%	2--400	4--401
														4	AAC27492	major capsid protein Vp49	266.16	1.38E-69	38%	54%	2--400	3--432
														5	NP_048787	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	248.05	3.90E-64	38%	53%	2--400	3--437
														6	BAA76601	major capsid protein MCP1	243.43	9.60E-63	37%	52%	2--400	3--437
														7	BAA76600	major capsid protein	242.66	1.64E-62	37%	53%	2--400	3--436
														8	BAA22198	major capsid protein Vp54	239.20	1.81E-61	36%	52%	2--400	3--437
														9	1M3Y_D	Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid Containing, Dna Virus	222.63	1.75E-56	36%	52%	24--400	1--413
														10	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	222.25	2.29E-56	36%	52%	24--400	1--413
C676L	284067--283435	211	23,244	10.31		No Hit Found								1	NP_048915	A559L	274.63	1.30E-72	70%	81%	1--211	1--213
C678L	285029--284070	320	36,583	8.22	1	smart00497	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished). .	42.43	7.26E-05	37%	60%	250--308	1--53	1	NP_048621	A267L	157.92	3.80E-37	33%	50%	13--288	36--302
					2	pfam07453	NUMOD1, NUMOD1 domain..	36.95	3.39E-03	36%	64%	250--283	1--34	2	NP_048846	Lys-, Glu-rich	76.64	1.11E-12	35%	45%	58--205	128--281
														3	NP_048834	Lys-, Arg-rich; contains eukaryotic putative RNA-binding region RNP-1 signature; similar to PBCV-1 ORF A267L, corresponds to GenBank Accession Number U42580	75.87	1.90E-12	33%	45%	56--214	126--290
														4	YP_142777	unknown	67.78	5.17E-10	46%	61%	132--205	268--342
														5	NP_048671	A315L	58.92	2.40E-07	40%	64%	243--312	183--246
C681L	287065--285113	651	71,617	10.08	1	pfam05887	Trypan_PARP, Procyclic acidic repetitive protein (PARP). This family consists of several Trypanosoma brucei procyclic acidic repetitive protein (PARP) like sequences. The procyclic acidic repetitive protein (parp) genes of Trypanosoma brucei encode a small family of abundant surface proteins whose expression is restricted to the procyclic form of the parasite. They are found at two unlinked loci, parpA and parpB; transcription of both loci is developmentally regulated	50.75	2.70E-07	22%	37%	252--317	33--98	1	BAA11342	DNA binding protein	628.63	1.98E-178	83%	88%	315--650	312--647
					2	pfam05335	DUF745, Protein of unknown function (DUF745). This family consists of several uncharacterised Drosophila melanogaster proteins of unknown function..	49.63	4.84E-07	26%	45%	117--271	21--187	2	NP_048917	similar to Chlorella virus CVK2 DNA binding protein, corresponds to GenBank Accession Number D78305	621.70	2.42E-176	82%	87%	315--651	314--649
					3	pfam01576	Myosin_tail_1, Myosin tail. The myosin molecule is a multi-subunit complex made up of two heavy chains and four light chains it is a fundamental contractile protein found in all eukaryote cell types. This family consists of the coiled-coil myosin heavy chain tail region. The coiled-coil is composed of the tail from two molecules of myosin. These can then assemble into the macromolecular thick filament. The coiled-coil region provides the structural backbone the thick filament	48.39	1.26E-06	22%	45%	58--275	609--828	3	NP_048921	A565R	221.09	9.54E-56	98%	98%	315--406	377--468
					4	pfam06519	TolA, TolA protein. This family consists of several bacterial TolA proteins as well as two eukaryotic proteins of unknown function. Tol proteins are involved in the translocation of group A colicins. Colicins are bacterial protein toxins, which are active against Escherichia coli and other related species (See pfam01024). TolA is anchored to the cytoplasmic membrane by a single membrane spanning segment near the N-terminus, leaving most of the protein exposed to the periplasm..	45.12	1.30E-05	27%	38%	95--265	102--260	4	BAA11343	DNA binding protein	219.16	3.62E-55	97%	97%	315--406	369--460
					5	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	44.75	1.53E-05	32%	42%	254--326	57--130	5	NP_048741	Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	109.00	5.28E-22	51%	58%	315--406	85--176
					6	COG1566	EmrA, Multidrug resistance efflux pump [Defense mechanisms].	43.79	2.73E-05	20%	41%	34--209	14--203	6	NP_048735	A378L	77.41	1.70E-12	34%	42%	325--400	139--245
					7	pfam02321	OEP, Outer membrane efflux protein. The OEP family (Outer membrane efflux protein) form trimeric channels that allow export of a variety of substrates in Gram negative bacteria. Each member of this family is composed of two repeats. The trimeric channel is composed of a 12 stranded all beta sheet barrel that spans the outer membrane, and a long all helical barrel that spans the periplasm	41.54	1.34E-04	23%	43%	83--235	21--171	7	BAB19127	vAL-1	69.32	4.64E-10	30%	44%	433--625	152--328
					8	pfam05616	Neisseria_TspB, Neisseria meningitidis TspB protein. This family consists of several Neisseria meningitidis TspB virulence factor proteins..	41.27	1.54E-04	35%	46%	279--331	335--387	8	BAA83789	alginate lyase	66.63	3.01E-09	30%	43%	433--625	136--312
					9	COG4372	COG4372, Uncharacterized protein conserved in bacteria with the myosin-like domain [Function unknown].	39.73	5.06E-04	20%	48%	130--271	97--225	9	NP_048562	PBCV-1 alginate lyase	65.08	8.75E-09	28%	43%	433--625	124--300
					10	COG0845	AcrA, Membrane-fusion protein [Cell envelope biogenesis, outer membrane].	39.33	5.85E-04	13%	31%	34--262	6--238	10	CAA64974	QI74 protein	57.38	1.82E-06	33%	46%	322--404	486--579
C685R	287109--289130	674	73,850	7.32	1	pfam05887	Trypan_PARP, Procyclic acidic repetitive protein (PARP). This family consists of several Trypanosoma brucei procyclic acidic repetitive protein (PARP) like sequences. The procyclic acidic repetitive protein (parp) genes of Trypanosoma brucei encode a small family of abundant surface proteins whose expression is restricted to the procyclic form of the parasite. They are found at two unlinked loci, parpA and parpB; transcription of both loci is developmentally regulated	54.98	1.39E-08	30%	34%	306--393	36--123	1	BAA11343	DNA binding protein	810.06	0.00E+00	76%	80%	23--549	23--540
					2	COG0419	SbcC, ATPase involved in DNA repair [DNA replication, recombination, and repair].	50.82	2.35E-07	19%	44%	34--255	248--465	2	NP_048921	A565R	606.29	1.10E-171	94%	97%	23--337	23--337
					3	pfam05616	Neisseria_TspB, Neisseria meningitidis TspB protein. This family consists of several Neisseria meningitidis TspB virulence factor proteins..	45.89	6.52E-06	34%	39%	320--384	310--377	3	NP_048917	similar to Chlorella virus CVK2 DNA binding protein, corresponds to GenBank Accession Number D78305	206.84	1.94E-51	98%	100%	383--467	314--398
					4	pfam01576	Myosin_tail_1, Myosin tail. The myosin molecule is a multi-subunit complex made up of two heavy chains and four light chains it is a fundamental contractile protein found in all eukaryote cell types. This family consists of the coiled-coil myosin heavy chain tail region. The coiled-coil is composed of the tail from two molecules of myosin. These can then assemble into the macromolecular thick filament. The coiled-coil region provides the structural backbone the thick filament	45.69	6.94E-06	25%	50%	33--182	680--822	4	BAA11342	DNA binding protein	205.68	4.33E-51	98%	98%	383--467	312--396
					5	pfam00769	ERM, Ezrin/radixin/moesin family. This family of proteins contain a band 4.1 domain (pfam00373), at their amino terminus. This family represents the rest of these proteins..	44.97	1.25E-05	20%	40%	34--240	97--305	5	NP_048741	Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	117.47	1.55E-24	34%	40%	286--465	9--167
					6	pfam04625	DEC-1_N, DEC-1 protein, N terminal region. The defective chorion-1 gene (dec-1) in Drosophila encodes follicle cell proteins necessary for proper eggshell assembly. Multiple products of the dec-1 gene are formed by alternative RNA splicing and proteolytic processing. Cleavage products include S80 (80 kDa) which is incorporated into the eggshell, and further proteolysis of S80 gives S60 (60 kDa)	42.95	4.86E-05	36%	40%	339--378	97--142	6	NP_048735	A378L	76.64	3.03E-12	34%	42%	393--467	139--244
