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 B: Gene Names B001L – B886R

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

William 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, William 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 B: Gene Names B001L – B886R" (2007). *Virology Papers*. 8.
https://digitalcommons.unl.edu/virologypub/8

Authors

Lisa A. Fitzgerald, Michael V. Graves, Xiao Li, Tamara Feldblyum, William 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 Copyright © 2006 Elsevier Inc. http://www.sciencedirect.com/science/journal/00426822
Used by permission. 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 B: Gene Names B001L – B886R

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
B001L	1156--878	93	10,443	4.71		No Hit Found									No Hit Found	No Hit Found						
B003R	1474--1707	78	8,909	4.55		No Hit Found									No Hit Found	No Hit Found						
B004R	1894--2199	102	11,766	6.88		No Hit Found								1	NP_049045	identical to Chlorella virus PBCV-1 terminal repeat ORF A3R, corresponds to GenBank Accession Number M55319	109.38	3.20E-23	60%	70%	1--101	1--102
B006L	2960--2385	192	22,528	4.34		No Hit Found									No Hit Found	No Hit Found						
B008R	3196--4188	331	37,021	7.24	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.	203.61	2.29E-53	33%	51%	4--326	1--314	1	CAD33713	putative DNA methylase	144.05	6.01E-33	32%	49%	4--315	1--293
					2	pfam00145	DNA methylase, C-5 cytosine-specific DNA methylase..	198.22	9.10E-52	33%	48%	4--326	1--323	2	CAD33712	putative DNA methylase	141.74	2.98E-32	30%	50%	4--331	1--312
					3	COG0270	Dcm, Site-specific DNA methylase [DNA replication, recombination, and repair].	180.27	2.62E-46	28%	49%	1--328	1--324	3	ZP_00504119	C-5 cytosine-specific DNA methylase	133.65	8.12E-30	33%	47%	1--320	9--309
														4	CAA59690	site-specific DNA-methyltransferase (cytosine-specific)	128.26	3.41E-28	30%	46%	3--324	44--365
														5	AAP07928	Type II restriction-modification system methylation subunit	123.64	8.40E-27	30%	49%	4--318	18--341
														6	BAB07227	cytosine-specific methyltransferase	123.64	8.40E-27	29%	45%	3--328	5--309
														7	AAF89681	cytosine-specific methyltransferase	122.87	1.43E-26	28%	45%	7--329	6--360
														8	CAA74996	Bpu10I (5m)cytosine-specific DNA modification methyltransferase (C1)	122.48	1.87E-26	26%	41%	2--329	6--397
														9	CAE76859	Cytosine-specific DNA-methyltransferase Sau96I	122.48	1.87E-26	30%	49%	5--318	6--324
														10	CAA68505	DdeI methylase	122.09	2.44E-26	30%	44%	4--324	1--366
B010R	4258--5166	303	35,232	8.42	1	pfam01555	N6_N4_Mtase, DNA methylase. Members of this family are DNA methylases. The family contains both N-4 cytosine-specific DNA methylases and N-6 Adenine-specific DNA methylases..	80.21	3.29E-16	31%	48%	178--286	89--197	1	ZP_00515061	DNA methylase N-4/N-6	241.12	3.14E-62	42%	60%	3--296	8--298
					2	COG0863	COG0863, DNA modification methylase [DNA replication, recombination, and repair].	68.81	8.15E-13	24%	41%	1--290	18--266	2	ZP_00728707	COG0863: DNA modification methylase	109.77	1.09E-22	26%	41%	3--288	13--344
					3	COG2189	COG2189, Adenine specific DNA methylase Mod [DNA replication, recombination, and repair].	37.81	1.82E-03	43%	57%	229--291	436--497	3	ZP_00712003	COG0863: DNA modification methylase	109.00	1.86E-22	25%	41%	3--290	13--346
														4	ZP_00723558	COG0863: DNA modification methylase	108.23	3.18E-22	25%	41%	3--290	13--346
														5	ZP_00709127	COG0863: DNA modification methylase	107.84	4.15E-22	25%	41%	3--290	13--346
														6	NP_599080	unknown	107.46	5.42E-22	25%	41%	3--290	13--346
														7	ZP_00703598	COG0863: DNA modification methylase	106.30	1.21E-21	25%	41%	3--291	13--347
														8	YP_310802	putative DNA adenine methyltransferase encoded by prophage	102.83	1.34E-20	25%	40%	3--285	13--341
														9	AAM88315	unknown	102.06	2.28E-20	25%	39%	3--285	13--341
														10	BAB35203	putative DNA methylase	102.06	2.28E-20	25%	39%	3--285	13--341
B013L	6175--5150	342	40,122	6.78		No Hit Found								1	AAX46848	hypothetical protein PSSM4_047	59.31	2.03E-07	31%	48%	1--164	1--155
B016L	7343--6240	368	42,159	6.71	1	COG4123	COG4123, Predicted O-methyltransferase [General function prediction only].	66.04	6.17E-12	25%	38%	36--163	39--176	1	AAC03124	DNA adenine methyltransferase	725.32	0.00E+00	96%	96%	1--368	1--368
					2	COG2890	HemK, Methylase of polypeptide chain release factors [Translation, ribosomal structure and biogenesis].	57.31	2.82E-09	25%	38%	41--190	110--264	2	CAA29835	unnamed protein product	570.85	2.33E-161	73%	83%	2--368	5--376
					3	COG0286	HsdM, Type I restriction-modification system methyltransferase subunit [Defense mechanisms].	50.41	2.82E-07	21%	40%	6--176	149--346	3	AAC57945	DNA adenine methyltransferase	367.47	3.92E-100	50%	69%	12--368	11--357
					4	COG2813	RsmC, 16S RNA G1207 methylase RsmC [Translation, ribosomal structure and biogenesis].	49.88	4.77E-07	23%	42%	43--157	160--266	4	P52284	Modification methylase CviRI (Adenine-specific methyltransferase CviRI) (M.CviRI)	218.39	2.94E-55	36%	54%	9--368	10--377
					5	smart00650	rADc, Ribosomal RNA adenine dimethylases; .	43.24	4.22E-05	28%	52%	44--114	16--87	5	AAC03125	DNA adenine methyltransferase	206.84	8.85E-52	33%	52%	9--368	10--380
					6	COG0030	KsgA, Dimethyladenosine transferase (rRNA methylation) [Translation, ribosomal structure and biogenesis].	40.61	2.98E-04	26%	48%	13--114	1--105	6	AAC57943	DNA adenine methyltransferase	206.07	1.51E-51	37%	57%	12--328	12--329
					7	COG1041	COG1041, Predicted DNA modification methylase [DNA replication, recombination, and repair].	40.33	3.33E-04	24%	40%	4--158	158--311	7	ZP_00518729	hypothetical protein CwatDRAFT_0727	82.03	3.29E-14	30%	50%	23--205	19--217
					8	COG4106	Tam, Trans-aconitate methyltransferase [General function prediction only].	37.97	1.47E-03	24%	39%	42--114	31--102	8	AAC60387	methyltransferase; M-AccI	80.88	7.32E-14	28%	47%	4--215	11--238
					9	COG0421	SpeE, Spermidine synthase [Amino acid transport and metabolism].	38.03	1.58E-03	29%	43%	41--155	76--188	9	AAA50500	AccI methylase	80.88	7.32E-14	28%	47%	4--215	11--238
					10	pfam03602	Cons_hypoth95, Conserved hypothetical protein 95..	36.76	3.53E-03	23%	40%	27--114	24--123	10	JU0470	site-specific DNA-methyltransferase (adenine-specific) (EC 2.1.1.72) type II - Acinetobacter calcoaceticus	80.88	7.32E-14	28%	47%	4--215	11--238
B018L	8812--7397	472	52,455	6.15	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	1.10E-26	46%	62%	231--360	1--126	1	EAL87814	NACHT domain protein, putative	209.53	1.92E-52	34%	51%	10--372	90--1481
					2	COG0666	Arp, FOG: Ankyrin repeat [General function prediction only].	79.93	3.49E-16	31%	45%	223--394	59--234	2	XP_783930	PREDICTED: similar to ankyrin 3, epithelial isoform b	207.99	5.58E-52	32%	54%	8--372	338--698
					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..	50.45	3.25E-07	53%	72%	302--334	1--33	3	XP_798405	PREDICTED: similar to Ankyrin-1 (Erythrocyte ankyrin) (Ankyrin R), partial	207.22	9.52E-52	32%	54%	6--381	38--1108
					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. .	43.80	3.06E-05	59%	72%	302--331	1--30	4	XP_788092	PREDICTED: similar to Ankyrin-2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid), partial	206.84	1.24E-51	31%	55%	10--381	06--1872
														5	XP_782809	PREDICTED: similar to ankyrin repeat domain 28	204.91	4.72E-51	31%	52%	10--383	420--788
														6	CAI56716	hypothetical protein	204.53	6.17E-51	33%	52%	13--381	269--632
														7	NP_066267	ankyrin 3 isoform 1	204.14	8.06E-51	33%	52%	13--381	275--638
														8	CAI40518	ankyrin 3, node of Ranvier (ankyrin G)	204.14	8.06E-51	33%	52%	13--381	275--638
														9	NP_733924	ankyrin 3, epithelial isoform a	204.14	8.06E-51	34%	51%	13--381	258--621
														10	NP_733789	ankyrin 3, epithelial isoform c	204.14	8.06E-51	34%	51%	13--381	258--621
B022R	8915--10117	401	47,051	5.68		No Hit Found								1	NP_048711	A354R	137.50	7.45E-31	39%	56%	156--368	2--207
B024L	11693--10098	532	58,481	7.46	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.	117.11	2.41E-27	46%	58%	255--381	1--126	1	EAL87814	NACHT domain protein, putative	207.22	1.11E-51	34%	51%	14--445	37--1462
					2	COG0666	Arp, FOG: Ankyrin repeat [General function prediction only].	74.92	1.15E-14	35%	55%	226--360	71--212	2	XP_788092	PREDICTED: similar to Ankyrin-2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid), partial	205.68	3.24E-51	34%	51%	15--443	89--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..	41.59	1.32E-04	48%	61%	360--391	1--32	3	XP_637214	SecG	204.53	7.21E-51	31%	46%	15--525	88--581
					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. .	36.48	4.51E-03	56%	78%	29--56	1--28	4	XP_783273	PREDICTED: similar to ankyrin 3, epithelial isoform b	200.29	1.36E-49	32%	48%	15--443	157--599
														5	XP_786227	PREDICTED: similar to ankyrin 3, epithelial isoform b	199.90	1.78E-49	33%	50%	11--440	27--455
														6	XP_787863	PREDICTED: similar to ankyrin 3, epithelial isoform b	198.36	5.17E-49	33%	47%	28--440	35--429
														7	XP_782887	PREDICTED: similar to Ankyrin-1 (Erythrocyte ankyrin) (Ankyrin R)	197.21	1.15E-48	36%	52%	29--399	209--578
														8	XP_792205	PREDICTED: similar to Ankyrin-3 (ANK-3) (Ankyrin G)	193.74	1.27E-47	32%	49%	28--443	189--597
														9	XP_658734	hypothetical protein AN1130.2	193.36	1.66E-47	31%	49%	15--446	523--947
														10	AAH95664	Hypothetical protein LOC553206	190.27	1.41E-46	30%	48%	12--453	188--632
B027L	12722--11790	311	34,170	4.05	1	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	42.05	1.04E-04	26%	34%	155--238	48--134		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
B031R	12817--13509	231	25,842	10.65		No Hit Found									No Hit Found	No Hit Found						
B033R	13725--14486	254	28,894	8.43		No Hit Found								1	NP_048357	A9R	313.15	4.80E-84	82%	94%	80--254	1--175
														2	NP_048807	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	107.07	5.23E-22	30%	48%	2--250	1--245
														3	NP_049005	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	105.92	1.16E-21	32%	48%	3--250	8--252
														4	NP_048629	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	102.83	9.86E-21	29%	50%	9--250	11--248
														5	NP_048427	A79R	93.20	7.81E-18	31%	51%	9--229	8--217
														6	NP_048525	A177R	80.88	4.01E-14	27%	48%	9--236	11--227
														7	AAU06304	hypothetical protein A275R	75.49	1.69E-12	31%	48%	84--250	1--167
														8	AAU06301	hypothetical protein A275R	74.33	3.76E-12	30%	48%	84--250	1--167
B034L	18948--14494	1485	154,849	5.34	1	COG2911	COG2911, Uncharacterized protein conserved in bacteria [Function unknown].	40.68	2.84E-04	20%	39%	490--820	743--1048	1	BAB83467	Vp260 like protein	1823.14	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	8.04E-03	18%	36%	677--796	10--112	2	BAB83469	Vp260 like protein	695.66	0.00E+00	33%	49%	11--1477	1--1458
														3	BAB83468	Vp260 like protein	672.93	0.00E+00	33%	49%	11--1469	1--1448
														4	BAB83470	Vp260 like protein	654.06	0.00E+00	34%	50%	11--1334	1--1306
														5	BAB83471	Vp260 like protein	594.73	8.04E-168	38%	54%	3--953	2--955
														6	NP_048470	PBCV-1 Vp260 protein	233.80	3.60E-59	27%	41%	7--824	17--871
														7	AAA86307	glycoprotein Vp260	207.22	3.61E-51	27%	39%	7--794	17--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	129.80	7.32E-28	21%	36%	119--1331	31--1233
														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
B039L	20232--18979	418	48,782	7.02	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	37.06	3.40E-03	32%	52%	34--104	13--79		No Hit Found	No Hit Found						
B040L	24708--20296	1471	152,143	5.01		No Hit Found								1	BAB83468	Vp260 like protein	2229.91	0.00E+00	76%	83%	10--1471	1--1462
														2	BAB83469	Vp260 like protein	2094.32	0.00E+00	72%	80%	10--1471	1--1464
														3	BAB83470	Vp260 like protein	1632.46	0.00E+00	65%	75%	10--1302	1--1291
														4	BAB83467	Vp260 like protein	613.99	1.27E-173	36%	53%	4--1088	5--1094
														5	BAB83471	Vp260 like protein	515.77	4.69E-144	34%	49%	1--938	1--947
														6	NP_048470	PBCV-1 Vp260 protein	211.08	2.47E-52	27%	39%	6--912	17--1025
														7	AAA86307	glycoprotein Vp260	196.44	6.30E-48	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	150.98	3.04E-34	24%	39%	57--808	13--759
														9	NP_048362	Asn/Thr/Ser/Val rich protein	137.12	4.54E-30	24%	38%	19--916	82--1137
														10	NP_048366	Asn/Thr/Ser/Ile rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	130.57	4.25E-28	22%	39%	109--1225	21--1020
B047L	29110--24752	1453	150,568	5.16		No Hit Found								1	BAB83469	Vp260 like protein	1658.27	0.00E+00	61%	72%	10--1452	1--1460
														2	BAB83468	Vp260 like protein	1646.71	0.00E+00	61%	70%	10--1452	1--1458
														3	BAB83470	Vp260 like protein	1586.62	0.00E+00	63%	73%	10--1317	1--1311
														4	BAB83467	Vp260 like protein	581.25	8.97E-164	34%	51%	4--1112	5--1121
														5	BAB83471	Vp260 like protein	569.31	3.53E-160	37%	52%	1--939	1--949
														6	NP_048470	PBCV-1 Vp260 protein	199.90	5.62E-49	27%	39%	6--856	17--973
														7	AAA86307	glycoprotein Vp260	185.65	1.10E-44	28%	39%	54--856	21--787
														8	NP_048377	Asn/Thr/Ser/Val rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	132.49	1.10E-28	23%	39%	57--785	13--745
														9	NP_048362	Asn/Thr/Ser/Val rich protein	130.18	5.47E-28	22%	36%	245--1310	34--1097
														10	NP_048366	Asn/Thr/Ser/Ile rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	120.17	5.66E-25	22%	37%	21--937	15--982
B055L	32010--29152	953	96,124	7.19		No Hit Found								1	BAB83471	Vp260 like protein	1522.29	0.00E+00	80%	87%	1--953	1--955
														2	BAB83467	Vp260 like protein	613.99	7.87E-174	38%	56%	2--953	3--954
														3	BAB83468	Vp260 like protein	613.61	1.03E-173	39%	56%	10--947	1--930
														4	BAB83469	Vp260 like protein	598.59	3.42E-169	38%	55%	10--947	1--930
														5	BAB83470	Vp260 like protein	595.89	2.22E-168	39%	55%	10--947	1--931
														6	NP_048470	PBCV-1 Vp260 protein	210.31	2.62E-52	27%	40%	6--912	17--1001
														7	AAA86307	glycoprotein Vp260	192.20	7.39E-47	28%	40%	51--912	15--815
														8	NP_048362	Asn/Thr/Ser/Val rich protein	158.30	1.18E-36	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	155.22	1.00E-35	23%	40%	22--840	18--800
														10	NP_048366	Asn/Thr/Ser/Ile rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	145.59	7.93E-33	25%	39%	74--927	28--952
B059R	32108--33340	411	46,391	6.18	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 NY2A the major capsid protein is a glycoprotein	281.81	6.40E-77	37%	53%	6--407	3--443	1	AAC27494	putative capsid protein	425.25	1.84E-117	83%	93%	22--251	1--230
														2	NP_048787	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	285.03	2.97E-75	39%	54%	6--411	3--437
														3	BAA76601	major capsid protein MCP1	283.49	8.64E-75	39%	54%	6--411	3--437
														4	AAC27492	major capsid protein Vp49	277.72	4.74E-73	40%	54%	6--411	3--432
														5	BAA76600	major capsid protein	277.72	4.74E-73	39%	54%	6--411	3--436
														6	BAA22198	major capsid protein Vp54	277.33	6.19E-73	38%	54%	6--411	3--437
														7	1M3Y_D	Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid Containing, Dna Virus	259.61	1.34E-67	38%	53%	28--411	1--413
														8	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	259.23	1.74E-67	38%	53%	28--411	1--413
														9	NP_048914	similar to Chlorella virus PBCV-1 ORF A11L, corresponds to GenBank Accession Number U42580	245.36	2.61E-63	36%	55%	6--411	2--400
														10	NP_048359	contains aminoacyl-tRNA synthetase class-II signature	232.65	1.75E-59	35%	54%	6--411	3--403
B061R	33410--34606	399	47,156	7.07		No Hit Found								1	NP_048711	A354R	155.22	3.42E-36	39%	55%	155--399	1--235
														2	NP_048779	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	77.80	6.95E-13	27%	45%	88--365	45--318
B068L	35422--34964	153	17,741	4.52	1	COG5201	SKP1, SCF ubiquitin ligase, SKP1 component [Posttranslational modification, protein turnover, chaperones].	104.76	1.20E-23	37%	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	168.32	5.78E-41	55%	71%	1--153	1--151
					2	pfam01466	Skp1, Skp1 family, dimerisation domain..	93.75	2.76E-20	57%	80%	69--146	1--77	2	AAT99735	SKP1	96.29	2.79E-19	36%	53%	3--146	7--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	6.76E-14	33%	55%	1--97	2--107	3	AAC63273	SKP1-like protein	95.13	6.22E-19	35%	52%	3--146	4--152
					4	pfam03931	Skp1_POZ, Skp1 family, tetramerisation domain..	54.52	1.70E-08	35%	57%	1--62	1--66	4	AAX83944	Skp1	94.74	8.12E-19	35%	53%	3--146	7--155
														5	XP_659906	hypothetical protein AN2302.2	93.97	1.38E-18	32%	52%	3--145	6--158
														6	AAB18274	sconCp	93.97	1.38E-18	32%	52%	3--145	6--158
														7	BAE55074	unnamed protein product	93.59	1.81E-18	32%	51%	2--145	5--158
														8	EAL91988	sulfur metabolism regulator SkpA, putative	93.20	2.36E-18	32%	51%	2--145	3--156
														9	AAO85510	SKP1	93.20	2.36E-18	34%	52%	3--146	4--152
														10	AAD34458	Skp1	91.66	6.87E-18	35%	53%	7--146	9--152
B070R	35493--36791	433	46,283	10.98	1	pfam05616	Neisseria_TspB, Neisseria meningitidis TspB protein. This family consists of several Neisseria meningitidis TspB virulence factor proteins..	57.83	1.72E-09	33%	37%	47--130	323--406	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	1.55E-162	91%	92%	119--433	98--412

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	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	53.99	2.72E-08	30%	40%	12--148	33--170	2	NP_048519	similar to PBCV-1 ORF A41R, corresponds to GenBank Accession Number U17055	497.28	4.11E-139	78%	84%	119--433	73--387
					3	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.	53.06	4.57E-08	33%	56%	30--94	59--123	3	BAD86968	hypothetical protein	72.40	3.26E-11	26%	43%	128--360	95--329
					4	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..	47.76	1.72E-06	18%	30%	39--127	166--254	4	NP_916095	P0481E12.18	72.40	3.26E-11	26%	43%	128--360	72--306
					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..	45.79	8.00E-06	25%	35%	11--113	103--206	5	AAM63817	unknown	65.47	3.99E-09	23%	44%	121--364	55--299
					6	COG3147	DedD, Uncharacterized protein conserved in bacteria [Function unknown].	44.63	1.79E-05	23%	32%	11--122	39--150	6	NP_191439	unknown protein	65.47	3.99E-09	23%	44%	121--364	55--299
					7	pfam05518	Totivirus coat, Totivirus coat protein..	44.20	2.03E-05	23%	27%	13--121	643--752	7	XP_464783	unknown protein	51.60	5.96E-05	24%	42%	145--372	74--311
B073L	38602--36794	603	68,715	6.80	1	COG0464	SpoVK, ATPases of the AAA+ class [Posttranslational modification, protein turnover, chaperones].	58.26	1.33E-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	1078.93	0.00E+00	88%	94%	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	1.95E-08	32%	50%	492--558	82--148	2	EAN88268	hypothetical protein, conserved	103.22	2.65E-20	25%	44%	212--542	267--562
					3	COG1222	RPT1, ATP-dependent 26S proteasome regulatory subunit [Posttranslational modification, protein turnover, chaperones].	51.38	1.65E-07	45%	67%	231--280	158--207	3	EAN97512	hypothetical protein, conserved	102.45	4.51E-20	24%	44%	212--542	267--562
					4	cd00009	AAA, AAA-superfamily of ATPases associated with a wide variety of cellular activities, including membrane fusion, proteolysis, and DNA replication..	48.05	1.38E-06	21%	45%	261--349	2--94	4	AAZ12359	hypothetical protein, conserved	100.91	1.31E-19	23%	42%	136--542	191--535
					5	COG0465	HflB, ATP-dependent Zn proteases [Posttranslational modification, protein turnover, chaperones].	43.39	3.99E-05	55%	70%	496--536	272--312	5	CAJ04397	hypothetical protein, conserved	100.14	2.24E-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	1.01E-04	55%	64%	260--307	545--1589	6	CAJ04393	hypothetical protein, conserved	100.14	2.24E-19	24%	43%	212--552	279--585
					7	COG0593	DnaA, ATPase involved in DNA replication initiation [DNA replication, recombination, and repair].	40.64	2.41E-04	28%	50%	247--335	103--185	7	CAC28656	related to human BCS1 protein	95.90	4.22E-18	26%	43%	248--558	371--710
					8	COG0714	COG0714, MoxR-like ATPases [General function prediction only].	39.72	5.29E-04	26%	44%	261--335	46--119	8	YP_142926	unknown	94.36	1.23E-17	31%	52%	429--591	18--188
					9	COG1223	COG1223, Predicted ATPase (AAA+ superfamily) [General function prediction only].	38.80	9.25E-04	39%	63%	261--304	154--203	9	NP_803852	ORF286	91.28	1.04E-16	25%	42%	226--549	198--460
					10	COG1484	DnaC, DNA replication protein [DNA replication, recombination, and repair].	38.45	1.06E-03	19%	39%	206--309	58--157	10	CAJ06263	hypothetical protein, conserved	75.10	7.72E-12	24%	40%	220--551	287--573
B074R	38665--39033	123	14,000	6.66		No Hit Found								1	NP_048396	A48R	223.79	1.19E-57	90%	91%	1--123	1--123
B075L	39713--39030	228	26,158	8.20	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
B076L	40053--39706	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	1.83E-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
B077R	40140--40430	97	11,095	10.05		No Hit Found									No Hit Found	No Hit Found						
B078L	40916--40458	153	17,497	7.71		No Hit Found								1	NP_048399	contains type I hydrophobic transmembrane region and ATP/GTP binding motif	286.57	1.46E-76	91%	97%	6--153	56--203
														2	ZP_00279033	hypothetical protein Bcep02006353	154.45	8.63E-37	46%	69%	6--152	58--210
														3	AAQ58722	conserved hypothetical protein	138.66	4.90E-32	47%	64%	6--152	47--198
														4	AAQ60907	conserved hypothetical protein	119.01	4.02E-26	42%	58%	6--152	43--197
														5	ZP_00902424	conserved hypothetical protein	102.06	5.09E-21	37%	58%	6--152	64--219
														6	NP_792595	hypothetical protein PSPTO2790	99.75	2.52E-20	37%	60%	6--153	114--270
														7	NP_746428	hypothetical protein PP4312	98.98	4.31E-20	37%	57%	6--152	64--219
														8	ABA75491	conserved hypothetical protein	98.60	5.62E-20	35%	56%	6--152	68--223
														9	YP_235595	hypothetical protein Psyr_2518	97.83	9.59E-20	33%	59%	6--152	70--225
														10	AAZ37859	conserved hypothetical protein	95.52	4.76E-19	34%	59%	6--152	70--225
B080L	42947--41013	645	74,191	10.63	1	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.	58.32	1.13E-09	34%	46%	566--638	1--69	1	AAU06281	putative transposase	218.39	6.11E-55	31%	49%	212--639	45--420
					2	COG0675	COG0675, Transposase and inactivated derivatives [DNA replication, recombination, and repair].	57.01	3.17E-09	20%	40%	324--638	62--345	2	NP_048981	similar to Synechocystis transposase, corresponds to GenBank Accession Number D90909	217.62	1.04E-54	31%	48%	212--639	39--414
					3	pfam01385	Transposase_2, Probable transposase. This family includes IS891, IS1136 and IS1341. .	37.20	2.94E-03	23%	42%	219--551	3--277	3	YP_143208	putative transposase	75.49	6.40E-12	23%	39%	175--640	101--524
														4	YP_143124	putative transposase	75.10	8.35E-12	24%	39%	202--640	122--517
														5	YP_142433	putative transposase	73.17	3.17E-11	22%	39%	175--640	94--517
														6	YP_142458	putative transposase	66.24	3.88E-09	25%	38%	431--638	317--535
														7	AAS54227	AGL264Wp	65.47	6.62E-09	23%	40%	216--643	71--456
														8	CAJ31329	insertion sequence IS606 transposase homolog A	61.62	9.56E-08	29%	51%	525--638	304--421
														9	BAD76101	transposase	60.85	1.63E-07	26%	40%	439--640	169--372
														10	ZP_00370996	ISCco1, transposase orfB	60.46	2.13E-07	31%	50%	526--635	277--390
B083L	42942--42259	228	25,945	9.23	1	COG2452	COG2452, Predicted site-specific integrase-resolvase [DNA replication, recombination, and repair].	140.82	1.87E-34	42%	56%	15--208	4--190	1	YP_143125	putative resolvase	113.62	4.52E-24	41%	61%	10--152	2--140
					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..	66.04	5.53E-12	30%	48%	75--201	2--132	2	YP_142434	putative resolvase	110.15	4.99E-23	39%	60%	10--152	2--140

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	cd01104	HTH_MlrA, Helix-turn-helix transcription regulator MlrA (merR-like regulator A). The MlrA protein, also known as YehV, has been shown to control cell-cell aggregation by co-regulating the expression of curli and extracellular matrix production in Escherichia coli and Salmonella typhimurium. Its close homolog, CarA from Myxococcus xanthus, is involved in activation of the carotenoid biosynthesis genes by light. These proteins belong to the MERR superfamily of transcription regulators that promote expression of several stress regulon genes by reconfiguring the spacer between the -35 and -10 promoter elements. Their conserved N-terminal domains contain predicted HTH (helix-turn-helix) motives that mediate DNA binding, while the dissimilar C-terminal domains bind specific coactivator molecules	41.85	1.12E-04	31%	42%	16--98	5--83	3	YP_142457	putative resolvase	105.15	1.61E-21	37%	56%	11--161	3--149
														4	AAK41573	First ORF in transposon ISC1904	96.29	7.46E-19	42%	59%	22--159	11--142
														5	BAD84844	predicted site-specific integrase/resolvase	95.90	9.74E-19	35%	57%	6--160	3--153
														6	AAK43255	First ORF in transposon ISC1904	95.90	9.74E-19	40%	58%	15--159	4--142
														7	AAK41585	First ORF in transposon ISC1904	95.90	9.74E-19	42%	59%	15--154	4--137
														8	ZP_00683312	regulatory protein, MerR:Resolvase, N-terminal	95.13	1.66E-18	38%	57%	11--151	10--150
														9	CAB49329	Resolvase related protein	95.13	1.66E-18	36%	60%	11--156	6--151
														10	AAK42026	First ORF in transposon ISC1904	94.74	2.17E-18	41%	58%	15--159	4--142
B086L	44356--43355	334	38,700	6.24		No Hit Found									No Hit Found	No Hit Found						
B087L	44805--44404	134	14,897	4.21		No Hit Found									No Hit Found	No Hit Found						
B088L	45914--44886	343	39,143	8.02	1	pfam00145	DNA methylase, C-5 cytosine-specific DNA methylase..	145.84	5.36E-36	28%	42%	5--288	1--289	1	NP_048873	M.CviAII cytosine DNA methyltransferase	322.01	1.69E-86	49%	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	138.51	9.89E-34	37%	52%	5--165	1--166	2	NP_048886	M.CviAIV cytosine DNA methyltransferase	288.89	1.58E-76	47%	61%	4--328	2--332
					3	COG0270	Dcm, Site-specific DNA methylase [DNA replication, recombination, and repair].	106.31	4.47E-24	30%	46%	4--164	3--169	3	AAC64006	cytosine methyltransferase	281.95	1.94E-74	43%	58%	5--339	3--355
														4	AAC55063	cytosine methyltransferase	270.40	5.83E-71	41%	57%	6--339	4--359
														5	NP_049039	nonfunctional M.CviAV cytosine DNA methyltransferase	268.86	1.70E-70	41%	57%	6--339	4--359
														6	AAV84097	CviPII m5C DNA methyltransferase	255.37	1.94E-66	41%	60%	3--339	11--356
														7	NP_818425	gp127	101.68	3.59E-20	40%	52%	5--157	3--165
														8	YP_358798	putative methylase	99.37	1.78E-19	31%	46%	5--227	1--261
														9	BAD65383	site-specific DNA-methyltransferase	98.60	3.04E-19	28%	46%	5--222	1--280
														10	AAV83360	DNA cytosine methylase	95.90	1.97E-18	35%	53%	5--160	99--258
B091L	46574--45945	210	23,885	4.58		No Hit Found									No Hit Found	No Hit Found						
B094R	46650--47408	253	29,734	5.72		No Hit Found									No Hit Found	No Hit Found						
B099R	47430--48365	312	34,095	4.08	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	42.27	9.05E-05	33%	36%	163--199	55--91	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	209.15	1.38E-52	71%	84%	31--167	1--138
B103R	48399--49457	353	42,124	7.74		No Hit Found								1	NP_048419	similar to Mycoplasma hypothetical protein MG366, corresponds to Swiss-Prot Accession Number P47606	605.13	1.05E-171	82%	91%	2--353	3--354
														2	NP_048411	A63L	153.30	1.09E-35	36%	53%	126--353	5--232
														3	AAZ58444	conserved hypothetical protein	90.51	8.66E-17	41%	63%	5--117	14--130
														4	ZP_00680654	conserved hypothetical protein	87.04	9.58E-16	38%	56%	4--117	6--126
														5	AAO29065	conserved hypothetical protein	85.50	2.79E-15	38%	57%	4--117	6--126
														6	BAE53285	hypothetical protein	83.19	1.38E-14	37%	58%	4--117	5--125
														7	ZP_00683615	conserved hypothetical protein	82.42	2.36E-14	38%	55%	6--117	1--119
														8	AAF84954	hypothetical protein XF2155	82.42	2.36E-14	38%	57%	6--117	1--119
														9	AAM40292	conserved hypothetical protein	82.03	3.08E-14	38%	54%	4--117	6--126
														10	BAE67710	conserved hypothetical protein	82.03	3.08E-14	38%	54%	4--117	6--126
B104R	49523--50464	314	36,074	7.66		No Hit Found								1	NP_048920	similar to Chlorella virus PBCV-1 ORF A154L, corresponds to GenBank Accession Number U42580	480.33	3.24E-134	72%	83%	3--314	41--350
														2	NP_048502	A154L	478.79	9.42E-134	70%	82%	3--314	37--347
														3	NP_048477	similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903	451.06	2.10E-125	60%	74%	3--314	5--356
														4	NP_077492	EsV-1-7	79.72	1.27E-13	29%	45%	5--202	41--261
B108L	50779--50579	67	8,323	11.05		No Hit Found									No Hit Found	No Hit Found						
B109R	50858--51082	75	8,714	11.72		No Hit Found									No Hit Found	No Hit Found						
B110L	51869--51339	177	20,853	4.90		No Hit Found									No Hit Found	No Hit Found						
B113R	51897--52241	115	13,351	8.91		No Hit Found									No Hit Found	No Hit Found						
B115L	52417--52163	85	9,670	7.69		No Hit Found								1	NP_048425	contains type 1 hydrophobic transmembrane region	90.51	1.57E-17	70%	83%	10--71	14--75
B116R	52496--53389	298	33,198	5.50	1	COG0388	COG0388, Predicted amidohydrolase [General function prediction only].	182.66	5.15E-47	34%	50%	3--297	1--271	1	NP_048426	contains ATP/GTP-binding site motif A; similar to rat beta-alanine synthetase, corresponds to Swiss-Prot Accession Number Q03248	577.40	1.80E-163	93%	95%	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..	142.40	5.48E-35	35%	53%	6--179	1--174	2	ZP_00417184	Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase	289.27	9.78E-77	48%	65%	1--296	1--283
					3	COG0815	Lnt, Apolipoprotein N-acyltransferase [Cell envelope biogenesis, outer membrane].	49.63	5.80E-07	23%	37%	18--217	233--451	3	ZP_00830003	COG0388: Predicted amidohydrolase	289.27	9.78E-77	49%	63%	4--296	3--285
														4	CAG77171	putative carbon-nitrogen hydrolase	288.89	1.28E-76	48%	63%	3--296	2--285
														5	AAS63658	putative carbon-nitrogen hydrolase	288.89	1.28E-76	49%	63%	4--296	3--285
														6	ZP_00831984	COG0388: Predicted amidohydrolase	288.50	1.67E-76	49%	63%	4--296	3--285
														7	ZP_00825332	COG0388: Predicted amidohydrolase	288.12	2.18E-76	49%	63%	4--296	3--285
														8	ZP_00140725	COG0388: Predicted amidohydrolase	285.03	1.84E-75	48%	62%	1--296	1--283
														9	ABA74114	Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase	283.49	5.36E-75	48%	65%	7--297	6--286
														10	YP_237998	Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase	280.80	3.48E-74	47%	63%	1--293	1--280
B117L	53903--53412	164	19,736	4.39		No Hit Found								1	NP_048628	encodes Asp/Lys rich sequence	56.23	3.84E-07	31%	49%	7--116	94--189
														2	NP_048438	contains phenyl group binding site (CAAX box)	48.91	6.13E-05	40%	67%	7--60	6--60
B118R	54289--55128	280	32,434	7.21		No Hit Found								1	NP_049003	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	189.50	9.53E-47	46%	63%	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	65%	85%	64--118	1--55
														3	NP_048807	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	83.96	5.64E-15	68%	95%	70--117	1--48
														4	NP_048525	A177R	67.40	5.46E-10	52%	67%	67--121	1--55
														5	NP_048427	A79R	67.01	7.13E-10	48%	80%	70--119	1--50
														6	NP_048629	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	65.47	2.08E-09	49%	78%	67--117	1--51

