University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Virology Papers | Virology, Nebraska Center for

February 2007

Supplementary Data for "Sequence and annotation of the 314-kb MT325 and the 321-kb FR483 viruses that infect *Chlorella* Pbi": Appendix D: Gene Names N003L through N847R

Lisa A. Fitzgerald
University of Nebraska-Lincoln, fitz918@msn.com

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

James Hartigan
Agencourt Bioscience Corporation, Beverly, MA

See next page for additional authors

Follow this and additional works at: http://digitalcommons.unl.edu/virologypub

 Part of the Virology Commons

Fitzgerald, Lisa A.; Graves, Michael V.; Li, Xiao; Feldblyum, Tamara; Hartigan, James; and Van Etten, James L., "Supplementary Data for "Sequence and annotation of the 314-kb MT325 and the 321-kb FR483 viruses that infect *Chlorella* Pbi": Appendix D: Gene Names N003L through N847R" (2007). *Virology Papers*. 5.
http://digitalcommons.unl.edu/virologypub/5

This Article is brought to you for free and open access by the Virology, Nebraska Center for at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Virology Papers by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors
Lisa A. Fitzgerald, Michael V. Graves, Xiao Li, Tamara Feldblyum, James Hartigan, and James L. Van Etten

This article is available at DigitalCommons@University of Nebraska - Lincoln: http://digitalcommons.unl.edu/virologypub/5

Published in *Virology* 358:2 (February 20, 2007), pp. 459–471. doi:10.1016/j.virol.2006.08.034 Copyright © 2006 Elsevier Inc. Used by permission. Submitted July 27, 2006; revised August 18, 2006; accepted August 23, 2006; published online October 4, 2006. Supplementary data originally published online at doi:10.1016/j.virol.2006.08.034.

SUPPLEMENTARY DATA FOR

Sequence and annotation of the 314-kb MT325 and the 321-kb FR483 viruses that infect *Chlorella* Pbi

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

[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]Agencourt Bioscience Corporation, 500 Cummings Center, Suite 2450, Beverly, MA 01915
[e]Deparment of Plant Pathology, University of Nebraska–Lincoln, Lincoln, NE 68583-0722
[f]Nebraska Center for Virology, University of Nebraska, Lincoln, NE 68588-0666
*Corresponding author. Email: jvanetten@unlnotes.unl.edu

Abstract: Viruses MT325 and FR483, members of the family Phycodnaviridae, genus *Chlorovirus*, infect the fresh water, unicellular, eukaryotic, chlorella-like green alga, *Chlorella* Pbi. The 314,335-bp genome of MT325 and the 321,240-bp genome of FR483 are the first viruses that infect *Chlorella* Pbi to have their genomes sequenced and annotated. Furthermore, these genomes are the two smallest chlorella virus genomes sequenced to date, MT325 has 331 putative protein-encoding and 10 tRNA-encoding genes and FR483 has 335 putative protein-encoding and 9 tRNA-encoding genes. The protein-encoding genes are almost evenly distributed on both strands, and intergenic space is minimal. Approximately 40% of the viral gene products resemble entries in public databases, including some that are the first of their kind to be detected in a virus. For example, these unique gene products include an aquaglyceroporin in MT325, a potassium ion transporter protein and an alkyl sulfatase in FR483, and a dTDP–glucose pyrophosphorylase in both viruses. Comparison of MT325 and FR483 protein-encoding genes with the prototype chlorella virus PBCV-1 indicates that approximately 82% of the genes are present in all three viruses.

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, Peptid e 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 m002R through m843L
Appendix B: Gene Names M001L through M807R
Appendix C: Gene Names n001L through n849R
Appendix D: Gene Names N003L through N847R