					7	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	42.82	5.13E-05	39%	50%	338--385	68--114	7	XP_645158	hypothetical protein DDB0216970	54.68	1.23E-05	23%	42%	36--229	435--1649
					8	COG4487	COG4487, Uncharacterized protein conserved in bacteria [Function unknown].	42.71	6.95E-05	20%	33%	31--183	69--223	8	AAS08950	hypothetical protein LJ_1128	53.14	3.59E-05	27%	41%	62--256	568--2788
					9	COG0711	AtpF, F0F1-type ATP synthase, subunit b [Energy production and conversion].	42.17	9.05E-05	19%	46%	4--180	10--144	9	XP_786840	PREDICTED: similar to Early endosome antigen 1 (Endosome-associated protein p162) (Zinc finger FYVE domain containing protein 2), partial	51.99	8.00E-05	24%	48%	35--180	377--521
C687L	289567--289133	145	16,664	9.88		No Hit Found								1	NP_048923	A567L	207.61	8.62E-53	68%	82%	1--145	1--152

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
C688L	290107--289571	179	21,641	6.04		No Hit Found								1	NP_048924	A568L	273.48	1.96E-72	73%	84%	1--176	1--176
C690L	290529--290134	132	15,213	4.58		No Hit Found								1	NP_048926	A570L	221.09	7.70E-57	90%	96%	1--106	1--106
C691R	290601--290945	115	12,977	12.12		No Hit Found								1	NP_048927	A571R	193.74	1.32E-48	86%	95%	4--115	5--116
														2	NP_048792	Arg-rich	57.77	1.13E-07	48%	63%	6--65	10--68
C693R	290960--291502	181	20,711	6.7		No Hit Found								1	NP_048928	A572R	337.42	1.14E-91	84%	95%	1--180	1--180
C694L	292243--291509	245	27,915	4.21	1	pfam00705	PCNA_N, Proliferating cell nuclear antigen, N-terminal domain. N-terminal and C-terminal domains of PCNA are topologically identical. Three PCNA molecules are tightly associated to form a closed ring encircling duplex DNA..	79.93	3.65E-16	30%	51%	1--116	1--124	1	NP_048930	similar to Periwinkle PCNA, corresponds to GenBank Accession Number X55052	417.93	1.29E-115	83%	93%	1--244	15--258
					2	pfam02747	PCNA_C, Proliferating cell nuclear antigen, C-terminal domain. N-terminal and C-terminal domains of PCNA are topologically identical. Three PCNA molecules are tightly associated to form a closed ring encircling duplex DNA..	59.56	5.02E-10	25%	45%	122--240	2--128	2	Q9MAY3	Proliferating cell nuclear antigen (PCNA)	144.82	2.10E-33	31%	50%	1--245	1--259
					3	COG0592	DnaN, DNA polymerase sliding clamp subunit (PCNA homolog) [DNA replication, recombination, and repair].	53.38	3.56E-08	22%	40%	13--241	72--323	3	XP_468284	SPATULA-like	140.20	5.18E-32	30%	52%	1--245	1--259
														4	CAA77062	PCNA protein	139.43	8.84E-32	30%	51%	1--245	1--259
														5	AAD10528	proliferating cell nuclear antigen	138.27	1.97E-31	29%	52%	1--245	1--259
														6	CAA55669	proliferative cell nuclear antigen	137.89	2.57E-31	29%	52%	1--245	1--259
														7	AAG24908	proliferating cell nuclear antigen	136.73	5.73E-31	31%	50%	1--245	1--259
														8	NP_180517	PCNA2 (PROLIFERATING CELL NUCLEAR 2); DNA binding / DNA polymerase processivity factor	136.73	5.73E-31	29%	51%	1--245	1--259
														9	AAC27992	proliferating cell nuclear antigen	136.73	5.73E-31	30%	50%	1--245	1--259
														10	CAA38893	proliferating cell nuclear antigen	136.35	7.48E-31	30%	50%	1--245	1--259
C696R	292341--293426	362	41,201	7.83	1	pfam00145	DNA methylase, C-5 cytosine-specific DNA methylase..	193.22	3.26E-50	30%	45%	3--331	1--316	1	AAC64006	cytosine methyltransferase	749.20	0.00E+00	98%	99%	1--362	1--362
					2	cd00315	Cyt_C5_DNA_methylase, Cytosine-C5 specific DNA methylases; Methyl transfer reactions play an important role in many aspects of biology. Cytosine-specific DNA methylases are found both in prokaryotes and eukaryotes. DNA methylation, or the covalent addition of a methyl group to cytosine within the context of the CpG dinucleotide, has profound effects on the mammalian genome. These effects include transcriptional repression via inhibition of transcription factor binding or the recruitment of methyl-binding proteins and their associated chromatin remodeling factors, X chromosome inactivation, imprinting and the suppression of parasitic DNA sequences. DNA methylation is also essential for proper embryonic development and is an important player in both DNA repair and genome stability.	188.97	5.60E-49	34%	48%	3--241	1--250	2	AAC55063	cytosine methyltransferase	507.68	2.36E-142	65%	78%	1--362	1--366
					3	COG0270	Dcm, Site-specific DNA methylase [DNA replication, recombination, and repair].	120.18	3.03E-28	28%	46%	1--194	2--200	3	NP_049039	nonfunctional M.CviAV cytosine DNA methyltransferase	503.83	3.41E-141	64%	77%	1--362	1--366
														4	AAV84097	CviPII m5C DNA methyltransferase	320.09	6.99E-86	46%	61%	1--362	14--363
														5	NP_048873	M.CviAII cytosine DNA methyltransferase	296.59	8.28E-79	45%	58%	3--356	2--342
														6	NP_048886	M.CviAIV cytosine DNA methyltransferase	271.94	2.18E-71	46%	57%	1--344	1--332
														7	CAD80133	gp9.1	102.83	1.75E-20	36%	50%	6--163	4--162
														8	BAD65383	site-specific DNA-methyltransferase	96.29	1.64E-18	33%	48%	3--198	1--190
														9	AAC98421	methyl transferase	93.20	1.39E-17	34%	52%	3--158	1--165
														10	ZP_00874816	C-5 cytosine-specific DNA methylase	93.20	1.39E-17	35%	49%	3--158	1--165
C698L	293937--293434	168	19,008	8.64		No Hit Found								1	NP_048931	A575L	313.92	1.11E-84	89%	95%	1--168	1--168
C699L	294397--293999	133	15,329	11.27		No Hit Found								1	NP_048933	A577L	151.37	7.46E-36	84%	92%	29--110	2--83
C701R	294232--295266	345	40,180	8.49	1	pfam02086	MethyltransfD12, D12 class N6 adenine-specific DNA methyltransferase..	182.84	4.61E-47	35%	53%	89--328	1--253	1	NP_048937	PBCV-1 M.CviAI methylase	478.79	1.09E-133	86%	92%	82--345	1--265
					2	COG0338	Dam, Site-specific DNA methylase [DNA replication, recombination, and repair].	172.40	5.81E-44	41%	58%	82--331	1--256	2	ZP_00510571	N6 adenine-specific DNA methyltransferase, D12 class	189.12	1.73E-46	40%	58%	84--344	8--277
														3	EAM94529	N6 adenine-specific DNA methyltransferase, D12 class	181.80	2.75E-44	42%	59%	85--330	31--282
														4	ZP_00886307	DNA adenine methylase	171.40	3.72E-41	37%	58%	85--343	14--287
														5	NP_394115	Site-specific DNA methylase	171.01	4.86E-41	40%	60%	85--330	11--262
														6	ABA22276	DNA adenine methylase	167.93	4.12E-40	41%	54%	85--334	13--267
														7	CAC11782	probable site-specific DNA-methyltransferase (adenine-specific)	167.55	5.37E-40	40%	60%	87--330	1--250
														8	BAC09192	tlr1640	165.62	2.04E-39	37%	55%	80--329	4--259
														9	ZP_00789107	putative DNA adenine methylase	165.62	2.04E-39	39%	58%	84--330	14--271
														10	EAM93174	N6 adenine-specific DNA methyltransferase, D12 class	165.24	2.67E-39	35%	55%	85--343	12--279
C705L	296426--295275	384	43,844	7.82		No Hit Found								1	NP_048502	A154L	507.29	3.35E-142	67%	80%	40--384	3--347
														2	NP_048920	similar to Chlorella virus PBCV-1 ORF A154L, corresponds to GenBank Accession Number U42580	501.52	1.84E-140	66%	79%	37--384	1--350
														3	NP_048477	similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903	459.14	1.05E-127	61%	76%	39--384	4--356
														4	NP_077492	EsV-1-7	102.45	2.49E-20	30%	44%	37--272	1--261
														5	AAG37861	variant-specific surface protein VSP136-4	67.40	8.87E-10	25%	37%	24--240	499--744
														6	AAF69839	variant-specific surface protein VSP136b	66.63	1.51E-09	25%	37%	24--240	99--344
														7	AAA74587	cysteine rich protein	66.63	1.51E-09	25%	37%	24--240	99--344
														8	CAG07044	unnamed protein product	59.69	1.85E-07	29%	38%	17--184	371--545
														9	CAA64974	QI74 protein	53.53	1.32E-05	27%	35%	4--250	195--431
														10	NP_077600	EsV-1-115	50.83	8.59E-05	25%	37%	27--198	163--338
C707L	299673--296488	1062	120,593	8.02	1	smart00433	TOP2c, TopoisomeraseII; Eukaryotic DNA topoisomerase II, GyrB, ParE .	557.10	9.52E-160	35%	52%	50--620	1--594	1	NP_048939	PBCV-1 DNA topoisomerase II	1891.70	0.00E+00	88%	93%	3--1061	2--1061