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
B120R	55378--56109	244	27,411	9.42		No Hit Found								1	NP_048629	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	389.81	3.72E-107	74%	84%	1--244	1--252
														2	NP_048807	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	381.72	1.01E-104	75%	84%	4--244	1--249
														3	NP_049005	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	377.87	1.46E-103	73%	82%	1--243	4--255
														4	NP_048427	A79R	335.50	8.33E-91	70%	82%	4--228	1--226
														5	AAU06304	hypothetical protein A275R	335.50	8.33E-91	95%	98%	74--244	1--171
														6	AAU06301	hypothetical protein A275R	334.34	1.86E-90	94%	98%	74--244	1--171
														7	NP_048525	A177R	306.61	4.15E-82	60%	76%	1--242	1--243
														8	AAU06302	hypothetical protein A275R	273.86	2.98E-72	99%	99%	110--244	1--135
														9	AAU06303	hypothetical protein A275R	116.70	6.08E-25	94%	96%	186--244	1--59
														10	AAU06299	hypothetical protein A275R	114.01	3.94E-24	93%	94%	186--244	5--63
B122L	56702--56127	192	22,335	7.62		No Hit Found								1	NP_048429	A81L	283.49	2.27E-75	71%	88%	4--190	1--187
B124L	57325--56765	187	22,310	8.19		No Hit Found								1	NP_048432	A84L	323.55	1.87E-87	82%	91%	1--186	1--186
B126R	57440--58129	230	26,650	4.87	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	116.71	3.27E-27	33%	45%	46--228	6--178	1	NP_048433	PBCV-1 prolyl 4-hydroxylase	345.13	9.40E-94	75%	80%	8--229	15--240
					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	47.42	2.15E-06	29%	39%	126--229	2--96	2	AAZ62310	Procollagen-proline,2-oxoglutarate-4-dioxygenase	84.73	2.29E-15	33%	46%	50--228	99--274
														3	ZP_00984285	hypothetical protein BdolA 01003928	84.73	2.29E-15	30%	44%	38--228	70--257
														4	ZP_00238502	prolyl 4-hydroxylase alpha subunit	81.26	2.53E-14	29%	50%	46--228	44--211
														5	AAT63151	prolyl 4-hydroxylase, alpha subunit	80.88	3.31E-14	29%	49%	46--228	60--227
														6	YP_142947	prolyl 4-hydroxylase	80.49	4.32E-14	26%	39%	1--228	1--237
														7	ZP_00508297	Procollagen-proline,2-oxoglutarate-4-dioxygenase	80.49	4.32E-14	31%	49%	48--228	62--237
														8	AAU16279	prolyl 4-hydroxylase, alpha subunit	79.34	9.62E-14	29%	49%	46--228	60--227
														9	ZP_00740932	Prolyl 4-hydroxylase alpha subunit	79.34	9.62E-14	30%	49%	50--228	80--243
														10	AAS43215	prolyl 4-hydroxylase, alpha subunit domain protein	78.18	2.14E-13	29%	49%	46--228	44--211
B130R	58211--58693	161	18,261	4.29		No Hit Found									No Hit Found	No Hit Found						
B133R	58776--60107	444	51,930	9.73	1	pfam01844	HNH, HNH endonuclease..	36.08	6.69E-03	35%	45%	324--380	3--52	1	NP_048435	A87R	572.39	1.04E-161	64%	78%	11--444	22--456
														2	NP_048779	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	136.73	1.46E-30	34%	49%	114--378	44--286
														3	NP_048711	A354R	53.91	1.25E-05	24%	42%	257--439	61--236
B136L	61382--60114	423	48,542	10.90		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	224.17	6.49E-57	85%	91%	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
B137L	62395--61436	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	5.13E-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	6.38E-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	7.30E-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	6.15E-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	2.23E-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	2.62E-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
B139R	62474--63997	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	6.98E-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	9.36E-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

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	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	37--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	92--1675
B143R	64124--65908	595	65,549	5.91	1	COG0449	GlmS, Glucosamine 6-phosphate synthetase, contains amidotransferase and phosphosugar isomerase domains [Cell envelope biogenesis, outer membrane].	626.44	1.35E-180	46%	63%	1--593	1--595	1	NP_048448	PBCV-1 glucosamine synthetase	1019.61	0.00E+00	85%	92%	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	266.62	2.68E-72	48%	69%	2--216	1--215	2	BAD15299	glutamine:fructose-6-phosphate amidotransferase GFAT	1018.45	0.00E+00	85%	92%	1--595	1--596
					3	COG2222	AgaS, Predicted phosphosugar isomerases [Cell envelope biogenesis, outer membrane].	186.25	3.85E-48	30%	47%	254--594	5--338	3	CAE39493	glucosamine--fructose-6-phosphate aminotransferase	548.51	2.34E-154	47%	65%	1--593	1--608
					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..	118.31	1.06E-27	28%	50%	2--228	1--224	4	CAE44992	glucosamine--fructose-6-phosphate aminotransferase	546.58	8.89E-154	47%	65%	1--593	1--608
					5	COG0034	PurF, Glutamine phosphoribosylpyrophosphate amidotransferase [Nucleotide transport and metabolism].	116.85	3.14E-27	28%	52%	1--232	4--233	5	ZP_00244599	COG0449: Glucosamine 6-phosphate synthetase, contains amidotransferase and phosphosugar isomerase domains	541.96	2.19E-152	46%	64%	1--593	1--617
					6	pfam00310	GATase 2, Glutamine amidotransferases class-II..	113.47	2.91E-26	41%	61%	2--136	1--136	6	ZP_00942990	Glucosamine--fructose-6-phosphate aminotransferase (isomerizing)	533.10	1.02E-149	45%	64%	1--593	1--610
					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 to the Ntn hydrolase superfamily and is found at the N-terminus of enzymes such as glucosamine-fructose 6-phosphate synthase (GLMS or GFAT), glutamine phosphoribosylpyrophosphate (Prpp) amidotransferase (GPATase) , asparagine synthetase B (AsnB), beta lactam synthetase (beta-LS) and glutamate synthase (GltS). GLMS catalyzes the formation of glucosamine 6-phosphate from fructose 6-phosphate and glutamine in amino sugar synthesis. GPATase catalyzes the first step in purine biosynthesis, an amide transfer from glutamine to PRPP, resulting in phosphoribosylamine, pyrophosphate and glutamate. Asparagine synthetase B synthesizes asparagine from aspartate and glutamine. Beta-LS catalyzes the formation of the beta-lactam ring in the beta-lactamase inhibitor clavulanic acid. GltS synthesizes L-glutamate from 2-	105.17	1.09E-23	28%	47%	2--221	1--220	7	CAD13706	PROBABLE GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE PROTEIN	531.18	3.87E-149	45%	64%	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.	98.44	9.73E-22	37%	53%	285--417	2--135	8	ZP_00509192	Glucosamine-fructose-6-phosphate aminotransferase, isomerising	529.64	1.13E-148	45%	64%	1--593	1--614
					9	COG0367	AsnB, Asparagine synthase (glutamine-hydrolyzing) [Amino acid transport and metabolism].	83.95	2.46E-17	22%	37%	1--352	1--345	9	ZP_00594232	Glucosamine-fructose-6-phosphate aminotransferase, isomerising	528.09	3.27E-148	45%	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.	75.64	7.50E-15	32%	50%	2--200	1--164	10	AAZ59612	Glucosamine-fructose-6-phosphate aminotransferase, isomerising	521.93	2.35E-146	45%	63%	1--593	1--610
B148R	65997--66959	321	37,157	7.61	1	pfam01331	mRNA_cap_enzyme, mRNA capping enzyme, catalytic domain. This family represents the ATP binding catalytic domain of the mRNA capping enzyme..	146.20	4.44E-36	27%	46%	51--227	1--192	1	NP_048451	PBCV-1 mRNA guanylyltransferase	532.72	5.71E-150	76%	90%	2--321	11--330
					2	COG5226	CEG1, mRNA capping enzyme, guanylyltransferase (alpha) subunit [RNA processing and modification].	93.95	2.47E-20	23%	39%	46--318	42--362	2	1CKN_B	Chain B, Structure Of Guanylylated Mrna Capping Enzyme Complexed With Gtp	530.41	2.83E-149	76%	90%	2--321	11--330
					3	pfam03919	mRNA_cap_C, mRNA capping enzyme, C-terminal domain..	56.14	5.72E-09	28%	43%	230--316	1--110	3	P78587	mRNA capping enzyme alpha subunit (mRNA guanylyltransferase) (GTP--RNA guanylyltransferase) (GTase)	90.89	5.74E-17	27%	42%	48--316	41--372
														4	NP_974263	mRNA guanylyltransferase/ phosphoprotein phosphatase/ protein tyrosine/serine/threonine phosphatase	89.74	1.28E-16	24%	46%	46--319	349--654
														5	1P16_B	Chain B, Structure Of An Mrna Capping Enzyme Bound To The Phosphorylated Carboxyl-Terminal Domain Of Rna Polymerase Ii	87.81	4.86E-16	27%	41%	48--316	41--372
														6	ABA92070	mRNA capping enzyme - like protein	82.42	2.04E-14	23%	45%	46--319	370--676
														7	CAG86747	unnamed protein product	82.03	2.66E-14	25%	41%	45--316	38--371
														8	ABA96061	mRNA capping enzyme - like protein	80.49	7.75E-14	25%	45%	46--316	343--636
														9	EAL46815	mRNA capping enzyme, putative	79.34	1.73E-13	25%	43%	41--315	358--662
														10	AAT68133	mRNA capping enzyme	78.95	2.26E-13	24%	46%	48--293	273--541
B150L	67869--67000	290	33,859	7.03	1	pfam00443	UCH, Ubiquitin carboxyl-terminal hydrolase..	123.19	3.41E-29	21%	40%	3--278	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	444.89	1.33E-123	68%	89%	1--284	1--284
					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.	79.20	6.40E-16	17%	33%	4--279	2--320	2	EAA08027	ENSANGP00000018711	71.25	3.99E-11	20%	39%	1--278	557--887
					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..	72.21	9.03E-14	23%	44%	5--278	5--303	3	AAN09567	CG14619-PB, isoform B	69.71	1.16E-10	23%	41%	1--280	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..	69.22	6.38E-13	17%	35%	5--279	3--335	4	AAN09566	CG14619-PC, isoform C	69.71	1.16E-10	23%	41%	1--280	612--937

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	cd02665	Peptidase_C19I, 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..	44.36	1.80E-05	41%	67%	222--259	401--440	5	AAN09565	CG14619-PE, isoform E	69.71	1.16E-10	23%	41%	1--280	530--855
					6	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..	43.70	2.94E-05	37%	57%	224--272	252--301	6	EAL32347	GA13118-PA	65.08	2.86E-09	22%	40%	1--280	512--837
					7	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..	41.07	1.71E-04	29%	52%	214--279	263--328	7	EAL45629	ubiquitin carboxyl-terminal hydrolase, putative	53.53	8.60E-06	42%	63%	224--280	585--640
					8	COG5560	UBP12, Ubiquitin C-terminal hydrolase [Posttranslational modification, protein turnover, chaperones].	39.23	6.65E-04	20%	39%	6--168	270--446	8	EAL23715	ubiquitin specific protease 42	52.37	1.92E-05	22%	42%	5--284	114--417
					9	cd02666	Peptidase_C19J, 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..	37.79	1.78E-03	31%	55%	224--272	473--522	9	NP_115548	ubiquitin specific protease 42	52.37	1.92E-05	22%	42%	5--284	114--417
					10	cd02668	Peptidase_C19L, 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..	37.21	2.79E-03	25%	39%	186--262	187--287	10	XP_527662	PREDICTED: similar to ubiquitin specific protease 42	52.37	1.92E-05	22%	42%	5--284	114--417
B154L	68926--67913	338	38,141	7.96	1	COG1405	SUA7, Transcription initiation factor TFIIIB, Brf1 subunit/Transcription initiation factor TFIIB [Transcription].	73.41	3.60E-14	21%	42%	63--318	7--263	1	NP_048455	similar to Pyrococcus woesei factor TFIIB homolog, corresponds to GenBank Accession Number X70668	415.23	1.43E-114	66%	83%	49--338	1--290
														2	YP_142604	putative transcription initiation factor IIB	62.39	2.36E-08	22%	42%	68--325	154--432
														3	XP_626996	transcription initiation factor TFIIB Sua7p; ZnR+2cyclins	62.00	3.08E-08	25%	44%	60--315	154--410
														4	XP_470574	Putative transcription initiation factor IIB	56.61	1.29E-06	21%	42%	60--338	9--306
														5	EAM94502	Transcription factor TFIIB	55.45	2.88E-06	22%	41%	63--301	18--272
														6	AAT12348	transcription initiation factor TFIIB	55.07	3.76E-06	21%	42%	64--325	13--280
														7	XP_635486	transcription initiation factor IIB	53.91	8.38E-06	21%	39%	61--323	26--294
														8	AAT43923	transcription initiation factor IIB	53.53	1.09E-05	22%	40%	63--301	17--271
														9	AAS51927	ADR007Cp	51.22	5.43E-05	22%	38%	78--318	48--310
														10	XP_327575	hypothetical protein	50.83	7.09E-05	22%	38%	39--335	13--328
B157L	69356--68868	163	19,371	9.10		No Hit Found								1	NP_048457	A109L	183.73	1.56E-45	84%	92%	61--162	1--102
B159R	69422--72022	867	99,781	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	1.55E-05	22%	39%	269--436	2--167	1	NP_048462	A114R	891.34	0.00E+00	87%	94%	387--867	5--485
					2	COG0463	WcaA, Glycosyltransferases involved in cell wall biogenesis [Cell envelope biogenesis, outer membrane].	40.51	3.05E-04	27%	48%	265--371	3--104	2	NP_048459	A111R	766.92	0.00E+00	94%	98%	8--386	1--379
					3	COG1216	COG1216, Predicted glycosyltransferases [General function prediction only].	37.42	2.77E-03	18%	34%	265--536	3--279	3	CAG34747	hypothetical protein	102.83	5.31E-20	31%	49%	30--242	23--234
					4	COG1215	COG1215, Glycosyltransferases, probably involved in cell wall biogenesis [Cell envelope biogenesis, outer membrane].	37.22	3.11E-03	25%	45%	261--373	50--160	4	ZP_00202013	COG0463: Glycosyltransferases involved in cell wall biogenesis	101.29	1.54E-19	32%	51%	267--479	48--264
					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%	748--840	23--100	5	AAU37201	unknown	98.98	7.66E-19	26%	44%	267--493	6--241
														6	AAP95345	possible glycosyltransferase	97.83	1.71E-18	31%	47%	262--482	1--225
														7	ZP_00154867	COG1216: Predicted glycosyltransferases	90.89	2.09E-16	29%	43%	267--493	29--261
														8	AAY61346	unknown	65.86	7.19E-09	23%	46%	10--213	53--254
														9	ZP_00340129	hypothetical protein RakaH01000503	65.08	1.23E-08	24%	47%	10--213	53--254
														10	EAN09672	hypothetical protein EfaeDRAFT_1144	62.77	6.09E-08	26%	46%	70--224	65--219
B163R	72054--73082	343	38,745	6.41	1	COG1089	Gmd, GDP-D-mannose dehydratase [Cell envelope biogenesis, outer membrane].	487.10	1.04E-138	59%	74%	2--343	1--341	1	NP_048466	PBCV-1 GDP-D-mannose dehydratase	622.08	7.90E-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..	143.04	4.15E-35	30%	46%	6--334	1--300	2	BAC93113	GDP-mannose-4,6-dehydratase	398.67	1.42E-109	57%	74%	1--343	1--354
					3	COG0451	WcaG, Nucleoside-diphosphate-sugar epimerases [Cell envelope biogenesis, outer membrane / Carbohydrate transport and metabolism].	130.85	1.88E-31	29%	46%	4--343	1--311	3	CAB63300	GDP-mannose-4,6-dehydratase	398.28	1.85E-109	57%	73%	3--343	2--353
					4	COG1088	RfbB, dTDP-D-glucose 4,6-dehydratase [Cell envelope biogenesis, outer membrane].	116.47	3.70E-27	27%	47%	6--337	3--313	4	ZP_00533296	GDP-mannose 4,6-dehydratase	393.66	4.57E-108	58%	73%	3--342	2--351
					5	COG1087	GalE, UDP-glucose 4-epimerase [Cell envelope biogenesis, outer membrane].	102.94	4.34E-23	25%	45%	6--334	3--315	5	BAB03208	putative GDP-mannose dehydratase	393.28	5.97E-108	56%	74%	4--343	2--337
					6	COG1091	RfbD, dTDP-4-dehydrorhamnose reductase [Cell envelope biogenesis, outer membrane].	64.90	1.27E-11	24%	42%	6--335	3--276	6	ZP_00826187	COG1089: GDP-D-mannose dehydratase	391.73	1.74E-107	56%	73%	4--343	2--352
					7	COG0702	COG0702, Predicted nucleoside-diphosphate-sugar epimerases [Cell envelope biogenesis, outer membrane / Carbohydrate transport and metabolism].	52.61	6.12E-08	35%	51%	6--101	3--91	7	YP_113616	GDP-mannose 4,6-dehydratase	389.81	6.60E-107	56%	72%	1--340	1--350
					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	2.55E-07	30%	51%	6--72	6--76	8	AAR38453	GDP-mannose 4,6-dehydratase	388.27	1.92E-106	55%	71%	4--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	7.60E-06	29%	45%	1--116	194--318	9	ZP_00826190	COG1089: GDP-D-mannose dehydratase	386.34	7.29E-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	1.57E-04	23%	40%	2--168	249--405	10	AAM30355	GDP-mannose 4,6 dehydratase	384.03	3.62E-105	56%	72%	3--342	2--342
B165R	73136--74512	459	53,302	9.29	1	pfam01844	HNH, HNH endonuclease..	36.85	3.38E-03	35%	45%	335--391	3--52	1	NP_048435	A87R	572.01	1.41E-161	62%	76%	18--458	22--453
														2	NP_048779	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	137.12	1.16E-30	29%	48%	98--389	18--286

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
B168R	74535--74846	104	12,461	11.18		No Hit Found								1	NP_048469	A121R	170.24	1.51E-41	75%	92%	8--104	1--97
														2	AAL73467	Tlr 6Fp protein	84.73	8.37E-16	44%	65%	13--98	14--97
														3	AAL73477	Tlr 6Fp protein	83.19	2.44E-15	43%	62%	13--98	14--97
														4	AAL73473	Tlr 6Fp protein	83.19	2.44E-15	43%	65%	13--98	14--97
B170R	74904--79061	1386	144,436	4.82		No Hit Found								1	NP_048470	PBCV-1 Vp260 protein	1026.16	0.00E+00	58%	76%	1--919	1--971
														2	AAA86307	glycoprotein Vp260	877.86	0.00E+00	59%	78%	1--754	1--795
														3	NP_048471	contains a leucine zipper motif; similar to Bacillus subtilis phage P2A preneck appendage protein, corresponds to Swiss-Prot Accession Number P07537	516.54	2.58E-144	78%	87%	1076--1386	1--311
														4	BAB83469	Vp260 like protein	285.42	9.66E-75	31%	45%	99--895	13--901
														5	BAB83468	Vp260 like protein	284.26	2.15E-74	31%	47%	99--821	13--811
														6	BAB83470	Vp260 like protein	277.72	2.01E-72	31%	46%	99--821	13--811
														7	BAB83467	Vp260 like protein	276.56	4.49E-72	29%	44%	76--962	3--879
														8	BAB83471	Vp260 like protein	253.45	4.07E-65	28%	45%	44--820	32--827
														9	NP_048362	Asn/Thr/Ser/Val rich protein	182.57	8.82E-44	25%	42%	33--943	69--1089
														10	NP_048377	Asn/Thr/Ser/Val rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	174.48	2.40E-41	24%	44%	160--913	23--814
B173L	80294--79053	414	48,628	10.15	1	pfam01844	HNH, HNH endonuclease..	35.69	7.33E-03	32%	42%	85--122	13--51	1	NP_048779	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	64.70	6.38E-09	27%	43%	72--308	58--285
														2	NP_048435	A87R	56.61	1.74E-06	25%	39%	87--305	158--386
														3	NP_047162	putative HNH homing endonuclease	56.23	2.27E-06	29%	53%	32--181	14--161
														4	NP_048711	A354R	54.30	8.62E-06	31%	46%	193--311	61--169
														5	YP_142599	HNH endonuclease	53.91	1.13E-05	24%	42%	34--304	14--264
														6	YP_142601	HNH endonuclease	53.53	1.47E-05	23%	44%	32--298	22--277
B175L	81274--80699	192	22,609	9.94	1	pfam01096	TFIIS, Transcription factor S-II (TFIIS)..	44.20	2.17E-05	70%	78%	141--164	1--24	1	NP_048472	contains a zinc ribbon domain; similar to Chlorella virus CVU1 TFIIS-like transcription factor, corresponds to GenBank Accession Number D29631	326.25	3.05E-88	91%	97%	1--164	1--164
					2	smart00440	ZnF_C2C2, C2C2 Zinc finger; Nucleic-acid-binding motif in transcriptional elongation factor TFIIS and RNA polymerases. .	40.27	3.12E-04	52%	83%	141--164	1--24	2	BAA04187	transcription elongation factor SII	325.87	3.99E-88	91%	96%	1--164	1--164
					3	COG1594	RPB9, DNA-directed RNA polymerase, subunit M/Transcription elongation factor TFIIS [Transcription].	38.09	1.57E-03	29%	52%	101--164	34--96	3	BAA04186	transcription elongation factor SII	321.63	7.52E-87	90%	96%	1--164	1--164
														4	S47662	transcription elongation factor TFIIS homolog - Chlorella virus CV-U1	321.63	7.52E-87	90%	96%	1--164	1--164
														5	CAG98214	unnamed protein product	68.94	8.79E-11	45%	59%	83--164	196--275
														6	CAG86394	unnamed protein product	68.94	8.79E-11	31%	51%	41--164	149--273
														7	AAS54500	AGR011Wp	67.78	1.96E-10	44%	59%	82--164	207--287
														8	CAG58585	unnamed protein product	63.93	2.83E-09	37%	52%	54--164	180--289
														9	AAF71710	transcription elongation factor TFIIS	57.38	2.65E-07	31%	50%	55--164	202--307
														10	XP_503724	hypothetical protein	57.00	3.46E-07	31%	52%	43--164	157--278
B177R	81304--82020	239	26,206	10.23		No Hit Found								1	NP_048475	A127R	418.31	9.49E-116	82%	89%	1--238	1--244
B179L	83145--82027	373	43,103	9.31	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	42.07	1.07E-04	30%	50%	20--105	3--83		No Hit Found	No Hit Found						
B181L	83646--83242	135	15,941	5.27		No Hit Found								1	NP_048479	A131L	206.07	2.53E-52	69%	84%	1--135	1--136
B183L	84405--83734	224	26,260	4.90		No Hit Found								1	AAK23092	hypothetical protein	90.12	5.21E-17	30%	52%	5--203	10--210
														2	ZP_00811430	Methyltransferase FkbM	76.64	5.97E-13	28%	49%	10--206	53--248
														3	ZP_00517616	Methyltransferase FkbM	70.86	3.27E-11	26%	46%	10--205	18--218
														4	ABA05654	methyltransferase FkbM	68.94	1.24E-10	28%	46%	10--206	40--235
														5	BAC88534	glr0593	67.01	4.73E-10	25%	48%	13--204	38--231
														6	BAA10459	slr0907	62.00	1.52E-08	24%	49%	10--206	04--1003
														7	ABA23290	Methyltransferase FkbM	60.08	5.78E-08	24%	49%	14--191	45--224
														8	BAA18280	sll1173	60.08	5.78E-08	27%	49%	16--180	41--216
														9	CAJ23679	hypothetical protein	57.77	2.87E-07	25%	43%	16--186	40--218
														10	BAA17872	sll1950	56.23	8.35E-07	22%	43%	10--205	32--233
B185L	85553--85062	164	19,413	10.13	1	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	46.22	5.43E-06	29%	45%	9--89	3--78	1	NP_048482	similar to bacteriophage T4 intron-associated endonuclease, corresponds to Swiss-Prot Accession Number P13299	256.14	2.53E-67	76%	82%	1--164	1--164
														2	NP_048671	A315L	57.38	1.72E-07	34%	55%	9--95	2--87
														3	YP_293795	putative endonuclease	53.14	3.25E-06	31%	55%	8--92	2--88
														4	NP_048641	PBCV-1 33kd peptide	51.22	1.24E-05	35%	57%	20--95	17--92
														5	NP_048895	A539R	50.83	1.61E-05	47%	71%	11--56	34--79
														6	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	49.68	3.59E-05	35%	54%	9--93	2--87
B187R	85628--86062	145	16,410	11.55		No Hit Found								1	NP_048484	a136R	232.65	2.50E-60	81%	90%	4--145	5--146
B188R	86122--86412	97	11,558	10.72		No Hit Found								1	NP_048485	A137R	84.34	1.12E-15	66%	84%	6--62	13--69
B190L	86605--86306	100	11,806	10.21		No Hit Found								1	NP_048487	A139L	145.21	5.30E-34	76%	84%	17--100	20--103
B192R	86869--90093	1075	117,503	11.12	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	52.29	8.05E-08	23%	44%	986--1068	43--125	1	NP_048488	PBCV-1 surface protein	1298.88	0.00E+00	66%	73%	1--996	1--1027
					2	pfam05616	Neisseria_TspB, Neisseria meningitidis TspB protein. This family consists of several Neisseria meningitidis TspB virulence factor proteins..	46.28	5.55E-06	37%	43%	1000--1049	334--383	2	BAD12236	surface protein	1294.64	0.00E+00	66%	73%	1--996	1--1027
					3	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	43.59	3.18E-05	36%	42%	978--1072	27--122	3	BAD22850	surface protein	1292.33	0.00E+00	66%	73%	1--996	1--1027
					4	pfam06735	DUF1210, Protein of unknown function (DUF1210). This family represents a conserved region within plant proline-rich proteins..	42.02	1.08E-04	42%	45%	1001--1067	136--202	4	BAE02830	surface protein	657.14	0.00E+00	42%	54%	1--875	1--882
					5	COG3147	DedD, Uncharacterized protein conserved in bacteria [Function unknown].	41.16	1.68E-04	37%	40%	1000--1052	96--148	5	T17636	proline-rich protein A145R - Chlorella virus PBCV-1	305.06	8.94E-81	69%	75%	770--996	1--236
					6	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..	41.17	1.89E-04	25%	39%	998--1072	108--195	6	NP_048762	Pro-, Lys-rich, PAPK (30x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	80.11	4.68E-13	58%	74%	935--996	3--64
					7	COG5373	COG5373, Predicted membrane protein [Function unknown].	40.73	2.32E-04	29%	33%	999--1072	38--116	7	NP_048519	similar to PBCV-1 ORF A41R, corresponds to GenBank Accession Number U17055	62.00	1.32E-07	62%	73%	875--919	3--47
B197L	90498--90106	131	14,999	4.67		No Hit Found								1	NP_048498	A150L	172.94	2.34E-42	79%	89%	1--106	1--107
B199R	90758--91786	343	40,227	9.93		No Hit Found								1	NP_048711	A354R	114.78	4.10E-24	34%	50%	119--343	1--235
														2	NP_048779	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	75.87	2.11E-12	26%	42%	32--299	25--317
														3	NP_048435	A87R	55.07	3.85E-06	26%	41%	26--270	99--390
B201R	92170--92535	122	13,528	8.88		No Hit Found								1	NP_048499	A151R	234.96	5.02E-61	94%	95%	3--122	16--135
														2	NP_048496	A148R	59.31	3.77E-08	35%	56%	3--97	8--107
B203R	92613--93992	460	51,878	8.58	1	COG1061	SSL2, DNA or RNA helicases of superfamily II [Transcription / DNA replication, recombination, and repair].	130.60	2.03E-31	25%	40%	82--449	28--432	1	NP_048501	similar to phage T5 helicase, corresponds to Swiss-Prot Accession Number P11107	730.32	0.00E+00	78%	88%	1--460	1--459
					2	smart00487	DEXDc, DEAD-like helicases superfamily; .	60.24	3.07E-10	23%	38%	97--247	13--184	2	YP_142750	VV A18 helicase	197.98	5.57E-49	30%	50%	7--433	9--453

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	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..	52.00	9.91E-08	22%	39%	113--233	2--143	3	NP_077551	EsV-1-66	188.35	4.41E-46	31%	49%	16--429	9--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..	51.63	1.31E-07	23%	41%	113--233	2--144	4	ZP_00754006	COG1061: DNA or RNA helicases of superfamily II	153.30	1.57E-35	32%	45%	81--433	412--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%	115--234	189--321	5	YP_437404	DNA or RNA helicase of superfamily II	145.98	2.51E-33	31%	45%	90--433	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.	43.49	3.86E-05	18%	36%	97--269	15--206	6	BAB34728	hypothetical protein	144.05	9.54E-33	31%	48%	84--434	417--778
					7	COG4889	COG4889, Predicted helicase [General function prediction only].	40.35	3.26E-04	33%	54%	351--409	523--586	7	NP_287072	putative helicase	144.05	9.54E-33	31%	48%	84--434	425--786
					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	6.32E-04	17%	36%	305--399	17--124	8	ZP_00665630	Type III restriction enzyme, res subunit:DEAD/DEAH box helicase, N-terminal	124.41	7.82E-27	29%	44%	70--433	469--849
														9	EAM63429	Type III restriction enzyme, res subunit:DEAD/DEAH box helicase, N-terminal	124.02	1.02E-26	30%	47%	84--433	430--790
														10	ZP_00798957	Helicase, C-terminal:Type III restriction enzyme, res subunit:DEAD/DEAH box helicase, N-terminal	122.48	2.97E-26	29%	46%	84--433	438--821
B206L	94835--93999	279	31,758	9.98	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-strand break in the DNA near the insertion site of that element to facilitate homing at that site. Class I homing endonucleases are sorted into four families based on the presence of these motifs in their respective N-termini: LAGLIDADG, His-Cys box, HNH, and GIY-YIG. This CD contains several but not all members of the GIY-YIG family. The C-terminus of GIY-YIG is a DNA-binding domain which is separated from the N-terminus by a long, flexible linker. The DNA-binding domain consists of a minor-groove binding alpha-helix, and a helix-turn-helix. Some also contain a zinc finger (i.e. I-TevI) which is not required for DNA binding or catalysis, but is a component of the linker and directs the catalytic domain to cleave the homing site at a fixed distance from the intron insertion site.	52.70	6.07E-08	40%	57%	114--182	1--69	1	NP_048671	A315L	216.47	7.23E-55	42%	54%	1--272	1--240
					2	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	45.84	6.41E-06	39%	50%	2--89	3--83	2	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	216.47	7.23E-55	43%	52%	1--274	1--225
					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.	42.84	5.30E-05	29%	45%	1--85	1--88	3	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	156.76	6.80E-37	64%	79%	1--108	1--110
					4	pfam07453	NUMOD1, NUMOD1 domain..	36.56	4.29E-03	41%	56%	220--252	1--33	4	NP_048641	PBCV-1 33kd peptide	150.21	6.37E-35	35%	49%	3--275	8--250
					5	smart00497	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished). .	36.26	5.68E-03	35%	50%	220--274	1--53	5	YP_293795	putative endonuclease	82.80	1.25E-14	41%	57%	2--107	3--112
														6	NP_048708	KKD (6X), mixed charge	54.30	4.75E-06	39%	54%	4--92	28--120
														7	NP_048895	A539R	54.30	4.75E-06	33%	48%	4--104	34--133
														8	NP_899393	SegD	52.76	1.38E-05	27%	43%	1--170	1--175
														9	CAA38804	GIY COII i1 grp IB protein	51.22	4.02E-05	29%	46%	15--173	86--234
B207L	95322--94990	111	12,206	3.97		No Hit Found								1	NP_048505	A157L	163.70	1.43E-39	78%	92%	1--101	1--101
B208L	95646--95362	95	11,379	3.37		No Hit Found								1	NP_048506	A158L	87.04	1.74E-16	53%	74%	10--89	18--97
B209R	95821--96129	103	11,747	11.19		No Hit Found								1	NP_048509	A161R	67.40	1.39E-10	45%	52%	4--103	14--123
B211L	96657--96322	112	12,740	5.72		No Hit Found								1	NP_048513	A165L	172.94	2.35E-42	74%	84%	1--112	168--279
B212L	97112--96669	148	17,373	10.10		No Hit Found								1	NP_048513	A165L	153.30	1.97E-36	65%	75%	8--120	22--134
B214R	97201--98004	268	31,318	4.54	1	COG5377	COG5377, Phage-related protein, predicted endonuclease [DNA replication, recombination, and repair].	36.95	3.22E-03	20%	36%	34--214	17--190	1	NP_048514	PBCV-1 exonuclease	470.70	1.97E-131	79%	90%	1--268	1--268
														2	AAG28903	F12A21.19	90.12	7.25E-17	31%	48%	19--211	84--276
														3	NP_176934	unknown protein	90.12	7.25E-17	31%	48%	19--211	105--297
														4	CAF27185	Exonuclease	77.03	6.35E-13	25%	43%	30--219	3--201
														5	CAF27488	Exonuclease	76.64	8.30E-13	25%	43%	30--212	3--195
														6	XP_480865	unknown protein	74.33	4.12E-12	29%	43%	19--209	124--317
														7	XP_472944	OSJNBa0081L15.20	73.56	7.02E-12	32%	46%	19--169	131--296
														8	AAR26918	FirrV-1-B43 precursor	73.56	7.02E-12	28%	48%	14--212	4--196
														9	NP_077549	EsV-1-64	68.17	2.95E-10	27%	50%	15--202	1--179
														10	YP_142708	Lambda-type exonuclease	66.63	8.59E-10	31%	46%	27--174	178--339
B215R	98046--98546	167	18,403	4.84		No Hit Found								1	NP_048516	A168R	246.13	2.79E-64	72%	77%	2--167	1--166
B218R	98605--99804	400	47,205	6.15		No Hit Found								1	NP_048711	A354R	155.22	3.44E-36	39%	56%	158--400	4--235
														2	NP_048779	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	68.17	5.53E-10	25%	45%	90--366	47--318
B222R	99830--100792	321	36,608	6.30	1	COG0540	PyrB, Aspartate carbamoyltransferase, catalytic chain [Nucleotide transport and metabolism].	305.99	3.46E-84	43%	60%	24--318	8--310	1	NP_048517	PBCV-1 aspartate transcarbamylase	554.67	1.40E-156	85%	92%	1--321	1--321
					2	COG0078	ArgF, Ornithine carbamoyltransferase [Amino acid transport and metabolism].	163.42	2.88E-41	31%	50%	24--320	7--310	2	CAC85728	aspartate carbamoyltransferase	287.73	3.20E-76	48%	67%	7--315	56--370
					3	pfam02729	OTCace_N, Aspartate/ornithine carbamoyltransferase, carbamoyl-P binding domain..	146.14	4.46E-36	45%	61%	24--163	1--143	3	Q43064	Aspartate carbamoyltransferase 3, chloroplast precursor (Aspartate transcarbamylase 3) (ATCase 3)	284.26	3.54E-75	50%	68%	26--315	88--384
					4	pfam00185	OTCace, Aspartate/ornithine carbamoyltransferase, Asp/Orn binding domain..	114.21	1.99E-26	33%	52%	166--315	1--156	4	Q43087	Aspartate carbamoyltransferase 2, chloroplast precursor (Aspartate transcarbamylase 2) (ATCase 2)	281.95	1.75E-74	50%	68%	26--315	82--378
														5	CAA50687	aspartate carbamoyltransferase	281.18	2.99E-74	50%	68%	26--315	87--383
														6	NP_188668	amino acid binding / aspartate carbamoyltransferase/ carboxyl- and carbamoyltransferase/ ornithine carbamoyltransferase	280.03	6.67E-74	50%	68%	26--315	87--383
														7	Q43086	Aspartate carbamoyltransferase 1, chloroplast precursor (Aspartate transcarbamylase 1) (ATCase 1)	276.94	5.64E-73	48%	69%	26--315	83--379
														8	AAL90999	AT3g20330/MQC12_8	275.79	1.26E-72	49%	68%	26--315	87--383
														9	XP_480986	aspartate carbamoyltransferase	267.70	3.42E-70	46%	65%	26--315	55--356
														10	CAA52201	aspartate carbamoyltransferase	264.62	2.90E-69	52%	70%	57--315	1--261
B224R	100805--101908	368	40,380	10.10		No Hit Found								1	NP_048519	similar to PBCV-1 ORF A41R, corresponds to GenBank Accession Number U17055	606.29	5.01E-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	515.38	1.16E-144	82%	88%	56--368	101--412
														3	NP_048488	PBCV-1 surface protein	65.86	2.44E-09	54%	64%	6--64	905--963
														4	BAD22850	surface protein	65.86	2.44E-09	54%	64%	6--64	905--963
														5	T17636	proline-rich protein A145R - Chlorella virus PBCV-1	65.86	2.44E-09	54%	64%	6--64	114--172
														6	BAD12236	surface protein	62.77	2.06E-08	52%	62%	6--64	905--963
														7	BAD86968	hypothetical protein	60.46	1.02E-07	23%	40%	62--314	95--340