Appendix D: Gene Names N003L through N847R

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
N003L	2472--490	661	74,180	9.20	1	COG2015	COG2015, Alkyl sulfatase and related hydrolases [Secondary metabolites biosynthesis. transoort. and catabolism].	700.55	0.00E+00	45%	64%	22--654	15--652	1	NP_896208	alkyl sulfatase family protein	752.67	0.00E+00	58%	73%	34--657	50--679
					2	pfam00753	Lactamase_B, Metallo-beta-lactamase superfamily.	48.00	3.49E-06	22%	36%	129--232	1--102	2	AAZ27425	metallo-beta-lactamase family protein	741.50	0.00E+00	57%	74%	34--657	45--675
					3	COG1237	COG1237, Metal-dependent hydrolases of the beta-lactamase superfamily II [General function prediction only]	40.68	6.24E-04	27%	42%	138--224	26--107	3	CAG19676	hypothetical beta-lactamase	731.48	0.00E+00	56%	73%	34--658	48--678
					4	COG4783	COG4783, Putative Zn-dependent protease, contains TPR repeats [General function prediction only].	38.01	3.33E-03	33%	54%	464--531	378--445	4	AAV96272	metallo-beta-lactamase family protein	721.47	0.00E+00	56%	73%	34--657	34--663
					5	COG0491	GloB, Zn-dependent hydrolases, including glyoxylases [General function prediction only].	35.99	1.66E-02	29%	44%	138--223	30--109	5	YP_593031	beta-lactamase-like	671.77	0.00E+00	52%	70%	34--658	38--667
					6	COG1234	ElaC, Metal-dependent hydrolases of the beta-lactamase superfamily III [General function prediction only].	35.80	1.93E-02	24%	48%	163--215	44--98	6	ZP_00753431	COG2015: Alkyl sulfatase and related hydrolases	620.93	4.82E-176	48%	67%	30--658	26--660
					7	cd00189	TPR, Tetratricopeptide repeat domain; typically contains 34 amino acids [WLF]-X(2)-[LIM]-[GAS]-X(2)-[YLF]-X(8)-[ASE]-X(3)-[FYL]-X(2)-[ASL]-X(4)-[PKE] is the consensus sequence; found in a variety of organisms including bacteria, cyanobacteria, yeast, fungi, plants, and humans in various subcellular locations; involved in a variety of functions including protein-protein interactions, but common features in the interaction partners have not been defined; involved in chaperone, cell-cycle, transciption, and protein transport complexes; the number of TPR motifs varies among proteins (1,3-11,13 15,16,19); 5-6 tandem repeats generate a right-handed helical structure with an amphipathic channel that is thought to accomodate an alpha-helix of a target protein; it has been proposed that TPR proteins preferably interact with WD-40 repeat proteins, but in many instances several TPR-proteins seem to aggregate to multi-protein complexes; examples of TPR-proteins include, Cdc16p, Cdc23p and Cdc27p components of the cyclosome/APC, the	33.90	7.00E-02	34%	50%	464--508	4--48	7	ZP_00750973	COG2015: Alkyl sulfatase and related hydrolases	620.54	6.29E-176	48%	67%	30--658	26--660
					8	COG2333	ComEC, Predicted hydrolase (metallo-beta-lactamase superfamily) [General function prediction only]	32.23	2.16E-01	25%	43%	137--224	57--138	8	ZP_00747297	COG2015: Alkyl sulfatase and related hydrolases	619.77	1.07E-175	48%	67%	31--658	27--660
					9	COG2813	RsmC, 16S RNA G1207 methylase RsmC [Translation, ribosomal structure and biogenesis].	31.39	3.98E-01	23%	45%	508--600	126--212	9	AAF93948	conserved hypothetical protein	619.00	1.83E-175	48%	67%	31--658	27--660
					10	COG0595	COG0595, Predicted hydrolase of the metallo-beta-lactamase superfamily [General function prediction only].	30.60	5.81E-01	23%	42%	135--189	23--83	10	ZP_00913757	Twin-arginine translocation pathway signal	605.52	2.09E-171	49%	66%	34--657	49--678
N007L	3122--2577	182	20,422	10.22		No Hit Found								1	AAC96449	A81L	108.61	9.58E-23	41%	62%	45--158	37--163
N010L	3666--3151	172	19,292	6.65		No Hit Found								1	AAC96452	A84L	71.63	1.10E-11	32%	55%	14--135	13--149
N012L	4563--3751	271	31,655	6.29	1	pfam03016	Exostosin, Exostosin family. The EXT family is a family of tumour suppressor genes. Mutations of EXT1 on 8q24.1, EXT2 on 11p11-13, and EXT3 on 19p have been associated with the autosomal dominant disorder known as hereditary multiple exostoses (HME). This is the most common known skeletal dysplasia. The chromosomal locations of other EXT genes suggest association with other forms of neoplasia. EXT1 and EXT2 have both been shown to encode a heparan sulphate polymerase with both D-glucuronyl (GlcA) and N-acetyl-D-glucosaminoglycan (GlcNAC) transferase activities. The nature of the defect in heparan sulphate biosynthesis in HME is unclear	45.82	6.04E-06	25%	51%	159--235	221--292	1	AAC96443	A75L	234.57	2.74E-60	41%	61%	8--268	7--276
					2	pfam01323	DSBA, DSBA-like thioredoxin domain. This family contains a diverse set of proteins with a thioredoxin-like structure pfam00085. This family also includes 2-hydroxychromene-2-carboxylate (HCCA) isomerase enzymes catalyse one step in prokaryotic polyaromatic hydrocarbon (PAH) catabolic pathways. This family also contains members with functions other than HCCA isomerisation, such as Kappa family GSTs, whose similarity to HCCA isomerases was not previously recognised. Some members of this family may have been mis-annotated in protein sequence databases	29.98	3.06E-01	17%	34%	43--161	8--138	2	NP_195517	catalytic	45.82	1.80E-03	27%	48%	162--268	284--384
														3	NP_195005	catalytic	44.67	4.02E-03	26%	49%	174--266	471--571
														4	BAC42936	unknown protein	44.67	4.02E-03	26%	49%	174--266	148--248
														5	ABA18110	exostosin family protein	43.51	8.95E-03	30%	44%	169--257	206--308
N014L	5025--4663	121	13,198	10.46	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	58.35	2.32E-10	37%	51%	4--106	1--94	1	AAC96568	contains cytidine and deoxycytidine deaminase Zn-binding region signature	159.84	2.40E-38	61%	83%	1--118	1--118
					2	COG0590	CumB, Cytosine/adenosine deaminases [Nucleotide transport and metabolism / Translation, ribosomal structure and biogenesis]	53.42	6.92E-09	33%	51%	2--106	10--105	2	AAR26853	FirrV-1-A29	50.83	1.57E-05	27%	50%	22--111	24--105
					3	pfam00383	dCMP_cyt_deam, Cytidine and deoxycytidylate deaminase zinc-binding region..	44.20	4.27E-06	31%	47%	4--106	7--101	3	AAX51127	tRNA-specific adenosine deaminase	50.06	2.67E-05	32%	49%	4--108	10--104
					4	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	37.92	3.09E-04	27%	45%	4--107	1--95	4	AAC68441	cytosine deaminase	49.68	3.49E-05	32%	49%	4--108	10--104
					5	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	31.78	2.25E-02	37%	51%	50--108	68--117	5	ZP_00679087	Cytidine/deoxycytidylate deaminase, zinc-binding region	45.44	6.58E-04	32%	44%	4--106	19--111
					6	COG0117	RibD, Pyrimidine deaminase [Coenzyme metabolism].	31.76	2.36E-02	24%	43%	7--109	13--104	6	GAA00608	unnamed protein product	44.67	1.12E-03	30%	46%	4--108	36--130
					7	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..	30.37	5.89E-02	20%	36%	6--108	3--95	7	NP_502546	JC8.4	44.28	1.47E-03	32%	48%	2--108	11--111
					8	pfam02041	Auxin_BP, Auxin binding protein..	28.08	3.02E-01	21%	50%	23--79	58--126	8	AAF73539	cytidine/deoxycytidylate deaminase family protein	44.28	1.47E-03	29%	46%	4--108	10--104
					9	COG2131	ComEB, Deoxycytidylate deaminase [Nucleotide transport and metabolism].	26.85	6.80E-01	45%	60%	88--108	106--126	9	ABB14795	cytidine/deoxycytidylate deaminase family protein	43.90	1.92E-03	29%	46%	6--108	9--101
														10	YP_428908	CMP/dCMP deaminase, zinc-binding	43.51	2.50E-03	31%	43%	6--108	9--101
N015L	5407--5111	99	10,616	10.54		No Hit Found								1	AAC96567	A199R	70.86	1.48E-11	47%	66%	1--71	1--72
N016R	5419--5907	163	18,543	8.04		No Hit Found								1	AAC96564	A196L	184.88	7.72E-46	61%	79%	23--156	18--151