					2	smart00434	TOP4c, DNA Topoisomerase IV; Bacterial DNA topoisomerase IV, GyrA, ParC .	444.63	5.71E-126	35%	53%	631--1055	1--456	2	AAU95770	topoisomerase II	1405.96	0.00E+00	65%	78%	1--1062	1--1057
					3	cd00187	TOP4c, DNA Topoisomerase, subtype IIA; domain A'; bacterial DNA topoisomerase IV (C subunit, ParC), bacterial DNA gyrases (A subunit, GyrA),mammalian DNA toposiomerases II. DNA topoisomerases are essential enzymes that regulate the conformational changes in DNA topology by catalysing the concerted breakage and rejoining of DNA strands during normal cellular growth.	419.19	3.08E-118	35%	53%	650--1061	2--445	3	CAD25222	DNA TOPOISOMERASE II	976.08	0.00E+00	48%	66%	5--1059	8--1067
					4	COG0187	GyrB, Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), B subunit [DNA replication, recombination, and repair].	381.40	7.12E-107	28%	45%	2--618	8--626	4	XP_467311	putative DNA topoisomerase II	927.93	0.00E+00	45%	60%	5--1060	32--1183
					5	pfam00521	DNA topoisoIV, DNA gyrase/topoisomerase IV, subunit A..	318.26	6.42E-88	28%	47%	651--1062	1--438	5	BAD86854	DNA topoisomerase II	927.55	0.00E+00	44%	61%	5--1062	103--1249
					6	COG0188	GyrA, Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), A subunit [DNA replication, recombination, and repair].	268.67	6.38E-73	24%	44%	630--1047	10--455	6	AAW40881	DNA topoisomerase II, putative	926.39	0.00E+00	45%	61%	5--1062	112--1235
					7	pfam00204	DNA_gyraseB, DNA gyrase B. This family represents the second domain of DNA gyrase B which has a ribosomal S5 domain 2-like fold. This family is structurally related to PF01119..	80.29	2.98E-16	29%	43%	236--343	1--125	7	CAA20107	SPBC1A4.03c	923.31	0.00E+00	44%	62%	1--1060	74--1197
					8	COG1389	COG1389, DNA topoisomerase VI, subunit B [DNA replication, recombination, and repair].	39.53	5.39E-04	34%	52%	50--144	36--125	8	NP_189031	TOPII (TOPOISOMERASE II); ATP binding / DNA binding / DNA topoisomerase (ATP-hydrolyzing)	922.54	0.00E+00	46%	62%	5--1061	35--1166
					9	pfam02518	HATPase_c, Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase. This family represents the structurally related ATPase domains of histidine kinase, DNA gyrase B and HSP90..	35.35	8.99E-03	17%	40%	51--149	6--92	9	CAA27857	unnamed protein product	921.77	0.00E+00	44%	62%	1--1060	20--1143
														10	BAE06274	topoisomerase II	921.00	0.00E+00	45%	60%	5--1061	35--1168
C714R	299913--300119	69	7,857	4.32		No Hit Found								1	NP_048948	A592R	111.69	6.46E-24	94%	94%	16--69	16--69
C715R	300242--301009	256	28,355	6.32	1	COG3889	COG3889, Predicted solute binding protein [General function prediction only].	36.17	6.12E-03	30%	43%	134--194	781--841	1	NP_048949	a593R	72.02	1.89E-11	61%	71%	87--143	1--57
C718R	301048--301473	142	16,088	7.95	1	cd01286	deoxycytidylate_deaminase, Deoxycytidylate deaminase domain. Deoxycytidylate deaminase catalyzes the deamination of dCMP to dUMP, providing the nucleotide substrate for thymidylate synthase. The enzyme binds Zn++, which is required for catalytic activity. The activity of the enzyme is allosterically regulated by the ratio of dCTP to dTTP not only in eukaryotic cells but also in T-even phage-infected Escherichia coli, with dCTP acting as an activator and dTTP as an inhibitor.	128.47	9.43E-31	40%	55%	5--133	2--131	1	NP_048952	similar to Vibrio fischeri dCMP deaminase, corresponds to Swiss-Prot Accession Number P33968	264.23	7.86E-70	86%	92%	1--142	1--142
					2	COG2131	ComEB, Deoxycytidylate deaminase [Nucleotide transport and metabolism].	102.73	4.77E-23	34%	53%	2--140	6--149	2	YP_223954	deoxycytidylate deaminase	121.32	8.23E-27	47%	64%	7--142	110--242

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					3	pfam00383	dCMP_cyt_deam, Cytidine and deoxycytidylate deaminase zinc-binding region..	82.72	5.46E-17	39%	55%	5--111	4--100	3	ZP_00052863	COG2131: Deoxycytidylate deaminase	102.83	3.03E-21	39%	57%	14--139	15--139
					4	cd00786	cytidine_deaminase-like, Cytidine and deoxycytidylate deaminase zinc-binding region. The family contains cytidine deaminases, nucleoside deaminases, deoxycytidylate deaminases and riboflavin deaminases. Also included are the apoBec family of mRNA editing enzymes. All members are Zn dependent. The zinc ion in the active site plays a central role in the proposed catalytic mechanism, activating a water molecule to form a hydroxide ion that performs a nucleophilic attack on the substrate..	62.34	7.16E-11	30%	47%	15--111	9--92	4	BAB80772	deoxycytidylate deaminase	102.06	5.17E-21	41%	54%	1--142	9--150
					5	cd01285	nucleoside_deaminase, Nucleoside deaminases include adenosine, guanine and cytosine deaminases. These enzymes are Zn dependent and catalyze the deamination of nucleosides. The zinc ion in the active site plays a central role in the proposed catalytic mec	52.57	7.06E-08	36%	59%	26--111	19--93	5	AAT75744	deoxycytidylate deaminase	101.68	6.75E-21	42%	53%	5--142	10--147
					6	cd01284	Riboflavin_deaminase-reductase, Riboflavin-specific deaminase. Riboflavin biosynthesis protein RibD (Diaminohydroxyphosphoribosylaminopyrimidine deaminase) catalyzes the deamination of 2,5-diamino-6-ribosylamino-4(3H)-pyrimidinone 5'-phosphate, which is an intermediate step in the biosynthesis of riboflavin.The ribG gene of Bacillus subtilis and the ribD gene of E. coli are bifunctional and contain this deaminase domain and a reductase domain which catalyzes the subsequent reduction of the ribosyl side chain	52.17	8.88E-08	32%	48%	10--113	3--95	6	BAE51501	Deoxycytidylate deaminase	100.91	1.15E-20	38%	57%	14--139	15--139
					7	COG0590	CumB, Cytosine/adenosine deaminases [Nucleotide transport and metabolism / Translation, ribosomal structure and biogenesis].	49.18	6.24E-07	39%	51%	30--111	34--104	7	XP_781375	PREDICTED: similar to Deoxycytidylate deaminase (dCMP deaminase)	100.91	1.15E-20	40%	55%	7--137	69--200
					8	COG0117	RibD, Pyrimidine deaminase [Coenzyme metabolism].	45.24	1.20E-05	29%	45%	5--140	7--126	8	NP_001006444	dCMP deaminase	100.52	1.50E-20	43%	57%	5--137	29--162
														9	AAR99137	RE06943p	100.52	1.50E-20	43%	56%	19--138	43--169
														10	BAC62535	putative deoxycytidylate deaminase	100.14	1.96E-20	40%	59%	4--142	6--141
C719L	302625--301480	382	43,898	7.74	1	COG0076	GadB, Glutamate decarboxylase and related PLP-dependent proteins [Amino acid transport and metabolism].	128.57	9.94E-31	25%	45%	72--341	94--383	1	NP_048954	similar to tomato histidine decarboxylase, corresponds to Swiss-Prot Accession Number P54772	666.00	0.00E+00	87%	94%	29--382	10--363
					2	pfam00282	Pyridoxal_deC, Pyridoxal-dependent decarboxylase conserved domain..	108.08	1.25E-24	24%	42%	89--338	88--372	2	NP_919502	putative histidine decarboxylase	220.32	8.12E-56	37%	55%	33--376	77--427
					3	COG1104	NifS, Cysteine sulfinate desulfinase/cysteine desulfurase and related enzymes [Amino acid transport and metabolism].	47.86	1.64E-06	30%	45%	72--259	39--219	3	ZP_00106716	COG0076: Glutamate decarboxylase and related PLP-dependent proteins	220.32	8.12E-56	35%	58%	41--380	25--368
					4	COG0520	CsdB, Selenocysteine lyase [Amino acid transport and metabolism].	43.78	2.67E-05	25%	47%	89--259	75--241	4	BAE07183	putative serine decarboxylase	207.22	7.11E-52	35%	55%	37--375	108--454
					5	COG1003	GcvP, Glycine cleavage system protein P (pyridoxal-binding), C-terminal domain [Amino acid transport and metabolism].	38.72	9.49E-04	24%	42%	108--379	134--413	5	BAA78331	serine decarboxylase	206.84	9.29E-52	36%	55%	37--375	112--458
														6	BAD28221	putative serine decarboxylase	205.30	2.70E-51	35%	55%	40--375	106--448
														7	XP_471202	OSJNBa0059H15.18	204.91	3.53E-51	37%	54%	41--374	69--410
														8	NP_175036	EMB1075; carboxy-lyase	202.99	1.34E-50	35%	55%	40--378	107--453
														9	CAA50719	histidine decarboxylase	202.60	1.75E-50	36%	54%	36--375	29--377
														10	ZP_00510529	Pyridoxal-dependent decarboxylase	197.59	5.64E-49	33%	55%	40--375	47--390