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	NP_916095	P0481E12.18	60.46	1.02E-07	23%	40%	62--314	72--317
														9	AAM63817	unknown	58.92	2.98E-07	23%	44%	63--288	63--289
														10	NP_191439	unknown protein	58.92	2.98E-07	23%	44%	63--288	63--289
B226L	102754--101918	279	30,632	8.06	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	4.08E-29	33%	49%	19--192	1--179	1	NP_048521	similar to E. coli hypothetical protein, corresponds to Swiss-Prot Accession Number P39407	512.69	4.88E-144	90%	96%	4--279	13--288
					2	COG1752	RssA, Predicted esterase of the alpha-beta hydrolase superfamily [General function prediction only].	87.06	2.59E-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	5.34E-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	81.26	3.63E-14	26%	46%	19--261	8--261
														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)	71.63	2.88E-11	23%	44%	5--276	48--326
B230L	103948--102803	382	43,399	6.59	1	COG4123	COG4123, Predicted O-methyltransferase [General function prediction only].	58.72	8.92E-10	27%	40%	48--162	45--171	1	AAC03125	DNA adenine methyltransferase	780.40	0.00E+00	100%	100%	1--382	1--382
					2	COG0286	HsdM, Type I restriction-modification system methyltransferase subunit [Defense mechanisms].	55.04	1.32E-08	21%	35%	3--220	145--386	2	P52284	Modification methylase CviRI (Adenine-specific methyltransferase CviRI) (M.CviRI)	617.46	2.28E-175	79%	88%	1--381	1--378
					3	COG2813	RsmC, 16S RNA G1207 methylase RsmC [Translation, ribosomal structure and biogenesis].	49.88	4.14E-07	27%	43%	40--118	151--234	3	CAA29835	unnamed protein product	216.47	1.17E-54	33%	54%	10--381	12--377
					4	COG2263	COG2263, Predicted RNA methylase [Translation, ribosomal structure and biogenesis].	48.69	1.12E-06	29%	50%	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	1.16E-05	23%	42%	33--129	93--203	5	AAC57945	DNA adenine methyltransferase	198.36	3.30E-49	31%	54%	8--380	6--357
					6	COG0421	SpeE, Spermidine synthase [Amino acid transport and metabolism].	42.65	6.09E-05	35%	47%	47--113	76--154	6	AAC57943	DNA adenine methyltransferase	196.44	1.26E-48	31%	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..	39.09	6.95E-04	20%	46%	25--118	5--107	7	YP_063409	cpp14	66.24	1.96E-09	30%	46%	14--169	273--429
					8	COG4106	Tam, Trans-aconitate methyltransferase [General function prediction only].	36.82	3.60E-03	25%	41%	38--162	21--130	8	AAW34165	unknown	66.24	1.96E-09	30%	46%	14--169	273--429
					9	COG0116	COG0116, Predicted N6-adenine-specific DNA methylase [DNA replication, recombination, and repair].	36.81	3.73E-03	27%	42%	29--118	218--309	9	AAR29548	cpp14	66.24	1.96E-09	30%	46%	14--169	273--429
					10	COG4262	COG4262, Predicted spermidine synthase with an N-terminal membrane domain [General function prediction only].	35.68	7.68E-03	26%	45%	57--175	299--416	10	ZP_00371038	helicase, Snf2 family	63.16	1.66E-08	29%	46%	14--169	195--351
B235L	104315--104007	103	12,059	9.39		No Hit Found								1	NP_049043	A687R	113.62	1.69E-24	71%	86%	26--101	1--75
B236L	105293--104322	324	37,286	7.70	1	COG3392	COG3392, Adenine-specific DNA methylase [DNA replication, recombination, and repair].	201.75	8.76E-53	38%	58%	4--303	1--311	1	AAC03127	DNA adenine methyltransferase	610.53	2.19E-173	94%	94%	1--324	1--324
					2	pfam02086	MethyltransfD12, D12 class N6 adenine-specific DNA methyltransferase..	150.87	1.65E-37	33%	47%	6--284	1--253	2	NP_048600	M.CviAII adenine DNA methyltransferase	493.04	5.09E-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	491.12	1.93E-137	76%	84%	1--321	1--324
														4	AAC57944	DNA adenine methyltransferase	489.19	7.35E-137	75%	83%	1--321	1--324
														5	YP_392646	Site-specific DNA-methyltransferase (adenine-specific)	177.18	6.18E-43	34%	52%	4--320	1--332
														6	ZP_00371322	ulcer associated adenine specific DNA methyltransferase	177.18	6.18E-43	36%	54%	4--305	1--316
														7	CAA38356	NlaIII methyltransferase	176.41	1.05E-42	35%	51%	4--316	1--327
														8	AAP77647	conserved hypothetical protein	172.56	1.52E-41	36%	49%	4--305	1--319
														9	AAC45814	methylase HpyI	172.56	1.52E-41	36%	53%	4--305	1--311
														10	AAC45818	methylase HpyI	171.79	2.60E-41	37%	54%	4--305	1--311
B239R	105405--107825	807	88,363	8.80	1	smart00637	CBD_II, CBD_II domain..	70.74	2.07E-13	27%	38%	11--105	3--101	1	BAA78554	vChti-1	1324.69	0.00E+00	79%	85%	1--806	1--835
					2	pfam00704	Glyco_hydro_18, Glycosyl hydrolases family 18..	64.00	2.45E-11	25%	37%	540--716	4--195	2	NP_048529	PBCV-1 chitinase	1319.68	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
B246R	108041--108646	202	23,666	10.26	1	smart00497	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished). .	49.75	5.05E-07	40%	54%	148--200	1--53	1	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	217.62	1.72E-55	49%	63%	1--199	1--224
					2	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-strand break in the DNA near the insertion site of that element to facilitate homing at that site. Class I homing endonucleases are sorted into four families based on the presence of these motifs in their respective N-termini: LAGLIDADG, His-Cys box, HNH, and GIY-YIG. This CD contains several but not all members of the GIY-YIG family. The C-terminus of GIY-YIG is a DNA-binding domain which is separated from the N-terminus by a long, flexible linker. The DNA-binding domain consists of a minor-groove binding alpha-helix, and a helix-turn-helix. Some also contain a zinc finger (i.e. I-TevI) which is not required for DNA binding or catalysis, but is a component of the linker and directs the catalytic domain to cleave the homing site at a fixed distance from the intron insertion site.	43.45	4.06E-05	42%	50%	95--198	23--113	2	NP_048671	A315L	215.31	8.53E-55	45%	59%	1--200	1--242
					3	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	39.29	6.26E-04	39%	50%	2--89	3--83	3	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	173.71	2.85E-42	52%	68%	1--150	1--159
					4	pfam07453	NUMOD1, NUMOD1 domain..	39.26	6.38E-04	48%	55%	148--181	1--34	4	NP_048641	PBCV-1 33kd peptide	134.81	1.47E-30	35%	50%	11--199	15--248
														5	YP_293795	putative endonuclease	80.11	4.28E-14	35%	54%	2--129	3--137
														6	NP_899393	SegD	62.00	1.21E-08	34%	52%	1--134	1--136
														7	ZP_00506764	Excinuclease ABC, C subunit, N-terminal	53.14	5.61E-06	39%	54%	2--92	310--408
														8	YP_024462	putative endonuclease	52.37	9.56E-06	27%	41%	68--197	2--163
														9	NP_048482	similar to bacteriophage T4 intron-associated endonuclease, corresponds to Swiss-Prot Accession Number P13299	51.99	1.25E-05	28%	47%	2--126	9--135
														10	CAA25939	unnamed protein product	51.99	1.25E-05	50%	61%	149--199	92--143
B249R	108892--111171	760	86,969	8.23	1	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 .	314.46	1.10E-86	32%	51%	181--649	1--475	1	P30320	DNA polymerase	1325.07	0.00E+00	94%	94%	1--698	1--693
					2	cd00145	POLBc, DNA polymerase type-B family; DNA directed DNA polymerase. Posseses DNA binding, polymerase and 3'-5' exonuclease activity..	295.30	5.74E-81	34%	50%	181--682	1--508	2	BAA35142	DNA polymerase	1211.05	0.00E+00	83%	90%	1--698	1--693
					3	COG0417	PolB, DNA polymerase elongation subunit (family B) [DNA replication, recombination, and repair].	252.28	5.43E-68	29%	48%	27--683	11--600	3	NP_048532	PBVC-1 DNA polymerase	1210.67	0.00E+00	83%	90%	1--698	1--693
					4	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..	227.68	1.29E-60	26%	41%	32--359	1--334	4	AAK28951	DNA polymerase	422.55	2.60E-116	100%	100%	473--682	1--210
					5	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..	223.33	2.86E-59	42%	57%	432--683	1--255	5	AAK28923	DNA polymerase	417.16	1.09E-114	98%	99%	473--682	1--210
														6	AAK28956	DNA polymerase	414.85	5.42E-114	96%	99%	473--682	1--210
														7	AAK28952	DNA polymerase	412.15	3.51E-113	96%	99%	473--682	1--210

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	AAK28963	DNA polymerase	396.74	1.53E-108	90%	97%	473--682	1--210
														9	AAK28965	DNA polymerase	396.36	1.99E-108	90%	97%	473--682	1--210
														10	AAK28930	DNA polymerase	395.20	4.44E-108	90%	96%	473--682	1--210
B253R	111273--111716	148	17,283	9.57	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..	72.33	8.64E-14	29%	52%	1--98	344--442	1	P30320	DNA polymerase	304.29	6.90E-82	100%	100%	1--148	766--913
					2	COG0417	PolB, DNA polymerase elongation subunit (family B) [DNA replication, recombination, and repair].	42.34	7.49E-05	33%	55%	1--98	681--775	2	NP_048532	PBVC-1 DNA polymerase	264.23	7.91E-70	86%	91%	1--148	766--913
														3	BAA35142	DNA polymerase	260.77	8.74E-69	84%	91%	1--148	766--913
														4	BAE06251	B-family DNA polymerase	82.03	5.57E-15	34%	59%	1--148	02--1143
														5	EAL49087	DNA polymerase delta catalytic subunit, putative	63.54	2.05E-09	38%	61%	6--104	861--958
														6	CAE75373	Hypothetical protein CBG23360	62.00	5.96E-09	32%	58%	6--133	853--978
														7	XP_623795	PREDICTED: similar to ENSANGP00000014184	62.00	5.96E-09	33%	51%	11--134	90--1009
														8	CAB04077	Hypothetical protein F10C2.4	61.62	7.78E-09	32%	59%	6--122	853--966
														9	EAA00051	ENSANGP00000014184	60.85	1.33E-08	31%	50%	19--148	859--969
														10	CAA43922	DNA polymerase III catalytic subunit	60.08	2.26E-08	36%	52%	1--100	884--978
B255R	111779--112036	86	10,382	6.53		No Hit Found									No Hit Found	No Hit Found						
B256R	112097--112876	260	29,718	5.07		No Hit Found								1	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	50.83	4.66E-05	28%	42%	81--233	52--218
B258R	112905--116738	1278	142,006	11.60	1	COG3064	TolA, Membrane protein involved in colicin uptake [Cell envelope biogenesis, outer membrane].	48.25	1.36E-06	27%	42%	999--1175	127--289	1	NP_048536	similar to SWI/SNF chromatin remodeling complex subunit OSA2	1289.63	0.00E+00	57%	63%	2--1275	12--1299
					2	COG1196	Smc, Chromosome segregation ATPases [Cell division and chromosome partitioning].	45.40	1.08E-05	14%	42%	1007--1262	169--403	2	T17682	hypothetical protein A192R - Chlorella virus PBCV-1	222.25	9.22E-56	44%	52%	948--1275	8--360
					3	COG5022	COG5022, Myosin heavy chain [Cytoskeleton].	41.89	9.96E-05	15%	37%	1011--1273	761--1024	3	T17681	hypothetical protein a191R - Chlorella virus PBCV-1	59.69	7.92E-07	57%	68%	876--935	2--60
B261L	117526--116741	262	29,688	4.72	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..	81.47	1.32E-16	28%	61%	7--129	1--122	1	NP_048540	similar to human PCNA, corresponds to Swiss-Prot Accession Number P12004	513.46	2.54E-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..	78.44	1.24E-15	32%	53%	137--259	4--127	2	XP_395519	PREDICTED: similar to ENSANGP00000012272	149.44	9.64E-35	29%	53%	7--259	533--785
					3	COG0592	DnaN, DNA polymerase sliding clamp subunit (PCNA homolog) [DNA replication, recombination, and repair].	62.62	6.33E-11	21%	41%	19--261	72--323	3	XP_514499	PREDICTED: proliferating cell nuclear antigen	147.90	2.81E-34	29%	56%	7--259	1--253
														4	AAX43349	proliferating cell nuclear antigen	147.90	2.81E-34	29%	56%	7--259	1--253
														5	AAX36355	proliferating cell nuclear antigen	147.90	2.81E-34	29%	56%	7--259	1--253
														6	AAB27811	PCNA	147.52	3.66E-34	30%	54%	7--259	1--253
														7	AAH64299	Pcna protein	147.52	3.66E-34	29%	56%	7--259	1--253
														8	XP_534355	PREDICTED: similar to proliferating cell nuclear antigen	147.52	3.66E-34	29%	55%	3--259	202--458
														9	BAE47145	proliferating cell nuclear antigen	147.13	4.79E-34	30%	56%	7--259	1--252
														10	1AXC_E	Chain E, Human Pcna	147.13	4.79E-34	29%	56%	8--259	2--253
B264L	118018--117563	152	17,348	7.64		No Hit Found								1	NP_048543	A196L	266.16	2.05E-70	79%	86%	1--152	1--152
B267R	118059--118358	100	11,011	10.76		No Hit Found								1	NP_048546	A199R	134.04	1.22E-30	82%	91%	1--83	1--84
B268L	118935--118366	190	21,806	8.56	1	pfam01753	zf-MYND, MYND finger..	40.41	3.13E-04	50%	65%	116--150	4--38	1	NP_849969	SDG37	55.84	7.48E-07	42%	57%	114--167	57--106
														2	AAD03568	putative SET-domain transcriptional regulator	55.84	7.48E-07	42%	57%	114--167	57--106
														3	CAG07000	unnamed protein product	54.30	2.18E-06	36%	56%	73--150	407--482
														4	XP_598182	PREDICTED: similar to SET and MYND domain-containing protein 3 (Zinc finger MYND domain-containing protein 1), partial	53.91	2.84E-06	45%	59%	107--150	23--66
														5	CAE59608	Hypothetical protein CBG03016	53.14	4.85E-06	38%	52%	118--173	32--99
														6	CAD43192	egg laying nine 1 protein	51.99	1.08E-05	44%	61%	112--158	13--59
														7	CAG09553	unnamed protein product	51.99	1.08E-05	36%	53%	112--176	7--66
														8	EAA74929	hypothetical protein FG06312.1	51.60	1.41E-05	52%	76%	117--150	40--1173
														9	CAA86783	Hypothetical protein R06F6.4	51.60	1.41E-05	40%	57%	106--163	19--76
														10	CAG04324	unnamed protein product	51.22	1.84E-05	60%	69%	118--150	47--79
B271R	119006--119440	145	16,487	10.35	1	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 mechanism, activating a water molecule to form a hydroxide ion that performs a nucleophilic attack on the substrate. The functional enzyme is a homodimer. Cytosine deaminase catalyzes the deamination of cytosine to uracil and ammonia and is a member of the pyrimidine salvage pathway. Cytosine deaminase is found in bacteria and fungi but is not present in mammals; for this reason, the enzyme is currently of interest for antimicrobial drug design and gene therapy applications against tumors. Some members of this family are tRNA-specific adenosine deaminases that generate inosine at the first position of their anticodon (position 34) of specific tRNAs; this modification is thought to enlarge the codon recognition capacity during protein synthesis. Other members of the family are guanine deaminases which deaminate guanine to xanthine as part of the utilization of guanine	53.34	4.30E-08	34%	49%	29--134	1--97	1	NP_048547	contains cytidine and deoxycytidine deaminase Zn-binding region signature	193.74	1.29E-48	75%	90%	26--143	1--118
					2	COG0590	CumB, Cytosine/adenosine deaminases [Nucleotide transport and metabolism / Translation, ribosomal structure and biogenesis].	51.49	1.60E-07	27%	49%	22--138	5--112	2	AAR26853	FirrV-1-A29	51.60	7.91E-06	29%	51%	47--132	24--101
B272L	119752--119453	100	10,751	10.64		No Hit Found								1	NP_048548	A201L	119.40	3.11E-26	61%	73%	1--90	1--93
B273L	120114--119779	112	12,175	5.05		No Hit Found								1	NP_048549	A202L	204.14	9.52E-52	85%	91%	1--112	1--112
B274R	120178--120825	216	24,068	5.13		No Hit Found								1	NP_048550	A203R	342.04	7.00E-93	92%	94%	35--216	35--216
B277R	120837--121454	206	22,639	12.01		No Hit Found								1	NP_048552	A205R	174.10	2.29E-42	51%	58%	17--205	10--205
B278R	121519--122634	372	41,925	4.58	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..	213.96	1.94E-56	39%	58%	25--260	4--246	1	NP_048554	PBCV-1 arginine decarboxylase	680.25	0.00E+00	86%	94%	1--372	1--372
					2	COG0019	LysA, Diaminopimelate decarboxylase [Amino acid transport and metabolism].	200.12	2.87E-52	25%	44%	19--372	29--394	2	AAD02222	ornithine decarboxylase	275.79	1.56E-72	39%	57%	19--372	60--428
					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. .	86.41	4.58E-18	30%	44%	263--365	1--103	3	P07805	Ornithine decarboxylase (ODC)	275.79	1.56E-72	39%	57%	19--372	38--406
					4	COG1166	SpeA, Arginine decarboxylase (spermidine biosynthesis) [Amino acid transport and metabolism].	43.70	2.85E-05	28%	43%	148--270	238--374	4	1NJJ_D	Chain D, Crystal Structure Determination Of T. Brucei Ornithine Decarboxylase Bound To D-Ornithine And To G418	275.79	1.56E-72	39%	57%	19--372	40--408
														5	1F3T_D	Chain D, Crystal Structure Of Trypanosoma Brucei Ornithine Decarboxylase (Odc) Complexed With Putrescine, Odc's Reaction Product.	275.79	1.56E-72	39%	57%	19--372	40--408
														6	1SZR_B	Chain B, A Dimer Interface Mutant Of Ornithine Decarboxylase Reveals Structure Of Gem Diamine Intermediate	274.63	3.48E-72	39%	57%	19--372	40--408
														7	AAV88093	ornithine decarboxylase 1	274.25	4.55E-72	38%	57%	18--372	39--408
														8	AAX36104	ornithine decarboxylase 1	274.25	4.55E-72	38%	57%	18--372	39--408
														9	EAA00421	ENSANGP00000020224	273.48	7.76E-72	39%	58%	3--370	19--389
														10	2TOD_D	Chain D, Ornithine Decarboxylase From Trypanosoma Brucei K69a Mutant In Complex With Alpha-Difluoromethylornithine	273.48	7.76E-72	39%	57%	19--372	40--408

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
B283R	122739--123407	223	25,833	4.82	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 chromosome 12p. This results in an extended polyglutamine region in atrophin-1, that is thought to confer toxicity to the protein, possibly through altering its interactions with other proteins. The expansion of a CAG repeat is also the underlying defect in six other neurodegenerative disorders, including Huntington's disease. One interaction of expanded polyglutamine repeats that is thought to be pathogenic is that with the short glutamine repeat in the transcriptional coactivator CREB binding protein, CBP. This interaction draws CBP away from its usual nuclear location to the expanded polyglutamine repeat protein aggregates that are characteristic of the polyglutamine neurodegenerative disorders. This interferes with CBP-mediated transcription and causes cytotoxicity..	42.59	5.95E-05	41%	70%	1--65	649--705	1	NP_048555	KAEKA (6X), SDDD (7X)	74.33	2.93E-12	32%	40%	1--155	30--180
B284L	123853--123410	148	16,593	4.71		No Hit Found								1	NP_048560	A213L	250.37	1.18E-65	80%	93%	1--146	1--146
B285L	124281--123883	133	15,174	7.19		No Hit Found								1	NP_048561	A214L	174.10	1.07E-42	79%	88%	1--110	1--110
B286R	124399--125487	363	42,043	5.78	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	4.16E-03	22%	46%	31--102	12--79		No Hit Found	No Hit Found						
B288L	126376--125459	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
														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
B289L	127526--126381	382	44,122	10.08	1	COG1233	COG1233, Phytoene dehydrogenase and related proteins [Secondary metabolites biosynthesis, transport, and catabolism].	43.54	3.27E-05	43%	52%	6--49	5--49	1	NP_048564	similar to bovine monoamine oxidase, corresponds to Swiss-Prot Accession Number P21398	615.15	1.13E-174	84%	89%	20--382	31--394
					2	COG1232	HemY, Protoporphyrinogen oxidase [Coenzyme metabolism].	40.29	3.13E-04	17%	31%	7--360	3--420	2	BAA83788	orf1	614.38	1.93E-174	84%	89%	20--382	34--397
B291L	127983--127570	138	16,027	10.02		No Hit Found								1	NP_048575	A227L	201.06	8.04E-51	78%	90%	18--138	17--137
B292L	128227--128003	75	8,400	7.94		No Hit Found								1	NP_048577	A229L	111.31	8.59E-24	66%	83%	1--75	1--77
B296L	128851--128618	78	9,780	9.05		No Hit Found									No Hit Found	No Hit Found						
B297R	128897--129988	364	41,805	5.18	1	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	35.44	9.95E-03	26%	48%	32--106	13--78		No Hit Found	No Hit Found						
B298L	131112--129997	372	43,124	10.16		No Hit Found								1	NP_048579	contains ATP/GTP-binding motif A	591.27	1.68E-167	75%	84%	5--372	9--383
														2	NP_048983	similar to Chlorella virus PBCV-1 ORF A231L, corresponds to GenBank Accession Number U42580	305.06	2.41E-81	45%	63%	6--330	103--427
B302R	131221--131553	111	13,385	10.24		No Hit Found								1	NP_048581	A233R	175.64	3.64E-43	73%	83%	1--111	1--112
B303L	131879--131556	108	12,654	10.59		No Hit Found								1	NP_048582	A234L	158.69	4.66E-38	66%	86%	1--107	1--107
B305R	131942--133462	507	57,119	9.49	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	894.03	0.00E+00	89%	92%	25--507	36--518
					2	COG1748	LYS9, Saccharopine dehydrogenase and related proteins [Amino acid transport and metabolism].	48.75	9.25E-07	22%	42%	32--463	2--369	2	ZP_00589757	Homospermidine synthase	313.15	1.35E-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	47.94	1.48E-06	26%	50%	34--190	1--138	3	AAM05046	homospermidine synthase	283.49	1.14E-74	35%	54%	33--479	15--459
														4	AAM29862	homospermidine synthase	278.10	4.81E-73	34%	54%	33--479	8--452
														5	ZP_00543002	Homospermidine synthase	277.72	6.28E-73	34%	54%	33--479	8--452
														6	ZP_00584810	Homospermidine synthase	243.82	1.00E-62	33%	52%	33--479	13--446
														7	AAU28555	homospermidine synthase	232.26	3.02E-59	33%	50%	29--494	11--460
														8	AAP94695	putative homospermidine synthase	231.88	3.95E-59	33%	51%	32--483	9--449
														9	CAH16656	hypothetical protein	231.49	5.16E-59	32%	50%	29--494	11--460
														10	BAE51070	Homospermidine synthase	230.34	1.15E-58	32%	50%	33--479	30--462
B310R	133532--134677	382	43,272	8.28		No Hit Found								1	NP_048920	similar to Chlorella virus PBCV-1 ORF A154L, corresponds to GenBank Accession Number U42580	516.15	7.16E-145	68%	81%	38--382	4--351
														2	NP_048502	A154L	502.29	1.07E-140	68%	80%	38--381	3--347
														3	NP_048477	similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903	455.68	1.15E-126	59%	71%	3--382	5--357
														4	NP_077492	EsV-1-7	109.00	2.64E-22	30%	45%	41--269	10--261
														5	AAG37861	variant-specific surface protein VSP136-4	56.23	2.03E-06	23%	36%	23--237	139--384
														6	AAF69839	variant-specific surface protein VSP136b	56.23	2.03E-06	23%	36%	23--237	59--304
														7	AAA74587	cysteine rich protein	56.23	2.03E-06	23%	36%	23--237	59--304
														8	XP_770190	variant-specific surface protein	51.22	6.52E-05	25%	35%	8--238	381--602
B314L	135118--134684	145	16,946	6.95		No Hit Found								1	NP_048587	A239L	125.95	3.30E-28	54%	75%	37--144	1--109
B316R	135269--137443	725	82,710	6.69	1	COG4581	COG4581, Superfamily II RNA helicase [DNA replication, recombination, and repair].	355.05	5.35E-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].	178.28	1.03E-45	30%	50%	25--410	47--436	2	CAD25317	ATP-DEPENDENT RNA HELICASE (SKI2 FAMILY)	318.93	3.82E-85	28%	47%	12--725	66--881
					3	COG1202	COG1202, Superfamily II helicase, archaea-specific [General function prediction only].	128.93	6.98E-31	28%	51%	29--398	236--577	3	XP_667944	ATP-dependent RNA helicase; ATP-dependent RNA 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	2.24E-25	24%	40%	9--192	9--206	4	XP_627152	Mtr4p like SKI family SFII helicase	317.78	8.51E-85	35%	54%	11--518	100--640
					5	smart00487	DEXDc, DEAD-like helicases superfamily; .	96.45	4.72E-21	24%	40%	3--184	1--195	5	NP_703683	ATP dependent RNA helicase, putative	315.46	4.22E-84	32%	53%	6--587	250--831
					6	COG1201	Lhr, Lhr-like helicases [General function prediction only].	81.06	1.69E-16	26%	47%	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	1.42E-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].	71.58	1.43E-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	2.12E-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

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
					10	COG0513	SrmB, Superfamily II DNA and RNA helicases [DNA replication, recombination, and repair / Transcription / Translation, ribosomal structure and biogenesis].	57.86	1.52E-09	20%	41%	15--417	56--414	10	NP_012485	Dead-box family ATP dependent helicase required for mRNA export from the nucleus; co-factor of the exosome complex, required for 3' end formation of 5.8S rRNA; Mtr4p	309.30	3.03E-82	37%	58%	6--432	145--612
B322R	137470--138396	309	33,067	10.94	1	pfam00967	Barwin, Barwin family..	154.41	1.52E-38	43%	54%	181--302	1--119	1	NP_048594	Pro-rich, PAPK (20X); similar to Arabidopsis anter-specific Pro-rich protein, corresponds to Swiss-Prot Accession Number P40602	316.24	7.87E-85	96%	98%	157--309	135--288
					2	pfam05616	Neisseria_TspB, Neisseria meningitidis TspB protein. This family consists of several Neisseria meningitidis TspB virulence factor proteins..	52.06	9.98E-08	33%	39%	79--158	323--402	2	NP_187123	PR4 (PATHOGENESIS-RELATED 4)	70.86	5.76E-11	35%	49%	161--303	60--192
					3	COG5373	COG5373, Predicted membrane protein [Function unknown].	46.89	3.24E-06	27%	40%	24--126	33--124	3	AAC33732	PR-4 type protein	68.94	2.19E-10	37%	46%	181--303	23--142
					4	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..	44.63	1.80E-05	25%	32%	27--133	109--214	4	BAC16357	hevein-like protein	67.40	6.37E-10	34%	44%	179--303	71--212
					5	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	43.59	3.71E-05	31%	39%	21--160	28--166	5	CAA42820	PR-4a protein	63.16	1.20E-08	34%	46%	181--302	27--145
					6	COG3115	ZipA, Cell division protein [Cell division and chromosome partitioning].	42.35	7.10E-05	22%	37%	22--148	60--185	6	CAA41437	pathogenesis-related protein 4A	62.39	2.05E-08	34%	46%	181--302	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.	41.12	1.76E-04	27%	54%	70--155	41--126	7	CAA41438	pathogenesis-related protein 4B	62.00	2.68E-08	34%	46%	181--302	19--137
					8	COG3087	FtsN, Cell division protein [Cell division and chromosome partitioning].	40.79	2.30E-04	24%	36%	16--158	54--190	8	CAA42821	PR-4b protein	62.00	2.68E-08	34%	46%	181--302	27--145
					9	pfam05518	Totivirus_coat, Totivirus coat protein..	39.58	5.56E-04	24%	29%	29--149	631--750	9	BAD11073	pathogenesis-related protein 4b	61.62	3.49E-08	33%	46%	181--303	70--189
					10	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..	38.90	8.01E-04	19%	28%	49--129	166--246	10	AAF61434	pathogenesis-related protein 4A	61.62	3.49E-08	34%	48%	174--302	21--145
B323R	138490--139629	380	43,579	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	4.12E-08	42%	65%	293--359	11--76	1	NP_048596	4 ankyrin repeats; similar to Drosophila melanogaster ankyrin, corresponds to GenBank Accession Number L35601	438.73	1.44E-121	60%	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	8.03E-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	94.74	5.10E-18	26%	45%	33--355	187--496
														4	XP_797753	PREDICTED: similar to ankyrin repeat domain protein 17 isoform b, partial	93.97	8.71E-18	25%	44%	17--359	'36--1098
														5	XP_782887	PREDICTED: similar to Ankyrin-1 (Erythrocyte ankyrin) (Ankyrin R)	92.43	2.53E-17	23%	46%	51--359	214--543
														6	ZP_00373467	ankyrin repeat domain protein	92.43	2.53E-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	90.12	1.26E-16	27%	43%	38--379	81--467
														9	AAM11327	GH01626p	89.74	1.64E-16	26%	44%	29--353	36--351
														10	AAN12046	CG7462-PC, isoform C	89.74	1.64E-16	26%	44%	29--353	186--501
B324L	141277--139793	495	57,851	8.80	1	pfam01844	HNH, HNH endonuclease..	38.00	1.67E-03	37%	45%	361--417	3--52	1	NP_048435	A87R	526.55	7.57E-148	60%	76%	47--479	22--454
														2	NP_048779	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	130.57	1.21E-28	33%	48%	151--446	46--307
														3	YP_142599	HNH endonuclease	53.53	1.88E-05	43%	60%	149--208	204--263
B331R	141226--142152	309	36,243	6.80	1	smart00220	S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-specific kinase subfamily. .	178.49	8.67E-46	32%	51%	48--305	1--256	1	AAU06280	protein kinase A248R	603.21	3.22E-171	96%	96%	1--309	1--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.	167.69	1.53E-42	32%	53%	47--305	1--256	2	AAU06282	protein kinase A248R	594.73	1.15E-168	94%	95%	1--309	1--309
					3	pfam00069	Pkinase, Protein kinase domain..	156.98	2.85E-39	30%	50%	48--305	1--258	3	AAU06275	protein kinase A248R	559.68	4.09E-158	95%	95%	22--309	1--288
					4	COG0515	SPS1, Serine/threonine protein kinase [General function prediction only / Signal transduction mechanisms / Transcription / DNA replication, recombination, and repair].	101.39	1.32E-22	26%	41%	47--305	1--278	4	AAU06274	protein kinase A248R	424.48	2.05E-117	93%	93%	89--309	22--242
					5	smart00219	TyrKc, Tyrosine kinase, catalytic domain; Phosphotransferases. Tyrosine-specific kinase subfamily. .	69.44	5.15E-13	23%	42%	49--300	2--255	5	NP_048597	PBCV-1 protein kinase	416.77	4.28E-115	71%	81%	25--309	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.	66.75	3.33E-12	23%	42%	46--298	7--261	6	AAA87065	serine/threonine protein kinase	412.15	1.05E-113	70%	81%	25--309	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.	39.09	6.74E-04	24%	42%	125--175	103--153	7	AAU06285	protein kinase A248R	399.82	5.41E-110	89%	90%	90--309	1--221
														8	AAU06286	protein kinase A248R	223.40	6.94E-57	68%	78%	155--309	1--156
														9	AAU06270	protein kinase A248R	207.61	3.94E-52	66%	77%	160--309	1--151
														10	AAU06273	protein kinase A248R	197.21	5.33E-49	66%	77%	168--309	1--143
B334R	142342--143637	432	49,998	10.77	1	COG0675	COG0675, Transposase and inactivated derivatives [DNA replication, recombination, and repair].	84.36	1.73E-17	23%	39%	37--423	1--356	1	AAU06281	putative transposase	827.40	0.00E+00	99%	99%	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.	73.73	2.92E-14	43%	56%	345--412	1--69	2	NP_048981	similar to Synechocystis transposase, corresponds to GenBank Accession Number D90909	720.31	0.00E+00	85%	92%	24--432	26--433
					3	pfam01385	Transposase_2, Probable transposase. This family includes IS891, IS1136 and IS1341. .	48.76	9.75E-07	23%	42%	42--314	1--261	3	YP_142458	putative transposase	110.92	8.25E-23	27%	43%	8--412	95--535
														4	AAS54227	AGL264Wp	104.38	7.72E-21	26%	43%	41--414	71--453
														5	ABA24789	Transposase, IS891/IS1136/IS1341	98.60	4.24E-19	26%	42%	6--414	105--533
														6	BAB78230	transposase	98.60	4.24E-19	26%	42%	6--414	61--489
														7	YP_238637	ORF021	85.50	3.71E-15	26%	44%	40--409	3--364
														8	ZP_00158267	COG0675: Transposase and inactivated derivatives	84.73	6.33E-15	28%	43%	114--414	43--359
														9	BAE47830	putative IS transposase (OrfB)	80.49	1.19E-13	24%	45%	42--409	6--384
														10	ZP_00766186	Transposase, IS605 OrfB	77.03	1.32E-12	28%	43%	36--421	2--367
B336R	143748--144029	94	10,518	8.12		No Hit Found								1	AAQ16140	potassium channel protein	194.51	7.50E-49	100%	100%	1--94	1--94
														2	AAQ16138	potassium channel protein	187.58	9.17E-47	95%	97%	1--94	1--94
														3	AAQ16141	potassium channel protein	185.27	4.55E-46	94%	96%	1--94	1--94
														4	NP_048599	PBCV-1 K+ ion channel protein	181.03	8.58E-45	90%	95%	1--94	1--94
														5	AAQ16135	potassium channel protein	178.72	4.26E-44	89%	94%	1--94	1--94
														6	AAQ16137	potassium channel protein	178.33	5.56E-44	88%	94%	1--94	1--94
														7	AAQ16142	potassium channel protein	177.95	7.26E-44	89%	93%	1--94	1--94
														8	ABA40764	potassium ion channel	82.80	3.18E-15	46%	64%	16--92	6--83
														9	NP_077708	EsV-1-223	55.07	7.10E-07	38%	59%	17--92	45--115
														10	BAD33183	putative outward-rectifying potassium channel KCO1	51.22	1.03E-05	37%	52%	14--91	159--231