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
N019R	5958--6755	266	29,728	5.00	1	cd00577	PCNA, Proliferating Cell Nuclear Antigen (PCNA) domain found in eukaryotes and archaea. These polymerase processivity factors play a role in DNA replication and repair. PCNA encircles duplex DNA in its central cavity, providing a DNA-bound platform for the attachment of the polymerase. The trimeric PCNA ring is structurally similar to the dimeric ring formed by the DNA polymerase processivity factors in bacteria (beta subunit DNA polymerase III holoenzyme) and in bacteriophages (catalytic subunits in T4 and RB69). This structural correspondence further substantiates the mechanistic connection between eukaryotic and prokaryotic DNA replication that has been suggested on biochemical grounds. PCNA is also involved with proteins involved in cell cycle processes such as DNA repair and apoptosis. Many of these proteins contain a highly conserved motif known as the PIP-box (PCNA interacting protein box) which contains the sequence Qxx[LIM]xxF[FY]. .	146.96	2.13E-36	30%	55%	12--261	1--245	1	AAC96561	similar to human PCNA, corresponds to Swiss-Prot Accession Number P12004	368.24	1.53E-100	68%	85%	9--261	7--259
					2	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 encirclino duplex DNA..	92.25	5.33E-20	28%	57%	9--132	1--125	2	XP_534355	PREDICTED: similar to proliferating cell nuclear antigen	158.69	1.84E-37	32%	55%	5--264	202--461
					3	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 encirclino duplex DNA..	79.98	2.97E-16	34%	55%	137--261	2--127	3	CAA55669	proliferative cell nuclear antigen	157.53	4.10E-37	34%	56%	9--264	1--256
					4	COG0592	DnaN, DNA polymerase sliding clamp subunit (PCNA homolog) [DNA replication, recombination, and repair]	56.08	4.36E-09	19%	37%	21--261	72--321	4	AAB27811	PCNA	156.76	6.99E-37	32%	57%	9--264	1--256
					5	COG1092	COG1092, Predicted SAM-dependent methyltransferases [General function prediction onlv].	31.44	1.30E-01	20%	31%	106--202	251--358	5	NP_172217	PCNA1 (PROLIFERATING CELLULAR NUCLEAR ANTIGEN); DNA binding / DNA polymerase processivity factor	156.76	6.99E-37	32%	57%	9--264	1--256
														6	AAG10077	proliferating cell nuclear antigen	155.99	1.19E-36	32%	55%	9--264	1--256
														7	CAA37243	unnamed protein product	155.61	1.56E-36	32%	55%	9--264	1--256
														8	BAB28355	unnamed protein product	155.61	1.56E-36	32%	55%	9--264	1--256
														9	NP_035175	proliferating cell nuclear antigen	155.61	1.56E-36	32%	55%	9--264	1--256
														10	NP_001029666	hypothetical protein LOC515499	155.22	2.04E-36	32%	55%	9--264	1--256
N020L	10628--6753	1292	141,764	11.55	1	cd03274	ABC_SMC4_euk, Eukaryotic SMC4 proteins; SMC proteins are large (approximately 110 to 170 kDa), and each is arranged into five recognizable domains. Amino-acid sequence homology of SMC proteins between species is largely confined to the amino- and carboxy-terminal globular domains. The amino-terminal domain contains a 'Walker A' nucleotide-binding domain (GxxGxGKS/T, in the single-letter amino-acid code), which by mutational studies has been shown to be essential in several proteins. The carboxy-terminal domain contains a sequence (the DA-box) that resembles a 'Walker B' motif, and a motif with homology to the signature sequence of the ATP-binding cassette (ABC) family of ATPases. The sequence homology within the carboxy-terminal domain is relatively high within the SMC1-SMC4 group, whereas SMC5 and SMC6 show some divergence in both of these sequences. In eukaryotic cells, the proteins are found as heterodimers of SMC1 paired with SMC3, SMC2 with SMC4, and SMC5 with SMC6 (formerly known as Rad18)..	33.04	2.36E-01	17%	51%	1073--1156	287--376	1	AAC96557	similar to SWI/SNF chromatin remodeling complex subunit OSA2	718.77	0.00E+00	39%	51%	2--1175	10--1139
					2	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	31.82	5.49E-01	34%	55%	1073--1155	772--846	2	T17682	hypothetical protein A192R - Chlorella virus PBCV-1	81.65	2.26E-13	43%	68%	1075--1175	103--200
					3	cd03275	ABC_SMC1_euk, Eukaryotic SMC1 proteins; SMC proteins are large (approximately 110 to 170 kDa), and each is arranged into five recognizable domains. Amino-acid sequence homology of SMC proteins between species is largely confined to the amino- and carboxy-terminal globular domains. The amino-terminal domain contains a 'Walker A' nucleotide-binding domain (GxxGxGKS/T, in the single-letter amino-acid code), which by mutational studies has been shown to be essential in several proteins. The carboxy-terminal domain contains a sequence (the DA-box) that resembles a 'Walker B' motif, and a motif with homology to the signature sequence of the ATP-binding cassette (ABC) family of ATPases. The sequence homology within the carboxy-terminal domain is relatively high within the SMC1-SMC4 group, whereas SMC5 and SMC6 show some divergence in both of these sequences. In eukaryotic cells, the proteins are found as heterodimers of SMC1 paired with SMC3, SMC2 with SMC4, and SMC5 with SMC6 (formerly known as Rad18)..	31.95	5.73E-01	19%	44%	1063--1170	245--353	3	AAK77699	ORF30, putative collagen	50.83	4.27E-04	23%	28%	98--482	666--1074
					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..	31.64	6.12E-01	29%	45%	1070--1168	67--171	4	NP_477523	wsv001	50.83	4.27E-04	23%	28%	98--482	666--1074
					5	cd03276	ABC_SMC6_euk, Eukaryotic SMC6 proteins; SMC proteins are large (approximately 110 to 170 kDa), and each is arranged into five recognizable domains. Amino-acid sequence homology of SMC proteins between species is largely confined to the amino- and carboxy-terminal globular domains. The amino-terminal domain contains a 'Walker A' nucleotide-binding domain (GxxGxGKS/T, in the single-letter amino-acid code), which by mutational studies has been shown to be essential in several proteins. The carboxy-terminal domain contains a sequence (the DA-box) that resembles a 'Walker B' motif, and a motif with homology to the signature sequence of the ATP-binding cassette (ABC) family of ATPases. The sequence homology within the carboxy-terminal domain is relatively high within the SMC1-SMC4 group, whereas SMC5 and SMC6 show some divergence in both of these sequences. In eukaryotic cells, the proteins are found as heterodimers of SMC1 paired with SMC3, SMC2 with SMC4, and SMC5 with SMC6 (formerly known as Rad18)..	31.16	8.12E-01	19%	43%	1073--1169	262--364	5	T17681	hypothetical protein a191R - Chlorella virus PBCV-1	59.31	1.20E-06	46%	65%	843--911	3--63
														6	NP_001015110	CG40323-PB.3	48.91	1.62E-03	23%	31%	96--472	3426--3778
														7	XP_783728	PREDICTED: similar to Protein transport protein Sec24C (SEC24-related protein C)	55.07	2.26E-05	25%	33%	190--498	6--311
														8	NP_571089	calymmin	52.37	1.47E-04	23%	32%	140--483	230--604
														9	AAH90694	Cmn protein	52.37	1.47E-04	23%	32%	140--483	222--596
														10	XP_394285	PREDICTED: similar to GA11046-PA	53.14	8.60E-05	27%	35%	192--481	124--394
N023L	13471--10658	938	106,276	7.86	1	COG0417	PolB, DNA polymerase elongation subunit (family B) [DNA replication, recombination. and reoairl.	346.65	6.28E-96	30%	48%	51--879	11--772	1	AAC96553	PBVC-1 DNA polymerase	1352.04	0.00E+00	72%	84%	30--932	1--913
					2	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..	337.35	4.53E-93	38%	55%	456--879	1--439	2	BAA35142	DNA polymerase	1347.03	0.00E+00	72%	83%	30--932	1--913
					3	smart00486	POLBc, DNA polymerase type-B family; DNA polymerase alpha, delta, epsilon and zeta chain (eukaryota), DNA polymerases in archaea, DNA polymerase II in e. coli, mitochondrial DNA polymerases and and virus DNA polvmerases .	292.12	1.89E-79	33%	51%	205--667	1--475	3	P30320	DNA polymerase	1341.25	0.00E+00	71%	83%	30--932	1--913
					4	cd00145	POLBc, DNA polymerase type-B family; DNA directed DNA polymerase. Posseses DNA binding, polymerase and 3'-5' exonuclease activitv..	270.64	5.27E-73	33%	49%	205--704	1--511	4	AAB49748	DNA polymerase	443.74	1.58E-122	96%	97%	489--718	1--230
					5	pfam03104	DNA_pol_B_exo, DNA polymerase family B, exonuclease domain. This domain has 3' to 5' exonuclease activity and adopts a ribonuclease H type fold..	194.94	3.39E-50	26%	41%	56--383	1--334	5	AAK28935	DNA polymerase	431.80	6.23E-119	100%	100%	497--711	1--215
														6	AAK28933	DNA polymerase	429.10	4.04E-118	99%	100%	497--711	1--215
														7	AAX86472	DNA polymerase	429.10	4.04E-118	99%	100%	498--711	1--214