C722R	302615--302884	90	10,220	10.93		No Hit Found								1	NP_048957	A601R	115.55	4.49E-25	60%	70%	3--90	1--101
C723L	303292--302891	134	15,454	4.66		No Hit Found								1	NP_048958	A602L	84.73	8.54E-16	73%	90%	1--52	62--113
C725R	303434--303745	104	12,366	6.1		No Hit Found								1	NP_048959	A603R	161.00	9.19E-39	71%	86%	1--104	1--105
C726L	304286--303942	115	13,052	9.9		No Hit Found								1	NP_048960	A604L	83.96	1.47E-15	35%	60%	6--115	20--134
C727L	304835--304347	163	18,132	10.59		No Hit Found								1	NP_048961	A605L	226.10	2.74E-58	75%	82%	1--151	1--151
C728R	304880--306052	391	45,712	6.45	1	cd00204	ANK, ankyrin repeats; ankyrin repeats mediate protein-protein interactions in very diverse families of proteins. The number of ANK repeats in a protein can range from 2 to over 20 (ankyrins, for example). ANK repeats may occur in combinations with other types of domains. The structural repeat unit contains two antiparallel helices and a beta-hairpin, repeats are stacked in a superhelical arrangement; this alignment contains 4 consecutive repeats.	58.17	1.31E-09	28%	48%	63--173	6--113	1	NP_048963	contains 3 ankyrin repeat-like elements; similar to Drosophila ankyrin, corresponds to GenBank Accession Number L35601	386.34	8.86E-106	79%	89%	1--224	1--224
					2	COG0666	Arp, FOG: Ankyrin repeat [General function prediction only].	38.72	1.01E-03	20%	35%	1--223	9--235	2	NP_048964	A608R	243.82	7.10E-63	72%	87%	240--390	1--151
														3	EAL45248	ankyrin repeat protein, putative	73.94	9.73E-12	24%	42%	20--292	100--354
														4	NP_048786	A429L	64.70	5.91E-09	19%	41%	68--391	164--474
														5	EAA14062	ENSANGP00000013300	60.08	1.45E-07	26%	43%	13--290	344--599
														6	BAC43653	unknown protein	59.69	1.90E-07	25%	44%	1--210	173--368
														7	CAB10219	hypothetical protei	59.31	2.48E-07	25%	44%	1--210	234--429
														8	BAD43172	unknown protein	59.31	2.48E-07	25%	44%	1--210	239--434
														9	NP_567430	ACD6 (ACCELERATED CELL DEATH 6); protein binding	59.31	2.48E-07	25%	44%	1--210	239--434
														10	NP_849381	ACD6 (ACCELERATED CELL DEATH 6); protein binding	59.31	2.48E-07	25%	44%	1--210	173--368
C729L	307376--306135	414	46,638	4.98	1	COG1004	Ugd, Predicted UDP-glucose 6-dehydrogenase [Cell envelope biogenesis, outer membrane].	367.98	7.32E-103	33%	52%	28--412	2--412	1	NP_048965	PBCV-1 UDP-glucose dehydrogenase	689.49	0.00E+00	89%	93%	26--414	1--389
					2	COG0677	WecC, UDP-N-acetyl-D-mannosaminuronate dehydrogenase [Cell envelope biogenesis, outer membrane].	126.88	2.86E-30	25%	44%	25--395	8--394	2	AAK02860	unknown	416.77	6.61E-115	55%	71%	26--412	1--387
					3	pfam03721	UDPG_MGDP_dh_N, UDP-glucose/GDP-mannose dehydrogenase family, NAD binding domain. The UDP-glucose/GDP-mannose dehydrogenaseses are a small group of enzymes which possesses the ability to catalyse the NAD-dependent 2-fold oxidation of an alcohol to an acid without the release of an aldehyde intermediate.	123.42	3.32E-29	33%	50%	28--194	2--179	3	CAG21035	putative UDP-glucose dehydrogenase	416.00	1.13E-114	55%	72%	28--414	2--388
					4	pfam00984	UDPG_MGDP_dh, UDP-glucose/GDP-mannose dehydrogenase family, central domain. The UDP-glucose/GDP-mannose dehydrogenaseses are a small group of enzymes which possesses the ability to catalyse the NAD-dependent 2-fold oxidation of an alcohol to an acid without the release of an aldehyde intermediate.	89.48	5.14E-19	33%	54%	217--310	1--96	4	AAC67251	UDP-glucose dehydrogenase	415.62	1.47E-114	55%	71%	26--412	1--387
					5	COG0240	GpsA, Glycerol-3-phosphate dehydrogenase [Energy production and conversion].	52.09	9.31E-08	28%	48%	26--148	1--126	5	ZP_00538387	UDP-glucose 6-dehydrogenase	415.23	1.92E-114	55%	72%	28--411	2--385
					6	pfam03720	UDPG_MGDP_dh_C, UDP-glucose/GDP-mannose dehydrogenase family, UDP binding domain. The UDP-glucose/GDP-mannose dehydrogenaseses are a small group of enzymes which possesses the ability to catalyse the NAD-dependent 2-fold oxidation of an alcohol to an acid without the release of an aldehyde intermediate.	46.43	5.35E-06	25%	45%	327--389	1--68	6	ABB42102	UDP-glucose/GDP-mannose dehydrogenase	411.38	2.78E-113	54%	72%	28--414	2--390
					7	pfam01210	NAD_Gly3P_dh, NAD-dependent glycerol-3-phosphate dehydrogenase..	42.51	6.90E-05	31%	52%	28--111	2--83	7	ZP_00851874	UDP-glucose 6-dehydrogenase	410.99	3.63E-113	54%	72%	28--414	2--388
					8	COG1893	ApbA, Ketopantoate reductase [Coenzyme metabolism].	37.99	1.65E-03	20%	36%	27--107	1--75	8	ZP_00697254	COG1004: Predicted UDP-glucose 6-dehydrogenase	410.61	4.74E-113	53%	72%	28--414	2--388
					9	COG1250	FadB, 3-hydroxyacyl-CoA dehydrogenase [Lipid metabolism].	36.78	3.97E-03	24%	45%	25--148	2--123	9	ZP_00579286	UDP-glucose 6-dehydrogenase	408.68	1.80E-112	53%	72%	28--412	10--394
					10	COG1748	LYS9, Saccharopine dehydrogenase and related proteins [Amino acid transport and metabolism].	36.81	3.99E-03	33%	51%	26--113	1--82	10	AAV74384	Ugd	408.68	1.80E-112	53%	72%	28--414	2--388
C731L	307736--307380	119	13,556	9.24	1	smart00317	SET, SET (Su(var)3-9, Enhancer-of-zeste, Trithorax) domain; Putative methyl transferase, based on outlier plant homologues .	68.50	9.71E-13	27%	42%	5--112	2--125	1	NP_048968	PBCV-1 histone H3-Lys 27 methyltransferase (vSET)	219.94	1.69E-56	84%	92%	1--119	1--119
					2	COG2940	COG2940, Proteins containing SET domain [General function prediction only].	52.03	9.77E-08	25%	40%	6--116	334--459	2	ZP_00661322	Nuclear protein SET	75.10	6.72E-13	34%	59%	5--113	37--149
					3	pfam00856	SET, SET domain. SET domains are protein lysine methyltransferase enzymes. SET domains appear to be protein-protein interaction domains. It has been demonstrated that SET domains mediate interactions with a family of proteins that display similarity with dual-specificity phosphatases (dsPTPases). A subset of SET domains have been called PR domains. These domains are divergent in sequence from other SET domains, but also appear to mediate protein-protein interaction..	51.60	1.41E-07	23%	39%	2--108	6--129	3	AAM72187	conserved hypothetical protein	71.63	7.43E-12	34%	58%	6--113	39--150
														4	ZP_00528743	Nuclear protein SET	70.86	1.27E-11	34%	55%	6--113	42--153
														5	ZP_00588496	Nuclear protein SET	69.71	2.82E-11	33%	55%	6--113	39--150
														6	NP_701503	hypothetical protein PFL0690c	65.08	6.95E-10	30%	44%	2--111	29--175
														7	EAM63832	Nuclear protein SET	65.08	6.95E-10	33%	53%	6--113	39--150
														8	ABB28752	Nuclear protein SET	65.08	6.95E-10	33%	57%	6--113	48--159
														9	ZP_00511449	Nuclear protein SET	64.70	9.08E-10	32%	56%	6--113	38--149
														10	ABB23988	Nuclear protein SET	63.16	2.64E-09	29%	57%	5--113	38--150

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
C733L	309296--307770	509	57,714	11.36	1	cd00180	S_TKc, Serine/Threonine protein kinases, catalytic domain. Phosphotransferases of the serine or threonine-specific kinase subfamily. The enzymatic activity of these protein kinases is controlled by phosphorylation of specific residues in the activation segment of the catalytic domain, sometimes combined with reversible conformational changes in the C-terminal autoregulatory tail	62.92	5.22E-11	28%	49%	62--239	6--152	1	NP_048970	RPQT-like (9x)	702.59	0.00E+00	62%	73%	1--509	1--577
					2	smart00220	S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-specific kinase subfamily. .	61.77	1.09E-10	29%	47%	62--239	5--151	2	NP_048632	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662	302.75	1.83E-80	47%	64%	6--330	10--344
					3	pfam00069	Pkinase, Protein kinase domain..	55.29	9.95E-09	31%	48%	62--236	5--148	3	NP_048636	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055	279.26	2.17E-73	40%	58%	6--377	24--406
					4	COG0515	SPS1, Serine/threonine protein kinase [General function prediction only / Signal transduction mechanisms / Transcription / DNA replication, recombination, and repair].	48.23	1.21E-06	15%	30%	105--431	23--380	4	XP_644812	hypothetical protein DDB0217139	90.89	1.09E-16	36%	63%	312--430	043--1166