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
B339L	144764--144036	243	27,092	9.42		No Hit Found								1	NP_048629	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	410.61	2.02E-113	78%	89%	1--243	1--252
														2	NP_048427	A79R	393.28	3.34E-108	83%	90%	4--227	1--226
														3	NP_048807	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	367.08	2.56E-100	71%	83%	4--243	1--249
														4	NP_049005	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	358.61	9.11E-98	68%	81%	1--242	4--255
														5	AAU06304	hypothetical protein A275R	335.50	8.26E-91	95%	98%	73--243	1--171
														6	AAU06301	hypothetical protein A275R	334.34	1.84E-90	94%	98%	73--243	1--171
														7	NP_048525	A177R	289.66	5.20E-77	57%	72%	1--241	1--243
														8	AAU06302	hypothetical protein A275R	275.40	1.01E-72	100%	100%	109--243	1--135
														9	AAU06303	hypothetical protein A275R	116.70	6.03E-25	94%	96%	185--243	1--59
														10	AAU06299	hypothetical protein A275R	114.01	3.91E-24	93%	94%	185--243	5--63
B341R	145096--145323	76	8,923	8.38		No Hit Found								1	ZP_00886073	conserved hypothetical protein	53.53	2.12E-06	40%	80%	17--68	35--86
B342R	145380--145877	166	19,571	10.15		No Hit Found								1	NP_048604	A253R	218.01	7.97E-56	70%	74%	16--165	1--150
														2	AAU06294	hypothetical protein A253R	214.93	6.75E-55	69%	76%	16--165	1--150
														3	AAU06289	hypothetical protein A253R	214.16	1.15E-54	69%	76%	16--165	1--150
														4	AAU06293	hypothetical protein A253R	119.78	2.95E-26	61%	65%	16--115	1--100
														5	AAU06291	hypothetical protein A253R	117.09	1.91E-25	58%	66%	16--113	1--98
														6	AAU06292	hypothetical protein A253R	101.68	8.33E-21	58%	63%	16--102	1--87
														7	AAU06290	hypothetical protein A253R	99.37	4.13E-20	56%	63%	16--101	1--86
B344R	145914--146522	203	24,959	9.40		No Hit Found								1	NP_048615	A261R	215.31	8.64E-55	92%	95%	35--145	23--135
B346L	147400--146654	249	28,614	10.48	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-strand break in the DNA near the insertion site of that element to facilitate homing at that site. Class I homing endonucleases are sorted into four families based on the presence of these motifs in their respective N-termini: LAGLIDADG, His-Cys box, HNH, and GIY-YIG. This CD contains several but not all members of the GIY-YIG family. The C-terminus of GIY-YIG is a DNA-binding domain which is separated from the N-terminus by a long, flexible linker. The DNA-binding domain consists of a minor-groove binding alpha-helix, and a helix-turn-helix. Some also contain a zinc finger (i.e. I-TevI) which is not required for DNA binding or catalysis, but is a component of the linker and directs the catalytic domain to cleave the homing site at a fixed distance from the intron insertion site	60.01	4.06E-10	45%	57%	103--237	2--113	1	NP_048641	PBCV-1 33kd peptide	261.92	1.22E-68	51%	65%	1--238	1--248
					2	smart00497	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished). .	49.75	5.43E-07	40%	54%	187--239	1--53	2	NP_048671	A315L	178.72	1.36E-43	41%	60%	17--237	13--240
					3	pfam07453	NUMOD1, NUMOD1 domain..	41.57	1.21E-04	45%	55%	187--220	1--34	3	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	168.32	1.84E-40	40%	60%	15--240	11--226
					4	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	41.60	1.46E-04	33%	49%	5--94	2--83	4	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	124.79	2.33E-27	42%	61%	17--175	15--171
					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	38.60	9.83E-04	28%	45%	4--91	1--89	5	NP_899393	SegD	89.35	1.08E-16	31%	51%	5--180	2--163
														6	CAA38804	GIY COII i1 grp IB protein	62.77	1.09E-08	27%	42%	3--234	70--303
														7	AAK09365	intron encoded BmoI	60.85	4.13E-08	29%	43%	8--180	6--201
														8	AAC49244	ORF301	59.31	1.20E-07	24%	41%	3--228	72--282
														9	YP_293795	putative endonuclease	58.15	2.68E-07	32%	52%	15--106	12--105
														10	AAU16837	GIY-YIG catalytic domain containing protein; possible intron encoded endonuclease	57.77	3.50E-07	31%	50%	5--138	3--137
B348L	148125--147481	215	24,358	8.82		No Hit Found								1	NP_048616	A262L	186.42	4.87E-46	83%	91%	100--190	1--91
														2	NP_048617	A263L	83.96	3.40E-15	53%	56%	1--81	43--123
B349R	148156--148356	67	7,721	9.56		No Hit Found								1	NP_048618	a264R	102.45	3.94E-21	68%	86%	1--61	1--61
B350L	149086--148340	249	27,962	9.68	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	45.99	5.87E-06	27%	42%	5--162	2--148	1	NP_048619	A265L	262.31	9.34E-69	72%	81%	80--249	1--170
B354L	149909--149118	264	30,217	6.68	1	COG2267	PldB, Lysophospholipase [Lipid metabolism].	88.94	6.95E-19	21%	40%	7--264	24--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	4.87E-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	9.20E-05	20%	35%	6--155	68--221	3	NP_568327	catalytic/ hydrolase	68.55	2.20E-10	25%	44%	18--262	82--334
					4	COG3571	COG3571, Predicted hydrolase of the alpha/beta-hydrolase fold [General function prediction only].	39.61	5.76E-04	34%	45%	8--108	5--106	4	ABB46702	lipase, putative	65.86	1.42E-09	24%	43%	8--262	93--356
					5	COG0596	MhpC, Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily) [General function prediction only].	38.14	1.51E-03	21%	35%	1--142	5--139	5	AAM60954	lysophospholipase isolog, putative	63.54	7.06E-09	24%	42%	14--260	117--371
														6	NP_177867	catalytic/ hydrolase	63.54	7.06E-09	24%	42%	14--260	117--371
														7	AAB89497	lysophospholipase	63.54	7.06E-09	24%	43%	18--264	23--266
														8	CAC01853	lipase-like protein	62.39	1.57E-08	25%	44%	37--262	90--323
														9	ZP_00486825	COG2267: Lysophospholipase	62.39	1.57E-08	23%	40%	18--264	26--276
														10	NP_191845	catalytic/ hydrolase	61.23	3.51E-08	25%	41%	10--243	26--268
B357L	150365--149949	139	15,824	10.68	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	49.96	4.49E-07	47%	57%	87--139	1--52	1	NP_048627	A273L	237.27	1.01E-61	86%	92%	4--139	3--138
					2	COG3544	COG3544, Uncharacterized protein conserved in bacteria [Function unknown].	42.35	8.51E-05	27%	43%	46--139	90--183	2	YP_142507	unknown	105.15	5.98E-22	42%	61%	9--139	7--141
														3	EAN05440	Protein of unknown function DUF305	54.30	1.21E-06	30%	49%	10--139	22--159
														4	EAN09897	Protein of unknown function DUF305	52.37	4.60E-06	27%	53%	9--139	8--142
														5	AAZ98258	hypothetical protein Tbd_2305	50.45	1.75E-05	31%	58%	43--139	82--175
														6	ZP_00333569	COG3544: Uncharacterized protein conserved in bacteria	50.45	1.75E-05	31%	58%	43--139	57--150
														7	EAN06281	Protein of unknown function DUF305	48.52	6.65E-05	34%	53%	43--139	57--150
														8	ZP_00577284	Protein of unknown function DUF305	48.14	8.68E-05	28%	48%	15--139	12--143
B358R	150699--151373	225	27,129	4.64		No Hit Found								1	NP_048628	encodes Asp/Lys rich sequence	219.55	5.76E-56	57%	67%	28--222	61--260
B359R	151439--152245	269	31,033	8.41	1	smart00650	rADc, Ribosomal RNA adenine dimethylases; .	36.30	5.30E-03	18%	38%	38--151	16--152	1	ZP_00579458	hypothetical protein SalaDRAFT_0836	166.39	8.02E-40	35%	58%	10--265	11--255
					2	COG2263	COG2263, Predicted RNA methylase [Translation, ribosomal structure and biogenesis].	35.60	7.79E-03	18%	33%	12--140	22--181	2	AAC03123	DNA adenine methyltransferase	108.23	2.59E-22	34%	60%	12--169	2--161
														3	NP_223729	putative TYPE II DNA MODIFICATION ENZYME (METHYLTRANSFERASE)	91.66	2.51E-17	32%	54%	8--168	24--190
														4	AAT27581	putative type II DNA methylase protein	91.28	3.28E-17	32%	57%	7--165	2--176
														5	ZP_00372070	putative type II DNA modification enzyme (methyltransferase)	73.17	9.24E-12	31%	52%	12--171	66--233
B361R	152245--153894	550	62,901	10.60	1	COG3889	COG3889, Predicted solute binding protein [General function prediction only].	38.87	8.31E-04	23%	40%	400--478	779--854	1	AAV84098	CviPII top-strand DNA nicking endonuclease	147.13	1.43E-33	36%	52%	1--293	5--285
					2	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	37.82	1.82E-03	25%	31%	415--486	59--130	2	AAC55064	restriction endonuclease	78.57	6.23E-13	30%	44%	22--268	41--257

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
B365L	155794--153980	605	68,918	11.01	1	pfam00069	Pkinase, Protein kinase domain..	57.22	2.38E-09	32%	54%	58--254	2--151	1	NP_048632	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662	1120.15	0.00E+00	91%	93%	1--602	1--607
					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	55.98	6.67E-09	28%	52%	58--254	3--152	2	NP_048636	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055	641.34	0.00E+00	57%	67%	1--602	15--566
					3	smart00220	S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-specific kinase subfamily. .	54.45	2.02E-08	28%	51%	58--254	2--151	3	NP_048970	RPQT-like (9x)	330.49	1.02E-88	48%	66%	10--341	6--337
					4	COG0661	AarF, Predicted unusual protein kinase [General function prediction only].	46.09	6.49E-06	32%	54%	221--278	285--341	4	NP_048441	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563	73.94	1.73E-11	44%	61%	509--602	1--84
					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	1.24E-04	39%	59%	219--287	139--195	5	NP_049032	similar to Chlorella virus PBCV-1 ORF A282L, corresponds to GenBank Accession Number U42580	68.94	5.55E-10	84%	94%	565--602	1--38
					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	2.13E-04	38%	52%	192--255	174--240	6	XP_757661	protein kinase Fuz7	52.76	4.12E-05	24%	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].	40.53	2.56E-04	15%	31%	106--481	25--370	7	Q99078	Dual specificity protein kinase FUZ7	52.76	4.12E-05	24%	44%	69--270	80--275
B368L	157355--155853	501	55,854	10.90	1	pfam00069	Pkinase, Protein kinase domain..	64.92	1.17E-11	29%	47%	85--283	6--183	1	NP_048636	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055	888.64	0.00E+00	89%	91%	1--499	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	62.92	5.77E-11	28%	47%	85--283	7--183	2	NP_048632	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662	610.53	4.05E-173	55%	63%	1--501	1--610
					3	smart00220	S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-specific kinase subfamily. .	61.77	1.23E-10	32%	49%	85--257	6--150	3	NP_048970	RPQT-like (9x)	293.51	1.10E-77	41%	59%	10--388	6--386
					4	smart00219	TyrKc, Tyrosine kinase, catalytic domain; Phosphotransferases. Tyrosine-specific kinase subfamily. .	44.40	1.98E-05	26%	47%	58--266	2--161	4	NP_048441	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563	84.73	7.73E-15	47%	66%	406--499	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	42.87	6.30E-05	28%	46%	58--255	10--161	5	NP_049032	similar to Chlorella virus PBCV-1 ORF A282L, corresponds to GenBank Accession Number U42580	70.86	1.16E-10	84%	94%	461--499	1--39
B370L	158428--157382	349	40,460	10.29		No Hit Found								1	NP_048711	A354R	284.26	4.01E-75	58%	74%	110--348	1--232
														2	NP_048779	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	74.33	6.31E-12	26%	47%	19--314	18--318
														3	NP_695069	putative endodeoxyribonuclease	68.17	4.52E-10	30%	46%	17--166	9--177
														4	AAM00817	HNH endonuclease I-TwoI	64.31	6.53E-09	36%	57%	16--103	11--97
														5	AAC41637	ORF168; putative	62.39	2.48E-08	29%	46%	18--175	13--166
														6	AAL98037	hypothetical phage protein	62.00	3.24E-08	32%	51%	19--152	11--151
														7	AAL73476	endonuclease	59.31	2.10E-07	31%	45%	16--168	22--165
														8	YP_189077	HNH endonuclease family protein	58.15	4.68E-07	38%	56%	19--98	30--112
														9	AAL73479	endonuclease	53.91	8.82E-06	21%	43%	16--178	15--184
														10	NP_047162	putative HNH homing endonuclease	52.37	2.57E-05	31%	53%	14--105	11--100
B371L	159338--158502	279	30,625	8.60	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	211.61	9.54E-56	34%	48%	2--276	1--297	1	NP_048638	PBCV-1 amidase	491.12	1.52E-137	86%	92%	1--279	1--279
					2	COG3049	COG3049, Penicillin V acylase and related amidases [Cell envelope biogenesis, outer membrane].	176.67	2.60E-45	29%	43%	1--279	22--323	2	ABB11030	Penicillin amidase	145.59	1.57E-33	32%	48%	1--277	1--313
					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	159.40	4.84E-40	33%	48%	2--278	1--300	3	AAU25651	Choloylglycine hydrolase	141.35	2.96E-32	31%	50%	1--279	1--301
					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..	122.61	5.83E-29	32%	44%	2--269	1--270	4	AAL51724	CHOLOYLGLYCINE HYDROLASE	139.04	1.47E-31	32%	47%	2--277	37--337
					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)..	77.74	1.91E-15	27%	45%	2--237	1--242	5	AAN30379	choloylglycine hydrolase family protein	139.04	1.47E-31	32%	47%	2--277	3--303
														6	YP_414856	Choloylglycine hydrolase	139.04	1.47E-31	32%	47%	2--277	3--303
														7	AAT60567	choloylglycine hydrolase	134.81	2.77E-30	30%	51%	1--271	1--293
														8	ZP_00238968	choloylglycine hydrolase family protein	134.04	4.72E-30	30%	51%	1--271	1--293
														9	CAG35911	related to choloylglycine hydrolase	132.11	1.79E-29	30%	49%	2--279	32--339
														10	AAT33011	choloylglycine hydrolase family protein	130.57	5.22E-29	29%	50%	1--271	1--293
B377R	159965--160174	70	8,309	11.47		No Hit Found									No Hit Found	No Hit Found						
B378L	162330--160396	645	74,187	10.63	1	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	58.32	1.12E-09	34%	46%	566--638	1--69	1	AAU06281	putative transposase	218.39	6.11E-55	31%	49%	212--639	45--420
					2	COG0675	COG0675, Transposase and inactivated derivatives [DNA replication, recombination, and repair].	57.01	3.05E-09	20%	40%	324--638	62--345	2	NP_048981	similar to Synechocystis transposase, corresponds to GenBank Accession Number D90909	217.62	1.04E-54	31%	48%	212--639	39--414
					3	pfam01385	Transposase_2, Probable transposase. This family includes IS891, IS1136 and IS1341. .	37.20	2.66E-03	23%	42%	219--551	3--277	3	YP_143208	putative transposase	75.49	6.40E-12	23%	39%	175--640	101--524
														4	YP_143124	putative transposase	75.10	8.35E-12	24%	39%	202--640	122--517
														5	YP_142433	putative transposase	73.17	3.17E-11	22%	39%	175--640	94--517
														6	YP_142458	putative transposase	66.24	3.88E-09	25%	38%	431--638	317--535
														7	AAS54227	AGL264Wp	65.47	6.62E-09	23%	40%	216--643	71--456
														8	CAJ31329	insertion sequence IS606 transposase homolog A	61.62	9.56E-08	29%	51%	525--638	304--421
														9	BAD76101	transposase	60.85	1.63E-07	26%	40%	439--640	169--372
														10	ZP_00370996	ISCco1, transposase orfB	60.46	2.13E-07	31%	50%	526--635	277--390
B381L	162325--161642	228	25,945	9.23	1	COG2452	COG2452, Predicted site-specific integrase-resolvase [DNA replication, recombination, and repair].	140.82	1.87E-34	42%	56%	15--208	4--190	1	YP_143125	putative resolvase	113.62	4.52E-24	41%	61%	10--152	2--140

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	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..	66.04	5.53E-12	30%	48%	75--201	2--132	2	YP_142434	putative resolvase	110.15	4.99E-23	39%	60%	10--152	2--140
					3	cd01104	HTH_MlrA, Helix-turn-helix transcription regulator MlrA (merR-like regulator A). The MlrA protein, also known as YehV, has been shown to control cell-cell aggregation by co-regulating the expression of curli and extracellular matrix production in Escherichia coli and Salmonella typhimurium. Its close homolog, CarA from Myxococcus xanthus, is involved in activation of the carotenoid biosynthesis genes by light. These proteins belong to the MERR superfamily of transcription regulators that promote expression of several stress regulon genes by reconfiguring the spacer between the -35 and -10 promoter elements. Their conserved N-terminal domains contain predicted HTH (helix-turn-helix) motives that mediate DNA binding, while the dissimilar C-terminal domains bind specific coactivator molecules.	41.85	1.12E-04	31%	42%	16--98	5--83	3	YP_142457	putative resolvase	105.15	1.61E-21	37%	56%	11--161	3--149
														4	AAK41573	First ORF in transposon ISC1904	96.29	7.46E-19	42%	59%	22--159	11--142
														5	BAD84844	predicted site-specific integrase/resolvase	95.90	9.74E-19	35%	57%	6--160	3--153
														6	AAK43255	First ORF in transposon ISC1904	95.90	9.74E-19	40%	58%	15--159	4--142
														7	AAK41585	First ORF in transposon ISC1904	95.90	9.74E-19	42%	59%	15--154	4--137
														8	ZP_00683312	regulatory protein, MerR:Resolvase, N-terminal	95.13	1.66E-18	38%	57%	11--151	10--150
														9	CAB49329	Resolvase related protein	95.13	1.66E-18	36%	60%	11--156	6--151
														10	AAK42026	First ORF in transposon ISC1904	94.74	2.17E-18	41%	58%	15--159	4--142
B385R	162793--163908	372	41,925	9.30		No Hit Found								1	NP_048640	A286R	575.47	9.55E-163	72%	85%	1--372	7--378
B388L	164705--163902	268	30,153	6.78	1	smart00220	S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-specific kinase subfamily. .	160.00	2.94E-40	31%	51%	17--265	3--256	1	NP_048643	similar to PBCV-1 serine/threonine protein kinase, corresponds to GenBank Accession Number U14660	374.40	1.93E-102	65%	80%	1--265	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.	159.60	4.01E-40	30%	50%	17--265	4--256	2	AAU06282	protein kinase A248R	136.35	8.82E-31	32%	56%	21--265	54--305
					3	pfam00069	Pkinase, Protein kinase domain..	152.75	4.47E-38	29%	48%	17--265	3--258	3	NP_048631	similar to PBCV-1 serine/threonine protein kinase, corresponds to GenBank Accession Number U14660	135.96	1.15E-30	31%	52%	21--266	25--277
					4	COG0515	SPS1, Serine/threonine protein kinase [General function prediction only / Signal transduction mechanisms / Transcription / DNA replication, recombination, and repair].	89.83	4.33E-19	26%	46%	16--266	3--279	4	AAU06280	protein kinase A248R	133.27	7.47E-30	31%	56%	21--265	54--305
					5	smart00219	TyrKc, Tyrosine kinase, catalytic domain; Phosphotransferases. Tyrosine-specific kinase subfamily. .	76.75	3.73E-15	23%	44%	17--261	3--253	5	AAU06275	protein kinase A248R	133.27	7.47E-30	31%	56%	21--265	33--284
					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.	72.91	5.78E-14	22%	45%	13--259	7--262	6	AAU06285	protein kinase A248R	124.41	3.47E-27	31%	56%	58--265	2--217
														7	AAU06274	protein kinase A248R	120.55	5.01E-26	31%	55%	56--265	22--238
														8	NP_048597	PBCV-1 protein kinase	117.09	5.54E-25	26%	52%	21--265	51--304
														9	XP_629689	putative CAM kinase (CAMK)	115.16	2.11E-24	31%	52%	19--267	113--363
														10	AAA87065	serine/threonine protein kinase	114.78	2.75E-24	27%	53%	21--265	47--300
B389R	164757--165629	291	33,848	9.10	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-strand break in the DNA near the insertion site of that element to facilitate homing at that site. Class I homing endonucleases are sorted into four families based on the presence of these motifs in their respective N-termini: LAGLIDADG, His-Cys box, HNH, and GIY-YIG. This CD contains several but not all members of the GIY-YIG family. The C-terminus of GIY-YIG is a DNA-binding domain which is separated from the N-terminus by a long, flexible linker. The DNA-binding domain consists of a minor-groove binding alpha-helix, and a helix-turn-helix. Some also contain a zinc finger (i.e. I-TevI) which is not required for DNA binding or catalysis, but is a component of the linker and directs the catalytic domain to cleave the homing site at a fixed distance from the intron insertion site.	71.19	1.63E-13	44%	59%	150--275	1--113	1	NP_048641	PBCV-1 33kd peptide	290.04	5.49E-77	55%	67%	24--278	1--250
					2	smart00497	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished). .	49.36	6.36E-07	37%	50%	225--277	1--53	2	NP_048671	A315L	212.62	1.11E-53	44%	66%	31--276	3--241
					3	pfam07453	NUMOD1, NUMOD1 domain..	44.65	1.67E-05	52%	61%	225--258	1--34	3	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	160.61	5.03E-38	39%	57%	31--277	3--225
					4	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	43.91	2.93E-05	35%	52%	27--117	1--83	4	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	103.61	7.29E-21	35%	52%	41--233	16--181
					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.	40.92	1.95E-04	28%	50%	27--114	1--89	5	NP_899393	SegD	83.19	1.02E-14	31%	49%	28--215	2--174
														6	CAA38804	GIY COII i1 grp IB protein	76.64	9.55E-13	29%	45%	8--222	56--237
														7	AAC49248	ORF211	68.94	1.99E-10	37%	51%	124--259	79--192
														8	AAK09365	intron encoded BmoI	68.55	2.60E-10	32%	50%	120--273	120--266
														9	AAC49244	ORF301	65.08	2.87E-09	49%	59%	144--214	176--236
														10	CAA73995	unnamed protein product	61.23	4.15E-08	28%	42%	42--246	127--320
B393L	166590--165625	322	36,662	9.75	1	pfam01374	Glyco_hydro_46, Glycosyl hydrolase family 46. This family are chitosanase enzymes..	254.20	1.22E-68	40%	54%	94--305	1--216	1	NP_048646	PBCV-1 chitosanase	539.27	6.13E-152	78%	86%	3--322	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..	213.28	2.95E-56	30%	46%	88--305	5--221	2	BAA20342	vChta-1	530.02	3.72E-149	76%	86%	3--322	1--328
														3	ABC17783	secreted chitosanase precursor	84.73	4.13E-15	29%	46%	88--305	55--275
														4	P48846	Chitosanase precursor	80.11	1.02E-13	28%	46%	91--295	53--255
														5	CAB14630	chitosanase	78.18	3.87E-13	29%	43%	96--321	49--276
														6	BAB19276	chitosanase	77.41	6.60E-13	28%	43%	96--321	50--277
														7	BAC06189	chitosanase	76.26	1.47E-12	27%	43%	91--309	41--255
														8	BAA94840	chitosanase	75.87	1.92E-12	27%	46%	94--295	63--263
														9	ZP_00908039	chitosanase	74.71	4.28E-12	30%	46%	94--305	38--247
														10	CAB61194	secreted chitosanase	73.17	1.24E-11	34%	54%	44--167	13--131
B395L	167521--166562	320	36,183	6.90	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

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	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	7.87E-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	2.62E-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	6.81E-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	4.79E-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
B397L	168970--167552	473	56,139	4.52		No Hit Found								1	NP_048725	Asp-rich	133.65	1.34E-29	25%	47%	63--432	1--371
B399R	169101--169880	260	30,966	6.72	1	pfam02086	MethyltransfD12, D12 class N6 adenine-specific DNA methyltransferase..	149.72	3.76E-37	31%	47%	7--237	1--251	1	AAC03126	DNA adenine methyltransferase	526.55	2.89E-148	96%	96%	1--260	1--261
					2	COG0338	Dam, Site-specific DNA methylase [DNA replication, recombination, and repair].	143.12	3.98E-35	30%	51%	1--260	2--265	2	Q01511	Modification methylase CviBI (Adenine-specific methyltransferase CviBI) (M.CviBI)	488.42	8.73E-137	85%	95%	1--260	1--260
														3	AAK08495	N.BstNBI methyltransferase	189.89	6.42E-47	38%	55%	1--260	1--283
														4	CAD92748	putative adenine methyltransferase	178.33	1.93E-43	38%	56%	1--260	1--283
														5	AAK27214	adenine methyltransferase M.PleI	171.40	2.36E-41	35%	53%	1--260	1--283
														6	AAK81290	Site-specific DNA methylase dam	167.55	3.41E-40	35%	55%	4--260	1--280
														7	ZP_00575937	N6 adenine-specific DNA methyltransferase, D12 class	135.96	1.10E-30	32%	51%	1--260	7--270
														8	BAD74251	type IIs modification methyltransferase	132.88	9.32E-30	34%	50%	2--253	16--276
														9	ABA22276	DNA adenine methylase	132.88	9.32E-30	33%	52%	2--225	12--243
														10	ZP_00886307	DNA adenine methylase	132.88	9.32E-30	34%	49%	3--260	14--290
B401L	170423--169944	160	19,027	10.12		No Hit Found									No Hit Found	No Hit Found						
B402L	171327--170515	271	32,144	6.33		No Hit Found									No Hit Found	No Hit Found						
B403R	171458--172435	326	37,811	9.11		No Hit Found								1	NP_048684	A328L	102.45	1.95E-20	25%	40%	7--326	4--350
B404R	172455--172850	132	15,151	12.05		No Hit Found								1	NP_048650	A296R	137.12	1.46E-31	77%	87%	47--132	72--156
B408R	172928--174235	436	48,636	8.47		No Hit Found								1	NP_048920	similar to Chlorella virus PBCV-1 ORF A154L, corresponds to GenBank Accession Number U42580	543.12	6.59E-153	69%	83%	78--436	4--350
														2	NP_048502	A154L	513.07	7.31E-144	67%	80%	78--436	3--347
														3	NP_048477	similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903	510.76	3.63E-143	65%	81%	77--436	4--356
														4	NP_077492	EsV-1-7	72.40	3.30E-11	28%	48%	41--261	4--222
														5	CAA36506	balbiani ring 3 (BR3)	65.86	3.09E-09	26%	35%	19--314	55--1413
														6	CAG01316	unnamed protein product	60.08	1.69E-07	27%	38%	27--275	324--534
														7	XP_698583	PREDICTED: similar to zinc finger protein	58.92	3.77E-07	21%	38%	17--262	38--283
														8	XP_707899	PREDICTED: similar to zinc finger protein 91 (HPF7, HTF10) isoform 8	58.92	3.77E-07	23%	38%	5--262	368--681
														9	XP_707898	PREDICTED: similar to zinc finger protein 91 (HPF7, HTF10) isoform 7	58.92	3.77E-07	23%	38%	5--262	272--585
														10	XP_685373	PREDICTED: similar to zinc finger protein 91 (HPF7, HTF10) isoform 1	58.92	3.77E-07	23%	38%	5--262	284--597
B411L	175347--174238	370	42,138	9.03	1	pfam00145	DNA_methylase, C-5 cytosine-specific DNA methylase..	139.29	5.55E-34	27%	41%	2--234	1--225	1	NP_048873	M.CviAII cytosine DNA methyltransferase	310.84	4.39E-83	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.	135.82	5.68E-33	31%	47%	2--173	1--178	2	AAC64006	cytosine methyltransferase	301.98	2.04E-80	45%	59%	2--367	3--356
					3	COG0270	Dcm, Site-specific DNA methylase [DNA replication, recombination, and repair].	109.01	6.96E-25	27%	44%	1--167	3--176	3	AAV84097	CviPII m5C DNA methyltransferase	296.59	8.56E-79	42%	62%	2--367	16--357
														4	AAC55063	cytosine methyltransferase	296.59	8.56E-79	42%	60%	4--367	5--360
														5	NP_049039	nonfunctional M.CviAV cytosine DNA methyltransferase	294.28	4.25E-78	42%	59%	4--368	5--361
														6	NP_048886	M.CviAIV cytosine DNA methyltransferase	280.41	6.35E-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
B416R	175415--176530	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	6.38E-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	8.90E-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	5.79E-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	6.78E-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	8.87E-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
B418R	176591--177742	384	44,217	8.41	1	COG2890	HemK, Methylase of polypeptide chain release factors [Translation, ribosomal structure and biogenesis].	56.16	5.27E-09	24%	44%	29--150	98--238	1	CAB92310	DNA methyltrasferase	273.09	1.06E-71	41%	59%	18--383	44--415
					2	COG1041	COG1041, Predicted DNA modification methylase [DNA replication, recombination, and repair].	54.20	2.43E-08	29%	45%	40--155	196--315	2	AAF74028	M.Hpy188I	270.78	5.28E-71	41%	59%	18--383	44--415
					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..	52.19	8.62E-08	29%	40%	39--152	26--142	3	ZP_00783145	reticulocyte binding protein	70.09	1.37E-10	34%	53%	36--163	694--827
					4	COG2813	RsmC, 16S RNA G1207 methylase RsmC [Translation, ribosomal structure and biogenesis].	51.42	1.39E-07	34%	44%	44--144	161--260	4	ZP_00780227	SNF2 family protein	70.09	1.37E-10	34%	53%	36--163	694--827
					5	COG0286	HsdM, Type I restriction-modification system methyltransferase subunit [Defense mechanisms].	46.56	4.41E-06	20%	36%	19--249	167--409	5	AAN00153	SNF2 family protein	70.09	1.37E-10	34%	53%	36--163	694--827
					6	COG4123	COG4123, Predicted O-methyltransferase [General function prediction only].	46.01	6.33E-06	28%	41%	44--152	47--172	6	ZP_00874800	SNF2-related	69.71	1.79E-10	34%	55%	36--163	694--827
					7	COG2519	GCD14, tRNA(1-methyladenosine) methyltransferase and related methyltransferases [Translation, ribosomal structure and biogenesis].	42.15	1.04E-04	23%	35%	40--153	93--198	7	ZP_00874780	SNF2-related:Helicase, C-terminal	69.32	2.33E-10	34%	53%	36--163	694--827
					8	COG2226	UbiE, Methylase involved in ubiquinone/menaquinone biosynthesis [Coenzyme metabolism].	38.33	1.22E-03	27%	47%	40--107	50--123	8	AAS71294	site-specific modification DNA-methyltransferase	66.24	1.98E-09	31%	52%	22--156	17--179
					9	COG4122	COG4122, Predicted O-methyltransferase [General function prediction only].	36.43	4.49E-03	26%	48%	65--148	82--164	9	ZP_00413234	Helicase, C-terminal	65.86	2.58E-09	30%	52%	42--182	263--406
					10	COG2230	Cfa, Cyclopropane fatty acid synthase and related methyltransferases [Cell envelope biogenesis, outer membrane].	36.40	5.31E-03	22%	42%	40--148	71--174	10	CAD47011	Unknown	64.31	7.51E-09	34%	52%	42--163	494--618
B419L	178256--177708	183	21,611	10.24		No Hit Found								1	NP_048651	A297L	299.67	2.72E-80	81%	91%	1--170	1--170
														2	BAE64611	unnamed protein product	54.30	1.99E-06	30%	48%	3--116	6--119

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
B421L	178974--178294	227	25,867	6.73		No Hit Found								1	NP_048652	A298L	368.62	7.71E-101	77%	89%	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
														6	CAD72673	hypothetical protein	49.68	7.93E-05	23%	44%	18--208	53--263
B424L	180056--179004	351	41,186	10.29		No Hit Found								1	NP_048711	A354R	265.77	1.49E-69	53%	70%	114--350	4--235
														2	NP_048779	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	64.31	6.59E-09	26%	44%	34--320	28--319
														3	NP_048435	A87R	62.77	1.92E-08	23%	40%	19--349	101--449
														4	AAL73479	endonuclease	60.85	7.28E-08	29%	43%	17--158	15--163
														5	AAL73476	endonuclease	58.15	4.72E-07	31%	47%	6--141	9--144
														6	YP_189077	HNH endonuclease family protein	52.37	2.59E-05	43%	53%	27--99	47--112
														7	AAM00817	HNH endonuclease I-TwoI	51.22	5.77E-05	32%	51%	17--99	11--92
														8	AAO93095	I-BasI	50.83	7.54E-05	38%	49%	18--99	14--93
B426L	180802--180173	210	24,079	5.19		No Hit Found								1	NP_048655	A301L	208.38	1.15E-52	52%	65%	1--201	1--239
B429R	180858--181100	81	9,853	7.54		No Hit Found								1	NP_048658	A304R	86.66	2.22E-16	60%	68%	4--79	1--76
B430L	181638--181096	181	20,418	10.79	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..	129.59	4.27E-31	33%	51%	24--160	1--139	1	NP_048659	similar to human protein Tyr-phosphatase, corresponds to GenBank Accession Number U27193	327.41	1.18E-88	92%	97%	1--169	1--169
					2	smart00195	DSPc, Dual specificity phosphatase, catalytic domain; .	124.62	1.32E-29	36%	56%	24--162	1--139	2	AAB88308	Lateral-signal-induced phosphatase protein 1	77.80	1.63E-13	33%	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..	123.83	2.37E-29	36%	54%	24--162	1--139	3	NP_998405	dual specificity phosphatase 16	73.56	3.07E-12	33%	50%	25--163	159--298
					4	COG2453	CDC14, Predicted protein-tyrosine phosphatase [Signal transduction mechanisms].	52.01	1.02E-07	28%	47%	65--144	71--147	4	XP_543810	PREDICTED: similar to Dual specificity protein phosphatase 16 (Mitogen-activated protein kinase phosphatase 7) (MAP kinase phosphatase 7) (MKP-7)	72.40	6.84E-12	33%	51%	25--163	159--298
					5	smart00404	PTPc_motif, Protein tyrosine phosphatase, catalytic domain motif; .	38.86	8.50E-04	22%	43%	73--138	11--74	5	AAH42101	DUSP16 protein	71.25	1.52E-11	33%	51%	25--163	159--298
														6	AAH31643	Unknown (protein for IMAGE:5176724)	71.25	1.52E-11	33%	51%	25--163	110--249
														7	AAI09236	Dual specificity phosphatase 16	71.25	1.52E-11	33%	51%	25--163	159--298
														8	AAI09235	Dual specificity phosphatase 16	71.25	1.52E-11	33%	51%	25--163	159--298
														9	XP_520751	PREDICTED: similar to KIAA1700 protein	71.25	1.52E-11	33%	51%	25--163	268--407
														10	BAB21791	KIAA1700 protein	71.25	1.52E-11	33%	51%	25--163	184--323
B431L	181923--181663	87	10,598	4.08		No Hit Found								1	NP_048660	A306L	124.79	7.48E-28	86%	93%	27--87	26--86
B432L	182286--181978	103	11,974	4.29		No Hit Found								1	NP_048663	A308L	90.12	2.00E-17	72%	92%	49--103	20--74
														2	NP_048664	a309L	70.86	1.26E-11	71%	90%	11--52	15--56
B433L	183229--182411	273	30,939	10.44	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-strand break in the DNA near the insertion site of that element to facilitate homing at that site. Class I homing endonucleases are sorted into four families based on the presence of these motifs in their respective N-termini: LAGLIDADG, His-Cys box, HNH, and GIY-YIG. This CD contains several but not all members of the GIY-YIG family. The C-terminus of GIY-YIG is a DNA-binding domain which is separated from the N-terminus by a long, flexible linker. The DNA-binding domain consists of a minor-groove binding alpha-helix, and a helix-turn-helix. Some also contain a zinc finger (i.e. I-TevI) which is not required for DNA binding or catalysis, but is a component of the linker and directs the catalytic domain to cleave the homing site at a fixed distance from the intron insertion site	54.24	2.10E-08	49%	67%	98--170	1--73	1	NP_048671	A315L	268.47	1.54E-70	52%	66%	1--269	1--240
					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	40.15	4.01E-04	27%	51%	1--86	1--89	2	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	211.46	2.24E-53	41%	57%	1--272	1--226
					3	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	39.29	6.62E-04	30%	49%	1--89	1--83	3	NP_048641	PBCV-1 33kd peptide	192.97	8.23E-48	40%	58%	13--270	17--248
					4	smart00497	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished). .	37.03	3.51E-03	42%	52%	215--271	1--53	4	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	159.84	7.72E-38	47%	68%	1--160	1--170
														5	NP_899393	SegD	78.57	2.26E-13	32%	50%	1--175	1--195
														6	YP_293795	putative endonuclease	77.03	6.58E-13	35%	59%	2--104	3--108
														7	AAC49248	ORF211	73.17	9.50E-12	32%	49%	119--268	55--208
														8	AAC49244	ORF301	71.63	2.76E-11	30%	49%	2--152	75--236
														9	AAK09365	intron encoded BmoI	70.48	6.16E-11	26%	43%	4--258	5--259
														10	CAA38804	GIY COII i1 grp IB protein	67.01	6.81E-10	30%	48%	15--202	86--248
B434L	183809--183300	170	18,455	8.41		No Hit Found								1	NP_048665	A310L	307.76	8.10E-83	85%	94%	1--170	1--170
B435L	184818--184075	248	27,585	5.08		No Hit Found								1	NP_048667	PBCV-1 33kDa translation peptide	421.78	9.14E-117	90%	92%	1--237	1--237
														2	NP_077561	EsV-1-76	73.94	4.68E-12	27%	42%	12--217	20--236
														3	AAR26966	FirrV-1-I1	63.16	8.25E-09	27%	42%	35--212	36--222
														4	AAR26885	FirrV-1-B10	56.23	1.01E-06	27%	41%	40--212	52--233
B437L	185347--185120	76	8,355	8.92		No Hit Found								1	NP_048669	A313L	95.13	6.35E-19	65%	73%	1--67	1--67
B438R	185480--185683	68	8,318	9.37		No Hit Found									No Hit Found	No Hit Found						
B440L	186653--185838	272	30,870	9.68	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-strand break in the DNA near the insertion site of that element to facilitate homing at that site. Class I homing endonucleases are sorted into four families based on the presence of these motifs in their respective N-termini: LAGLIDADG, His-Cys box, HNH, and GIY-YIG. This CD contains several but not all members of the GIY-YIG family. The C-terminus of GIY-YIG is a DNA-binding domain which is separated from the N-terminus by a long, flexible linker. The DNA-binding domain consists of a minor-groove binding alpha-helix, and a helix-turn-helix. Some also contain a zinc finger (i.e. I-TevI) which is not required for DNA binding or catalysis, but is a component of the linker and directs the catalytic domain to cleave the homing site at a fixed distance from the intron insertion site	51.15	1.64E-07	45%	58%	98--184	1--77	1	NP_048671	A315L	254.60	2.29E-66	49%	65%	1--268	1--240
					2	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	42.37	7.27E-05	34%	55%	1--89	1--83	2	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	208.76	1.44E-52	44%	57%	1--270	1--225
					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	40.53	2.78E-04	28%	47%	1--86	1--89	3	NP_048641	PBCV-1 33kd peptide	180.64	4.20E-44	41%	58%	11--270	15--249