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	AAK28936	DNA polymerase	427.56	1.17E-117	99%	99%	497–711	1–215
														9	XP_757605	hypothetical protein UM01458.1	411.38	8.71E-113	32%	49%	46–880	83–929
														10	XP_961558	hypothetical protein	408.30	7.37E-112	32%	52%	46–892	123–980
N029R	13538–14146	203	23,119	6.74	1	COG5283	COG5283, Phage-related tail protein [Function unknown].	31.97	5.05E-02	16%	36%	46–198	936–1079	1	AAC96971	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	139.43	6.66E-32	37%	57%	1–199	1–185
					2	COG5325	COG5325, t-SNARE complex subunit, syntaxin [Intracellular trafficking and secretion].	28.06	7.54E-01	27%	49%	41–86	161–206	2	AAC96545	A177R	65.86	9.36E-10	59%	70%	1–47	4–50
														3	AAC96972	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	62.39	1.04E-08	54%	69%	2–47	8–53
														4	AAC96643	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	58.92	1.14E-07	47%	59%	1–57	4–60
														5	AAC96818	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	57.38	3.33E-07	48%	62%	1–56	1–51
														6	AAC96447	A79R	56.61	5.68E-07	48%	68%	1–47	1–47
N030R	14295–15038	248	27,869	8.33		No Hit Found								1	AAC96377	A9R	222.63	9.12E-57	57%	79%	73–243	3–173
														2	AAC96972	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	133.65	5.57E-30	32%	50%	4–242	10–253
														3	AAC96818	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	131.34	2.77E-29	31%	52%	1–244	1–248
														4	AAC96545	A177R	131.34	2.77E-29	33%	54%	1–227	4–227
														5	AAC96643	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	127.87	3.06E-28	31%	51%	1–241	4–248
														6	AAC96447	A79R	122.48	1.29E-26	33%	52%	1–221	1–218
														7	AAU06304	hypothetical protein A275R	78.95	1.63E-13	30%	52%	78–241	4–167
														8	AAU06301	hypothetical protein A275R	77.80	3.63E-13	30%	52%	78–241	4–167
														9	AAU06302	hypothetical protein A275R	64.31	4.15E-09	31%	54%	110–242	1–132
														10	AAC96971	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	56.23	1.13E-06	46%	68%	1–47	1–47
N031L	15533–15027	169	19,307	4.42		No Hit Found								0	No Hit Found	No Hit Found						
N033L	17512–15860	551	63,615	6.12	1	cd00204	ANK, ankyrin repeats; ankyrin repeats mediate protein-protein interactions in very diverse families of proteins. The number of ANK repeats in a protein can range from 2 to over 20 (ankyrins, for example). ANK repeats may occur in combinations with other types of domains. The structural repeat unit contains two antiparallel helices and a beta-hairpin. repeats are stacked in a superhelical arrangement; this alignment contains 4 consecutive repeats	58.55	2.03E-09	40%	57%	395–466	54–126	1	XP_797753	PREDICTED: similar to ankyrin repeat domain protein 17 isoform b, partial	61.62	9.02E-08	20%	40%	22–479	451–898
					2	COG0666	Arp, FOG: Ankyrin repeat [General function prediction only].	39.10	1.28E-03	29%	42%	405–482	139–216	2	XP_788092	PREDICTED: similar to Ankyrin-2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid) partial	47.75	1.35E-03	31%	53%	22–109	1117–1198
					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	36.19	1.12E-02	44%	69%	412–444	1–33	3	XP_786001	PREDICTED: similar to Ankyrin-1 (Erythrocyte ankyrin) (Ankyrin R)	71.63	8.72E-11	20%	38%	22–469	1116–1524
					4	pfam04851	ResIII, Type III restriction enzyme, res subunit. This family represents the res subunit of type III restriction enzymes (EC:3.1.21.5).	30.32	5.73E-01	20%	28%	80–176	787–888	4	XP_682036	hypothetical protein AN8767.2	71.25	1.14E-10	25%	47%	237–477	580–786
					5	COG1381	RecO, Recombinational DNA repair protein (RecF pathway) [DNA replication, recombination, and repair].	30.33	6.27E-01	14%	30%	199–305	62–184	5	XP_787823	PREDICTED: similar to Ankyrin-1 (Erythrocyte ankyrin) (Ankyrin R)	48.52	7.90E-04	32%	52%	22–110	626–708
														6	AAS13899	ankyrin repeat domain protein	50.06	2.72E-04	24%	39%	236–466	732–947
														7	AAY54249	ankyrin domain protein	67.78	1.26E-09	28%	44%	263–479	291–505
														8	ZP_01314584	hypothetical protein Wendoof 01000602	67.78	1.26E-09	28%	44%	263–479	613–827
														9	XP_780328	PREDICTED: similar to Ankyrin-3 (ANK-3) (Ankyrin G)	45.44	6.69E-03	25%	43%	263–476	698–901
														10	AAY61232	Guanosine polyphosphate pyrophosphohydrolases/synthetases homolog	50.45	2.08E-04	26%	40%	263–481	718–934
N034L	18223–17576	216	24,527	8.64	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..	201.73	4.40E-53	37%	55%	1–210	1–216	1	AAC96983	similar to Synechocystis ORF s111635, corresponds to GenBank Accession Number D90903	275.02	1.17E-72	59%	79%	1–213	1–215
					2	COG1351	THY1, Predicted alternative thymidylate synthase [Nucleotide transport and metabolism].	135.92	2.82E-33	31%	44%	16–213	29–237	2	ZP_00588145	Thymidylate synthase complementing protein	221.09	2.00E-56	52%	68%	1–214	21–236
														3	YP_195228	thymidilate synthase	220.71	2.62E-56	53%	68%	1–214	3–213
														4	AAZ59097	Thymidylate synthase complementing protein ThyX	217.62	2.22E-55	54%	70%	12–213	6–210
														5	ZP_01006033	thymidylate synthase	216.47	4.94E-55	55%	72%	19–213	13–210
														6	ABB27663	Thymidylate synthase complementing protein ThyX	214.54	1.88E-54	51%	67%	1–214	1–216
														7	YP_214756	Td	214.16	2.45E-54	58%	71%	19–213	12–210
														8	ZP_00531300	Thymidylate synthase (FAD)	212.62	7.13E-54	52%	66%	1–213	1–215
														9	AAM72779	thymidylate synthase, flavin-dependent	212.23	9.31E-54	51%	67%	1–214	1–216
														10	ABB49306	Thymidylate synthase complementing protein ThyX	212.23	9.31E-54	53%	74%	19–213	13–210
N035R	18207–20141	645	70,502	6.81	1	COG0449	GlmS, Glucosamine 6-phosphate synthetase, contains amidotransferase and phosphosugar isomerase domains [Cell envelope biogenesis, outer membrane].	622.20	5.19E-179	44%	63%	51–643	1–597	1	BAD15299	glutamine:fructose-6-phosphate amidotransferase GFAT	731.10	0.00E+00	61%	78%	51–641	1–594
					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..	253.52	4.70E-68	47%	65%	52–263	1–215	2	AAC96468	PBCV-1 glucosamine synthetase	727.63	0.00E+00	60%	77%	51–641	1–593
					3	COG2222	AgaS, Predicted phosphosugar isomerases [Cell envelope biogenesis, outer membrane].	182.79	9.22E-47	29%	49%	298–633	2–329	3	CAE35165	glucosamine--fructose-6-phosphate aminotransferase	512.69	1.79E-143	46%	63%	51–643	1–610
					4	pfam00310	GATase 2, Glutamine amidotransferases class-II..	129.65	1.01E-30	36%	59%	52–183	1–136	4	CAE44992	glucosamine--fructose-6-phosphate aminotransferase	509.61	1.52E-142	46%	63%	51–643	1–610
					5	COG0034	PurF, Glutamine phosphoribosylpyrophosphate amidotransferase [Nucleotide transport and metabolism]	128.79	1.93E-30	31%	53%	48–278	1–232	5	YP_582342	glucosamine--fructose-6-phosphate aminotransferase, isomerizing	508.45	3.38E-142	45%	64%	51–643	1–612
					6	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..	124.86	2.45E-29	32%	53%	52–275	1–224	6	ZP_00942990	Glucosamine--fructose-6-phosphate aminotransferase (isomerizing)	508.06	4.41E-142	45%	64%	51–643	1–612
					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. GlcS synthesizes L-glutamate from ?-	112.11	1.89E-25	34%	48%	52–246	1–203	7	CAD13706	probable glucosamine--fructose-6-phosphate aminotransferase	504.21	6.37E-141	45%	64%	51–643	1–612
					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.	91.12	3.78E-19	37%	55%	333–460	3–131	8	YP_438548	glucosamine--fructose-6-phosphate aminotransferase, isomerizing	499.59	1.57E-139	44%	65%	51–643	1–605
					9	COG0367	AsnB, Asparagine synthase (glutamine-hydrolyzing) [Amino acid transport and metabolism].	76.63	8.19E-15	25%	44%	51–279	1–195	9	YP_547531	glucosamine--fructose-6-phosphate aminotransferase, isomerizing	497.66	5.96E-139	44%	63%	51–643	1–616