					5	pfam01636	APH, Phosphotransferase enzyme family. This family consists of bacterial antibiotic resistance proteins, which confer resistance to various aminoglycosides they include:- aminoglycoside 3'-phosphotransferase or kanamycin kinase / neomycin-kanamycin phosphotransferase and streptomycin 3''-kinase or streptomycin 3''-phosphotransferase. The aminoglycoside phosphotransferases inactivate aminoglycoside antibiotics via phosphorylation. This family also includes homoserine kinase. This family is related to fructosamine kinase pfam03881	41.67	1.37E-04	32%	68%	207--234	171--199	5	ZP_00788171	pathogenicity protein, putative	87.81	9.26E-16	25%	59%	303--447	116--259
					6	COG3087	FtsN, Cell division protein [Cell division and chromosome partitioning].	40.79	2.37E-04	18%	45%	314--407	94--190	6	XP_641859	hypothetical protein DDB0205029	86.27	2.70E-15	50%	74%	326--409	359--443
					7	smart00219	TyrKc, Tyrosine kinase, catalytic domain; Phosphotransferases. Tyrosine-specific kinase subfamily. .	37.85	1.72E-03	23%	41%	62--239	5--153	7	ZP_00783686	pathogenicity protein, putative	85.89	3.52E-15	28%	65%	303--409	580--688
					8	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	37.05	3.09E-03	25%	44%	318--412	40--133	8	ZP_00789197	Gram positive anchor domain protein	85.50	4.60E-15	29%	63%	303--410	111--216
					9	COG0478	COG0478, RIO-like serine/threonine protein kinase fused to N-terminal HTH domain [Signal transduction mechanisms].	36.76	3.65E-03	22%	43%	153--252	169--269	9	ZP_00780097	surface protein PspC	84.73	7.84E-15	34%	72%	325--409	403--486
					10	COG2334	COG2334, Putative homoserine kinase type II (protein kinase fold) [General function prediction only].	36.86	4.05E-03	46%	62%	207--233	200--226	10	XP_580236	PREDICTED: hypothetical protein XP_580236	83.96	1.34E-14	42%	69%	324--427	32--132
C735R	309366--310322	319	37,122	9.68		No Hit Found								1	NP_048973	similar to Variola virus orf E10L, corresponds to Swiss-Prot Accession Number P33801	543.89	2.45E-153	83%	93%	1--318	1--318
														2	YP_142754	S/T protein kinase, similar to Paramecium bursaria chlorella virus 1 A617R	90.51	7.41E-17	27%	44%	11--303	107--412
														3	NP_149843	380R	53.14	1.31E-05	29%	42%	42--247	182--378
C736L	310705--310331	125	14,495	3.94		No Hit Found								1	NP_048974	A618L	174.48	8.22E-43	72%	78%	1--123	1--129
C737L	311378--310725	218	25,446	4.41		No Hit Found								1	NP_048975	A619L	219.94	4.10E-56	49%	54%	1--218	1--237
C739L	311670--311422	83	9,614	9.35		No Hit Found								1	NP_048976	similar to Synechocystis orf 90, corresponds to GenBank Accession Number D90902	157.15	1.33E-37	89%	92%	1--83	1--83
														2	NP_048991	A635R	72.79	3.30E-12	44%	66%	1--77	1--82
C740L	312048--311698	117	12,936	9.89		No Hit Found								1	NP_048977	A621L	230.72	9.64E-60	99%	100%	1--117	1--117
C741L	313667--312108	520	58,224	5.52	1	pfam04451	Capsid_Iridovir, Iridovirus major capsid protein. This family includes the major capsid protein of iridoviruses, chlorella virus and Spodoptera ascovirus, which are all dsDNA viruses with no RNA stage. This is the most abundant structural protein and can account for up to 45% of virion protein. In Chlorella virus AR158 the major capsid protein is a glycoprotein	366.17	2.45E-102	43%	56%	174--516	86--443	1	NP_048978	similar to Simulium iridescent virus capsid protein, corresponds to Swiss-Prot Accession Number P22166	1035.02	0.00E+00	95%	97%	1--520	1--520
														2	AAC27493	putative capsid protein	763.84	0.00E+00	71%	81%	1--520	1--521
														3	BAE06835	hypothetical major capsid protein	266.93	1.15E-69	40%	59%	180--520	94--440
														4	BAA76601	major capsid protein MCP1	202.60	2.66E-50	35%	50%	188--520	92--437
														5	BAA76600	major capsid protein	200.68	1.01E-49	35%	49%	188--520	92--436
														6	NP_048787	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	199.52	2.25E-49	35%	49%	188--520	92--437
														7	1M3Y_D	Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid Containing, Dna Virus	199.52	2.25E-49	35%	49%	188--520	68--413
														8	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	198.36	5.01E-49	34%	49%	188--520	68--413
														9	AAC27492	major capsid protein Vp49	197.21	1.12E-48	35%	52%	188--520	89--432
														10	BAA22198	major capsid protein Vp54	196.44	1.90E-48	34%	49%	188--520	92--437
C742L	313920--313720	67	7,583	8.73	1	pfam01428	zf-AN1, AN1-like Zinc finger. Zinc finger at the C-terminus of An1, a ubiquitin-like protein in Xenopus laevis. The following pattern describes the zinc finger. C-X2-C-X(9-12)-C-X(1-2)-C-X4-C-X2-H-X5-H-X-C Where X can be any amino acid, and numbers in brackets indicate the number of residues	56.12	5.81E-09	53%	60%	12--52	1--41	1	NP_048979	similar to frog ubiquitin-like fusion protein, corresponds to Accession Number JN0673	135.58	4.20E-31	95%	95%	1--67	1--67
					2	smart00154	ZnF_AN1, AN1-like Zinc finger; Zinc finger at the C-terminus of An1, a ubiquitin-like protein in Xenopus laevis. .	45.75	7.04E-06	57%	62%	12--48	1--38	2	NP_194268	DNA binding / zinc ion binding	60.46	1.72E-08	62%	74%	11--45	70--104
														3	XP_469956	putative zinc finger protein	57.38	1.46E-07	47%	57%	10--67	107--169
														4	CAI76168	hypothetical protein, conserved	56.61	2.48E-07	49%	61%	10--63	110--168
														5	XP_469958	putative zinc finger protein	56.61	2.48E-07	44%	57%	8--67	173--237
														6	NP_957243	zinc finger, A20 domain containing 2, like	56.61	2.48E-07	45%	64%	10--63	151--209
														7	XP_466086	putative multiple stress-responsive zinc-finger protein	55.84	4.23E-07	51%	68%	4--44	86--126
														8	XP_482578	putative zinc finger protein	55.84	4.23E-07	55%	67%	10--52	163--204
														9	NP_565844	DNA binding / zinc ion binding	55.45	5.53E-07	57%	71%	10--44	100--134
														10	1WFH_A	Chain A, Solution Structrue Of The Zf-An1 Domain From Arabidopsis Thaliana At2g36320 Protein	55.45	5.53E-07	57%	71%	10--44	16--50
C744R	313951--314313	121	13,585	9.7	1	COG4852	COG4852, Predicted membrane protein [Function unknown].	42.98	5.33E-05	30%	44%	14--120	13--124	1	NP_048980	A624R	212.62	2.68E-54	88%	89%	1--121	1--121
														2	ZP_00234461	conserved hypothetical protein	53.91	1.59E-06	26%	45%	4--120	3--125
														3	CAC95830	lin0598	53.14	2.71E-06	26%	45%	4--120	3--125
														4	ZP_00231099	conserved hypothetical protein	52.76	3.54E-06	25%	45%	4--120	3--125
														5	CAC98668	lmo0589	51.99	6.05E-06	25%	45%	4--120	3--125
C745R	314385--315680	432	49,946	10.88	1	COG0675	COG0675, Transposase and inactivated derivatives [DNA replication, recombination, and repair].	86.29	4.85E-18	24%	39%	38--415	1--348	1	NP_048981	similar to Synechocystis transposase, corresponds to GenBank Accession Number D90909	845.88	0.00E+00	100%	100%	22--432	23--433
					2	pfam07282	Transposase_35, Putative transposase DNA-binding domain. This putative domain is found at the C-terminus of a large number of transposase proteins. This domain contains four conserved cysteines suggestive of a zinc binding domain. Given the need for transposases to bind DNA as well as the large number of DNA-binding zinc fingers we hypothesise this domain is DNA-binding	74.50	1.92E-14	41%	56%	345--412	1--69	2	AAU06281	putative transposase	721.08	0.00E+00	85%	91%	24--432	31--439
					3	pfam01385	Transposase_2, Probable transposase. This family includes IS891, IS1136 and IS1341. .	51.46	1.45E-07	22%	40%	43--334	1--278	3	ABA24789	Transposase, IS891/IS1136/IS1341	100.52	1.12E-19	27%	42%	49--414	155--533
														4	BAB78230	transposase	100.52	1.12E-19	27%	42%	49--414	111--489
														5	AAS54227	AGL264Wp	98.21	5.54E-19	25%	43%	42--414	71--453
														6	YP_142458	putative transposase	97.83	7.23E-19	26%	41%	39--412	130--535
														7	YP_238637	ORF021	92.05	3.97E-17	26%	44%	41--409	3--364
														8	ZP_00158267	COG0675: Transposase and inactivated derivatives	90.51	1.15E-16	34%	48%	229--414	167--359