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	smart00497	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished). .	37.03	2.92E-03	40%	56%	217--270	1--53	4	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	166.39	8.19E-40	49%	67%	1--171	1--171
														5	YP_293795	putative endonuclease	82.03	2.03E-14	39%	60%	2--106	3--110
														6	NP_899393	SegD	72.40	1.61E-11	29%	47%	1--220	1--205
														7	AAC49244	ORF301	63.93	5.72E-09	28%	50%	2--142	75--226
														8	CAA73995	unnamed protein product	61.62	2.84E-08	29%	40%	10--243	122--325
														9	CAA38804	GIY COII i1 grp IB protein	57.77	4.10E-07	30%	49%	15--162	86--246
														10	AAU16837	GIY-YIG catalytic domain containing protein; possible intron encoded endonuclease	57.77	4.10E-07	31%	57%	4--134	5--137
B441R	186800--186994	65	7,380	10.89		No Hit Found								1	NP_048437	A89R	124.79	7.47E-28	92%	96%	1--65	1--65
B443R	187137--188426	430	48,141	10.73	1	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	57.85	1.90E-09	29%	38%	7--117	24--133	1	NP_048674	A318R	423.32	7.46E-117	91%	95%	220--430	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.	54.60	1.88E-08	33%	53%	30--94	59--123	2	NP_048672	PAPK (17X); similar to PBCV-1 ORF A41R, corresponds to Genbank Accession Number U17055	110.54	1.07E-22	73%	84%	118--183	126--194
B445L	188739--188443	99	11,787	4.32		No Hit Found									No Hit Found	No Hit Found						
B446R	188488--189723	412	48,721	7.29		No Hit Found								1	NP_048711	A354R	150.21	1.15E-34	37%	53%	157--405	4--234
														2	NP_048779	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	62.77	2.41E-08	25%	44%	84--382	42--327
														3	AAL73456	endonuclease	50.83	9.46E-05	30%	52%	59--169	84--191
B450R	189733--190077	115	12,814	10.38		No Hit Found								1	NP_048676	A320R	130.57	1.37E-29	57%	69%	1--115	25--139
B451R	190094--190447	118	12,805	8.36		No Hit Found								1	NP_048677	A321R	141.74	5.87E-33	60%	75%	6--118	8--119
B453L	190988--190455	178	20,649	5.40		No Hit Found								1	NP_048678	A322L	201.06	1.21E-50	57%	68%	1--178	1--176
B455L	192365--191067	433	48,402	4.69		No Hit Found								1	NP_048680	A324L	550.44	4.09E-155	68%	74%	1--433	1--453
														2	AAR26897	FirrV-1-B22	79.72	2.04E-13	27%	55%	96--267	113--278
														3	NP_077588	EsV-1-103	72.40	3.26E-11	22%	41%	30--289	145--391
														4	YP_142803	unknown	60.85	9.83E-08	28%	47%	145--288	225--361
														5	YP_294142	hypothetical protein EhV 384	55.07	5.39E-06	23%	41%	62--256	125--309
B457L	193000--192404	199	23,227	9.22		No Hit Found								1	NP_048682	A326L	331.64	7.86E-90	74%	90%	1--198	1--208
B458L	194109--193036	358	42,420	9.55		No Hit Found								1	NP_048684	A328L	569.70	4.98E-161	74%	87%	1--358	1--355
B460R	194141--194464	108	12,108	8.39		No Hit Found								1	NP_048685	A329R	142.51	3.45E-33	68%	90%	13--107	1--95
tRNAs																						
Leu	194698--947781	81 bs					anticodon TAA															
Leu	194921--195005	85 bs					anticodon CAA															
Arg	195029--195100	72 bs					anticodon TCT															
Asn	195201--195272	72 bs					anticodon GTT															
Tyr	195295--195379	85 bs					anticodon GTA															
Intron	195331--195343	13 bs					Intron (195331-195343)															
Lys	195382--195456	75 bs					anticodon CTT															
Val	195482--195554	73 bs					anticodon AAC															
B461R	195038--195235	66	7,351	10.82		No Hit Found									No Hit Found	No Hit Found						
B462R	195636--195842	69	8,082	6.45		No Hit Found									No Hit Found	No Hit Found						
B463L	195924--195721	68	7,966	9.99		No Hit Found									No Hit Found	No Hit Found						
B465R	196060--197196	379	41,981	5.26	1	COG0677	WecC, UDP-N-acetyl-D-mannosaminuronate dehydrogenase [Cell envelope biogenesis, outer membrane].	212.01	5.99E-56	30%	48%	3--361	9--399	1	BAE48155	UDP-glucose dehydrogenase	731.87	0.00E+00	94%	97%	1--379	1--379
					2	COG1004	Ugd, Predicted UDP-glucose 6-dehydrogenase [Cell envelope biogenesis, outer membrane].	135.70	5.64E-33	26%	43%	5--351	2--373	2	EAA78333	hypothetical protein FG06548.1	231.49	3.48E-59	38%	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.	84.13	2.24E-17	25%	44%	5--171	2--190	3	EAA72911	hypothetical protein FG03171.1	214.54	4.40E-54	36%	54%	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.	70.60	2.24E-13	28%	42%	182--256	2--76	4	BAE60327	unnamed protein product	207.61	5.38E-52	36%	54%	6--367	61--435
					5	COG2084	MmsB, 3-hydroxyisobutyrate dehydrogenase and related beta-hydroxyacid dehydrogenases [Lipid metabolism].	55.23	1.04E-08	24%	40%	4--237	1--231	5	EAA78448	hypothetical protein FG11489.1	204.14	5.95E-51	38%	55%	3--319	41--366
					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.	39.89	4.04E-04	29%	51%	293--361	2--74	6	EAA70861	hypothetical protein FG04144.1	193.74	8.04E-48	37%	52%	6--339	79--417
					7	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..	38.31	1.37E-03	24%	41%	5--92	2--87	7	AAU22397	UDP-glucose 6-dehydrogenase	180.64	7.04E-44	29%	54%	6--373	11--405
					8	COG0039	Mdh, Malate/lactate dehydrogenases [Energy production and conversion].	37.85	1.65E-03	29%	42%	5--76	2--80	8	BAD63054	UDP-N-acetyl-D-mannosaminuronate dehydrogenase	179.10	2.05E-43	32%	51%	4--369	15--410
					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.	37.36	2.67E-03	31%	51%	5--71	1--73	9	AAM23919	UDP-N-acetyl-D-mannosaminuronate dehydrogenase	177.56	5.96E-43	32%	52%	6--378	25--427
					10	cd00650	LDH_MDH, NAD-dependent 2-hydroxycarboxylate dehydrogenases. Members of this family include such ubiquitous enzymes, like L-lactate dehydrogenases (LDH) and malate dehydrogenases (MDH). 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..	35.60	8.24E-03	28%	45%	5--76	1--77	10	AAK81242	Predicted UDP-glucose 6-dehydrogenase	173.71	8.61E-42	31%	50%	3--376	14--411
B468R	197225--197977	251	27,538	8.33		No Hit Found								1	BAE48156	hypothetical methionyl-tRNA synthetase	466.85	2.54E-130	88%	93%	1--251	1--251
														2	NP_048562	PBCV-1 alginate lyase	127.10	4.77E-28	36%	51%	9--247	87--315
														3	BAA83789	alginate lyase	124.02	4.04E-27	35%	51%	9--247	99--327
														4	BAB19127	vAL-1	121.32	2.62E-26	35%	50%	9--247	115--343
														5	BAC87758	alginate lyase	80.88	3.92E-14	32%	44%	32--250	55--263
														6	BAE45131	alginase	80.88	3.92E-14	32%	44%	32--250	55--263
														7	EAL19065	hypothetical protein CNBH1670	68.94	1.54E-10	29%	46%	55--247	315--491
														8	AAW45360	expressed protein	68.94	1.54E-10	29%	46%	55--247	315--491
														9	EAL19679	hypothetical protein CNBG3070	59.31	1.22E-07	28%	40%	38--246	155--361

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
														10	AAW44587	hypothetical protein CNG01710	59.31	1.22E-07	28%	40%	38--246	145--351
B469L	199194--197986	403	46,109	4.65	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..	81.48	1.22E-16	31%	43%	74--188	5--118	1	BAE48157	hypothetical chitooligosaccharide deacetylase	707.21	0.00E+00	89%	94%	36--403	2--369
														2	EAL33604	GA14716-PA	150.98	6.56E-35	30%	47%	55--350	230--529
														3	EAA12484	ENSANGP00000011077	150.60	8.56E-35	29%	47%	55--374	141--463
														4	AAF53561	CG17905-PA	147.13	9.47E-34	29%	46%	55--364	219--533
														5	CAE68839	Hypothetical protein CBG14800	136.35	1.67E-30	29%	46%	54--350	329--635
														6	XP_624655	PREDICTED: similar to CG31973-PA, isoform A	135.96	2.18E-30	29%	46%	55--347	478--765
														7	EAA06323	ENSANGP00000021951	133.65	1.08E-29	29%	48%	54--344	104--395
														8	AAF51568	CG31973-PB, isoform B	130.18	1.20E-28	26%	46%	55--350	77--2767
														9	AAF51567	CG31973-PA, isoform A	130.18	1.20E-28	26%	46%	55--350	683--973
														10	EAL34164	GA16591-PA	128.26	4.55E-28	26%	46%	55--350	661--951
B472R	199284--200864	527	60,493	9.36	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	2.35E-14	26%	43%	77--435	28--424	1	BAB83509	chitin synthase	985.33	0.00E+00	93%	97%	18--527	7--516
					2	COG1215	COG1215, Glycosyltransferases, probably involved in cell wall biogenesis [Cell envelope biogenesis, outer membrane].	46.47	5.22E-06	21%	36%	22--517	3--430	2	BAE48158	chitin synthase	392.12	2.41E-107	91%	95%	1--210	1--210
														3	BAE48153	chitin synthase	308.53	3.50E-82	36%	56%	4--514	8--507
														4	EAA78335	hypothetical protein FG06550.1	182.96	2.22E-44	28%	47%	4--509	131--649
														5	EAA68628	hypothetical protein FG10619.1	177.95	7.13E-43	27%	46%	22--527	192--702
														6	BAE60326	unnamed protein product	147.90	7.90E-34	26%	42%	29--509	132--580
														7	NP_077569	ESV-1-84	127.10	1.44E-27	24%	44%	2--509	1--484
														8	XP_503779	hypothetical protein	82.80	3.12E-14	24%	40%	80--511	88--1228
														9	EAA72910	hypothetical protein FG03170.1	82.03	5.33E-14	23%	42%	76--507	255--674
														10	XP_502222	hypothetical protein	79.72	2.64E-13	23%	42%	72--509	00--1167
B477R	200923--202251	443	51,081	7.43	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	87.83	1.70E-18	43%	59%	6--100	5--99	1	NP_048686	4 ankyrin repeats; similar to Drosophila melanogaster ankyrin, encoded by GenBank Accession Number L35601	839.34	0.00E+00	94%	95%	1--443	1--432
					2	COG0666	Arp, FOG: Ankyrin repeat [General function prediction only].	50.66	2.58E-07	32%	48%	6--223	38--205	2	XP_395788	PREDICTED: similar to CG7462-PB, isoform B	132.88	2.10E-29	27%	48%	7--317	341--661
					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	2.06E-05	52%	66%	166--195	3--32	3	AAC37208	ankyrin	127.87	6.76E-28	28%	49%	7--317	365--684
					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. .	37.63	1.90E-03	59%	74%	166--193	3--30	4	AAN06551	CG1651-PD, isoform D	127.10	1.15E-27	28%	48%	7--317	365--684
														5	XP_392578	PREDICTED: similar to CG7462-PB, isoform B	127.10	1.15E-27	26%	48%	7--327	403--733
														6	XP_697279	PREDICTED: similar to Ankyrin-1 (Erythrocyte ankyrin), partial	124.41	7.48E-27	30%	47%	10--299	472--738
														7	EAA14062	ENSANGP00000013300	123.25	1.67E-26	28%	50%	7--317	330--650
														8	XP_699321	PREDICTED: similar to ankyrin 1 isoform 4	122.87	2.18E-26	29%	47%	10--299	460--726
														9	XP_689727	PREDICTED: similar to ankyrin 1 isoform 4	122.87	2.18E-26	29%	47%	10--299	442--708
														10	EAL87814	NACHT domain protein, putative	122.48	2.84E-26	30%	48%	3--322	80--1476
B480L	203450--202290	387	43,026	8.21	1	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	3.86E-04	45%	52%	214--265	152--208	1	NP_048689	PLPRNLLL (4X), SPPPSKP (3X)	417.93	2.71E-115	85%	92%	1--213	1--213
					2	smart00494	ChtBD2, Chitin-binding domain type 2; .	39.34	7.21E-04	45%	53%	278--321	9--49	2	NP_048688	a332L	228.41	3.04E-58	96%	98%	280--387	1--109
					3	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	38.68	1.02E-03	22%	29%	215--347	634--769	3	EAA01148	ENSANGP00000018413	78.95	2.98E-13	31%	43%	23--210	9--202
					4	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..	38.47	1.32E-03	25%	35%	214--273	144--207	4	AAM50982	RE24790p	64.31	7.59E-09	26%	41%	1--212	36--250
					5	COG3147	DedD, Uncharacterized protein conserved in bacteria [Function unknown].	37.31	2.98E-03	26%	33%	214--271	88--145	5	AAF46012	CG15786-PA	64.31	7.59E-09	26%	41%	1--212	17--231
					6	pfam01044	Vinculin, Vinculin family..	36.86	3.68E-03	26%	38%	153--303	558--685	6	EAA06469	ENSANGP00000012390	63.16	1.69E-08	27%	41%	26--212	1--197
					7	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 chromosome 12p. This results in an extended polyglutamine region in atrophin-1, that is thought to confer toxicity to the protein, possibly through altering its interactions with other proteins. The expansion of a CAG repeat is also the underlying defect in six other neurodegenerative disorders, including Huntington's disease. One interaction of expanded polyglutamine repeats that is thought to be pathogenic is that with the short glutamine repeat in the transcriptional coactivator CREB binding protein, CBP. This interaction draws CBP away from its usual nuclear location to the expanded polyglutamine repeat protein aggregates that are characteristic of the polyglutamine neurodegenerative disorders. This interferes with CBP-mediated transcription and causes cytotoxicity..	36.43	4.32E-03	34%	42%	214--271	247--306	7	EAL32472	GA13958-PA	59.31	2.44E-07	25%	41%	23--212	11--210
					8	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	36.57	4.42E-03	38%	66%	290--321	18--47	8	EAL29007	GA18137-PA	54.68	6.02E-06	25%	40%	23--210	4--195
					9	pfam06735	DUF1210, Protein of unknown function (DUF1210). This family represents a conserved region within plant proline-rich proteins..	36.24	6.24E-03	35%	37%	213--272	139--199	9	EAL29006	GA18133-PA	53.53	1.34E-05	26%	40%	23--209	18--210
					10	pfam05210	Sprouty, Sprouty protein (Spry). This family consists of eukaryotic Sprouty protein homologues. Sprouty proteins have been revealed as inhibitors of the Ras/mitogen-activated protein kinase (MAPK) cascade, a pathway crucial for developmental processes initiated by activation of various receptor tyrosine kinases. The sprouty gene has found to be expressed in the the brain, cochlea, nasal organs, teeth, salivary gland, lungs, digestive tract, kidneys and limb buds in mice	36.15	6.60E-03	23%	32%	215--272	54--111	10	EAA00829	ENSANGP00000011567	52.76	2.29E-05	26%	37%	22--209	16--215

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
B485L	204187--203474	238	28,073	10.43	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 chromosome 12p. This results in an extended polyglutamine region in atrophin-1, that is thought to confer toxicity to the protein, possibly through altering its interactions with other proteins. The expansion of a CAG repeat is also the underlying defect in six other neurodegenerative disorders, including Huntington's disease. One interaction of expanded polyglutamine repeats that is thought to be pathogenic is that with the short glutamine repeat in the transcriptional coactivator CREB binding protein, CBP. This interaction draws CBP away from its usual nuclear location to the expanded polyglutamine repeat protein aggregates that are characteristic of the polyglutamine neurodegenerative disorders. This interferes with CBP-mediated transcription and causes cytotoxicity..	46.45	5.01E-06	27%	52%	2--85	606--694	1	NP_048695	A339L	104.38	3.01E-21	41%	49%	1--147	1--156
														2	NP_048693	A337L	103.22	6.70E-21	64%	80%	162--238	1--77
B487L	204708--204292	139	16,082	4.11		No Hit Found								1	NP_048697	A341L	229.95	1.61E-59	82%	90%	8--139	4--135
B488L	206441--204789	551	60,640	8.88		No Hit Found								1	NP_048699	A342L	952.97	0.00E+00	85%	90%	1--537	22--559
B492R	206482--206814	111	13,209	10.12		No Hit Found									No Hit Found	No Hit Found						
B494R	206855--207400	182	20,795	9.20		No Hit Found								1	NP_048705	A348R	219.16	4.60E-56	71%	80%	23--179	1--156
B496L	207712--207416	99	11,354	10.16		No Hit Found								1	BAA22202	URF14.2	168.32	5.86E-41	88%	88%	4--99	16--111
														2	BAA22201	URF14.2	162.16	4.20E-39	83%	87%	4--99	16--111
														3	BAA22199	URF14.2	161.00	9.36E-39	82%	87%	4--99	16--111
														4	BAA22200	URF14.2	159.46	2.72E-38	82%	86%	4--99	16--111
														5	NP_048706	A349L	96.29	2.83E-19	93%	100%	4--50	87--133
B497L	208940--208188	251	28,364	10.32		No Hit Found								1	NP_048708	KKD (6X), mixed charge	85.50	1.59E-15	44%	63%	7--106	26--123
														2	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	65.08	2.23E-09	27%	41%	8--250	3--224
														3	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	55.07	2.30E-06	33%	49%	9--125	4--110
														4	YP_293795	putative endonuclease	55.07	2.30E-06	36%	52%	9--101	5--94
														5	NP_048671	A315L	53.53	6.70E-06	23%	36%	8--248	3--239
														6	NP_048641	PBCV-1 33kd peptide	52.76	1.14E-05	24%	39%	4--250	3--248
B498R	209126--209491	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
B499L	210510--209476	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	5.38E-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
B503L	211246--210626	207	23,365	3.66		No Hit Found								1	NP_048709	Asp/Glu rich; DAEDDDIYxxET (2X) negative charge cluster	355.53	5.62E-97	84%	86%	1--207	1--207
														2	YP_142843	unknown	53.14	5.96E-06	22%	41%	6--207	8--221
B506L	212324--211314	337	38,566	4.25		No Hit Found								1	NP_048714	A357L	328.18	2.31E-88	67%	77%	88--337	25--276
														2	NP_048716	a359L	92.43	2.14E-17	61%	75%	1--79	1--81
B508R	212392--217227	1612	179,400	10.81	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	41.30	1.56E-04	22%	42%	1432--1556	842--960	1	NP_048720	similar to chicken vitellogenin II, corresponds to Swiss-Prot Accession Number P02845	393.66	2.96E-107	37%	54%	830--1456	35--666
					2	pfam04484	DUF566, Family of unknown function (DUF566). Family of related proteins that is plant specific..	40.07	4.19E-04	22%	37%	1426--1581	47--194	2	YP_444182	putative replicative DNA helicase, intein-containing	176.02	9.70E-42	33%	47%	437--837	229--582
					3	pfam03999	MAP65_ASE1, Microtubule associated protein (MAP65/ASE1 family)..	36.54	4.26E-03	24%	37%	1442--1611	428--591	3	YP_224270	putative helicase	146.36	8.23E-33	29%	45%	401--837	5--381
					4	smart00490	HELICc, helicase superfamily c-terminal domain; .	36.36	4.73E-03	24%	39%	1235--1311	7--81	4	CAC86344	vacuolar membrane H-ATPase	100.14	6.76E-19	24%	38%	439--829	19--472
														5	BAC66646	vacuolar membrane ATPase subunit a precursor	94.74	2.84E-17	22%	39%	439--829	259--712
														6	AAB63978	transmembrane ATPase-like protein	94.74	2.84E-17	22%	39%	439--829	244--697
														7	CAA58261	ORF D1286	94.74	2.84E-17	22%	39%	439--829	284--737
														8	CAA98762	PI-SceI	94.74	2.84E-17	22%	39%	439--829	1--454
														9	BAC66640	vacuolar membrane ATPase subunit a precursor	94.74	2.84E-17	22%	39%	439--829	259--712
														10	1DFA_A	Chain A, Crystal Structure Of Pi-Scei In C2 Space Group	94.74	2.84E-17	22%	39%	439--829	1--454
B519L	217629--217234	132	15,166	4.89		No Hit Found								1	NP_048723	A366L	67.40	1.42E-10	30%	47%	9--124	2--132
B522L	218432--217677	252	28,633	8.88		No Hit Found								1	NP_048723	A366L	328.56	1.09E-88	61%	77%	1--252	4--255
B524L	219443--218679	255	28,698	9.81		No Hit Found								1	NP_048735	A378L	349.75	4.67E-95	73%	76%	19--239	21--245
														2	BAA11342	DNA binding protein	83.96	4.78E-15	35%	44%	133--244	322--402
														3	NP_048921	A565R	83.57	6.24E-15	35%	44%	133--244	387--467
														4	NP_048917	similar to Chlorella virus CVK2 DNA binding protein, corresponds to GenBank Accession Number D78305	83.57	6.24E-15	35%	44%	133--244	324--404
														5	BAA11343	DNA binding protein	83.19	8.15E-15	35%	44%	133--244	379--459
														6	CAA64974	QI74 protein	73.17	8.43E-12	35%	44%	116--253	206--326
														7	NP_048741	Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	64.70	3.00E-09	26%	33%	79--246	22--177
														8	AAA99804	220 kDa silk protein	51.60	2.63E-05	27%	35%	123--248	45--1296
B525L	220100--219474	209	24,566	10.53		No Hit Found								1	NP_048736	A379L	379.02	4.86E-104	86%	93%	1--209	1--207
B529R	220244--221698	485	55,624	6.62	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 NY2A the major capsid protein is a glycoprotein	177.43	1.71E-45	30%	47%	95--445	3--387	1	NP_048740	similar to PBCV-1 major capsid protein, corresponds to Swiss-Prot Accession Number P30328	481.87	2.08E-134	89%	95%	162--418	1--257
														2	NP_048739	cytochrome C family heme-binding site signature	177.18	1.10E-42	62%	70%	1--147	1--147
														3	BAA22198	major capsid protein Vp54	154.07	9.93E-36	31%	44%	95--443	3--372
														4	BAE06835	hypothetical major capsid protein	153.30	1.69E-35	28%	46%	95--484	3--438
														5	BAA76601	major capsid protein MCP1	151.37	6.43E-35	30%	45%	95--443	3--372
														6	BAA76600	major capsid protein	150.21	1.43E-34	30%	45%	95--443	3--371
														7	NP_048787	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	149.83	1.87E-34	30%	44%	95--443	3--372
														8	AAC27492	major capsid protein Vp49	142.51	2.99E-32	30%	45%	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	127.10	1.30E-27	81%	94%	416--485	387--456
														10	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	124.41	8.42E-27	28%	42%	117--443	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
B532R	221970--222644	225	23,311	7.37		No Hit Found								1	NP_048743	Gly-rich, AGLG (9x); similar to herpesvirus hypothetical protein 5, corresponds to GenBank Accession Number S43071	163.70	3.75E-39	66%	68%	99--225	130--256
B533L	223579--221984	532	57,888	7.64	1	pfam05616	Neisseria_TspB, Neisseria meningitidis TspB protein. This family consists of several Neisseria meningitidis TspB virulence factor proteins..	42.43	6.81E-05	31%	46%	446--508	335--396	1	NP_048741	Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	428.33	3.07E-118	76%	83%	183--436	8--261
					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	42.27	7.51E-05	27%	46%	439--480	71--112	2	BAA11343	DNA binding protein	162.16	4.10E-38	38%	47%	144--349	244--458
					3	pfam01213	CAP, CAP protein..	40.76	2.14E-04	31%	41%	356--480	230--326	3	NP_048921	A565R	157.92	7.74E-37	37%	45%	144--349	244--466
					4	pfam04610	TrbL, TrbL/VirB6 plasmid conjugal transfer protein..	39.89	4.96E-04	16%	28%	1--94	21--127	4	NP_048917	similar to Chlorella virus CVK2 DNA binding protein, corresponds to GenBank Accession Number D78305	104.76	7.77E-21	46%	56%	249--349	304--403
					5	COG0341	SecF, Preprotein translocase subunit SecF [Intracellular trafficking and secretion].	39.41	6.36E-04	25%	45%	24--95	80--156	5	BAA11342	DNA binding protein	103.99	1.33E-20	46%	56%	249--349	302--401
					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)	38.32	1.43E-03	30%	43%	444--481	98--135	6	NP_048747	Asn-rich	94.36	1.05E-17	60%	66%	48--125	59--134
					7	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	38.20	1.60E-03	54%	59%	442--482	84--123	7	NP_048688	a332L	65.86	4.00E-09	59%	78%	491--532	2--43
					8	COG1971	COG1971, Predicted membrane protein [Function unknown].	36.69	3.99E-03	25%	43%	45--107	34--102	8	BAE02830	surface protein	55.45	5.41E-06	40%	59%	376--436	569--629
					9	pfam02355	SecD_SecF, Protein export membrane protein. This family consists of various prokaryotic SecD and SecF protein export membrane proteins. This SecD and SecF proteins are part of the multimeric protein export complex comprising SecA, D, E, F, G, Y, and YajC. SecD and SecF are required to maintain a proton motive force	36.32	4.99E-03	28%	45%	19--95	67--131	9	CAA64974	QI74 protein	54.30	1.20E-05	25%	37%	95--345	253--501
B536R	223657--224418	254	29,147	9.51	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. .	47.05	3.41E-06	17%	39%	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	395.20	9.61E-109	78%	82%	1--253	1--253
														2	BAB69884	AGB-1	162.16	1.37E-38	43%	64%	33--229	39--238
														3	AAR26836	FirrV-1-A12	135.96	1.05E-30	36%	55%	33--230	37--243
														4	NP_077511	EsV-1-26	124.02	4.13E-27	33%	54%	33--230	48--254
														5	NP_149538	075L	123.25	7.05E-27	33%	57%	33--230	46--246
														6	YP_293826	putative DNA-binding protein	117.09	5.05E-25	34%	55%	33--214	35--215
														7	AAS18149	ATPase	112.85	9.53E-24	31%	57%	33--229	47--245
														8	YP_003858	ATPase	112.85	9.53E-24	32%	57%	33--230	44--243
														9	AAV91100	ATPase-like protein	112.85	9.53E-24	31%	57%	33--229	47--245
														10	AAT09674	AAA-ATPase	112.46	1.24E-23	32%	57%	33--230	44--243
B537R	224436--224639	68	7,583	6.49		No Hit Found									No Hit Found	No Hit Found						
B539R	224639--225016	126	14,884	3.81		No Hit Found								1	NP_048751	A394R	179.10	3.33E-44	65%	86%	5--124	2--121
B540R	225075--225320	82	9,411	10.31		No Hit Found								1	NP_048752	A395R	138.27	6.41E-32	81%	89%	1--82	1--82
B541L	225909--225460	150	18,221	6.81		No Hit Found								1	NP_048753	A396L	94.74	8.23E-19	69%	81%	83--149	1--69
B542R	226129--227022	298	34,691	10.38		No Hit Found									No Hit Found	No Hit Found						
B543L	227791--226985	269	31,138	8.32	1	smart00650	rADc, Ribosomal RNA adenine dimethylases; .	47.09	3.10E-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
B546L	228364--228011	118	12,964	10.01		No Hit Found								1	NP_048755	A398L	203.76	1.26E-51	84%	92%	1--118	1--118
B547R	228437--229015	193	21,772	8.12	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..	79.57	5.61E-16	31%	39%	27--157	1--124	1	NP_048756	A399R	288.12	9.35E-77	73%	86%	2--191	3--193
B548R	229041--229403	121	13,904	9.87		No Hit Found								1	NP_048757	similar to PBCV-1 ORF A214, corresponds to GenBank Accession Number U42580	171.40	6.84E-42	75%	81%	1--117	1--117
B549R	229428--230261	278	31,748	6.07		No Hit Found								1	NP_048758	A401R	485.34	8.29E-136	86%	91%	4--278	2--277
														2	BAC51116	bll5851	173.33	6.98E-42	38%	56%	36--276	3--249
														3	ZP_00863599	conserved hypothetical protein	164.08	4.23E-39	35%	56%	36--276	7--253
														4	YP_190685	hypothetical protein GOX0246	139.81	8.55E-32	30%	53%	40--275	11--253
														5	ZP_00577151	conserved hypothetical protein	96.29	1.08E-18	27%	49%	40--274	14--260
														6	YP_162362	hypothetical protein ZMO0627	54.30	4.72E-06	24%	42%	38--275	9--207
														7	YP_162363	hypothetical protein ZMO0628	52.37	1.79E-05	22%	42%	64--269	13--204
B550R	230410--231105	232	26,043	7.97		No Hit Found								1	NP_048759	A402R	402.91	3.87E-111	85%	94%	6--232	1--227
B552R	231143--231433	97	11,479	9.68		No Hit Found								1	NP_048760	A403R	173.71	1.41E-42	92%	96%	7--97	3--93
B553R	231469--232029	187	21,800	3.09	1	COG5271	MDN1, AAA ATPase containing von Willebrand factor type A (vWA) domain [General function prediction only].	38.15	1.48E-03	29%	48%	52--172	920--4046	1	NP_048761	Glu-, Asn-rich	103.99	2.33E-21	84%	94%	3--59	6--64
B554R	232061--233629	523	55,631	10.60	1	pfam05616	Neisseria_TspB, Neisseria meningitidis TspB protein. This family consists of several Neisseria meningitidis TspB virulence factor proteins..	50.52	2.51E-07	32%	37%	102--177	323--398	1	NP_048762	Pro-, Lys-rich, PAPK (30x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	442.58	1.54E-122	75%	86%	249--523	218--496
B556R	233711--234832	374	44,109	7.49		No Hit Found									No Hit Found	No Hit Found						
B559L	235477--234848	210	23,552	8.92		No Hit Found								1	NP_048764	A407L	335.11	8.13E-91	77%	81%	1--210	1--210
B560L	236307--235510	266	29,777	4.73		No Hit Found								1	NP_048765	A408L	346.67	4.24E-94	81%	88%	15--228	18--233
														2	NP_048767	A410L	83.19	8.74E-15	40%	61%	36--134	8--107
														3	NP_077527	EsV-1-42	80.49	5.67E-14	31%	51%	35--181	6--164
														4	AAR26867	FirrV-1-A43	68.17	2.91E-10	26%	54%	31--169	1--144
B562L	236642--236268	125	14,946	4.61		No Hit Found								1	NP_048767	A410L	207.22	1.14E-52	90%	95%	16--125	1--110
														2	NP_048765	A408L	82.80	3.26E-15	37%	59%	21--122	39--139
														3	NP_077527	EsV-1-42	50.83	1.37E-05	28%	46%	20--121	4--110
														4	AAR26867	FirrV-1-A43	48.14	8.89E-05	25%	51%	23--124	6--111
B564R	236732--237265	178	20,261	6.52		No Hit Found								1	NP_048769	A412R	320.86	1.05E-86	86%	90%	1--178	1--179