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	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	71.40	3.39E-13	29%	48%	52--246	1--163	10	ZP_00685730	Glucosamine-fructose-6-phosphate aminotransferase, isomerising	497.66	5.96E-139	44%	64%	51--643	1--605
N040R	20182--23058	959	96,991	6.54		No Hit Found								1	BAB83471	Vp260 like protein	1210.67	0.00E+00	65%	75%	1--957	1--955
														2	BAB83467	Vp260 like protein	598.59	3.95E-169	38%	55%	2--957	3--954
														3	BAB83468	Vp260 like protein	148.29	1.41E-33	27%	41%	27--568	346--997
														4	BAB83469	Vp260 like protein	529.64	2.25E-148	37%	52%	10--951	1--930
														5	BAB83470	Vp260 like protein	65.47	1.20E-08	28%	47%	1--179	640--798
														6	AAC96490	PBCV-1 Vp260 protein	61.23	2.27E-07	26%	42%	595--839	15--252
														7	AAA86307	glycoprotein Vp260	48.91	1.17E-03	32%	48%	751--863	32--153
														8	AAC96397	Asn/Thr/Ser/Val rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	74.71	1.98E-11	25%	40%	484--872	14--387
														9	AAC96382	Asn/Thr/Ser/Val rich protein	110.15	4.26E-22	22%	38%	102--858	12--779
														10	AAC96386	Asn/Thr/Ser/Ile rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	109.00	9.49E-22	22%	37%	24--831	153--950
N045R	23093--27454	1454	149,210	6.85		No Hit Found								1	BAB83469	Vp260 like protein	1417.52	0.00E+00	53%	65%	10--1452	1--1461
														2	BAB83468	Vp260 like protein	1399.42	0.00E+00	52%	64%	10--1452	1--1459
														3	BAB83470	Vp260 like protein	1367.44	0.00E+00	56%	68%	10--1290	1--1283
														4	BAB83467	Vp260 like protein	576.63	2.54E-162	35%	51%	4--1131	5--1141
														5	BAB83471	Vp260 like protein	545.81	4.79E-153	36%	52%	1--932	1--940
														6	AAC96490	PBCV-1 Vp260 protein	101.68	2.39E-19	28%	40%	27--464	546--1011
														7	AAA86307	glycoprotein Vp260	84.34	3.95E-14	26%	40%	455--855	8--398
														8	AAC96382	Asn/Thr/Ser/Val rich protein	81.26	3.34E-13	20%	35%	16--599	573--1231
														9	AAC96386	Asn/Thr/Ser/Ile rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	82.03	1.96E-13	21%	39%	88--708	622--1242
														10	AAC96397	Asn/Thr/Ser/Val rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	57.77	3.96E-06	23%	38%	27--282	551--812
N052R	27493--31881	1463	149,949	6.74	1	pfam03961	DUF342, Protein of unknown function (DUF342). This family of bacterial proteins has no known function. The proteins are in the region of 500-600 amino acid residues in length..	33.39	2.37E-01	29%	46%	441--595	189--318	1	BAB83469	Vp260 like protein	1630.92	0.00E+00	58%	71%	10--1461	1--1464
					2	COG5520	COG5520, O-Glycosyl hydrolase [Cell envelope biogenesis, outer membrane].	31.16	9.33E-01	30%	44%	518--584	351--412	2	BAB83468	Vp260 like protein	1603.57	0.00E+00	56%	70%	10--1461	1--1462
					3	COG1315	COG1315, Predicted polymerase, most proteins contain PALM domain, HD hydrolase domain and Zn-ribbon domain [DNA replication, recombination. and repair].	31.06	9.75E-01	26%	42%	23--128	256--347	3	BAB83470	Vp260 like protein	1542.32	0.00E+00	60%	72%	10--1329	1--1328
														4	BAB83467	Vp260 like protein	602.44	4.35E-170	35%	52%	4--1128	5--1135
														5	BAB83471	Vp260 like protein	567.77	1.19E-159	38%	53%	1--940	1--949
														6	AAC96490	PBCV-1 Vp260 protein	203.37	5.88E-50	28%	40%	8--883	42--999
														7	AAA86307	glycoprotein Vp260	58.92	1.79E-06	22%	35%	749--1411	43--677
														8	AAC96397	Asn/Thr/Ser/Val rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	60.46	6.15E-07	23%	36%	469--956	9--565
														9	AAC96382	Asn/Thr/Ser/Val rich protein	54.30	4.41E-05	21%	37%	746--1451	79--765
														10	AAC96386	Asn/Thr/Ser/Ile rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	49.29	1.42E-03	21%	35%	745--1461	24--768
N059R	31996--36417	1474	151,244	6.52	1	pfam03961	DUF342, Protein of unknown function (DUF342). This family of bacterial proteins has no known function. The proteins are in the region of 500-600 amino acid residues in length..	31.85	5.63E-01	29%	46%	441--595	189--318	1	BAB83468	Vp260 like protein	1989.16	0.00E+00	67%	79%	10--1472	1--1462
					2	COG5520	COG5520, O-Glycosyl hydrolase [Cell envelope biogenesis, outer membrane].	31.93	5.74E-01	30%	44%	518--584	351--412	2	BAB83469	Vp260 like protein	1955.64	0.00E+00	67%	77%	10--1472	1--1464