														9	ZP_00766186	Transposase, IS605 OrfB	89.74	1.97E-16	27%	43%	37--421	2--367
														10	AAS40029	transposase, IS605 family	77.41	1.01E-12	23%	45%	38--414	1--375
C747R	315897--317213	439	48,966	11.18		No Hit Found								1	NP_048983	similar to Chlorella virus PBCV-1 ORF A231L, corresponds to GenBank Accession Number U42580	816.99	0.00E+00	92%	93%	1--439	1--441
														2	NP_048579	contains ATP/GTP-binding motif A	301.98	2.58E-80	43%	63%	101--425	10--338
C748R	317282--319582	767	86,005	7.1	1	cd01679	RNR_1, RNR, class I. Ribonucleotide reductase (RNR) catalyzes the reductive synthesis of deoxyribonucleotides from their corresponding ribonucleotides. It provides the precursors necessary for DNA synthesis. RNRs are separated into three classes based on	723.92	0.00E+00	50%	67%	177--747	1--578	1	NP_048985	similar to Schizosaccharomyces ribonucleotide reductase M1 chain, corresponds to Swiss-Prot Accession Number P36602	1404.04	0.00E+00	91%	94%	2--767	6--771
					2	pfam02867	Ribonuc_red_lgC, Ribonucleotide reductase, barrel domain..	650.43	0.00E+00	50%	66%	225--749	1--532	2	CAA79574	Hypothetical protein T23G5.1	756.13	0.00E+00	48%	68%	12--763	9--763
					3	COG0209	NrdA, Ribonucleotide reductase, alpha subunit [Nucleotide transport and metabolism].	543.07	1.68E-155	37%	53%	66--766	7--696	3	CAB98233	ribonucleoside-diphosphate reductase large chain (un-24)	755.36	0.00E+00	49%	67%	10--763	1--758
					4	cd02888	RNR_1_like, RNR, class I_like family. Ribonucleotide reductase (RNR) catalyzes the reductive synthesis of deoxyribonucleotides from their corresponding ribonucleotides. It provides the precursors necessary for DNA synthesis. RNRs are separated into three	321.38	7.14E-89	35%	53%	197--744	2--521	4	AAD49743	ribonucleotide reductase large subunit	755.36	0.00E+00	49%	67%	10--763	1--758

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					5	pfam00317	Ribonuc red lgN, Ribonucleotide reductase, all-alpha domain..	84.52	1.57E-17	42%	58%	151--223	1--78	5	CAE65263	Hypothetical protein CBG10154	753.82	0.00E+00	49%	68%	12--763	9--763
					6	cd00576	RNR_PFL, RNR_PFL. Ribonucleotide reductase (RNR) and pyruvate formate lyase (PFL) have a structurally similar ten-stranded alpha-beta barrel active site domain and are believed to have diverged from a common ancestor. RNRs are found in all organisms and provide the only mechanism by which nucleotides are converted to deoxynucleotides, while PFL, an essential enzyme in anaerobic bacteria, catalyzes the conversion of pyruvate and CoA to acteylCoA and formate. Both RNR and PFL are glycyl radical enzymes.	61.28	1.70E-10	22%	38%	250--645	70--432	6	XP_370503	hypothetical protein MG07000.4	752.28	0.00E+00	48%	67%	10--763	1--758
					7	pfam03477	ATP-cone, ATP cone domain..	54.22	2.07E-08	31%	48%	10--111	1--89	7	EAL90119	ribonucleotide reductase large subunit (Rnr1), putative	749.58	0.00E+00	49%	67%	10--763	1--758
														8	CAA91952	cdc22	748.81	0.00E+00	49%	67%	10--759	1--752
														9	NP_001026008	ribonucleoside-diphosphate reductase M1 chain	748.04	0.00E+00	49%	68%	10--763	1--757
														10	BAE59411	unnamed protein product	744.58	0.00E+00	49%	67%	10--763	1--758
C753R	319620--319979	120	13,219	8.43		No Hit Found								1	NP_048989	A633R	204.14	9.56E-52	82%	87%	1--120	1--120
C754L	320387--319986	134	15,587	9.64		No Hit Found								1	NP_048990	A634L	259.23	2.53E-68	85%	96%	1--134	1--134
C755R	320430--320684	85	9,924	9.8		No Hit Found								1	NP_048991	A635R	173.33	1.85E-42	97%	98%	1--85	1--85
														2	NP_048976	similar to Synechocystis orf 90, corresponds to GenBank Accession Number D90902	72.79	3.40E-12	43%	63%	1--82	1--77
C756R	320741--321019	93	10,922	5.28		No Hit Found								1	NP_048992	A636R	119.40	3.08E-26	68%	75%	1--93	1--96
C757R	321093--321515	141	16,474	8.62		No Hit Found								1	NP_048993	A637R	238.04	6.06E-62	83%	85%	1--141	1--141
C759R	321562--322638	359	40,854	5.87	1	pfam04371	PAD_porph, Porphyromonas-type peptidyl-arginine deiminase. Peptidyl-arginine deiminase (PAD) enzymes catalyse the deimination of the guanidino group from carboxy-terminal arginine residues of various peptides to produce ammonia. PAD from Porphyromonas gi	474.32	7.33E-135	55%	73%	8--358	1--329	1	NP_048994	PBCV-1 Agmatine iminohydrolase	728.78	0.00E+00	96%	98%	1--359	1--359
					2	COG2957	COG2957, Peptidylarginine deiminase and related enzymes [Amino acid transport and metabolism].	397.35	1.17E-111	49%	68%	2--358	9--344	2	CAC98253	lmo0038	386.34	7.84E-106	52%	68%	4--358	9--362
														3	ZP_00232718	peptidyl-arginine deiminase-like protein	385.57	1.34E-105	52%	68%	4--358	9--362
														4	ZP_00229955	peptidyl-arginine deiminase-like protein	384.80	2.28E-105	52%	68%	4--358	9--362
														5	ZP_00322658	COG2957: Peptidylarginine deiminase and related enzymes	377.10	4.76E-103	51%	68%	2--358	7--362
														6	AAL98713	LabD	372.09	1.53E-101	50%	67%	2--358	7--362
														7	CAI54369	Putative peptidylarginine deiminase (Amidinotransferase)	372.09	1.53E-101	50%	67%	2--358	17--372
														8	YP_424617	peptidylarginine deiminase-related protein, putative	371.32	2.61E-101	51%	66%	2--358	8--361
														9	AAG03681	agmatine deiminase	369.78	7.60E-101	50%	65%	2--358	7--363
														10	ZP_00972430	COG2957: Peptidylarginine deiminase and related enzymes	368.62	1.69E-100	50%	65%	2--358	7--363
C763R	322660--324090	477	54,458	11.48	1	COG4487	COG4487, Uncharacterized protein conserved in bacteria [Function unknown].	45.80	6.47E-06	22%	44%	255--401	64--216	1	NP_048999	Gln-rich; KQQ (6X)	298.52	3.18E-79	62%	64%	197--477	1--269
					2	COG0488	Uup, ATPase components of ABC transporters with duplicated ATPase domains [General function prediction only].	44.09	2.49E-05	16%	43%	327--408	236--317	2	NP_048998	A642R	119.78	2.03E-25	100%	100%	1--56	1--56
C765R	324132--324650	173	19,150	7.3		No Hit Found								1	NP_049000	A644R	154.45	1.22E-36	88%	94%	87--173	1--87
														2	YP_142763	unknown	50.45	2.47E-05	27%	49%	52--171	61--184
C766R	324772--325140	123	14,332	7.32		No Hit Found								1	NP_049001	A645R	217.24	1.12E-55	82%	90%	1--122	1--122
C767L	325747--325157	197	22,631	5.68	1	pfam00583	Acetyltransf_1, Acetyltransferase (GNAT) family. This family contains proteins with N-acetyltransferase functions..	39.47	5.68E-04	29%	43%	97--163	12--82	1	NP_049010	A654L	355.53	5.04E-97	86%	94%	1--197	1--197
					2	COG0456	RimI, Acetyltransferases [General function prediction only].	36.52	4.56E-03	23%	39%	110--182	94--168	2	ZP_00675288	hypothetical protein TeryDRAFT 0585	59.31	7.46E-08	28%	46%	23--184	28--202
														3	XP_641184	hypothetical protein DDB0205209	52.37	9.12E-06	23%	41%	7--184	14--204
														4	AAF57583	CG18607-PA	49.29	7.72E-05	28%	46%	65--174	89--200
C771L	326193--325789	135	15,462	4.58		No Hit Found								1	NP_049012	A656L	175.64	3.66E-43	71%	85%	17--130	1--114
C772L	326898--326347	184	20,616	3.45		No Hit Found								1	NP_049015	A659L	191.05	1.39E-47	75%	78%	1--122	1--128
C773L	327432--326920	171	19,547	10.86		No Hit Found								1	NP_049018	A662L	274.25	1.01E-72	74%	90%	1--171	1--171
														2	BAD87006	unknown protein	71.25	1.30E-11	30%	49%	47--170	115--239
														3	NP_565983	unknown protein	64.31	1.59E-09	25%	51%	47--170	108--232
														4	AAM62733	contains similarity to 22 kDa peroxisomal membrane protein	55.84	5.66E-07	23%	50%	20--165	95--245
														5	NP_568621	unknown protein	55.84	5.66E-07	23%	50%	20--165	94--244
														6	XP_635703	hypothetical protein DDB0189006	55.45	7.39E-07	27%	45%	4--158	21--172
														7	BAB08278	unnamed protein product	55.07	9.65E-07	25%	52%	20--145	94--224
														8	CAB80107	putative protein	54.30	1.65E-06	25%	50%	20--151	98--234
														9	NP_567940	unknown protein	52.76	4.79E-06	25%	49%	20--147	98--230
														10	NP_957459	MpV17 transgene, murine homolog, glomerulosclerosis	52.76	4.79E-06	26%	49%	24--158	14--153