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
B566R	237261--238307	349	39,454	8.41	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	204.77	9.10E-54	33%	52%	19--347	1--314	1	AAP07928	Type II restriction-modification system methylation subunit	151.75	3.11E-35	34%	50%	19--348	18--350
					2	COG0270	Dcm, Site-specific DNA methylase [DNA replication, recombination, and repair].	196.84	2.77E-51	34%	50%	16--348	1--326	2	CAD33713	putative DNA methylase	150.21	9.05E-35	34%	53%	19--335	1--292
					3	pfam00145	DNA_methylase, C-5 cytosine-specific DNA methylase..	194.76	9.92E-51	32%	50%	19--347	1--323	3	AAC97192	modification methylase M.NspHI	148.29	3.44E-34	34%	49%	22--335	63--359
														4	AAC97190	modification methylase M.NspI	146.75	1.00E-33	33%	49%	22--335	63--359
														5	AAX14650	BbvCI methyltransferase 1	145.59	2.23E-33	31%	47%	5--337	5--402
														6	YP_208922	putative 5-methylcytosine methyltransferase	145.59	2.23E-33	31%	50%	15--347	1--328
														7	CAA74996	Bpu10I (5m)cytosine-specific DNA modification methyltransferase (C1)	141.74	3.22E-32	29%	46%	16--347	5--393
														8	AAT40769	putative DNA methylase	138.27	3.56E-31	34%	51%	15--330	1--293
														9	CAA68505	DdeI methylase	138.27	3.56E-31	31%	45%	19--336	1--357
														10	ABB52090	Mod	135.96	1.77E-30	32%	47%	21--336	7--353
B567L	239412--238300	371	42,589	8.66	1	COG4123	COG4123, Predicted O-methyltransferase [General function prediction only].	59.49	5.19E-10	25%	37%	43--154	45--171	1	AAC57943	DNA adenine methyltransferase	649.43	0.00E+00	85%	94%	1--368	1--368
					2	COG2890	HemK, Methylase of polypeptide chain release factors [Translation, ribosomal structure and biogenesis].	51.92	9.50E-08	26%	45%	31--114	99--187	2	P52284	Modification methylase CviRI (Adenine-specific methyltransferase CviRI) (M.CviRI)	208.38	3.08E-52	34%	54%	3--368	4--378
					3	COG0286	HsdM, Type I restriction-modification system methyltransferase subunit [Defense mechanisms].	51.18	1.79E-07	20%	37%	18--236	161--408	3	AAC03124	DNA adenine methyltransferase	204.53	4.45E-51	33%	55%	1--367	1--368
					4	COG2813	RsmC, 16S RNA G1207 methylase RsmC [Translation, ribosomal structure and biogenesis].	49.88	4.10E-07	29%	49%	45--152	161--265	4	AAC57945	DNA adenine methyltransferase	204.14	5.81E-51	33%	53%	7--367	6--357
					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..	47.57	2.43E-06	25%	43%	44--113	30--107	5	CAA29835	unnamed protein product	193.36	1.03E-47	32%	54%	11--368	14--377
					6	COG0421	SpeE, Spermidine synthase [Amino acid transport and metabolism].	44.96	1.28E-05	28%	44%	38--151	72--188	6	AAC03125	DNA adenine methyltransferase	187.58	5.63E-46	29%	53%	3--368	4--381
					7	COG1041	COG1041, Predicted DNA modification methylase [DNA replication, recombination, and repair].	42.64	6.38E-05	27%	39%	44--154	199--311	7	AAC60387	methyltransferase; M-AccI	78.57	3.68E-13	30%	47%	12--209	19--237
					8	COG2230	Cfa, Cyclopropane fatty acid synthase and related methyltransferases [Cell envelope biogenesis, outer membrane].	41.40	1.75E-04	24%	43%	32--151	61--174	8	AAA50500	AccI methylase	78.57	3.68E-13	30%	47%	12--209	19--237
					9	COG2263	COG2263, Predicted RNA methylase [Translation, ribosomal structure and biogenesis].	39.06	7.14E-04	24%	44%	41--117	44--123	9	JU0470	site-specific DNA-methyltransferase (adenine-specific) (EC 2.1.1.72) type II - Acinetobacter calcoaceticus	78.57	3.68E-13	30%	47%	12--209	19--237
					10	COG2264	PrmA, Ribosomal protein L11 methylase [Translation, ribosomal structure and biogenesis].	37.20	3.14E-03	23%	41%	42--110	162--236	10	P25201	Modification methylase AccI (Adenine-specific methyltransferase AccI) (M.accI)	76.26	1.83E-12	30%	47%	12--209	19--237
B568L	240166--239459	236	26,384	8.98		No Hit Found								1	NP_048770	Gln-rich, QQQQM(4x); similar to human transcription factor TFIID, corresponds to Swiss-Prot Accession Number P20226	277.72	1.95E-73	59%	70%	1--236	1--244
B570R	240239--240454	72	8,003	10.71		No Hit Found								1	NP_048771	A414R	130.57	1.33E-29	87%	95%	1--72	22--93
B571L	241762--240461	434	49,993	6.78	1	COG2256	MGS1, ATPase related to the helicase subunit of the Holliday junction resolvase [DNA replication, recombination, and repair].	36.74	3.75E-03	21%	39%	54--244	52--252	1	NP_048774	A417L	650.20	0.00E+00	76%	87%	19--427	19--427
														2	XP_724804	replication factor C, 140 kDa subunit	51.22	7.81E-05	18%	40%	55--351	397--724
B573L	242013--241789	75	8,628	7.17		No Hit Found								1	NP_048777	A420L	114.78	7.77E-25	77%	94%	5--74	1--70
B575R	242046--242339	98	11,060	9.78		No Hit Found								1	NP_048778	A421R	135.96	3.24E-31	63%	75%	1--98	1--98
														2	NP_048487	A139L	51.22	1.05E-05	35%	53%	14--90	11--77
B576R	242579--243034	152	17,929	4.82		No Hit Found								1	NP_048780	A423R	226.48	1.80E-58	71%	81%	1--152	1--157
B579R	243061--243312	84	9,919	10.99		No Hit Found									No Hit Found	No Hit Found						
B580R	243363--243710	116	13,441	4.87		No Hit Found								1	NP_048783	A426R	190.66	1.11E-47	79%	89%	3--116	1--114
B581L	244069--243713	119	13,722	6.48		No Hit Found								1	NP_048784	contains thioredoxin active site-like sequence; similar to Synechocystis thioredoxin-like protein, corresponds to Swiss-Prot Accession Number P52232	194.90	5.82E-49	75%	87%	1--118	1--118
														2	EAL29786	GA21460-PA	50.06	2.32E-05	24%	48%	9--111	34--140
														3	ZP_00533498	Thioredoxin	49.68	3.02E-05	28%	52%	22--109	7--98
														4	AAF47638	CG8993-PA	48.14	8.80E-05	25%	51%	9--100	34--129
B582L	244350--244123	76	8,409	4.10		No Hit Found									No Hit Found	No Hit Found						
B583L	245741--244383	453	53,286	5.29	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	3.10E-05	28%	45%	140--250	11--118	1	NP_048786	A429L	741.11	0.00E+00	77%	87%	1--452	24--475
														2	NP_192255	protein binding	75.49	4.11E-12	24%	41%	34--364	128--472
														3	CAB77831	hypothetical protein	70.09	1.73E-10	22%	43%	34--364	128--432
														4	BAD29430	ankyrin-like protein	62.77	2.76E-08	23%	39%	37--353	138--470
														5	CAB10219	hypothetical protei	56.23	2.58E-06	22%	39%	13--353	74--434
														6	BAD94307	hypothetical protein	56.23	2.58E-06	22%	39%	13--353	79--439
														7	BAD43172	unknown protein	56.23	2.58E-06	22%	39%	13--353	79--439
														8	BAC43653	unknown protein	56.23	2.58E-06	22%	39%	13--353	13--373
														9	NP_567430	ACD6 (ACCELERATED CELL DEATH 6); protein binding	56.23	2.58E-06	22%	39%	13--353	79--439
														10	NP_849381	ACD6 (ACCELERATED CELL DEATH 6); protein binding	56.23	2.58E-06	22%	39%	13--353	13--373
B585L	247180--245870	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 NY2A the major capsid protein is a glycoprotein	477.50	8.14E-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
B587R	247296--247760	155	17,172	8.45		No Hit Found								1	NP_048789	Lys-rich	182.96	2.32E-45	74%	85%	9--124	1--116
B590R	247745--247954	70	7,812	11.50		No Hit Found									No Hit Found	No Hit Found						
B591L	248300--247992	103	11,196	11.01		No Hit Found								1	NP_048794	similar to Methanothrix chromosomal protein MC1A, corresponds to Swiss-Prot Accession Number P15251	133.27	2.06E-30	66%	74%	1--102	1--103
B592L	248563--248330	78	9,165	10.77	1	COG0695	GrxC, Glutaredoxin and related proteins [Posttranslational modification, protein turnover, chaperones].	44.95	1.26E-05	28%	45%	1--78	1--76	1	NP_048795	similar to E. coli glutaredoxin, corresponds to Swiss-Prot Accession Number P37687	133.65	1.60E-30	79%	91%	1--78	1--78
B593R	248586--248924	113	13,309	7.36		No Hit Found								1	NP_048796	A439R	180.26	1.47E-44	72%	85%	1--112	1--112

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
B595L	249499--249089	137	15,536	4.42		No Hit Found								1	NP_048798	A441L	208.38	5.06E-53	73%	83%	1--137	1--137
B596R	249630--250553	308	35,495	5.35		No Hit Found								1	NP_048800	A443R	525.78	6.57E-148	85%	91%	1--308	1--308
B597L	250971--250657	105	11,899	5.02		No Hit Found								1	NP_048801	A444L	181.03	8.86E-45	89%	92%	1--105	1--104
B598L	252214--251018	399	46,694	6.53		No Hit Found								1	NP_048711	A354R	159.84	1.39E-37	37%	56%	155--399	1--235
														2	NP_048779	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	82.03	3.68E-14	29%	48%	87--366	44--319
B602L	253070--252429	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-strand break in the DNA near the insertion site of that element to facilitate homing at that site. Class I homing endonucleases are sorted into four families based on the presence of these motifs in their respective N-termini: LAGLIDADG, His-Cys box, HNH, and GIY-YIG. This CD contains several but not all members of the GIY-YIG family. The C-terminus of GIY-YIG is a DNA-binding domain which is separated from the N-terminus by a long, flexible linker. The DNA-binding domain consists of a minor-groove binding alpha-helix, and a helix-turn-helix. Some also contain a zinc finger (i.e. I-TevI) which is not required for DNA binding or catalysis, but is a component of the linker and directs the catalytic domain to cleave the homing site at a fixed distance from the intron insertion site	49.23	6.11E-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	8.21E-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	8.66E-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
B606L	254543--253158	462	52,742	6.71	1	COG0661	AarF, Predicted unusual protein kinase [General function prediction only].	181.69	8.89E-47	27%	43%	42--436	56--477	1	NP_048802	similar to Clostridium pasteurianum ORF, corresponds to GenBank Accession Number Z28353	812.37	0.00E+00	89%	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	1.78E-28	40%	56%	94--208	6--119	2	BAB66733	488aa long conserved hypothetical protein	140.20	1.39E-31	28%	48%	54--383	50--399
														3	AAY80974	universally conserved protein	135.96	2.61E-30	27%	46%	42--383	38--399
														4	AAK42726	ABC transporter, ABC1 family, putative	135.58	3.41E-30	26%	48%	24--382	19--393
														5	ZP_00779577	2-polyprenylphenol 6-hydroxylase	132.11	3.78E-29	30%	45%	39--385	50--420
														6	XP_450284	ABC1 family protein-like	129.80	1.87E-28	25%	47%	42--383	207--575
														7	AAF13088	unknown protein	128.64	4.17E-28	24%	46%	53--402	199--573
														8	AAF21180	unknown protein	127.87	7.12E-28	25%	48%	53--383	199--554
														9	AAL10497	At3g07700/F17A17.4	127.87	7.12E-28	25%	48%	53--383	199--554
														10	AAM67100	unknown	127.87	7.12E-28	25%	48%	53--383	199--554
B611L	254928--254611	106	12,541	10.78	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..	49.45	5.59E-07	29%	51%	31--104	33--105	1	NP_048805	contains cytochrome C family heme-binding site signature; similar to maize protein disulphide isomerase, correspond to Swiss-Prot Accession Number P52588	189.12	3.24E-47	82%	92%	1--106	1--106
					2	COG3118	COG3118, Thioredoxin domain-containing protein [Posttranslational modification, protein turnover, chaperones].	36.04	5.95E-03	22%	45%	23--105	46--129	2	XP_532876	PREDICTED: similar to Protein disulfide-isomerase A6 precursor (Thioredoxin domain containing protein 7)	53.53	2.13E-06	40%	57%	31--99	55--124
														3	EAN99322	protein disulfide isomerase, putative	53.14	2.78E-06	30%	54%	31--103	49--121
														4	CAD99203	protein disulfide isomerase1-2	52.76	3.63E-06	31%	54%	22--102	40--122
														5	EAN94802	protein disulfide isomerase, putative	52.37	4.73E-06	30%	53%	31--103	49--121
														6	AAF78087	protein disulfide isomerase ER-60	51.22	1.05E-05	31%	48%	25--104	43--120
														7	AAQ23042	transglutaminase	50.83	1.38E-05	30%	52%	14--103	8--102
														8	BAC86977	unnamed protein product	50.83	1.38E-05	37%	55%	31--99	107--176
														9	NP_005733	protein disulfide isomerase-associated 6	50.83	1.38E-05	37%	55%	31--99	55--124
														10	XP_515706	PREDICTED: protein disulfide isomerase-related protein	50.83	1.38E-05	37%	55%	31--99	3--72
B612R	255088--255651	188	22,207	7.75		No Hit Found								1	NP_048806	PBCV-1 RNA triphosphatase	322.40	4.16E-87	82%	89%	2--188	7--193
B614R	255923--256660	246	29,191	10.70		No Hit Found									No Hit Found	No Hit Found						
B617L	258269--256959	437	48,352	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 NY2A the major capsid protein is a glycoprotein	467.48	8.58E-133	51%	64%	1--433	1--443	1	BAA76601	major capsid protein MCP1	776.16	0.00E+00	87%	89%	1--437	1--437
														2	NP_048787	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	769.62	0.00E+00	87%	88%	1--437	1--437
														3	BAA22198	major capsid protein Vp54	760.76	0.00E+00	85%	87%	1--437	1--437
														4	BAA76600	major capsid protein	757.29	0.00E+00	85%	88%	1--437	1--436
														5	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	722.62	0.00E+00	86%	88%	25--437	1--413
														6	1M3Y_D	Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid Containing, Dna Virus	722.62	0.00E+00	86%	88%	25--437	1--413
														7	AAC27492	major capsid protein Vp49	632.48	8.29E-180	73%	79%	1--437	1--432
														8	BAE06835	hypothetical major capsid protein	318.16	3.46E-85	41%	53%	1--437	1--440
														9	NP_048359	contains aminoacyl-tRNA synthetase class-II signature	248.83	2.58E-64	36%	51%	1--437	1--403
														10	NP_048914	similar to Chlorella virus PBCV-1 ORF A11L, corresponds to GenBank Accession Number U42580	229.95	1.24E-58	35%	50%	3--437	2--400
B618R	258376--259185	270	32,681	5.62	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	91.17	1.44E-19	26%	48%	4--239	84--311	1	CAE79544	putative N-acetylglucosaminyltransferase	128.26	2.44E-28	30%	49%	6--240	2--246
														2	EAA75195	hypothetical protein FG05624.1	77.41	4.93E-13	28%	41%	4--240	89--334
														3	XP_362071	hypothetical protein MG04516.4	73.94	5.45E-12	26%	40%	4--257	79--341
														4	XP_466520	glycosyl transferase-like protein	64.70	3.31E-09	23%	44%	2--239	79--307
														5	XP_466518	glycosyl transferase-like protein	63.93	5.64E-09	22%	44%	2--239	112--340
														6	XP_472878	OSJNBa0044K18.7	62.39	1.64E-08	22%	42%	2--239	109--337
														7	NP_178963	acetylglucosaminyltransferase/ transferase, transferring glycosyl groups	60.85	4.78E-08	25%	41%	4--239	81--311

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	XP_327741	predicted protein	56.61	9.01E-07	42%	60%	4--71	78--148
														9	NP_186811	acetylglucosaminyltransferase/ transferase, transferring glycosyl groups	54.68	3.42E-06	20%	42%	4--239	110--337
														10	AAH53040	Mannoside acetylglucosaminyltransferase 3	51.60	2.90E-05	32%	57%	4--74	210--282
B619L	260121--259261	287	31,185	4.73		No Hit Found								1	NP_048811	A454L	501.52	1.17E-140	85%	91%	1--287	1--289
B623L	262114--260153	654	75,534	5.56	1	COG3378	COG3378, Predicted ATPase [General function prediction only].	67.31	2.32E-12	24%	40%	283--533	146--390	1	NP_048813	contains ATP/GTP-binding site motif A	1232.62	0.00E+00	91%	96%	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..	66.04	6.76E-12	27%	44%	292--524	3--228	2	NP_077594	EsV-1-109	281.95	4.58E-74	31%	49%	36--618	18--575
														3	AAR26902	FirrV-1-B27	263.85	1.29E-68	32%	50%	132--619	93--574
														4	YP_294217	putative nucleic acid independent nucleoside triphosphatase	148.29	7.90E-34	24%	43%	191--643	170--672
														5	ZP_00123428	COG3378: Predicted ATPase	65.86	5.16E-09	27%	48%	340--532	310--502
														6	ZP_00593388	Phage/plasmid primase P4, C-terminal	57.77	1.40E-06	25%	40%	332--523	318--508
														7	BAE05402	putative DNA primase-phage associated	57.00	2.40E-06	25%	45%	307--525	110--315
														8	ZP_00503756	Phage/plasmid primase P4, C-terminal	56.61	3.13E-06	23%	40%	239--501	246--492
														9	AAF27348	phage phi-R73 primase-like protein	53.91	2.03E-05	22%	41%	251--532	186--465
														10	CAG34908	hypothetical protein, probably cold-shock inducible	53.53	2.65E-05	26%	40%	354--530	361--539
B628R	262453--263256	268	30,485	5.78	1	COG0571	Rnc, dsRNA-specific ribonuclease [Transcription].	167.75	1.33E-42	33%	54%	41--266	5--235	1	NP_048820	similar to Bacillus ribonuclease III, corresponds to Swiss-Prot Accession Number P51833	489.57	4.11E-137	88%	95%	1--267	8--274
					2	smart00535	RIBOc, Ribonuclease III family; .	128.05	1.25E-30	38%	58%	59--185	1--126	2	YP_445467	ribonuclease III	139.81	7.98E-32	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	3.63E-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	5.68E-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	1.33E-11	36%	55%	194--260	1--68	5	ZP_00591208	Ribonuclease III	129.80	8.26E-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	118.24	2.49E-25	30%	52%	36--257	20--252
B629R	263456--264316	287	34,606	9.63	1	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	46.61	4.69E-06	41%	62%	45--124	12--83	1	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	71.25	3.91E-11	31%	50%	35--197	3--165
					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	5.56E-06	31%	49%	39--120	4--88	2	NP_048671	A315L	69.32	1.49E-10	31%	45%	47--213	13--178
														3	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	67.40	5.65E-10	45%	56%	47--133	13--98
														4	NP_048641	PBCV-1 33kd peptide	65.86	1.64E-09	31%	52%	39--213	9--189
														5	NP_048482	similar to bacteriophage T4 intron-associated endonuclease, corresponds to Swiss-Prot Accession Number P13299	57.00	7.64E-07	38%	60%	38--120	10--93
														6	YP_293795	putative endonuclease	51.22	4.19E-05	35%	53%	38--128	4--97
														7	AAU16837	GIY-YIG catalytic domain containing protein; possible intron encoded endonuclease	50.06	9.34E-05	29%	48%	39--165	5--132
B630R	264343--264693	117	13,610	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	82.65	6.71E-17	33%	49%	18--112	1--91	1	NP_048821	PBCV-1 thiol oxidoreductase	194.13	1.00E-48	72%	84%	1--117	1--118
					2	COG5054	ERV1, Mitochondrial sulfhydryl oxidase involved in the biogenesis of cytosolic Fe/S proteins [Posttranslational modification, protein turnover, chaperones].	55.45	1.03E-08	30%	49%	1--111	69--175	2	YP_142722	putative thiol oxidoreductase	79.72	2.75E-14	38%	54%	11--105	3--97
														3	YP_142950	thiol oxidoreductase E10R	65.08	7.01E-10	37%	52%	12--95	38--123
														4	CAG59329	unnamed protein product	63.16	2.66E-09	28%	53%	15--111	76--168
														5	XP_503294	hypothetical protein	62.39	4.54E-09	32%	52%	10--112	88--186
														6	CAA48192	ERV1	57.38	1.46E-07	27%	51%	15--111	19--111
														7	CAA97017	ERV1	57.38	1.46E-07	27%	51%	15--111	91--183
														8	CAH02199	unnamed protein product	57.00	1.91E-07	29%	49%	4--112	74--178
														9	CAE74303	Hypothetical protein CBG22010	55.07	7.25E-07	31%	47%	15--108	59--152
														10	EAL40090	ENSANGP00000025411	54.68	9.47E-07	30%	51%	10--98	59--144
B631L	265646--264711	312	36,748	6.77		No Hit Found								1	NP_048823	A467L	580.48	2.27E-164	91%	94%	1--312	1--312
														2	NP_065022	Hypothetical protein	58.92	2.30E-07	26%	44%	42--223	39--218
B633R	265781--267109	443	50,874	9.00		No Hit Found								1	NP_048824	A468R	752.67	0.00E+00	80%	90%	1--443	1--443
														2	AAR26870	FirrV-1-A46	57.77	8.59E-07	22%	41%	1--276	1--256
B636R	267189--267788	200	22,684	4.50		No Hit Found								1	NP_048826	A470R	295.82	4.84E-79	73%	83%	1--199	1--203
														2	BAB69883	UKCH-2	91.28	1.81E-17	36%	61%	15--139	57--179
B638R	267825--268343	173	20,873	9.59		No Hit Found								1	NP_048827	A471R	309.30	2.95E-83	84%	91%	1--173	1--173
														2	YP_142861	unknown	120.94	1.49E-26	44%	62%	17--156	43--191
														3	AAR26829	FirrV-1-A5	84.73	1.18E-15	32%	58%	3--149	14--161
														4	NP_077626	EsV-1-141	75.87	5.49E-13	31%	56%	7--143	23--160
B641R	268408--269379	324	37,498	4.61	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, many prokaryotes, several viruses, and few archaea. The catalytically active form of RNR is a proposed alpha2-beta2 tetramer. The homodimeric alpha subunit (R1) contains the active site and redox active cysteines as well as the allosteric binding sites. The beta subunit (R2) contains a diiron cluster that, in its reduced state, reacts with dioxygen to form a stable tyrosyl radical and a diiron(III) cluster. This essential tyrosyl radical is proposed to generate a thiyl radical, located on a cysteine residue in the R1 active site that initiates ribonucleotide reduction. The beta subunit is composed of 10-13 helices, the 8 longest helices form an alpha-helical bundle; some have 2 addition beta strands. Yeast is unique in that it assembles both homodimers and heterodimers of RNRR2. The yeast heterodimer	298.70	5.55E-82	49%	64%	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	604.75	1.20E-171	89%	94%	1--324	1--324
					2	pfam00268	Ribonuc red sm, Ribonucleotide reductase, small chain..	297.94	9.39E-82	45%	62%	5--285	1--281	2	AAO62422	ribonucleotide reductase small subunit	395.20	1.44E-108	57%	72%	4--324	6--333
					3	COG0208	NrdF, Ribonucleotide reductase, beta subunit [Nucleotide transport and metabolism].	237.50	1.45E-63	32%	48%	7--324	18--346	3	NP_189342	ribonucleoside-diphosphate reductase	394.05	3.22E-108	57%	73%	4--324	6--332

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	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 oxidation-hydroxylation reactions within diiron centers; one exception is manganese catalase, which catalyzes peroxide-dependent oxidation-reduction within a dimanganese center. Diiron-carboxylate proteins are further characterized by the presence of duplicate metal ligands, glutamates and histidines (ExxH) and two additional glutamates within a four-helix bundle. Outside of these conserved residues there is little obvious homology. Members include bacterioferritin, ferritin, rubrerythrin, aromatic and alkene monooxygenase hydroxylases (AAMH), ribonucleotide reductase R2 (RNRR2), acyl-ACP-desaturases (Acyl_ACP_Desat), manganese (Mn) catalases, demethoxyubiquinone hydroxylases (DMQH), DNA protecting proteins (DPS), and ubiquinol oxidases (AOX). Additional members include the Fe-containing subunit of the aerobic cyclase system (ACSF), the ferritin-like	53.95	2.78E-08	21%	31%	66--213	1--140	4	XP_500466	hypothetical protein	390.58	3.56E-107	59%	73%	5--324	77--403
														5	CAF24073	probable ribonucleoside-diphosphate reductase small chain	385.96	8.76E-106	58%	73%	1--324	1--323
														6	CAA63194	ribonucleotide reductase R2	384.03	3.33E-105	57%	73%	4--324	6--329
														7	NP_619839	CPXV051 protein	381.33	2.16E-104	58%	73%	5--324	16--333
														8	BAD46317	putative ribonucleotide reductase R2	381.33	2.16E-104	55%	73%	4--324	14--339
														9	XP_739266	ribonucleotide reductase small subunit	380.18	4.81E-104	58%	76%	5--324	31--349
														10	XP_550581	putative ribonucleotide reductase R2	379.41	8.20E-104	55%	73%	4--324	20--345
B644L	269768--269493	92	9,801	10.58		No Hit Found								1	NP_048836	Gly-rich	112.08	4.93E-24	87%	94%	1--56	1--56
B645L	270471--269800	224	26,000	4.84	1	COG5540	COG5540, RING-finger-containing ubiquitin ligase [Posttranslational modification, protein turnover, chaperones].	37.76	2.17E-03	33%	49%	145--194	324--373	1	NP_048837	A481L	422.17	5.79E-117	85%	94%	1--224	1--224
B647R	270545--271189	215	24,901	10.03		No Hit Found								1	NP_048838	A482R	390.58	1.70E-107	85%	95%	1--212	1--214
														2	AAR26970	FirrV-1-I5	65.08	1.63E-09	30%	44%	30--126	32--146
														3	AAR26889	FirrV-1-B14	60.85	3.08E-08	32%	45%	33--126	32--145
														4	YP_294196	hypothetical protein EhV 438	58.54	1.53E-07	28%	43%	33--159	77--201
														5	NP_077581	EsV-1-96	58.15	2.00E-07	30%	46%	30--126	67--181
B650L	271671--271207	155	18,688	9.78		No Hit Found								1	NP_048840	A484L	278.87	3.10E-74	83%	92%	1--155	1--155
B651R	271706--272137	144	17,056	10.89		No Hit Found								1	NP_048841	A485R	242.28	3.17E-63	83%	90%	3--144	6--147
B652R	272197--273147	317	34,878	5.46		No Hit Found								1	NP_048844	Pro-,Gln-rich	446.82	4.03E-124	75%	80%	1--317	1--317
														2	YP_142771	unknown	57.77	5.27E-07	25%	43%	8--177	5--191
B655R	273188--273415	76	8,424	4.55		No Hit Found								1	NP_048847	A491R	126.33	2.57E-28	77%	92%	1--76	1--76
B656L	273983--273393	197	23,266	7.16		No Hit Found								1	NP_048848	A492L	309.30	4.14E-83	80%	85%	1--189	1--189
														2	NP_048849	A493L	197.98	1.34E-49	57%	70%	10--187	1--178
B657L	274583--274026	186	22,128	9.66		No Hit Found								1	NP_048849	A493L	305.45	5.20E-82	81%	84%	1--185	1--186
														2	NP_048848	A492L	228.41	8.08E-59	62%	75%	1--177	10--187
B659L	274938--274711	76	8,978	3.97		No Hit Found									No Hit Found	No Hit Found						
B662R	275157--275582	142	16,228	7.78		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	174.10	1.07E-42	54%	71%	1--142	167--308
														2	NP_048846	Lys-, Glu-rich	166.39	2.23E-40	53%	69%	1--142	167--308
														3	YP_142777	unknown	90.12	2.03E-17	38%	56%	1--121	332--457
														4	NP_048621	A267L	53.91	1.61E-06	30%	44%	6--116	128--239
B663R	275618--276697	360	42,234	8.31		No Hit Found								1	NP_048850	A494R	597.43	2.25E-169	80%	89%	6--360	7--360
														2	YP_142783	PBCV1-A494R-like protein	149.44	1.62E-34	40%	60%	170--360	236--431
														3	NP_077586	EsV-1-101	142.12	2.59E-32	39%	57%	168--359	120--318
														4	AAR26895	FirrV-1-B20	119.40	1.80E-25	32%	55%	171--359	105--300
														5	YP_294161	hypothetical protein EhV 403	104.76	4.58E-21	32%	48%	155--356	156--362
														6	NP_149745	282R	63.93	8.95E-09	23%	40%	153--357	155--380
														7	YP_142784	unknown	63.16	1.53E-08	24%	40%	89--358	5--251
B666R	276730--277146	139	15,131	10.43		No Hit Found								1	NP_048853	Gly-rich, KGKDIGGG (4x)	81.65	7.08E-15	34%	39%	1--139	1--146
B667L	278216--277209	336	36,877	5.34	1	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	56.69	4.35E-09	40%	48%	124--209	51--137	1	NP_048857	A501L	91.28	4.74E-17	92%	92%	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	4.53E-09	32%	58%	115--186	54--125	2	NP_048856	Pro-, Ser-rich	57.38	7.58E-07	51%	59%	259--313	1--62
B670L	278536--278252	95	11,120	9.12		No Hit Found								1	NP_048858	A502L	139.43	2.96E-32	89%	94%	23--95	23--95
B671L	279393--278557	279	32,354	9.57		No Hit Found								1	NP_048859	contains prokaryotic membrane lipoprotein lipid attachment site	486.88	2.87E-136	83%	92%	1--276	24--299
B672L	280914--279460	485	55,252	5.87		No Hit Found								1	NP_048861	A505L	794.65	0.00E+00	79%	88%	1--482	1--480
B677R	281091--281477	129	15,213	7.88	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).	41.23	1.68E-04	34%	57%	80--129	1--45	1	XP_793752	PREDICTED: similar to tripartite motif-containing 33, partial	52.37	4.64E-06	37%	58%	70--127	8--64
					2	smart00184	RING, Ring finger; E3 ubiquitin-protein ligase activity is intrinsic to the RING domain of c-Cbl and is likely to be a general function of this domain; Various RING fingers exhibit binding activity towards E2 ubiquitin-conjugating enzymes (Ubc' s) .	36.30	4.77E-03	45%	60%	81--125	1--41	2	AAH85684	Tripartite motif protein 50	50.83	1.35E-05	35%	59%	69--129	3--60
														3	AAL91073	tripartite motif protein 50	50.83	1.35E-05	35%	59%	69--129	3--60
														4	XP_791398	PREDICTED: similar to Transcription intermediary factor 1-gamma (TIF1-gamma) (RET-fused gene 7 protein) (Rfg7 protein) (Tripartite motif protein 33), partial	50.45	1.76E-05	38%	58%	65--127	5--64
														5	XP_799164	PREDICTED: similar to tripartite motif-containing 33, partial	50.45	1.76E-05	36%	56%	70--127	8--64
														6	XP_798828	PREDICTED: similar to tripartite motif-containing 59	50.06	2.30E-05	36%	58%	70--127	8--64
														7	AAN14328	CG10981-PB, isoform B	49.68	3.01E-05	43%	56%	79--129	257--302
														8	AAF51944	CG10981-PA, isoform A	49.68	3.01E-05	43%	56%	79--129	264--309
														9	XP_691234	PREDICTED: similar to Tripartite motif protein 32 (Zinc-finger protein HT2A) (72 kDa Tat-interacting protein)	49.68	3.01E-05	29%	47%	36--127	262--346
														10	XP_792036	PREDICTED: similar to Transcription intermediary factor 1-gamma (TIF1-gamma) (RET-fused gene 7 protein) (Rfg7 protein) (Tripartite motif protein 33)	49.29	3.93E-05	34%	58%	70--127	8--64
B680L	282061--281747	105	12,432	10.87		No Hit Found								1	NP_048876	A520L	170.63	1.20E-41	86%	92%	10--105	5--100

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
B681L	283318--282269	350	39,717	8.02	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	145.45	6.86E-36	32%	46%	3--264	1--270	1	AAV84097	CviPII m5C DNA methyltransferase	683.72	0.00E+00	92%	97%	1--350	14--363
					2	pfam00145	DNA methylase, C-5 cytosine-specific DNA methylase..	132.74	4.86E-32	37%	49%	3--179	1--178	2	NP_048873	M.CviAII cytosine DNA methyltransferase	319.32	1.13E-85	47%	63%	3--344	2--342
					3	COG0270	Dcm, Site-specific DNA methylase [DNA replication, recombination, and repair].	108.24	1.09E-24	22%	41%	1--264	2--279	3	AAC64006	cytosine methyltransferase	317.78	3.29E-85	46%	61%	1--350	1--362
														4	AAC55063	cytosine methyltransferase	273.09	9.29E-72	40%	57%	1--350	1--366
														5	NP_049039	nonfunctional M.CviAV cytosine DNA methyltransferase	268.86	1.75E-70	40%	56%	1--350	1--366
														6	NP_048886	M.CviAIV cytosine DNA methyltransferase	258.07	3.09E-67	41%	58%	1--332	1--332
														7	ZP_00783160	C-5 cytosine-specific DNA methylase	90.12	1.12E-16	31%	50%	3--168	1--175
														8	ZP_00874816	C-5 cytosine-specific DNA methylase	89.74	1.46E-16	31%	50%	3--168	1--175
														9	CAD47029	unknown	87.43	7.23E-16	33%	51%	3--162	1--169
														10	ZP_00393966	COG0270: Site-specific DNA methylase	86.27	1.61E-15	29%	45%	3--226	9--259
B684L	283920--283378	181	20,210	11.40		No Hit Found								1	NP_048877	contains Gln-rich, neutral zinc metallopeptidase, zinc binding region signature	261.92	6.09E-69	76%	80%	1--181	212--391
B685L	284553--283948	202	22,943	6.27		No Hit Found								1	NP_048877	contains Gln-rich, neutral zinc metallopeptidase, zinc binding region signature	298.13	1.00E-79	76%	85%	1--193	1--192
														2	YP_142679	metal-dependent hydrolase	59.31	7.82E-08	27%	48%	29--179	33--191
														3	NP_149599	136R	52.37	9.56E-06	36%	57%	98--160	98--163
B687R	284606--285109	168	18,788	9.68		No Hit Found								1	NP_048879	A523R	288.50	5.01E-77	83%	90%	4--168	7--171
B688L	286084--285122	321	36,796	8.00		No Hit Found								1	NP_048920	similar to Chlorella virus PBCV-1 ORF A154L, corresponds to GenBank Accession Number U42580	488.42	1.23E-136	65%	76%	5--321	4--350
														2	NP_048502	A154L	458.37	1.37E-127	66%	80%	3--321	35--347
														3	NP_048477	similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903	453.75	3.37E-126	60%	72%	1--321	1--356
														4	NP_077492	EsV-1-7	69.71	1.37E-10	29%	45%	4--209	40--261
B692R	286188--286625	146	16,278	7.30		No Hit Found								1	NP_048882	A526R	228.79	3.60E-59	84%	91%	18--146	1--129
B694R	286652--286936	95	11,330	11.07		No Hit Found								1	NP_048883	A527R	100.52	1.52E-20	85%	93%	1--60	5--63
B696R	287026--287250	75	8,207	9.39		No Hit Found									No Hit Found	No Hit Found						
B697R	287217--288323	369	41,930	6.98	1	pfam00145	DNA methylase, C-5 cytosine-specific DNA methylase..	152.00	7.22E-38	28%	44%	3--244	1--246	1	NP_048886	M.CviAIV cytosine DNA methyltransferase	591.27	1.68E-167	85%	92%	1--331	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	143.91	2.38E-35	28%	42%	3--287	1--269	2	NP_048873	M.CviAII cytosine DNA methyltransferase	421.39	2.30E-116	58%	73%	2--343	1--342
					3	COG0270	Dcm, Site-specific DNA methylase [DNA replication, recombination, and repair].	107.86	1.63E-24	32%	46%	3--162	4--169	3	AAC64006	cytosine methyltransferase	298.13	2.93E-79	46%	59%	1--343	1--356
														4	AAC55063	cytosine methyltransferase	291.58	2.74E-77	43%	59%	1--343	1--360
														5	NP_049039	nonfunctional M.CviAV cytosine DNA methyltransferase	290.81	4.68E-77	43%	58%	1--344	1--361
														6	AAV84097	CviPII m5C DNA methyltransferase	280.80	4.84E-74	44%	60%	1--343	14--357
														7	ABA25040	C-5 cytosine-specific DNA methylase	93.59	1.10E-17	35%	50%	6--157	5--160
														8	AAN07998	gp80	88.58	3.53E-16	33%	49%	3--157	1--159
														9	NP_945098	gp67	87.81	6.02E-16	29%	48%	1--160	4--190
														10	BAB77127	site-specific DNA-methyltransferase	87.81	6.02E-16	34%	48%	6--157	5--160
B699L	288473--288240	78	8,752	7.69	1	cd01803	Ubiquitin, Ubiquitin (includes Ubq/RPL40e and Ubq/RPS27a fusions as well as homopolymeric multiubiquitin protein chains).	131.27	1.25E-31	93%	97%	1--76	1--76	1	AAQ07453	ubiquitin	144.44	9.07E-34	93%	97%	1--77	305--381
					2	pfam00240	ubiquitin, Ubiquitin family. This family contains a number of ubiquitin-like proteins: SUMO (smt3 homologue), Nedd8, Elongin B, Rub1..	93.77	2.83E-20	71%	84%	1--74	1--74	2	CAA52290	polyubiquitin	142.51	3.44E-33	94%	97%	1--76	1--76
					3	smart00213	UBQ, Ubiquitin homologues; Ubiquitin-mediated proteolysis is involved in the regulated turnover of proteins required for controlling cell cycle progression .	89.54	4.55E-19	68%	80%	1--72	1--72	3	CAA43216	ubiquitin extension protein (UbCEP52)	142.51	3.44E-33	94%	97%	1--76	1--76
					4	cd01806	Nedd8, Nedd8 (also known as Rub1) has a single conserved ubiquitin-like domain that is part of a protein modification pathway similar to that of ubiquitin. Nedd8 modifies a family of molecular scaffold proteins called cullins that are responsible for assembling the ROC1/Rbx1 RING-based E3 ubiquitin ligases, of which several play a direct role in tumorigenesis..	87.71	1.65E-18	55%	77%	1--76	1--76	4	BAE48510	polyubiquitin	142.51	3.44E-33	94%	97%	1--76	70--145
					5	cd01769	UBL, UBLs function by remodeling the surface of their target proteins, changing their target's half-life, enzymatic activity, protein-protein interactions, subcellular localization or other properties. At least 10 different ubiquitin-like modifications exist in mammals, and attachment of different ubls to a target leads to different biological consequences. Ubl-conjugation cascades are initiated by activating enzymes, which also coordinate the ubls with their downstream pathways	79.18	6.39E-16	63%	76%	4--72	1--69	5	P14624	Ubiquitin	142.51	3.44E-33	94%	97%	1--76	1--76
					6	cd01802	AN1_N, AN1 (also known as ANUBL1 and RSD-7) is ubiquitin-like protein with a testis-specific expression in rats that has an N-terminal ubiquitin-like domain and a C-terminal zinc-binding domain. Unlike ubiquitin polyproteins and most ubiquitin fusion proteins, the N-terminal ubiquitin-like domain of An1 does not undergo proteolytic processing. The function of AN1 is unknown	76.96	3.06E-15	49%	68%	1--76	28--103	6	AAC49014	ubiquitin	142.12	4.50E-33	92%	97%	1--77	305--381
					7	cd01807	GDX_N, GDX contains an N-terminal ubiquitin-like domain as well as an uncharacterized c-terminal domain. The function of GDX is unknown..	60.04	4.03E-10	37%	63%	1--72	1--72	7	CAA40325	hexaubiquitin protein	141.74	5.88E-33	92%	97%	1--77	381--457
					8	cd01809	Scythe_N, Scythe protein (also known as Bat3) is an apoptotic regulator that is highly conserved in eukaryotes and contains a ubiquitin-like domain near its N-terminus. Scythe binds reaper, a potent apoptotic inducer, and Scythe/Reaper are thought to signal apoptosis, in part through regulating the folding and activity of apoptotic signaling molecules	59.45	5.07E-10	45%	69%	1--72	1--72	8	AAF31707	polyubiquitin	141.74	5.88E-33	92%	97%	1--77	139--215
					9	cd01805	RAD23_N, RAD23 belongs to a family of adaptor molecules having affinity for both the proteasome and ubiquitinylated proteins and thought to shuttle these ubiquitinylated proteins to the proteasome for destruction. RAD23 interacts with ubiquitin through its C-terminal ubiquitin-associated domains (UBA) and with the proteasome through its N-terminal ubiquitin-like domain (UBL)	59.46	5.51E-10	36%	66%	1--72	1--74	9	AAL25813	polyubiquitin	141.35	7.67E-33	93%	97%	1--76	2--77
					10	cd01798	parkin_N, parkin_N parkin protein is a RING-type E3 ubiquitin ligase with an amino-terminal ubiquitin-like (Ubl) domain and an RBR signature consisting of two RING finger domains separated by an IBR/DRIL domain. Naturally occurring mutations in parkin are thought to cause the disease AR_JP (autosomal-recessive juvenile parkinsonism). Parkin binds the Rpn10 subunit of 26S proteasomes through its Ubl domain..	54.50	1.86E-08	35%	65%	3--72	1--70	10	AAX40652	polyubiquitin	141.35	7.67E-33	93%	97%	1--76	77--152
B700L	288718--288524	65	7,519	6.52		No Hit Found								1	NP_048887	A531L	98.98	4.39E-20	75%	81%	1--65	1--66
B701L	288990--288754	79	8,694	10.13		No Hit Found								1	NP_048888	A532L	93.97	1.40E-18	58%	60%	1--79	1--79