														3	BAB83470	Vp260 like protein	1516.52	0.00E+00	59%	72%	10--1336	1--1323
														4	BAB83467	Vp260 like protein	632.10	5.16E-179	37%	54%	4--1097	5--1103
														5	BAB83471	Vp260 like protein	588.57	6.55E-166	38%	55%	1--940	1--949
														6	AAC96490	PBCV-1 Vp260 protein	192.97	8.01E-47	29%	42%	91--941	15--839
														7	AAA86307	glycoprotein Vp260	48.14	3.19E-03	21%	35%	749--1472	43--664
														8	AAC96397	Asn/Thr/Ser/Val rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	65.47	1.93E-08	23%	39%	17--282	472--812
														9	AAC96382	Asn/Thr/Ser/Val rich protein	95.13	2.27E-17	25%	40%	4--599	622--1231
														10	AAC96386	Asn/Thr/Ser/Ile rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	122.09	1.74E-25	23%	38%	21--939	15--983
N067R	36531--41027	1499	156,160	6.02	1	COG1664	CcmA, Integral membrane protein CcmA involved in cell shape determination [Cell envelope biogenesis. outer membrane].	33.75	1.66E-01	19%	41%	492--610	20--120	1	BAB83467	Vp260 like protein	1314.67	0.00E+00	59%	68%	1--1169	1--1163
					2	COG1315	COG1315, Predicted polymerase, most proteins contain PALM domain, HD hydrolase domain and Zn-ribbon domain [DNA replication, recombination. and repair].	31.83	5.64E-01	27%	40%	498--617	185--317	2	BAB83469	Vp260 like protein	167.93	2.81E-39	29%	42%	10--603	327--842
														3	BAB83468	Vp260 like protein	113.62	6.29E-23	29%	42%	3--336	397--783
														4	BAB83470	Vp260 like protein	81.26	3.46E-13	35%	50%	3--194	641--813
														5	BAB83471	Vp260 like protein	444.12	2.03E-122	33%	47%	3--944	2--944
														6	AAC96490	PBCV-1 Vp260 protein	56.61	9.13E-06	31%	46%	10--166	856--1018
														7	AAA86307	glycoprotein Vp260	54.68	3.47E-05	30%	46%	10--166	670--832
														8	AAC96397	Asn/Thr/Ser/Val rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	46.60	9.44E-03	31%	46%	10--148	670--805
														9	AAC96382	Asn/Thr/Ser/Val rich protein	56.23	1.19E-05	20%	35%	756--1490	145--828
														10	AAC35172	cell surface antigen rOmpA	53.14	1.01E-04	24%	38%	388--627	230--470
N074R	41091--42326	412	46,415	6.11	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 PBCV-1 the major capsid protein is a glycoprotein	280.66	2.04E-76	36%	54%	5--406	3--443	1	AAC27494	putative capsid protein	476.09	1.04E-132	97%	99%	21--250	1--230
					2	COG0770	MurF, UDP-N-acetylmuramyl pentapeptide synthase [Cell envelope biogenesis. outer membrane].	37.60	3.15E-03	32%	47%	10--68	382--441	2	AAC96798	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	294.28	5.63E-78	40%	55%	5--410	3--437
														3	BAA76601	major capsid protein MCP1	291.97	2.80E-77	40%	55%	5--410	3--437
														4	BAA22198	major capsid protein Vp54	286.96	8.99E-76	40%	54%	5--410	3--437
														5	BAA76600	major capsid protein	285.42	2.62E-75	40%	54%	5--410	3--436
														6	AAC27492	major capsid protein Vp49	281.57	3.78E-74	41%	54%	5--410	3--432
														7	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	265.00	3.66E-69	39%	54%	27--410	1--413
														8	1M3Y_D	Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid Containing, Dna Virus	265.00	3.66E-69	39%	54%	27--410	1--413
														9	AAC96379	contains aminoacyl-tRNA synthetase class-II signature	229.57	1.70E-58	34%	53%	5--410	3--403
														10	AAC96917	similar to Chlorella virus PBCV-1 ORF A11L, corresponds to GenBank Accession Number U42580	229.18	2.22E-58	33%	53%	5--410	2--400
N079R	42805--43311	169	18,821	11.55		No Hit Found								0	No Hit Found	No Hit Found						
N081L	43657--43367	97	10,019	11.01		No Hit Found								1	AAC96569	A201L	124.41	1.14E-27	78%	91%	25--95	24--94
N082R	43922--44515	198	22,100	8.10		No Hit Found								1	AAC96669	A301L	59.69	6.39E-08	30%	52%	3--132	2--135
N084R	44545--45186	214	23,739	4.57		No Hit Found								1	AAC96666	A298L	150.60	3.25E-35	40%	60%	1--208	1--220
N086L	45704--45258	149	16,563	11.87	1	COG5406	COG5406, Nucleosome binding factor SPN, SPT16 subunit [Transcription / DNA replication, recombination, and repair / Chromatin structure and dynamics].	28.91	2.47E-01	23%	48%	68--110	442--482	1	AAC96664	A296R	80.88	1.44E-14	64%	78%	62--126	72--135
N087R	45698--46732	345	39,362	9.75	1	pfam01374	Glyco_hydro_46, Glycosyl hydrolase family 46. This family are chitosanase enzymes	238.79	6.89E-64	40%	51%	117--324	1--215	1	BAA20342	vChta-1	429.10	1.12E-118	66%	77%	18--335	5--321