C774L	328659--327490	390	10,635	11.27		No Hit Found								1	NP_048920	similar to Chlorella virus PBCV-1 ORF A154L, corresponds to GenBank Accession Number U42580	532.72	7.63E-150	70%	80%	43--389	4--350
														2	NP_048477	similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903	504.60	2.22E-141	61%	71%	1--389	1--356
														3	NP_048502	A154L	502.29	1.10E-140	67%	78%	43--389	3--347
														4	NP_077492	EsV-1-7	89.74	1.71E-16	29%	49%	12--214	13--222
														5	AAG37861	variant-specific surface protein VSP136-4	67.01	1.19E-09	25%	37%	26--245	139--384
														6	AAF69839	variant-specific surface protein VSP136b	67.01	1.19E-09	25%	37%	26--245	59--304
														7	AAA74587	cysteine rich protein	67.01	1.19E-09	25%	37%	26--245	59--304
														8	XP_605581	PREDICTED: similar to laminin, beta 4	58.15	5.51E-07	24%	34%	7--281	690--957
														9	AAA99804	220 kDa silk protein	54.68	6.09E-06	27%	35%	5--227	115--1370
														10	CAA36506	balbiani ring 3 (BR3)	54.68	6.09E-06	23%	35%	4--208	087--1276
C781L	329188--328736	151	17,077	7.18		No Hit Found								1	NP_049020	similar to Chlorella virus CVK2 unknown ORF, corresponds to GenBank Accession Number D16505	171.79	5.27E-42	63%	80%	15--149	1--136
														2	NP_049021	similar to Chlorella virus CVK2 unknown ORF, corresponds to GenBank Accession Number D16505	142.12	4.47E-33	48%	71%	1--150	21--169
C783L	329707--329255	151	17,165	6.9		No Hit Found								1	NP_049021	similar to Chlorella virus CVK2 unknown ORF, corresponds to GenBank Accession Number D16505	266.93	1.20E-70	86%	94%	1--151	21--171
														2	NP_049020	similar to Chlorella virus CVK2 unknown ORF, corresponds to GenBank Accession Number D16505	136.73	1.88E-31	51%	69%	15--150	1--138
C785L	332366--329757	870	95,770	6.32	1	COG0474	MgtA, Cation transport ATPase [Inorganic ion transport and metabolism].	528.01	5.47E-151	30%	49%	12--861	27--909	1	XP_636219	hypothetical protein DDB0188438	511.92	3.81E-143	34%	54%	36--855	66--977
					2	COG2217	ZntA, Cation transport ATPase [Inorganic ion transport and metabolism].	237.82	1.06E-63	30%	51%	121--682	201--678	2	AAL73984	type IIB calcium ATPase	497.28	9.70E-139	35%	54%	27--863	131--1012
					3	pfam00122	E1-E2 ATPase, E1-E2 ATPase..	166.83	2.59E-42	35%	56%	99--332	1--223	3	XP_483341	putative calcium-transporting ATPase 8, plasma membrane-type	493.81	1.07E-137	36%	53%	42--850	182--1053
					4	COG2216	KdpB, High-affinity K+ transport system, ATPase chain B [Inorganic ion transport and metabolism].	161.52	1.06E-40	28%	49%	127--701	100--607	4	XP_473800	OSJNBb0015N08.12	490.35	1.19E-136	35%	53%	31--853	162--1044
					5	pfam00702	Hydrolase, haloacid dehalogenase-like hydrolase. This family are structurally different from the alpha/ beta hydrolase family (pfam00561). This family includes L-2-haloacid dehalogenase, epoxide hydrolases and phosphatases. The structure of the family consists of two domains. One is an inserted four helix bundle, which is the least well conserved region of the alignment. The rest of the fold is composed of the core alpha/beta domain.	60.38	3.06E-10	22%	45%	487--632	68--197	5	NP_191292	calcium-transporting ATPase/ calmodulin binding	488.03	5.89E-136	35%	55%	29--863	132--1003
					6	pfam00689	Cation_ATPase_C, Cation transporting ATPase, C-terminus. Members of this families are involved in Na+/K+, H+/K+, Ca++ and Mg++ transport..	58.33	1.11E-09	21%	41%	727--851	1--147	6	NP_851200	ACA8 (AUTOINHIBITED CA2+ -ATPASE, ISOFORM 8); calcium-transporting ATPase/ calmodulin binding	488.03	5.89E-136	35%	53%	31--850	157--1032
					7	COG4087	COG4087, Soluble P-type ATPase [General function prediction only].	44.58	1.81E-05	30%	48%	558--645	47--138	7	NP_181687	ACA4 (AUTO-INHIBITED CA(2+)-ATPASE, ISOFORM 4); calcium-transporting ATPase/ calmodulin binding	484.18	8.50E-135	35%	54%	29--863	132--1006
					8	COG0561	Cof, Predicted hydrolases of the HAD superfamily [General function prediction only].	43.55	3.40E-05	33%	55%	606--657	208--259	8	CAA68234	calmodulin-stimulated calcium-ATPase	483.80	1.11E-134	34%	54%	29--863	132--1003
					9	pfam00690	Cation_ATPase_N, Cation transporter/ATPase, N-terminus. Members of this families are involved in Na+/K+, H+/K+, Ca++ and Mg++ transport..	42.50	6.66E-05	23%	47%	21--80	21--81	9	AAL17949	type IIB calcium ATPase	483.41	1.45E-134	35%	53%	12--863	102--1009
														10	NP_188755	ACA9; calcium-transporting ATPase/ calmodulin binding	480.72	9.40E-134	35%	53%	31--853	171--1052

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
C788L	335290--332537	918	103,098	6.2	1	COG0488	Uup, ATPase components of ABC transporters with duplicated ATPase domains [General function prediction only].	243.63	2.24E-65	27%	42%	314--666	2--425	1	NP_049022	Chlorella virus CVK2 translation elongation factor-3 homolog, refer to GenBank Accession Number D16505	1728.76	0.00E+00	93%	95%	1--918	1--918
					2	pfam00005	ABC_tran, ABC transporter. ABC transporters for a large family of proteins responsible for translocation of a variety of compounds across biological membranes. ABC transporters are the largest family of proteins in many completely sequenced bacteria. ABC transporters are composed of two copies of this domain and two copies of a transmembrane domain pfam00664. These four domains may belong to a single polypeptide, or belong in different polypeptide chains	121.17	1.41E-28	33%	50%	341--505	1--181	2	A48779	translation elongation factor EF-3 homolog - Chlorella virus CVK2	1649.80	0.00E+00	92%	94%	1--892	218--1110
					3	cd00267	ABC_ATPase, ABC (ATP-binding cassette) transporter nucleotide-binding domain; ABC transporters are a large family of proteins involved in the transport of a wide variety of different compounds, like sugars, ions, peptides and more complex organic molecules. The nucleotide binding domain shows the highest similarity between all members of the family. ABC transporters are a subset of nucleotide hydrolases that contain a signature motif, Q-loop, and H-loop/switch region in addition to the Walker A motif/P-loop andWalker B motif commonly found in a number of ATP- and GTP-binding and hydrolyzing proteins..	115.30	9.22E-27	29%	48%	322--502	7--204	3	CAA78282	translation elongation factor 3	739.18	0.00E+00	44%	61%	10--890	96--1021
					4	COG1123	COG1123, ATPase components of various ABC-type transport systems, contain duplicated ATPase [General function prediction only].	105.65	6.51E-24	24%	43%	314--618	9--349	4	XP_711404	translation elongation factor 3	738.80	0.00E+00	43%	61%	10--890	96--1021
					5	COG1131	CcmA, ABC-type multidrug transport system, ATPase component [Defense mechanisms].	104.28	1.62E-23	30%	49%	326--502	16--208	5	XP_711356	translation elongation factor 3	738.41	0.00E+00	43%	61%	10--890	96--1021
					6	COG1121	ZnuC, ABC-type Mn/Zn transport systems, ATPase component [Inorganic ion transport and metabolism].	101.06	1.56E-22	30%	50%	321--503	10--211	6	CAA77567	elongation factor 3	735.33	0.00E+00	44%	61%	10--890	96--1020
					7	COG1129	MglA, ABC-type sugar transport system, ATPase component [Carbohydrate transport and metabolism].	97.90	1.72E-21	24%	43%	325--617	18--316	7	CAG58023	unnamed protein product	722.62	0.00E+00	43%	60%	11--890	94--1015
					8	COG1124	DppF, ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component [Amino acid transport and metabolism / Inorganic ion transport and metabolism].	93.76	3.04E-20	28%	46%	322--502	10--213	8	CAG89810	unnamed protein product	722.24	0.00E+00	43%	60%	18--890	105--1021
					9	COG1120	FepC, ABC-type cobalamin/Fe3+-siderophores transport systems, ATPase components [Inorganic ion transport and metabolism / Coenzyme metabolism].	93.70	3.14E-20	27%	50%	322--502	9--210	9	BAA33959	translation elongation factor3	721.47	0.00E+00	43%	60%	11--890	94--1015
					10	COG2274	SunT, ABC-type bacteriocin/lantibiotic exporters, contain an N-terminal double-glycine peptidase domain [Defense mechanisms].	87.97	1.67E-18	27%	46%	318--489	476--666	10	AAS50338	AAL028Wp	719.15	0.00E+00	43%	60%	18--890	101--1015