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
B702L	290554--289259	432	50,008	10.75	1	COG0675	COG0675, Transposase and inactivated derivatives [DNA replication, recombination, and repair].	86.67	3.41E-18	24%	39%	37--423	1--356	1	AAU06281	putative transposase	782.33	0.00E+00	93%	96%	24--432	32--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	73.35	3.30E-14	41%	54%	345--412	1--69	2	NP_048981	similar to Synechocystis transposase, corresponds to GenBank Accession Number D90909	731.48	0.00E+00	86%	92%	24--432	26--433
					3	pfam01385	Transposase_2, Probable transposase. This family includes IS891, IS1136 and IS1341. .	46.06	6.86E-06	22%	41%	42--314	1--261	3	YP_142458	putative transposase	113.24	1.66E-23	28%	42%	10--412	104--535
														4	AAS54227	AGL264Wp	103.61	1.32E-20	26%	45%	41--414	71--453
														5	ABA24789	Transposase, IS891/IS1136/IS1341	98.60	4.24E-19	27%	42%	20--414	123--533
														6	BAB78230	transposase	98.60	4.24E-19	27%	42%	20--414	79--489
														7	YP_238637	ORF021	87.81	7.48E-16	27%	46%	40--409	3--364
														8	ZP_00158267	COG0675: Transposase and inactivated derivatives	85.11	4.85E-15	28%	42%	114--414	43--359
														9	BAE47830	putative IS transposase (OrfB)	82.80	2.41E-14	23%	44%	42--409	6--384
														10	ZP_00766186	Transposase, IS605 OrfB	79.72	2.04E-13	28%	43%	36--421	2--367
B704L	290797--290591	69	8,738	10.15		No Hit Found								1	ZP_00673812	hypothetical protein TeryDRAFT_2410	50.45	1.77E-05	40%	66%	4--57	214--265
														2	NP_701794	hypothetical protein PFL2170c	48.52	6.71E-05	44%	61%	4--58	585--643
B705R	290861--292450	530	57,739	4.46		No Hit Found								1	NP_048889	A533R	659.45	0.00E+00	86%	90%	1--374	1--374
														2	AAA66400	unknown protein	657.52	0.00E+00	86%	90%	1--374	1--374
														3	NP_048890	a534R	186.42	2.02E-45	95%	97%	441--530	16--105
B710L	292665--292453	71	8,274	4.48		No Hit Found								1	NP_048891	A535L	140.58	1.29E-32	90%	95%	1--71	1--71
B711L	294182--292998	395	46,128	10.49	1	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	56.78	3.35E-09	34%	46%	316--388	1--69	1	AAU06281	putative transposase	204.14	6.33E-51	31%	51%	12--389	52--420
					2	COG0675	COG0675, Transposase and inactivated derivatives [DNA replication, recombination, and repair].	55.09	1.29E-08	20%	40%	74--388	62--345	2	NP_048981	similar to Synechocystis transposase, corresponds to GenBank Accession Number D90909	200.68	7.00E-50	33%	53%	66--389	102--414
					3	pfam01385	Transposase_2, Probable transposase. This family includes IS891, IS1136 and IS1341. .	36.43	5.47E-03	22%	41%	74--301	63--277	3	YP_143208	putative transposase	70.48	1.09E-10	27%	41%	90--390	231--524
														4	YP_143124	putative transposase	70.48	1.09E-10	27%	41%	90--390	224--517
														5	YP_142433	putative transposase	68.55	4.15E-10	26%	40%	90--390	224--517
														6	YP_142458	putative transposase	66.24	2.06E-09	25%	38%	181--388	317--535
														7	AAS54227	AGL264Wp	62.39	2.98E-08	24%	42%	57--393	137--456
														8	CAJ31329	insertion sequence IS606 transposase homolog A	61.62	5.08E-08	29%	51%	275--388	304--421
														9	BAD76101	transposase	60.85	8.66E-08	26%	40%	189--390	169--372
														10	ZP_00370996	ISCco1, transposase orfB	60.46	1.13E-07	31%	50%	276--385	277--390
B712R	293043--293300	86	9,916	10.96		No Hit Found									No Hit Found	No Hit Found						
B714L	294931--294134	266	29,813	10.81		No Hit Found									No Hit Found	No Hit Found						
B715L	294926--294243	228	25,945	9.23	1	COG2452	COG2452, Predicted site-specific integrase-resolvase [DNA replication, recombination, and repair].	140.82	1.87E-34	42%	56%	15--208	4--190	1	YP_143125	putative resolvase	113.62	4.52E-24	41%	61%	10--152	2--140
					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..	66.04	5.53E-12	30%	48%	75--201	2--132	2	YP_142434	putative resolvase	110.15	4.99E-23	39%	60%	10--152	2--140
					3	cd01104	HTH_MlrA, Helix-turn-helix transcription regulator MlrA (merR-like regulator A). The MlrA protein, also known as YehV, has been shown to control cell-cell aggregation by co-regulating the expression of curli and extracellular matrix production in Escherichia coli and Salmonella typhimurium. Its close homolog, CarA from Myxococcus xanthus, is involved in activation of the carotenoid biosynthesis genes by light. These proteins belong to the MERR superfamily of transcription regulators that promote expression of several stress regulon genes by reconfiguring the spacer between the -35 and -10 promoter elements. Their conserved N-terminal domains contain predicted HTH (helix-turn-helix) motives that mediate DNA binding, while the dissimilar C-terminal domains bind specific coactivator molecules.	41.85	1.12E-04	31%	42%	16--98	5--83	3	YP_142457	putative resolvase	105.15	1.61E-21	37%	56%	11--161	3--149
														4	AAK41573	First ORF in transposon ISC1904	96.29	7.46E-19	42%	59%	22--159	11--142
														5	BAD84844	predicted site-specific integrase/resolvase	95.90	9.74E-19	35%	57%	6--160	3--153
														6	AAK43255	First ORF in transposon ISC1904	95.90	9.74E-19	40%	58%	15--159	4--142
														7	AAK41585	First ORF in transposon ISC1904	95.90	9.74E-19	42%	59%	15--154	4--137
														8	ZP_00683312	regulatory protein, MerR:Resolvase, N-terminal	95.13	1.66E-18	38%	57%	11--151	10--150
														9	CAB49329	Resolvase related protein	95.13	1.66E-18	36%	60%	11--156	6--151
														10	AAK42026	First ORF in transposon ISC1904	94.74	2.17E-18	41%	58%	15--159	4--142
B718L	295930--294971	320	37,235	7.16		No Hit Found								1	NP_048711	A354R	86.27	1.41E-15	33%	48%	117--294	4--197
														2	NP_048779	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	71.25	4.68E-11	27%	44%	81--297	77--314
														3	NP_048435	A87R	69.32	1.78E-10	26%	40%	33--263	104--390
														4	NP_695069	putative endodeoxyribonuclease	50.83	6.54E-05	31%	51%	20--131	9--118
B719L	296234--295992	81	9,172	11.05		No Hit Found								1	NP_048892	A536L	78.57	6.05E-14	66%	74%	23--81	19--73
B720L	297005--296262	248	28,197	9.20		No Hit Found								1	NP_048723	A366L	216.85	4.47E-55	61%	79%	73--239	86--252
B723L	297952--297176	259	30,657	6.19		No Hit Found								1	NP_048893	A537L	323.55	3.69E-87	60%	75%	2--258	9--263
B724R	298088--298606	173	19,989	10.30		No Hit Found								1	NP_048895	A539R	291.97	4.88E-78	80%	91%	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
B725L	302168--298596	1191	121,563	6.40		No Hit Found								1	NP_048896	A540L	1056.59	0.00E+00	48%	60%	81--1186	1--1175
														2	AAA66404	unknown protein	371.32	1.14E-100	64%	73%	885--1186	1--291
														3	ZP_00532602	Hep_Hag	147.52	2.67E-33	28%	41%	257--799	285--865
														4	AAX44675	possible T4-like proximal tail fiber	139.04	9.48E-31	28%	39%	280--817	147--706
														5	ZP_00950302	outer membrane protein	134.04	3.05E-29	25%	39%	256--920	201--797
														6	NP_048899	A543L	125.56	1.08E-26	79%	87%	1--78	1--78
														7	ZP_00533161	Hep_Hag	119.78	5.95E-25	24%	35%	251--804	104--752
														8	CAI76745	nucleoporin, putative	102.06	1.28E-19	26%	39%	361--799	190--610
														9	CAH35630	putative outer membrane protein	98.60	1.42E-18	25%	39%	271--790	389--934
														10	ABB10557	outer membrane protein, Haemagluttinin-like	97.44	3.16E-18	20%	36%	230--1016	58--2402
B734R	302287--303255	323	37,263	7.25	1	pfam01068	DNA_ligase_A_M, ATP dependent DNA ligase domain. This domain belongs to a more diverse superfamily, including pfam01331 and pfam01653..	93.83	2.24E-20	24%	42%	31--213	1--201	1	1P8L_A	Chain A, New Crystal Structure Of Chlorella Virus Dna Ligase- Adenylate	509.22	6.83E-143	80%	89%	20--322	1--303
					2	COG1793	CDC9, ATP-dependent DNA ligase [DNA replication, recombination, and repair].	66.18	5.94E-12	21%	38%	31--318	119--417	2	NP_048900	PBCV-1 DNA ligase	507.68	1.99E-142	81%	90%	26--322	1--297
														3	1FVI_A	Chain A, Crystal Structure Of Chlorella Virus Dna Ligase-Adenylate	502.67	6.39E-141	81%	90%	28--322	2--296
														4	ABA50091	PBCV-1 DNA ligase	180.64	5.56E-44	38%	56%	30--322	9--305
														5	CAI10149	DNA ligase, ATP-dependent	82.80	1.58E-14	28%	46%	48--318	69--301
														6	YP_189099	DNA ligase, ATP-dependent	75.10	3.29E-12	25%	41%	52--322	160--427
														7	ZP_00838729	ATP-dependent DNA ligase	74.71	4.30E-12	27%	48%	49--318	88--319
														8	AAZ12125	DNA ligase, putative	73.17	1.25E-11	25%	39%	49--322	215--509
														9	AAZ80807	ATP-dependent DNA ligase	72.40	2.13E-11	26%	46%	48--321	53--284
														10	ZP_00550374	ATP-dependent DNA ligase	72.02	2.79E-11	23%	44%	48--315	48--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
B736L	304417--303203	405	45,549	6.51	1	COG0438	RfaG, Glycosyltransferase [Cell envelope biogenesis, outer membrane].	63.50	3.20E-11	20%	38%	1--392	2--379	1	NP_048902	similar to Streptomyces glucosyltransferase protein, corresponds to GenBank Accession Number AB005901	532.33	1.05E-149	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	51.87	9.87E-08	26%	46%	178--369	9--172	2	ZP_00687600	hypothetical protein BambDRAFT_3460	57.77	7.61E-07	31%	46%	185--320	54--179
					3	COG0297	GlgA, Glycogen synthase [Carbohydrate transport and metabolism].	36.01	6.08E-03	25%	42%	173--307	283--402	3	NP_835600	putative glycosyltransferases	57.77	7.61E-07	31%	47%	173--313	194--342
														4	AAZ58719	putative Glycosyltransferase	55.45	3.78E-06	25%	47%	126--311	581--756
B738L	305824--304451	458	53,265	8.01	1	COG0553	HepA, Superfamily II DNA/RNA helicases, SNF2 family [Transcription / DNA replication, recombination, and repair].	150.30	2.32E-37	28%	45%	5--433	337--846	1	NP_048904	similar to Caenorhabditis transcription activator, corresponds to Swiss-Prot Accession Number P41877	807.36	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).	122.70	5.82E-29	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	1.59E-14	20%	34%	2--196	4--202	3	AAK75613	Snf2 family protein	173.33	1.46E-41	29%	47%	7--439	83--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	2.50E-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	83--1016
					5	COG1061	SSL2, DNA or RNA helicases of superfamily II [Transcription / DNA replication, recombination, and repair].	66.27	4.78E-12	21%	41%	1--412	31--394	5	AAX71394	phage-related DNA helicase	171.79	4.25E-41	27%	49%	10--444	89--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..	63.95	2.85E-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	89--1025
					7	smart00490	HELICc, helicase superfamily c-terminal domain; .	62.94	5.66E-11	27%	46%	314--395	3--82	7	AAZ50907	SWF/SNF family helicase	171.40	5.55E-41	27%	49%	10--444	89--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	1.20E-10	29%	46%	317--395	2--78	8	NP_296465	helicase, Snf2 family	171.01	7.25E-41	30%	49%	2--437	12--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	2.39E-08	17%	35%	29--164	2--144	9	ZP_00519830	SNF2-related:Helicase, C-terminal:SWIM Zn-finger	171.01	7.25E-41	29%	47%	3--441	15--1065
					10	COG0513	SrmB, Superfamily II DNA and RNA helicases [DNA replication, recombination, and repair / Transcription / Translation, ribosomal structure and biogenesis].	48.62	9.16E-07	19%	39%	280--446	255--420	10	AAL97089	putative SNF helicase	170.63	9.46E-41	27%	49%	10--444	88--1024
B741L	306328--305906	141	14,861	4.73	1	pfam00692	dUTPase, dUTPase. dUTPase hydrolyses dUTP to dUMP and pyrophosphate..	132.28	7.05E-32	52%	67%	12--140	2--129	1	NP_048907	similar to tomato dUTP pyrophosphatase, corresponds to GenBank Accesssion Number S40549	218.01	6.49E-56	83%	92%	1--131	1--131
					2	COG0717	Dcd, Deoxycytidine deaminase [Nucleotide transport and metabolism].	57.20	2.45E-09	31%	49%	33--118	74--158	2	AAW51452	deoxyuridine triphosphatase	211.85	4.65E-54	80%	90%	1--131	1--131
														3	AAW51453	deoxyuridine triphosphatase	210.31	1.35E-53	81%	90%	1--131	1--131
														4	EAL88001	dUTPase	154.84	6.74E-37	58%	78%	4--131	113--240
														5	XP_657875	hypothetical protein AN0271.2	152.14	4.37E-36	55%	79%	3--131	69--197
														6	BAE55800	unnamed protein product	149.06	3.70E-35	57%	77%	4--131	57--184
														7	AAB22611	deoxyuridine triphosphatase; dUTPase; P18	145.21	5.34E-34	56%	77%	6--131	34--159
														8	NP_190278	dUTP diphosphatase/ hydrolase	144.82	6.98E-34	57%	76%	6--131	31--156
														9	XP_469212	putative deoxyuridine triphosphatase	144.82	6.98E-34	55%	76%	4--131	88--215
														10	EAA67244	hypothetical protein FG00904.1	144.05	1.19E-33	57%	75%	3--131	31--159
B743R	306447--307397	317	36,299	9.43	1	pfam00352	TBP, Transcription factor TFIID (or TATA-binding protein, TBP)..	37.03	2.88E-03	26%	56%	187--283	3--85	1	NP_048908	similar to Sulfolobus TATA-binding protein, corresponds to GenBank Accession Number S55311'	456.45	5.09E-127	80%	89%	48--317	1--270
B744L	308889--307378	504	59,723	7.03	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.	114.56	1.38E-26	27%	49%	200--385	2--185	1	NP_048910	similar to MesJ cell cycle protein	932.17	0.00E+00	88%	92%	1--497	1--497
					2	pfam01171	ATP_bind_3, PP-loop family. This family of proteins belongs to the PP-loop superfamily..	108.41	1.07E-24	31%	53%	204--385	6--186	2	T18059	hypothetical protein A557L - Chlorella virus PBCV-1	198.36	4.84E-49	85%	93%	24--124	12--112
					3	COG0037	MesJ, Predicted ATPase of the PP-loop superfamily implicated in cell cycle control [Cell division and chromosome partitioning].	93.37	3.11E-20	25%	45%	184--385	6--210	3	T18058	hypothetical protein A556L - Chlorella virus PBCV-1	187.96	6.54E-46	86%	89%	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.	71.04	1.84E-13	26%	44%	200--371	2--180	4	CAD84933	conserved hypothetical protein	75.49	4.73E-12	25%	49%	195--400	24--227
					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..	38.67	1.15E-03	28%	39%	200--365	1--148	5	ZP_00144172	Cell cycle protein MesJ	75.49	4.73E-12	24%	45%	176--484	8--304
														6	AAZ50630	tRNA(Ile)-lysidine synthetase	75.49	4.73E-12	28%	45%	200--401	23--211
														7	AAT86146	tRNA(Ile)-lysidine synthetase	75.49	4.73E-12	28%	45%	200--401	23--211
														8	AAX71126	tRNA(Ile)-lysidine synthetase	75.49	4.73E-12	28%	45%	200--401	23--211
														9	ZP_00366567	COG0037: Predicted ATPase of the PP-loop superfamily implicated in cell cycle control	75.49	4.73E-12	28%	45%	200--401	23--211
														10	AAG12421	MesJ	75.10	6.18E-12	27%	46%	200--403	44--246
B747L	309892--308966	309	37,316	9.56		No Hit Found								1	NP_048621	A267L	188.73	1.90E-46	41%	55%	55--289	11--250
														2	NP_048846	Lys-, Glu-rich	117.09	7.01E-25	27%	48%	5--275	21--280
														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	109.38	1.46E-22	26%	46%	3--305	23--305
														4	YP_142777	unknown	97.06	7.50E-19	24%	46%	2--269	180--443
B748L	311175--309976	400	45,367	5.20	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 NY2A the major capsid protein is a glycoprotein.	425.49	4.16E-120	40%	57%	1--395	2--442	1	NP_048914	similar to Chlorella virus PBCV-1 ORF A11L, corresponds to GenBank Accession Number U42580	768.46	0.00E+00	93%	97%	1--400	1--400
														2	NP_048359	contains aminoacyl-tRNA synthetase class-II signature	417.93	2.84E-115	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	264.62	4.02E-69	38%	54%	2--400	3--432
														5	NP_048787	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	246.51	1.13E-63	37%	53%	2--400	3--437
														6	BAA76601	major capsid protein MCP1	241.89	2.79E-62	36%	52%	2--400	3--437
														7	BAA76600	major capsid protein	241.12	4.76E-62	36%	53%	2--400	3--436
														8	BAA22198	major capsid protein Vp54	237.65	5.27E-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	221.09	5.10E-56	36%	52%	24--400	1--413
														10	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	220.71	6.66E-56	36%	52%	24--400	1--413
B751L	311864--311232	211	23,214	10.31		No Hit Found								1	NP_048915	A559L	274.63	1.30E-72	70%	81%	1--211	1--213

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
B753L	313024--311885	380	43,547	8.86	1	smart00497	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished). .	48.21	1.24E-06	35%	54%	253--305	1--53	1	NP_048621	A267L	152.91	1.58E-35	31%	49%	7--300	31--311
					2	pfam07453	NUMOD1, NUMOD1 domain..	36.56	4.09E-03	36%	61%	253--286	1--34	2	NP_048846	Lys-, Glu-rich	99.75	1.59E-19	33%	49%	60--229	128--299
														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	98.21	4.62E-19	29%	45%	2--229	80--299
														4	YP_142777	unknown	87.04	1.06E-15	25%	46%	4--225	241--472
														5	CAA25939	unnamed protein product	76.26	1.88E-12	30%	48%	243--368	81--225
														6	CAA25938	unnamed protein product	76.26	1.88E-12	30%	48%	243--368	118--262
														7	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	63.93	9.65E-09	52%	66%	243--304	162--224
														8	NP_048671	A315L	58.15	5.29E-07	44%	67%	311--374	180--246
B756L	315055--313109	649	71,424	10.04	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	50.75	2.41E-07	22%	37%	252--317	33--98	1	BAA11342	DNA binding protein	613.22	8.59E-174	82%	87%	315--648	312--647
					2	pfam05335	DUF745, Protein of unknown function (DUF745). This family consists of several uncharacterised Drosophila melanogaster proteins of unknown function..	48.47	1.13E-06	25%	45%	117--271	21--187	2	NP_048917	similar to Chlorella virus CVK2 DNA binding protein, corresponds to GenBank Accession Number D78305	610.14	7.27E-173	81%	86%	315--649	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	48.00	1.75E-06	22%	45%	58--275	609--828	3	NP_048921	A565R	206.45	2.42E-51	97%	97%	315--400	377--462
					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..	47.81	1.87E-06	28%	39%	95--265	102--260	4	BAA11343	DNA binding protein	204.53	9.20E-51	96%	96%	315--400	369--454
					5	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	44.75	1.58E-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	102.45	4.92E-20	51%	59%	315--397	85--167
					6	COG1566	EmrA, Multidrug resistance efflux pump [Defense mechanisms].	43.79	2.89E-05	20%	41%	34--209	14--203	6	NP_048735	A378L	76.64	2.89E-12	34%	42%	325--400	139--245
					7	pfam05616	Neisseria_TspB, Neisseria meningitidis TspB protein. This family consists of several Neisseria meningitidis TspB virulence factor proteins..	41.27	1.66E-04	35%	46%	279--331	335--387	7	BAB19127	vAL-1	70.09	2.71E-10	30%	44%	431--623	152--328
					8	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	41.16	1.77E-04	23%	43%	83--235	21--171	8	BAA83789	alginate lyase	67.40	1.76E-09	29%	44%	431--623	136--312
					9	COG0845	AcrA, Membrane-fusion protein [Cell envelope biogenesis, outer membrane].	40.87	2.18E-04	17%	34%	31--262	22--238	9	NP_048562	PBCV-1 alginate lyase	66.63	2.99E-09	28%	43%	431--623	124--300
B759R	315098--317122	675	73,893	7.46	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	54.21	2.35E-08	33%	36%	338--393	68--123	1	BAA11343	DNA binding protein	817.77	0.00E+00	77%	81%	23--550	23--540
					2	COG0419	SbcC, ATPase involved in DNA repair [DNA replication, recombination, and repair].	48.51	1.28E-06	18%	40%	34--255	248--465	2	NP_048921	A565R	578.56	2.46E-163	89%	95%	23--337	23--337
					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	46.85	3.24E-06	24%	43%	33--180	680--838	3	BAA11342	DNA binding protein	222.63	3.43E-56	98%	98%	383--475	312--404
					4	pfam05616	Neisseria_TspB, Neisseria meningitidis TspB protein. This family consists of several Neisseria meningitidis TspB virulence factor proteins..	46.28	5.88E-06	33%	39%	320--384	310--377	4	NP_048917	similar to Chlorella virus CVK2 DNA binding protein, corresponds to GenBank Accession Number D78305	219.94	2.22E-55	96%	97%	383--475	314--406
					5	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	43.98	2.50E-05	38%	51%	338--380	78--123	5	NP_048741	Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	120.17	2.40E-25	34%	40%	286--474	9--176
					6	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..	43.81	3.07E-05	19%	39%	34--240	97--305	6	NP_048735	A378L	76.64	3.04E-12	34%	42%	393--468	139--245
					7	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	4.31E-05	35%	39%	339--379	97--143	7	CAA64974	QI74 protein	58.15	1.12E-06	38%	48%	390--472	501--579
					8	COG1196	Smc, Chromosome segregation ATPases [Cell division and chromosome partitioning].	41.94	1.13E-04	14%	35%	31--221	345--553	8	AAF58977	CG8055-PA	52.37	6.14E-05	27%	43%	30--227	11--221
					9	pfam00430	ATP-synt_B, ATP synthase B/B' CF(0). Part of the CF(0) (base unit) of the ATP synthase. The base unit is thought to translocate protons through membrane (inner membrane in mitochondria, thylakoid membrane in plants, cytoplasmic membrane in bacteria). The B subunits are thought to interact with the stalk of the CF(1) subunits. This domain should not be confused with the ab CF(1) proteins (in the head of the ATP synthase) which are found in pfam00006	41.00	1.94E-04	19%	40%	4--136	3--132	9	XP_645158	hypothetical protein DDB0216970	51.99	8.02E-05	23%	41%	36--229	135--1649
B761L	317559--317125	145	16,641	10.22		No Hit Found								1	NP_048923	A567L	206.45	1.92E-52	67%	82%	1--145	1--152
B762L	318099--317563	179	21,601	5.68		No Hit Found								1	NP_048924	A568L	271.94	5.69E-72	73%	84%	1--176	1--176
B763L	318503--318126	126	14,497	5.07		No Hit Found								1	NP_048926	A570L	229.95	1.64E-59	82%	89%	1--125	1--121
B765R	318578--318922	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
B766R	318937--319479	181	20,767	7.07		No Hit Found								1	NP_048928	A572R	343.20	2.08E-93	86%	95%	1--180	1--180
B767L	320220--319486	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	3.66E-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	5.03E-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

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
														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
B769R	320318--321403	362	41,127	7.83	1	pfam00145	DNA methylase, C-5 cytosine-specific DNA methylase..	199.38	4.54E-52	31%	45%	3--331	1--316	1	AAC64006	cytosine methyltransferase	753.05	0.00E+00	99%	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	190.51	2.27E-49	35%	48%	3--241	1--250	2	AAC55063	cytosine methyltransferase	507.29	3.09E-142	65%	78%	1--362	1--366
					3	COG0270	Dcm, Site-specific DNA methylase [DNA replication, recombination, and repair].	121.34	1.45E-28	26%	41%	1--279	2--280	3	NP_049039	nonfunctional M.CviAV cytosine DNA methyltransferase	503.44	4.46E-141	64%	77%	1--362	1--366
														4	AAV84097	CviPII m5C DNA methyltransferase	326.25	9.76E-88	46%	62%	1--362	14--363
														5	NP_048873	M.CviAII cytosine DNA methyltransferase	297.75	3.72E-79	45%	58%	3--356	2--342
														6	NP_048886	M.CviAIV cytosine DNA methyltransferase	276.56	8.87E-73	46%	57%	1--344	1--332
														7	CAD80133	gp9.1	102.83	1.75E-20	35%	50%	6--163	4--162
														8	BAD65383	site-specific DNA-methyltransferase	97.83	5.64E-19	33%	48%	3--198	1--190
														9	ZP_00874816	C-5 cytosine-specific DNA methylase	94.36	6.24E-18	35%	49%	3--158	1--165
														10	AAC98421	methyl transferase	93.97	8.15E-18	34%	52%	3--158	1--165
B771L	321914--321411	168	19,008	8.64		No Hit Found								1	NP_048931	A575L	313.92	1.11E-84	89%	95%	1--168	1--168
B772L	322374--321976	133	15,329	11.27		No Hit Found								1	NP_048933	A577L	151.37	7.46E-36	84%	92%	29--110	2--83
B774R	322209--323243	345	40,186	8.48	1	pfam02086	MethyltransfD12, D12 class N6 adenine-specific DNA methyltransferase..	180.92	1.48E-46	35%	53%	89--328	1--253	1	NP_048937	PBCV-1 M.CviAI methylase	478.40	1.42E-133	86%	92%	82--345	1--265
					2	COG0338	Dam, Site-specific DNA methylase [DNA replication, recombination, and repair].	171.63	8.76E-44	41%	57%	82--331	1--256	2	ZP_00510571	N6 adenine-specific DNA methyltransferase, D12 class	188.35	2.94E-46	40%	58%	84--344	8--277
														3	EAM94529	N6 adenine-specific DNA methyltransferase, D12 class	180.26	8.01E-44	42%	58%	85--330	31--282
														4	NP_394115	Site-specific DNA methylase	170.63	6.35E-41	40%	60%	85--330	11--262
														5	ZP_00886307	DNA adenine methylase	169.86	1.08E-40	36%	58%	85--343	14--287
														6	CAC11782	probable site-specific DNA-methyltransferase (adenine-specific)	167.16	7.02E-40	40%	60%	87--330	1--250
														7	ABA22276	DNA adenine methylase	166.39	1.20E-39	41%	54%	85--334	13--267
														8	EAM93174	N6 adenine-specific DNA methyltransferase, D12 class	164.85	3.48E-39	35%	55%	85--343	12--279
														9	ZP_00789107	putative DNA adenine methylase	164.08	5.94E-39	38%	58%	84--330	14--271
														10	BAC09192	tlr1640	162.93	1.32E-38	36%	55%	80--329	4--259
B777L	324310--323252	353	40,156	7.67		No Hit Found								1	NP_048920	similar to Chlorella virus PBCV-1 ORF A154L, corresponds to GenBank Accession Number U42580	511.15	2.06E-143	67%	77%	1--353	1--350
														2	NP_048477	similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903	485.72	9.25E-136	64%	77%	1--353	1--356
														3	NP_048502	A154L	463.00	6.42E-129	62%	75%	4--353	3--347
														4	NP_077492	EsV-1-7	81.65	4.02E-14	33%	51%	32--183	30--190
														5	AAG37861	variant-specific surface protein VSP136-4	61.62	4.31E-08	26%	36%	5--209	571--784
														6	AAF69839	variant-specific surface protein VSP136b	60.85	7.35E-08	26%	36%	5--209	611--824
														7	AAA74587	cysteine rich protein	60.46	9.60E-08	26%	36%	5--209	51--264
														8	CAA64974	QI74 protein	50.45	9.93E-05	27%	38%	26--219	59--254
B781L	327552--324370	1061	120,400	8.57	1	smart00433	TOP2c, TopoisomeraseII; Eukaryotic DNA topoisomerase II, GyrB, ParE .	558.64	3.05E-160	35%	52%	50--620	1--594	1	NP_048939	PBCV-1 DNA topoisomerase II	1910.58	0.00E+00	88%	93%	3--1061	2--1061
					2	smart00434	TOP4c, DNA Topoisomerase IV; Bacterial DNA topoisomerase IV, GyrA, ParC .	445.40	4.16E-126	35%	53%	631--1055	1--456	2	AAU95770	topoisomerase II	1415.21	0.00E+00	66%	78%	1--1061	1--1056
					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	419.96	1.64E-118	35%	52%	650--1061	2--445	3	CAD25222	DNA TOPOISOMERASE II	969.53	0.00E+00	47%	66%	5--1059	8--1067
					4	COG0187	GyrB, Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), B subunit [DNA replication, recombination, and repair].	383.33	1.82E-107	28%	45%	2--618	8--626	4	CAA20107	SPBC1A4.03c	921.00	0.00E+00	44%	62%	1--1060	74--1197
					5	pfam00521	DNA topoisoIV, DNA gyrase/topoisomerase IV, subunit A..	305.55	4.83E-84	28%	45%	651--1060	1--436	5	CAA27857	unnamed protein product	919.46	0.00E+00	44%	62%	1--1060	20--1143
					6	COG0188	GyrA, Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), A subunit [DNA replication, recombination, and repair].	261.73	8.20E-71	25%	46%	630--1047	10--455	6	BAD86854	DNA topoisomerase II	919.07	0.00E+00	43%	61%	5--1058	03--1245
					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	2.88E-16	29%	43%	236--343	1--125	7	XP_467311	putative DNA topoisomerase II	919.07	0.00E+00	45%	60%	5--1060	32--1183
					8	COG1389	COG1389, DNA topoisomerase VI, subunit B [DNA replication, recombination, and repair].	41.07	2.06E-04	34%	52%	50--144	36--125	8	AAW40881	DNA topoisomerase II, putative	913.68	0.00E+00	45%	61%	5--1060	12--1233
					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..	36.89	3.76E-03	17%	40%	51--149	6--92	9	NP_189031	TOPII (TOPOISOMERASE II); ATP binding / DNA binding / DNA topoisomerase (ATP-hydrolyzing)	912.14	0.00E+00	45%	62%	5--1061	35--1166
														10	AAN85208	DNA topoisomerase II	911.75	0.00E+00	44%	62%	5--1061	27--1150
B791R	327785--327991	69	7,857	4.32		No Hit Found								1	NP_048948	A592R	111.69	6.46E-24	94%	94%	16--69	16--69
B792R	328107--328850	248	27,556	6.17		No Hit Found								1	NP_048949	a593R	72.79	1.04E-11	37%	47%	87--194	1--109
B795R	328889--329314	142	16,028	7.69	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	127.70	1.59E-30	39%	56%	5--133	2--131	1	NP_048952	similar to Vibrio fischeri dCMP deaminase, corresponds to Swiss-Prot Accession Number P33968	252.29	3.09E-66	83%	90%	1--142	1--142
					2	COG2131	ComEB, Deoxycytidylate deaminase [Nucleotide transport and metabolism].	103.12	4.30E-23	34%	55%	2--140	6--149	2	YP_223954	deoxycytidylate deaminase	123.25	2.17E-27	47%	64%	7--142	110--242
					3	pfam00383	dCMP_cyt_deam, Cytidine and deoxycytidylate deaminase zinc-binding region..	83.11	4.41E-17	39%	55%	5--111	4--100	3	ZP_00052863	COG2131: Deoxycytidylate deaminase	103.22	2.32E-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..	62.72	5.81E-11	30%	47%	15--111	9--92	4	BAC62535	putative deoxycytidylate deaminase	102.83	3.03E-21	41%	59%	4--142	6--141
					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 mechanism, activating a water molecule to form a hydroxide ion that performs a nucleophilic attack on the substrate. The functional enzyme is a homodimer. Cytosine deaminase catalyzes the deamination of cytosine to uracil and ammonia and is a member of the pyrimidine salvage pathway. Cytosine deaminase is found in bacteria and fungi but is not present in mammals; for this reason, the enzyme is currently of interest for antimicrobial drug design and gene therapy applications against tumors. Some members of this family are tRNA-specific adenosine deaminases that generate inosine at the first position of their anticodon (position 34) of specific tRNAs; this modification is thought to enlarge the codon recognition capacity during protein synthesis. Other members of the family are guanine deaminases which deaminate guanine to xanthine as part of the utilization of guanine	52.57	6.82E-08	36%	59%	26--111	19--93	5	AAR99137	RE06943p	101.29	8.81E-21	43%	56%	19--138	43--169