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	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..	203.65	2.15E-53	31%	45%	107--324	1--220	2	AAC96660	PBCV-1 chitosanase	426.40	7.28E-118	65%	77%	18--335	5--321
					3	cd00985	Maf_Ham1, Maf_Ham1. Maf, a nucleotide binding protein, has been implicated in inhibition of septum formation in eukaryotes, bacteria and archaea. A Ham1-related protein from Methanococcus jannaschii is a novel NTPase that has been shown to hydrolyze nonstandard nucleotides, such as hypoxanthine/xanthine NTP, but not standard nucleotides	31.38	1.98E-01	21%	34%	56--168	8--120	3	AAA19865	chitosanase precursor	86.27	1.78E-15	28%	44%	93--329	17--266
					4	pfam06798	PrkA, PrkA serine protein kinase. This is a family of PrkA bacterial and archaeal serine kinases approximately 630 residues long. PrkA possesses the A-motif of nucleotide-binding proteins and exhibits distant homology to eukaryotic protein kinases. Note that many family members are hypothetical	30.66	3.07E-01	25%	45%	20--117	522--610	4	BAA94840	chitosanase	86.27	1.78E-15	29%	47%	96--304	41--252
					5	COG0632	RuvA, Holliday junction resolvasome, DNA-binding subunit [DNA replication, recombination, and repair].	30.21	3.79E-01	25%	46%	39--121	102--181	5	CAB14630	chitosanase	85.50	3.04E-15	29%	41%	107--328	37--264
														6	1CHK_B	Chain B, Streptomyces N174 Chitosanase Ph5.5 298k	83.96	8.86E-15	29%	46%	114--329	12--226
														7	P48846	Chitosanase precursor	82.80	1.97E-14	29%	46%	101--329	39--266
														8	ZP_00908039	chitosanase	82.42	2.58E-14	31%	48%	117--324	38--246
														9	ABC17783	secreted chitosanase precursor	81.65	4.40E-14	28%	46%	107--324	51--274
														10	BAB19276	chitosanase	80.49	9.79E-14	27%	41%	107--328	38--265
N089L	48009--46927	361	39,928	8.72		No Hit Found								1	AAC96654	A286R	362.46	1.39E-98	51%	69%	17--344	22--362
N091R	48048--48560	171	18,329	8.00	1	pfam00080	Sod_Cu, Copper/zinc superoxide dismutase (SODC). superoxide dismutases (SODs) catalyse the conversion of superoxide radicals to hydrogen peroxide and molecular oxygen. Three evolutionarily distinct families of SODs are known, of which the copper/zinc-binding family is one. Defects in the human SOD1 gene cause familial amyotrophic lateral sclerosis (Lou Gehrig's disease). Structure is an eight-stranded beta sandwich. similar to the immunoglobulin fold..	177.74	5.41E-46	58%	73%	16--165	1--152	1	AAC96613	contains Cu/Zn superoxide dismutase signatures 1 and 2; similar to Neurospora Cu/Zn-superoxide dismutase, corresponds to Swiss-Prot Accession Number P07509	261.15	9.84E-69	76%	85%	10--169	28--187
					2	cd00305	Cu-Zn_Superoxide_Dismutase, Copper/zinc superoxide dismutase (SOD). superoxide dismutases catalyse the conversion of superoxide radicals to molecular oxygen. Three evolutionarily distinct families of SODs are known, of which the copper/zinc-binding family is one. Defects in the human SOD1 gene causes familial amyotrophic lateral sclerosis (Lou Gehrig's disease). Cytoplasmic and periplasmic SODs exist as dimers, whereas chloroplastic and extracellular enzymes exist as tetramers. Structure supports independent functional evolution in prokaryotes (P-class) and eukaryotes (E-class) [PMID:.8176730]..	163.51	1.16E-41	53%	69%	18--158	2--144	2	Q751L8	Superoxide dismutase	177.56	1.43E-43	57%	71%	16--164	1--153
														3	NP_012638	Cu, Zn superoxide dismutase; some mutations are analogous to those that cause ALS (amvotrophic lateral sclerosis) in humans; Sod1p	176.79	2.44E-43	58%	72%	16--162	1--151
														4	XP_503850	hypothetical protein	176.02	4.16E-43	59%	70%	16--162	1--151
														5	P09670	Superoxide dismutase	175.64	5.44E-43	58%	71%	18--164	3--150
														6	1YSO	Yeast Cu, Zn Superoxide Dismutase With The Reduced Bridge Broken	174.87	9.27E-43	58%	72%	17--162	1--150
														7	XP_445379	unnamed protein product	174.87	9.27E-43	57%	70%	16--162	1--151
														8	AAS54170	AGL321Wp	174.10	1.58E-42	56%	71%	17--164	21--172
														9	AAX84946	Cu/Zn superoxide dismutase	174.10	1.58E-42	57%	70%	18--165	4--152
														10	1F1A_A	Chain A, Crystal Structure Of Yeast H48q Cuznsod Fals Mutant Analog	173.71	2.07E-42	57%	71%	16--162	1--151
N092L	49150--48557	198	23,569	8.76	1	COG0406	GpmB, Fructose-2,6-bisphosphatase [Carbohydrate transport and metabolism].	35.18	6.20E-03	28%	42%	4--82	1--80	1	AAC96665	A297L	195.67	7.46E-49	53%	72%	8--176	3--171
					2	COG2609	AceE, Pyruvate dehydrogenase complex, dehydrogenase (E1) component [Energy production and conversion].	31.41	7.99E-02	38%	57%	15--56	19--61	2	BAE64611	unnamed protein product	57.00	4.14E-07	27%	50%	8--121	6--119
					3	pfam03071	GNT-I, GNT-I family. Alpha-1,3-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase (GNT-I, GLCNAC-T I) EC:2.4.1.101 transfers N-acetyl-D-glucosamine from UDP to high-mannose glycoprotein N-oligosaccharide. This is an essential step in the synthesis of complex or hybrid-type N-linked oligosaccharides. The enzyme is an integral membrane protein localised to the Golgi apparatus, and is probably distributed in all tissues. The catalytic domain is located at the C-terminus..	28.39	6.84E-01	18%	39%	13--92	60--142	3	XP_750408	hypothetical protein Afu1g06590	52.37	1.02E-05	26%	50%	8--121	26--140
					4	COG2062	SixA, Phosphohistidine phosphatase SixA [Signal transduction mechanisms].	27.94	7.67E-01	21%	36%	5--106	1--107	4	BAB49344	mlr2143	49.68	6.61E-05	29%	47%	8--121	9--123
														5	XP_748836	hypothetical protein Afu7g06750	48.14	1.92E-04	24%	43%	8--168	22--174
N093R	49178--50020	281	31,162	6.76	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..	185.80	4.42E-48	32%	47%	2--275	1--297	1	AAC96652	PBCV-1 amidase	314.69	2.23E-84	54%	71%	1--278	1--279
					2	COG3049	COG3049, Penicillin V acylase and related amidases [Cell envelope biogenesis, outer membrane].	167.43	1.47E-42	28%	45%	1--278	22--323	2	ZP_00318651	COG3049: Penicillin V acylase and related amidases	141.74	2.59E-32	33%	51%	39--276	61--304
					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	141.29	9.63E-35	31%	48%	2--275	1--298	3	AAP08002	Choloylglycine hydrolase	141.35	3.38E-32	35%	51%	49--275	68--298
					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..	107.59	1.40E-24	28%	39%	2--268	1--270	4	ZP_01094011	choloylglycine hydrolase-like protein	140.20	7.52E-32	33%	47%	2--276	22--327
					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)	75.04	9.35E-15	25%	44%	2--261	1--270	5	AAT60567	choloylglycine hydrolase	138.27	2.86E-31	33%	51%	49--275	68--298
														6	ZP_01187466	Choloylglycine hydrolase	138.27	2.86E-31	33%	51%	49--275	68--298
														7	ZP_00238968	choloylglycine hydrolase family protein	136.73	8.32E-31	33%	51%	49--275	68--298
														8	NP_980098	choloylglycine hydrolase family protein	132.49	1.57E-29	32%	50%	49--275	68--298
														9	ABB11030	Penicillin amidase	132.11	2.05E-29	28%	45%	1--278	1--315
														10	AAL51724	CHOLOYLGLYCINE HYDROLASE	131.34	3.49E-29	30%	44%	2--276	37--337
N095L	51088--50195	298	33,002	6.91	1	COG0388	COG0388, Predicted amidohydrolase [General function prediction only].	182.28	5.50E-47	35%	52%	1--296	1--272	1	AAC96446	contains ATP/GTP-binding site motif A; similar to rat beta-alanine synthetase, corresponds to Swiss-Prot Accession Number Q03248	421.01	2.43E-116	68%	79%	2--295	4--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..	145.10	9.22E-36	35%	53%	4--177	1--174	2	ZP_01112744	probable hydratase	280.80	3.93E-74	49%	64%	2--295	3--283
					3	COG0815	Lnt, Apolipoprotein N-acyltransferase [Cell envelope biogenesis, outer membrane].	51.94	9.70E-08	21%	35%	35--215	269--451	3	AAK75046	carbon-nitrogen hydrolase family protein	278.87	1.49E-73	46%	65%	5--294	6--289
					4	pfam01457	Peptidase_M8, Leishmanolysin..	31.87	9.54E-02	27%	47%	147--198	167--212	4	AAK99627	Beta-alanine synthase or beta-ureidopropionase	276.94	5.67E-73	46%	65%	5--294	6--289
					5	pfam03639	Glyco_hydro_81, Glycosyl hydrolase family 81. Family of eukaryotic beta-1,3-glucanases..	31.17	1.79E-01	21%	35%	173--278	243--353	5	BAC60037	putative carbon-nitrogen hydrolase	275.40	1.65E-72	49%	65%	6--291	8--280
														6	ZP_01259147	putative carbon-nitrogen hydrolase	275.40	1.65E-72	49%	64%	6--291	8--280
														7	ZP_00831984	COG0388: Predicted amidohydrolase	273.48	6.27E-72	46%	63%	2--294	3--285
														8	CAB45873	beta-alanine synthase	273.09	8.19E-72	47%	63%	4--295	9--291