C793R	335539--336186	216	24,894	7.17	1	pfam02511	Thy1, Thymidylate synthase complementing protein. Thymidylate synthase complementing protein (Thy1) complements the thymidine growth requirement of the organisms in which it is found, but shows no homology to thymidylate synthase..	231.39	9.83E-62	41%	54%	1--212	1--216	1	NP_049030	similar to Synechocystis ORF s111635, corresponds to GenBank Accession Number D90903	370.93	1.41E-101	83%	90%	1--216	1--216
					2	COG1351	THY1, Predicted alternative thymidylate synthase [Nucleotide transport and metabolism].	135.15	9.64E-33	31%	47%	22--215	35--237	2	CAF34258	thymidilate synthase	238.42	1.09E-61	57%	70%	1--212	3--209
														3	ZP_01006033	thymidylate synthase	238.04	1.42E-61	58%	75%	19--215	13--210
														4	CAE18702	possible Thy1 protein homolog	229.57	5.07E-59	55%	73%	19--215	13--211
														5	AAX46996	Td	227.25	2.51E-58	57%	73%	19--215	12--210
														6	AAX44700	Td	226.87	3.28E-58	50%	69%	1--216	1--211
														7	AAZ59097	Thymidylate synthase complementing protein ThyX	225.71	7.32E-58	53%	71%	9--216	3--211
														8	ABB49306	Thymidylate synthase complementing protein ThyX	219.55	5.24E-56	55%	71%	19--215	13--210
														9	AAP99319	Predicted alternative thymidylate synthase	215.70	7.57E-55	53%	70%	19--216	13--211
														10	ZP_00529968	Thymidylate synthase complementing protein	211.85	1.09E-53	50%	66%	1--215	1--215
C796L	336950--336189	254	28,418	9.42		No Hit Found								1	NP_048629	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	394.82	1.26E-108	75%	84%	1--254	1--252
														2	NP_049005	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	378.64	9.31E-104	70%	83%	1--252	4--254
														3	NP_048807	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	371.32	1.49E-101	71%	82%	4--254	1--249
														4	NP_048427	A79R	325.09	1.22E-87	69%	83%	4--230	1--218
														5	AAU06304	hypothetical protein A275R	317.39	2.55E-85	88%	94%	83--254	1--171
														6	AAU06301	hypothetical protein A275R	316.24	5.67E-85	88%	94%	83--254	1--171
														7	NP_048525	A177R	286.19	6.29E-76	57%	73%	1--249	1--240
														8	AAU06302	hypothetical protein A275R	243.43	4.67E-63	88%	91%	119--254	1--135
														9	AAU06303	hypothetical protein A275R	121.71	2.05E-26	96%	98%	196--254	1--59
														10	AAU06299	hypothetical protein A275R	119.01	1.33E-25	94%	96%	196--254	5--63
C798R	337196--338287	364	41,450	10.65		No Hit Found								1	NP_049032	similar to Chlorella virus PBCV-1 ORF A282L, corresponds to GenBank Accession Number U42580	647.12	0.00E+00	85%	92%	1--358	1--368
														2	NP_048636	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055	73.17	1.50E-11	42%	52%	2--127	414--525
														3	NP_048632	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662	70.48	9.73E-11	84%	92%	1--39	570--608
														4	NP_048441	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563	64.31	6.97E-09	69%	80%	2--43	48--89
														5	NP_149841	378R	57.00	1.11E-06	29%	42%	60--202	5--148
C800R	338414--338908	165	18,627	4.59		No Hit Found									No Hit Found	No Hit Found						
C802L	340185--339088	366	40,120	5.85	1	cd00204	ANK, ankyrin repeats; ankyrin repeats mediate protein-protein interactions in very diverse families of proteins. The number of ANK repeats in a protein can range from 2 to over 20 (ankyrins, for example). ANK repeats may occur in combinations with other types of domains. The structural repeat unit contains two antiparallel helices and a beta-hairpin, repeats are stacked in a superhelical arrangement; this alignment contains 4 consecutive repeats.	111.33	1.25E-25	39%	60%	65--189	1--125	1	NP_049038	contains 10 ankyrin-like repeats; similar to human ankyrin, corresponds to Swiss-Prot Accession Number P16157	528.48	1.32E-148	69%	81%	1--366	1--368
					2	COG0666	Arp, FOG: Ankyrin repeat [General function prediction only].	79.93	4.43E-16	31%	52%	4--166	39--209	2	NP_048353	contains 4 ankyrin repeats; similar to reticulocyte ankyrin, corresponds to Swiss-Prot Accession Number P16157	258.84	1.95E-67	55%	67%	108--350	8--252
					3	pfam00023	Ank, Ankyrin repeat. There's no clear separation between noise and signal on the HMM search Ankyrin repeats generally consist of a beta, alpha, alpha, beta order of secondary structures. The repeats associate to form a higher order structure..	44.28	2.21E-05	50%	70%	137--167	2--32	3	EAL87814	NACHT domain protein, putative	179.87	1.15E-43	37%	56%	11--290	126--1405
					4	smart00248	ANK, ankyrin repeats; Ankyrin repeats are about 33 amino acids long and occur in at least four consecutive copies. They are involved in protein protein interactions. The core of the repeat seems to be an helix-loop-helix structure. .	39.17	7.48E-04	45%	72%	136--165	1--30	4	XP_637214	SecG	173.33	1.08E-41	36%	55%	14--288	212--487
														5	AAV85825	ankyrin domain protein	171.01	5.34E-41	35%	50%	21--290	73--368
														6	ZP_00373467	ankyrin repeat domain protein	162.54	1.90E-38	36%	51%	21--290	87--381
														7	EAL84954	ankyrin repeat protein	161.38	4.23E-38	33%	55%	8--288	430--710
														8	XP_681288	hypothetical protein AN8019.2	161.00	5.52E-38	35%	55%	8--289	834--1115
														9	NP_065209	ankyrin 1 isoform 1	159.84	1.23E-37	33%	54%	11--294	246--529
														10	NP_065210	ankyrin 1 isoform 2	159.84	1.23E-37	33%	54%	11--294	246--529
C806L	341500--340301	400	47,173	7.39		No Hit Found								1	NP_048711	A354R	162.54	2.15E-38	38%	55%	158--399	4--234
														2	NP_048779	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	85.89	2.56E-15	27%	47%	68--367	25--319
C809L	342273--341569	235	25,882	6.63	1	cd00204	ANK, ankyrin repeats; ankyrin repeats mediate protein-protein interactions in very diverse families of proteins. The number of ANK repeats in a protein can range from 2 to over 20 (ankyrins, for example). ANK repeats may occur in combinations with other types of domains. The structural repeat unit contains two antiparallel helices and a beta-hairpin, repeats are stacked in a superhelical arrangement; this alignment contains 4 consecutive repeats.	112.48	5.70E-26	43%	58%	30--154	1--126	1	NP_048353	contains 4 ankyrin repeats; similar to reticulocyte ankyrin, corresponds to Swiss-Prot Accession Number P16157	300.44	2.78E-80	64%	71%	1--219	1--252
					2	COG0666	Arp, FOG: Ankyrin repeat [General function prediction only].	79.16	7.47E-16	34%	50%	19--154	56--200	2	NP_049038	contains 10 ankyrin-like repeats; similar to human ankyrin, corresponds to Swiss-Prot Accession Number P16157	258.84	9.28E-68	59%	73%	10--235	145--368
					3	pfam00023	Ank, Ankyrin repeat. There's no clear separation between noise and signal on the HMM search Ankyrin repeats generally consist of a beta, alpha, alpha, beta order of secondary structures. The repeats associate to form a higher order structure..	40.43	2.69E-04	45%	61%	36--67	2--33	3	XP_637214	SecG	112.08	1.40E-23	40%	55%	6--154	338--487
														4	CAH65056	hypothetical protein	102.06	1.45E-20	38%	51%	2--141	338--478
														5	XP_684802	PREDICTED: similar to ankyrin repeat and SOCS box-containing protein 13	101.68	1.90E-20	32%	49%	6--217	93--303
														6	EAL87814	NACHT domain protein, putative	101.29	2.48E-20	38%	57%	10--154	093--1238

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
														7	1N0R_A	Chain A, 4ank: A Designed Ankyrin Repeat Protein With Four Identical Consensus Repeats	100.91	3.23E-20	47%	61%	35--154	1--121
														8	XP_637278	hypothetical protein DDB0187458	99.37	9.41E-20	37%	57%	1--145	273--419
														9	XP_912160	PREDICTED: similar to ankyrin repeat domain 28 isoform 10	98.98	1.23E-19	36%	51%	2--141	338--478
														10	AAW23170	ankyrin domain protein	98.60	1.61E-19	40%	55%	10--154	185--330
C815L	344037--343804	78	8,840	10.61		No Hit Found									No Hit Found	No Hit Found						