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	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	9.96E-08	32%	48%	10--113	3--95	6	BAE51501	Deoxycytidylate deaminase	101.29	8.81E-21	38%	57%	14--139	15--139
					7	COG0590	CumB, Cytosine/adenosine deaminases [Nucleotide transport and metabolism / Translation, ribosomal structure and biogenesis].	49.18	6.36E-07	39%	51%	30--111	34--104	7	BAB80772	deoxycytidylate deaminase	100.91	1.15E-20	40%	53%	1--142	9--150
					8	COG0117	RibD, Pyrimidine deaminase [Coenzyme metabolism].	44.08	2.58E-05	29%	46%	5--111	7--100	8	NP_001006444	dCMP deaminase	100.52	1.50E-20	42%	56%	5--137	29--162
														9	XP_781375	PREDICTED: similar to Deoxycytidylate deaminase (dCMP deaminase)	100.52	1.50E-20	40%	56%	7--137	69--200
														10	AAT75744	deoxycytidylate deaminase	99.75	2.56E-20	41%	53%	5--142	10--147
B796L	330418--329321	366	41,991	7.35	1	COG0076	GadB, Glutamate decarboxylase and related PLP-dependent proteins [Amino acid transport and metabolism].	130.50	2.61E-31	26%	45%	56--325	94--383	1	NP_048954	similar to tomato histidine decarboxylase, corresponds to Swiss-Prot Accession Number P54772	658.29	0.00E+00	84%	92%	4--366	1--363
					2	pfam00282	Pyridoxal_deC, Pyridoxal-dependent decarboxylase conserved domain..	110.01	3.65E-25	24%	42%	73--322	88--372	2	NP_919502	putative histidine decarboxylase	222.25	2.02E-56	37%	56%	16--361	78--428
					3	COG1104	NifS, Cysteine sulfinate desulfinase/cysteine desulfurase and related enzymes [Amino acid transport and metabolism].	48.24	1.25E-06	31%	46%	56--243	39--219	3	ZP_00106716	COG0076: Glutamate decarboxylase and related PLP-dependent proteins	220.71	5.87E-56	36%	58%	25--359	25--363
					4	COG0520	CsdB, Selenocysteine lyase [Amino acid transport and metabolism].	46.09	6.47E-06	25%	47%	73--243	75--241	4	BAA78331	serine decarboxylase	209.92	1.04E-52	35%	54%	5--362	98--461
														5	XP_471202	OSJNBa0059H15.18	208.76	2.31E-52	37%	54%	25--361	69--413
														6	BAE07183	putative serine decarboxylase	208.76	2.31E-52	35%	56%	24--361	111--456
														7	NP_175036	EMB1075; carboxy-lyase	207.99	3.94E-52	34%	54%	5--362	90--453
														8	BAD28221	putative serine decarboxylase	207.22	6.72E-52	35%	55%	24--361	106--450
														9	CAA50719	histidine decarboxylase	203.76	7.43E-51	35%	55%	20--359	29--377
														10	BAC87908	probable acinetobactin biosynthesis protein	199.13	1.83E-49	35%	55%	24--359	26--369
B798R	330531--331469	313	36,636	8.14	1	smart00497	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished). .	48.98	8.00E-07	40%	62%	255--307	1--53	1	NP_048621	A267L	182.57	1.39E-44	35%	51%	2--285	21--293
					2	pfam07453	NUMOD1, NUMOD1 domain..	41.95	1.17E-04	58%	70%	255--288	1--34	2	NP_048846	Lys-, Glu-rich	111.31	3.93E-23	34%	50%	3--217	67--288
														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	105.53	2.16E-21	31%	48%	1--219	57--290
														4	YP_142777	unknown	73.94	6.95E-12	25%	44%	50--216	281--456
														5	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	69.32	1.71E-10	52%	75%	248--307	165--225
														6	NP_048671	A315L	68.55	2.92E-10	49%	70%	248--308	183--243
														7	NP_048641	PBCV-1 33kd peptide	59.31	1.77E-07	42%	62%	244--307	187--249
B800R	331566--331835	90	10,181	10.23		No Hit Found								1	NP_048957	A601R	117.86	9.05E-26	61%	70%	3--90	1--101
B801L	332243--331842	134	15,468	4.66		No Hit Found								1	NP_048958	A602L	84.73	8.54E-16	73%	90%	1--52	62--113
B802R	332353--332679	109	13,104	6.11		No Hit Found								1	NP_048959	A603R	163.70	1.44E-39	73%	85%	6--109	1--105
B803L	333220--332876	115	13,080	9.95		No Hit Found								1	NP_048960	A604L	82.80	3.27E-15	34%	60%	6--115	20--134
B805R	333334--334353	340	40,540	6.21	1	COG0419	SbcC, ATPase involved in DNA repair [DNA replication, recombination, and repair].	43.12	4.31E-05	20%	44%	10--124	450--564	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	304.68	2.75E-81	45%	62%	1--339	1--306
					2	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..	39.34	5.69E-04	23%	55%	21--125	73--170	2	NP_048846	Lys-, Glu-rich	300.44	5.19E-80	44%	60%	1--339	1--306
					3	pfam01442	Apolipoprotein, Apolipoprotein A1/A4/E family. These proteins contain several 22 residue repeats which form a pair of alpha helices. This family includes: Apolipoprotein A-I. Apolipoprotein A-IV. Apolipoprotein E..	38.29	1.18E-03	18%	51%	36--143	100--195	3	YP_142777	unknown	157.15	7.10E-37	29%	51%	1--320	135--457
														4	NP_048621	A267L	85.89	2.01E-15	25%	39%	45--322	4--246
														5	NP_701067	hypothetical protein PF11_0207	64.70	4.79E-09	32%	60%	25--128	486--595
														6	CAI94257	putative transmembrane protein	64.31	6.26E-09	37%	72%	25--99	31--110
														7	CAI94306	hypothetical protein	63.16	1.39E-08	37%	71%	25--99	10--89
														8	NP_703225	ring-infected erythrocyte surface antigen precursor	60.85	6.92E-08	34%	60%	29--120	41--1035
														9	XP_728106	hypothetical protein PY07276	56.61	1.31E-06	29%	57%	24--119	193--288
														10	XP_646809	WD40 repeat-containing protein	55.84	2.23E-06	33%	58%	21--131	42--1157
B811R	334668--334862	65	7,886	10.70		No Hit Found									No Hit Found	No Hit Found						
B812R	334878--336050	391	45,703	6.07	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	60.48	3.01E-10	29%	49%	27--151	5--125	1	NP_048963	contains 3 ankyrin repeat-like elements; similar to Drosophila ankyrin, corresponds to GenBank Accession Number L35601	389.81	8.01E-107	82%	92%	1--217	1--217
														2	NP_048964	A608R	239.20	1.75E-61	72%	85%	240--390	1--151
														3	EAL45248	ankyrin repeat protein, putative	71.25	6.31E-11	24%	42%	20--292	100--354
														4	BAC43653	unknown protein	58.92	3.24E-07	28%	47%	31--210	198--368
														5	CAB10219	hypothetical protei	58.54	4.23E-07	28%	47%	31--210	259--429
														6	BAD43172	unknown protein	58.54	4.23E-07	28%	47%	31--210	264--434
														7	NP_567430	ACD6 (ACCELERATED CELL DEATH 6); protein binding	58.54	4.23E-07	28%	47%	31--210	264--434
														8	NP_849381	ACD6 (ACCELERATED CELL DEATH 6); protein binding	58.54	4.23E-07	28%	47%	31--210	198--368
														9	EAA14062	ENSANGP00000013300	57.77	7.22E-07	24%	40%	13--309	344--615
														10	BAD94307	hypothetical protein	57.00	1.23E-06	27%	47%	31--210	264--434
B813L	336508--336152	119	13,441	9.25	1	smart00317	SET, SET (Su(var)3-9, Enhancer-of-zeste, Trithorax) domain; Putative methyl transferase, based on outlier plant homologues .	71.19	1.80E-13	29%	40%	5--110	2--123	1	NP_048968	PBCV-1 histone H3-Lys 27 methyltransferase (vSET)	206.07	2.52E-52	78%	89%	1--119	1--119
					2	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..	58.92	8.93E-10	25%	39%	2--107	6--128	2	ZP_00661322	Nuclear protein SET	70.48	1.66E-11	33%	55%	5--114	37--150
					3	COG2940	COG2940, Proteins containing SET domain [General function prediction only].	52.03	9.79E-08	27%	41%	1--117	329--460	3	ABB28752	Nuclear protein SET	65.08	6.95E-10	34%	53%	1--113	43--159
														4	AAM72187	conserved hypothetical protein	64.31	1.19E-09	33%	53%	6--113	39--150
														5	ZP_00588496	Nuclear protein SET	64.31	1.19E-09	32%	51%	6--114	39--151
														6	ZP_00592010	Nuclear protein SET	62.77	3.45E-09	36%	53%	6--113	39--150
														7	ABB23988	Nuclear protein SET	62.39	4.51E-09	30%	53%	5--113	38--150
														8	NP_701503	hypothetical protein PFL0690c	61.62	7.69E-09	28%	45%	2--110	29--174
														9	ZP_00511449	Nuclear protein SET	61.23	1.00E-08	32%	52%	8--113	40--149
														10	ZP_00528743	Nuclear protein SET	60.85	1.31E-08	31%	50%	8--113	44--153
B816L	338000--336525	492	55,746	11.39	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	60.22	3.31E-10	27%	48%	62--239	6--152	1	NP_048970	RPQT-like (9x)	690.65	0.00E+00	62%	72%	1--492	1--577
					2	smart00220	S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases, Serine or threonine-specific kinase subfamily. .	59.07	7.13E-10	28%	47%	62--239	5--151	2	NP_048632	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662	303.91	7.91E-81	47%	65%	6--330	10--344

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	pfam00069	Pkinase, Protein kinase domain..	54.91	1.27E-08	31%	48%	62--236	5--148	3	NP_048636	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055	273.48	1.15E-71	41%	58%	6--365	24--407
					4	COG0515	SPS1, Serine/threonine protein kinase [General function prediction only / Signal transduction mechanisms / Transcription / DNA replication, recombination, and repair].	49.00	8.41E-07	16%	31%	84--430	3--376	4	XP_644812	hypothetical protein DDB0217139	94.74	7.29E-18	54%	88%	338--409	061--1130
					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	41.29	1.65E-04	32%	68%	207--234	171--199	5	XP_641859	hypothetical protein DDB0205029	90.89	1.05E-16	47%	69%	324--416	348--443
					6	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	39.58	5.19E-04	18%	46%	328--408	37--117	6	ZP_00783686	pathogenicity protein, putative	87.43	1.16E-15	29%	64%	303--418	580--690
					7	smart00219	TyrKc, Tyrosine kinase, catalytic domain; Phosphotransferases. Tyrosine-specific kinase subfamily. .	38.23	1.49E-03	23%	41%	62--239	5--153	7	ZP_00780097	surface protein PspC	86.66	1.99E-15	32%	66%	325--424	403--502
					8	COG0478	COG0478, RIO-like serine/threonine protein kinase fused to N-terminal HTH domain [Signal transduction mechanisms].	36.76	3.42E-03	22%	43%	153--252	169--269	8	ZP_00789197	Gram positive anchor domain protein	86.27	2.59E-15	31%	60%	303--418	111--217
					9	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	36.66	3.81E-03	20%	38%	318--426	40--152	9	ZP_00788171	pathogenicity protein, putative	85.50	4.42E-15	29%	61%	303--418	116--222
					10	COG3087	FtsN, Cell division protein [Cell division and chromosome partitioning].	36.56	4.40E-03	18%	45%	314--406	94--190	10	ABA89280	ribonuclease, Rne/Rng family protein	81.65	6.39E-14	21%	53%	329--469	672--812
B818R	338069--339019	317	36,800	9.92		No Hit Found								1	NP_048973	similar to Variola virus orf E10L, corresponds to Swiss-Prot Accession Number P33801	519.24	6.39E-146	79%	90%	2--316	4--318
														2	YP_142754	S/T protein kinase, similar to Paramecium bursaria chlorella virus 1 A617R	87.04	8.11E-16	27%	46%	9--250	107--359
														3	NP_149843	380R	57.00	8.99E-07	26%	41%	2--245	157--378
B819L	339378--339028	117	13,544	3.92		No Hit Found								1	NP_048974	A618L	163.31	1.89E-39	77%	86%	14--115	20--129
B820L	340051--339416	212	24,652	4.35		No Hit Found								1	NP_048975	A619L	213.77	2.75E-54	49%	55%	1--212	1--237
B823L	340341--340093	83	9,641	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	74.71	8.67E-13	44%	65%	1--79	1--84
B824L	340719--340369	117	12,939	10.38		No Hit Found								1	NP_048977	A621L	216.85	1.44E-55	88%	96%	1--117	1--117
B825L	342338--340779	520	58,195	5.75	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 NY2A the major capsid protein is a glycoprotein	366.56	2.24E-102	43%	56%	174--516	86--443	1	NP_048978	similar to Simulium iridescent virus capsid protein, corresponds to Swiss-Prot Accession Number P22166	1032.32	0.00E+00	94%	97%	1--520	1--520
														2	AAC27493	putative capsid protein	764.22	0.00E+00	71%	81%	1--520	1--521
														3	BAE06835	hypothetical major capsid protein	265.39	3.34E-69	39%	59%	180--520	94--440
														4	BAA76601	major capsid protein MCP1	205.30	4.10E-51	36%	50%	188--520	92--437
														5	BAA76600	major capsid protein	203.37	1.56E-50	35%	49%	188--520	92--436
														6	NP_048787	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	202.22	3.47E-50	35%	49%	188--520	92--437
														7	1M3Y_D	Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid Containing, Dna Virus	202.22	3.47E-50	35%	49%	188--520	68--413
														8	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	201.06	7.73E-50	35%	49%	188--520	68--413
														9	AAC27492	major capsid protein Vp49	199.90	1.72E-49	36%	53%	188--520	89--432
														10	BAA22198	major capsid protein Vp54	199.13	2.94E-49	35%	49%	188--520	92--437
B826L	342592--342392	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	5.83E-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	7.05E-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
B828R	342623--342985	121	13,557	10.01	1	COG4852	COG4852, Predicted membrane protein [Function unknown].	43.75	3.08E-05	30%	45%	14--120	13--124	1	NP_048980	A624R	207.22	1.12E-52	86%	87%	1--121	1--121
														2	ZP_00234461	conserved hypothetical protein	55.84	4.19E-07	26%	46%	4--120	3--125
														3	CAC95830	lin0598	55.07	7.14E-07	26%	45%	4--120	3--125
														4	ZP_00231099	conserved hypothetical protein	54.68	9.33E-07	25%	45%	4--120	3--125
														5	CAC98668	lmo0589	53.91	1.59E-06	25%	45%	4--120	3--125
B829R	343057--344352	432	49,946	10.88	1	COG0675	COG0675, Transposase and inactivated derivatives [DNA replication, recombination, and repair].	86.29	4.86E-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	1.46E-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
B831R	344451--345755	435	48,558	11.14		No Hit Found								1	NP_048983	similar to Chlorella virus PBCV-1 ORF A231L, corresponds to GenBank Accession Number U42580	812.76	0.00E+00	92%	94%	1--434	1--436
														2	NP_048579	contains ATP/GTP-binding motif A	301.21	4.35E-80	44%	63%	101--425	10--338
B832R	345824--349132	1103	124,556	8.18	1	pfam02867	Ribonuc_red_lgC, Ribonucleotide reductase, barrel domain..	375.78	3.58E-105	47%	63%	767--1085	207--532	1	NP_048985	similar to Schizosaccharomyces ribonucleotide reductase M1 chain, corresponds to Swiss-Prot Accession Number P36602	778.86	0.00E+00	88%	91%	2--438	6--442

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	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 their metallocofactor usage. Class I RNRs, found in eukaryotes, bacteria, and many viruses, use a diiron-tyrosyl radical, Class II RNRs, found in bacteria, and bacteriophages, use coenzyme B12 (adenosylcobalamin, AdoCbl). Class III RNRs, found in anaerobic bacteria, bacteriophages, and archaea, use an FeS cluster and S-adenosylmethionine to generate a glycyl radical. Many organisms have more than one class of RNR present in their genomes. All three RNRs have a ten-stranded alpha-beta barrel domain that is structurally similar to the domain of PFL (pyruvate formate lyase). Class I RNR is oxygen-dependent and can be subdivided into classes Ia (eukaryotes, prokaryotes, viruses and phages) and Ib (which is found in prokaryotes only). It is a tetrameric enzyme of two alpha and two beta subunits; this model covers the major part of the alpha or large subunit, c	364.91	6.45E-102	50%	67%	773--1083	264--578	2	ZP_00514074	Protein splicing (intein) site	462.23	4.51E-128	30%	48%	12--1102	35--1115
					3	COG0209	NrdA, Ribonucleotide reductase, alpha subunit [Nucleotide transport and metabolism].	298.85	5.33E-82	33%	50%	66--495	7--437	3	NP_149548	085L	460.69	1.31E-127	33%	50%	157--1089	18--900
					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 classes based on their metallocofactor usage. Class I RNRs, found in eukaryotes, bacteria, and many viruses, use a diiron-tyrosyl radical, Class II RNRs, found in bacteria, and bacteriophages, use coenzyme B12 (adenosylcobalamin, AdoCbl). Class III RNRs, found in anaerobic bacteria, bacteriophages, and archaea, use an FeS cluster and S-adenosylmethionine to generate a glycyl radical. Many organisms have more than one class of RNR present in their genomes. All three RNRs have a ten-stranded alpha-beta barrel domain that is structurally similar to the domain of PFL (pyruvate formate lyase). This family appears similar to class I RNRs, as judged by sequence similarity and the predicted active site..	169.23	5.54E-43	32%	51%	787--1080	232--521	4	EAL90119	ribonucleotide reductase large subunit (Rnr1), putative	402.13	5.54E-110	46%	63%	10--437	1--429
					5	pfam00317	Ribonuc_red_lgN, Ribonucleotide reductase, all-alpha domain..	84.14	2.03E-17	42%	58%	151--223	1--78	5	BAE59411	unnamed protein product	401.36	9.44E-110	47%	63%	10--437	1--429
					6	cd00081	Hint, Hedgehog/Intein domain, found in Hedgehog proteins as well as proteins which contain inteins and undergo protein splicing (e.g. DnaB, RIR1-2, GyrA and Pol). In protein splicing an intervening polypeptide sequence - the intein - is excised from a protein, and the flanking polypeptide sequences - the exteins - are joined by a peptide bond. In addition to the autocatalytic splicing domain, many inteins contain an inserted endonuclease domain, which plays a role in spreading inteins. Hedgehog proteins are a major class of intercellular signaling molecules, which control inductive interactions during animal development. The mature signaling forms of hedgehog proteins are the N-terminal fragments, which are covalently linked to cholesterol at their C-termini. This modification is the result of an autoprocessing step catalyzed by the C-terminal fragments, which are aligned here.	59.00	7.51E-10	14%	29%	437--770	1--338	6	CAB98233	ribonucleoside-diphosphate reductase large chain (un-24)	400.98	1.23E-109	46%	63%	10--437	1--429
					7	smart00306	HintN, Hint (Hedgehog/Intein) domain N-terminal region; Hedgehog/Intein domain, N-terminal region. Domain has been split to accommodate large insertions of endonucleases. .	58.03	1.54E-09	24%	46%	437--531	1--96	7	AAD49743	ribonucleotide reductase large subunit	400.98	1.23E-109	46%	63%	10--437	1--429
					8	pfam03477	ATP-cone, ATP cone domain..	53.83	2.67E-08	31%	48%	10--111	1--89	8	XP_370503	hypothetical protein MG07000.4	399.82	2.75E-109	45%	64%	10--437	1--429
					9	COG1372	COG1372, Intein/homing endonuclease [DNA replication, recombination, and repair].	53.11	5.09E-08	20%	37%	482--675	64--280	9	XP_661984	ribonucleoside-diphosphate reductase large chain	397.90	1.04E-108	46%	63%	7--437	12--443
					10	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.	43.94	2.48E-05	21%	35%	799--1033	270--502	10	ZP_00310043	COG0209: Ribonucleotide reductase, alpha subunit	395.20	6.77E-108	46%	64%	10--437	1--429
B839R	349163--349522	120	13,219	8.43		No Hit Found								1	NP_048989	A633R	204.14	9.56E-52	82%	87%	1--120	1--120
B840L	349933--349529	135	15,757	8.50		No Hit Found								1	NP_048990	A634L	260.77	8.66E-69	91%	95%	1--135	1--134
B841R	349976--350230	85	9,902	9.80		No Hit Found								1	NP_048991	A635R	172.56	3.15E-42	97%	98%	1--85	1--85
														2	NP_048976	similar to Synechocystis orf 90, corresponds to GenBank Accession Number D90902	72.02	5.79E-12	43%	63%	1--82	1--77
B842R	350288--350500	71	8,173	5.88		No Hit Found								1	NP_048992	A636R	82.03	5.44E-15	54%	58%	1--71	1--96
B843R	350574--350996	141	16,476	9.16		No Hit Found								1	NP_048993	A637R	237.27	1.03E-61	82%	85%	1--141	1--141
B844R	351043--352119	359	40,963	5.54	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 gingivalis (PPAD) appears to be evolutionarily unrelated to mammalian PAD (pfam03068), which is a metalloenzyme. PPAD is thought to belong to the same superfamily as aminotransferase and arginine deiminase, and to form an alpha/beta propeller structure. This family has previously been named PPADH (Porphyromonas peptidyl-arginine deiminase homologues). The predicted catalytic residues in PPAD are Asp130, Asp187, His236, Asp238 and Cys351. These are absolutely conserved with the exception of Asp187 which is absent in two family members. PPAD is also able to catalyse the deimination of free L-arginine, but has primarily peptidyl-arginine specificity. It may have a FMN cofactor..	467.77	6.14E-133	55%	72%	8--358	1--329	1	NP_048994	PBCV-1 Agmatine iminohydrolase	727.63	0.00E+00	96%	98%	1--359	1--359
					2	COG2957	COG2957, Peptidylarginine deiminase and related enzymes [Amino acid transport and metabolism].	385.41	4.55E-108	47%	67%	2--358	9--344	2	CAC98253	lmo0038	380.95	3.29E-104	52%	67%	4--358	9--362
														3	ZP_00232718	peptidyl-arginine deiminase-like protein	380.18	5.62E-104	51%	67%	4--358	9--362
														4	ZP_00229955	peptidyl-arginine deiminase-like protein	379.41	9.59E-104	51%	67%	4--358	9--362
														5	ZP_00322658	COG2957: Peptidylarginine deiminase and related enzymes	368.24	2.21E-100	50%	67%	2--358	7--362
														6	AAL98713	LabD	366.70	6.43E-100	50%	66%	2--358	7--362
														7	CAI54369	Putative peptidylarginine deiminase (Amidinotransferase)	366.70	6.43E-100	50%	66%	2--358	17--372
														8	YP_424617	peptidylarginine deiminase-related protein, putative	363.61	5.44E-99	51%	65%	2--358	8--361
														9	AAG03681	agmatine deiminase	363.61	5.44E-99	50%	65%	2--358	7--363
														10	ZP_00972430	COG2957: Peptidylarginine deiminase and related enzymes	362.46	1.21E-98	50%	65%	2--358	7--363
B847R	352141--353598	486	55,509	11.46	1	COG4487	COG4487, Uncharacterized protein conserved in bacteria [Function unknown].	47.34	2.55E-06	22%	45%	255--401	64--216	1	NP_048999	Gln-rich; KQQ (6X)	298.13	4.27E-79	62%	64%	197--486	1--269
					2	COG0488	Uup, ATPase components of ABC transporters with duplicated ATPase domains [General function prediction only].	42.17	9.95E-05	16%	44%	327--408	236--317	2	NP_048998	A642R	119.78	2.08E-25	100%	100%	1--56	1--56
B849R	353640--354152	171	19,008	7.30		No Hit Found								1	NP_049000	A644R	141.35	1.02E-32	80%	91%	87--170	1--84
														2	YP_142763	unknown	57.00	2.54E-07	31%	52%	52--163	61--176
B850L	355333--354161	391	44,883	10.11	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.	48.62	1.12E-06	29%	45%	17--92	3--78	1	NP_048641	PBCV-1 33kd peptide	55.45	3.58E-06	27%	44%	17--252	6--248
					2	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	45.84	6.35E-06	38%	56%	15--100	1--81	2	AAU16837	GIY-YIG catalytic domain containing protein; possible intron encoded endonuclease	55.07	4.68E-06	33%	48%	14--124	1--115
					3	COG2827	COG2827, Predicted endonuclease containing a URI domain [DNA replication, recombination, and repair].	36.82	3.44E-03	36%	55%	17--60	6--48	3	NP_049849	I-TevI homing endonuclease	52.76	2.32E-05	31%	47%	14--126	1--118
														4	NP_899393	SegD	51.22	6.76E-05	35%	50%	17--107	3--89

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
B852R	355481--355849	123	14,342	7.32		No Hit Found								1	NP_049001	A645R	214.54	7.23E-55	81%	89%	1--122	1--122
B853L	356456--355866	197	22,619	6.50	1	pfam00583	Acetyltransf_1, Acetyltransferase (GNAT) family. This family contains proteins with N-acetyltransferase functions..	38.31	1.37E-03	27%	41%	97--163	12--82	1	NP_049010	A654L	360.53	1.57E-98	86%	94%	1--197	1--197
														2	ZP_00675288	hypothetical protein TeryDRAFT_0585	59.69	5.71E-08	26%	46%	23--184	28--202
														3	AAL68103	AT19803p	49.68	5.91E-05	22%	41%	7--173	6--191
														4	XP_641184	hypothetical protein DDB0205209	49.68	5.91E-05	21%	41%	7--184	14--204
														5	CAB72290	EG:BACR25B3.5	49.68	5.91E-05	22%	41%	7--173	6--191
B857L	356897--356490	136	15,549	4.58		No Hit Found								1	NP_049012	A656L	175.64	3.65E-43	71%	85%	17--130	1--114
B858L	357602--357051	184	20,646	3.45		No Hit Found								1	NP_049015	A659L	189.50	4.03E-47	74%	77%	1--122	1--128
B859L	358136--357624	171	19,591	10.82		No Hit Found								1	NP_049018	A662L	280.80	1.08E-74	77%	90%	1--171	1--171
														2	BAD87006	unknown protein	73.94	2.01E-12	30%	50%	42--170	110--239
														3	NP_565983	unknown protein	66.63	3.21E-10	25%	52%	44--170	105--232
														4	AAM62733	contains similarity to 22 kDa peroxisomal membrane protein	58.15	1.14E-07	25%	49%	26--165	101--245
														5	XP_635703	hypothetical protein DDB0189006	58.15	1.14E-07	30%	45%	28--158	40--172
														6	NP_568621	unknown protein	58.15	1.14E-07	25%	49%	26--165	100--244
														7	BAB08278	unnamed protein product	57.38	1.95E-07	28%	52%	26--145	100--224
														8	CAB80107	putative protein	51.99	8.17E-06	25%	49%	26--151	104--234
														9	XP_326388	hypothetical protein	50.83	1.82E-05	26%	45%	26--167	82--227
														10	NP_567940	unknown protein	50.45	2.38E-05	25%	48%	26--147	104--230
B860L	358655--358194	154	17,456	6.36		No Hit Found								1	NP_049020	similar to Chlorella virus CVK2 unknown ORF, corresponds to GenBank Accession Number D16505	169.86	1.98E-41	58%	79%	15--154	1--141
														2	NP_049021	similar to Chlorella virus CVK2 unknown ORF, corresponds to GenBank Accession Number D16505	142.12	4.42E-33	47%	71%	1--151	21--170
B862L	359162--358710	151	17,199	6.90		No Hit Found								1	NP_049021	similar to Chlorella virus CVK2 unknown ORF, corresponds to GenBank Accession Number D16505	265.39	3.50E-70	86%	93%	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
B865R	359447--360094	216	24,863	7.99	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..	231.78	7.09E-62	40%	54%	1--212	1--216	1	NP_049030	similar to Synechocystis ORF s111635, corresponds to GenBank Accession Number D90903	373.24	2.84E-102	83%	91%	1--216	1--216
					2	COG1351	THY1, Predicted alternative thymidylate synthase [Nucleotide transport and metabolism].	136.30	3.72E-33	31%	47%	22--215	35--237	2	CAF34258	thymidilate synthase	238.81	8.35E-62	57%	70%	1--214	3--211
														3	ZP_01006033	thymidylate synthase	234.19	2.06E-60	57%	75%	19--215	13--210
														4	AAX46996	Td	230.72	2.27E-59	58%	73%	19--215	12--210
														5	CAE18702	possible Thy1 protein homolog	228.79	8.64E-59	54%	73%	19--215	13--211
														6	AAX44700	Td	225.71	7.32E-58	50%	69%	1--216	1--211
														7	AAZ59097	Thymidylate synthase complementing protein ThyX	224.94	1.25E-57	53%	71%	9--216	3--211
														8	ABB49306	Thymidylate synthase complementing protein ThyX	216.08	5.80E-55	54%	71%	19--215	13--210
														9	AAP99319	Predicted alternative thymidylate synthase	214.93	1.29E-54	52%	70%	19--216	13--211
														10	ZP_00531300	Thymidylate synthase (FAD)	210.69	2.44E-53	50%	66%	1--215	1--215
B867L	360849--360097	251	28,390	9.40		No Hit Found								1	NP_048629	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	417.16	2.31E-115	79%	88%	1--249	1--251
														2	NP_048807	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	397.13	2.47E-109	75%	87%	4--249	1--248
														3	NP_049005	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	355.53	8.24E-97	66%	80%	1--248	4--254
														4	NP_048427	A79R	333.18	4.38E-90	68%	84%	4--234	1--226
														5	AAU06304	hypothetical protein A275R	329.72	4.84E-89	92%	96%	80--249	1--170
														6	AAU06301	hypothetical protein A275R	328.56	1.08E-88	92%	96%	80--249	1--170
														7	NP_048525	A177R	296.59	4.54E-79	59%	75%	1--245	1--240
														8	AAU06302	hypothetical protein A275R	261.15	2.12E-68	95%	96%	116--248	1--133
														9	AAU06303	hypothetical protein A275R	117.47	3.78E-25	96%	96%	192--249	1--58
														10	AAU06299	hypothetical protein A275R	114.78	2.45E-24	94%	94%	192--249	5--62
B869R	361094--362185	364	41,460	10.63		No Hit Found								1	NP_049032	similar to Chlorella virus PBCV-1 ORF A282L, corresponds to GenBank Accession Number U42580	644.04	0.00E+00	85%	92%	1--358	1--368
														2	NP_048636	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055	72.79	1.96E-11	42%	51%	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	55.45	3.24E-06	28%	42%	60--202	5--148
B871R	362347--362841	165	18,579	4.52		No Hit Found									No Hit Found	No Hit Found						
B874L	364117--363020	366	40,137	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.	113.25	3.98E-26	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	537.34	2.83E-151	70%	82%	1--366	1--368
					2	COG0666	Arp, FOG: Ankyrin repeat [General function prediction only].	79.93	3.91E-16	31%	53%	4--166	39--209	2	NP_048353	contains 4 ankyrin repeats; similar to reticulocyte ankyrin, corresponds to Swiss-Prot Accession Number P16157	264.62	3.55E-69	57%	68%	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..	45.82	7.33E-06	53%	77%	137--167	2--32	3	EAL87814	NACHT domain protein, putative	181.03	5.16E-44	37%	57%	11--290	26--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. .	40.71	2.37E-04	52%	76%	136--165	1--30	4	XP_637214	SecG	173.71	8.23E-42	32%	50%	14--288	212--520
														5	AAV85825	ankyrin domain protein	172.17	2.40E-41	34%	51%	21--290	73--368
														6	XP_681288	hypothetical protein AN8019.2	164.85	3.82E-39	35%	55%	11--289	04--1082
														7	ZP_00373467	ankyrin repeat domain protein	163.70	8.52E-39	36%	52%	21--290	87--381
														8	XP_637278	hypothetical protein DDB0187458	160.61	7.21E-38	32%	53%	11--307	216--514
														9	EAL84954	ankyrin repeat protein	160.23	9.42E-38	33%	56%	8--288	430--710
														10	NP_065209	ankyrin 1 isoform 1	158.69	2.74E-37	32%	55%	11--294	246--529
B878L	365432--364233	400	47,080	6.92		No Hit Found								1	NP_048711	A354R	155.22	3.44E-36	37%	55%	158--399	4--234
														2	NP_048779	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	85.11	4.37E-15	27%	46%	68--367	25--319
B880L	366205--365501	235	25,796	5.73	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.	110.94	1.92E-25	42%	59%	30--154	1--126	1	NP_048353	contains 4 ankyrin repeats; similar to reticulocyte ankyrin, corresponds to Swiss-Prot Accession Number P16157	291.20	1.69E-77	62%	69%	1--219	1--252
					2	COG0666	Arp, FOG: Ankyrin repeat [General function prediction only].	79.55	5.26E-16	35%	50%	19--154	56--200	2	NP_049038	contains 10 ankyrin-like repeats; similar to human ankyrin, corresponds to Swiss-Prot Accession Number P16157	249.60	5.63E-65	57%	71%	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	3.19E-04	48%	65%	36--67	2--33	3	XP_637214	SecG	110.54	4.08E-23	34%	52%	10--201	176--373
														4	AAO91105	ankyrin repeat family protein	101.68	1.90E-20	28%	53%	10--202	102--299
														5	XP_237153	PREDICTED: similar to hypothetical protein DKFZp434D2328	101.68	1.90E-20	35%	52%	7--173	788--964
														6	XP_912160	PREDICTED: similar to ankyrin repeat domain 28 isoform 10	100.14	5.52E-20	35%	52%	7--173	638--814

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	XP_925269	PREDICTED: similar to ankyrin repeat domain 28 isoform 16	100.14	5.52E-20	35%	52%	7--173	584--760
														8	XP_925237	PREDICTED: similar to ankyrin repeat domain 28 isoform 12	100.14	5.52E-20	35%	52%	7--173	656--832
														9	1N0R_A	Chain A, 4ank: A Designed Ankyrin Repeat Protein With Four Identical Consensus Repeats	99.75	7.21E-20	46%	61%	35--154	1--121
														10	XP_637278	hypothetical protein DDB0187458	99.37	9.41E-20	30%	51%	10--154	216--395
B883L	366677--366372	102	11,766	6.88		No Hit Found								1	NP_049045	identical to Chlorella virus PBCV-1 terminal repeat ORF A3R, corresponds to GenBank Accession Number M55319	109.38	3.20E-23	60%	70%	1--101	1--102
B884L	367168--366935	78	8,909	4.55		No Hit Found									No Hit Found	No Hit Found						
B886R	367486--367764	93	10,443	4.71		No Hit Found									No Hit Found	No Hit Found						