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	CAG77171	putative carbon-nitrogen hydrolase	272.71	1.07E-71	47%	62%	2--294	3--285
														10	ABB08352	Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase	272.32	1.40E-71	48%	62%	4--295	7--294
N099R	51143--52117	325	38,567	10.78	1	COG5077	COG5077, Ubiquitin carboxyl-terminal hydrolase [Posttranslational modification, protein turnover, chaperones].	29.20	8.06E-01	21%	40%	67--256	529--722	1	AAC96611	A243R	464.54	2.21E-129	73%	84%	22--323	1--302
N101L	52394--52128	89	10,192	7.21		No Hit Found								0	No Hit Found	No Hit Found						
N102R	52475--52927	151	17,027	9.24	1	pfam04925	SHQ1, SHQ1 protein. S. cerevisiae SHQ1 protein is required for SnoRNAs of the box H/ACA Quantitative accumulation (unpublished)..	27.60	6.83E-01	27%	42%	78--137	45--105	1	AAU06294	hypothetical protein A253R	78.18	9.24E-14	35%	56%	25--147	20--146
														2	AAC96621	A253R	78.18	9.24E-14	35%	57%	25--147	20--146
														3	AAU06289	hypothetical protein A253R	77.80	1.21E-13	34%	56%	25--149	20--148
														4	AAU06293	hypothetical protein A253R	55.45	6.41E-07	39%	61%	25--97	20--91
														5	AAU06292	hypothetical protein A253R	55.07	8.38E-07	40%	62%	25--93	20--87
														6	AAU06291	hypothetical protein A253R	52.76	4.16E-06	38%	58%	25--97	20--91
														7	AAU06290	hypothetical protein A253R	50.45	2.06E-05	39%	60%	25--90	20--84
N103R	52977--53915	313	35,392	10.08	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.	64.64	1.55E-11	41%	61%	172--307	1--113	1	AAA88832	unknown	181.03	4.57E-44	35%	51%	1--309	1--242
					2	smart00497	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished). .	39.35	5.75E-04	40%	63%	254--309	1--53	2	AAC96973	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	175.64	1.92E-42	36%	50%	1--308	1--224
					3	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	38.52	1.23E-03	29%	49%	1--71	1--70	3	AAC96655	PBCV-1 33kd peptide	116.32	1.38E-24	40%	55%	11--179	15--192
					4	pfam07453	NUMOD1, NUMOD1 domain..	36.95	3.61E-03	48%	67%	254--288	1--34	4	AAC96862	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	64.70	4.77E-09	37%	58%	170--280	97--196
														5	AAC49248	ORF211	62.77	1.81E-08	50%	69%	120--178	93--151
														6	NP_899393	SeqD	77.03	9.29E-13	26%	43%	1--300	1--231
														7	AAC49244	ORF301	59.31	2.00E-07	44%	63%	190--247	176--240
														8	AAK09365	intron encoded BmoI	45.44	2.99E-03	34%	47%	169--251	102--195
														9	AAT53588	group I intron GIY-YIG endonuclease	62.00	3.06E-08	30%	49%	139--276	100--215
														10	ZP_00391680	COG0532: Translation initiation factor 2 (IF-2; GTPase)	62.00	3.09E-08	30%	49%	139--276	75--190
N106L	54593--53910	228	25,018	8.41		No Hit Found								1	AAC96630	A262L	120.94	3.20E-26	47%	66%	111--215	2--103
														2	AAC96631	A263L	43.51	6.50E-03	41%	50%	25--91	66--123
N107L	54836--54615	74	7,911	10.56		No Hit Found								1	AAC96633	A265L	90.12	2.30E-17	62%	74%	2--71	101--170
N109L	55366--54983	128	14,609	7.93	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	32.12	1.87E-02	33%	48%	3--46	2--54	1	AAC96633	A265L	68.55	7.31E-11	66%	80%	79--120	1--42
N110R	55489--57474	662	73,068	9.97	1	pfam02705	K_trans, K+ potassium transporter. This is a family of K+ potassium transporters that are conserved across phyla, having both bacterial (KUP), yeast (HAK), and plant (AtKT) sequences as members..	375.38	1.13E-104	36%	57%	31--594	2--587	1	NP_567404	HAK5; potassium ion transporter	324.32	9.33E-87	33%	53%	9--592	38--640
					2	COG3158	Kup, K+ transporter [Inorganic ion transport and metabolism].	369.16	6.95E-103	34%	53%	27--660	14--627	2	NP_194095	TRH1 (TINY ROOT HAIR 1); potassium ion transporter	318.93	3.92E-85	32%	52%	28--598	11--605
					3	COG2971	COG2971, Predicted N-acetylglucosamine kinase [Carbohydrate transport and metabolism].	37.56	4.67E-03	17%	34%	89--213	6--133	3	BAD87337	putative potassium transporter	310.46	1.39E-82	31%	48%	26--660	73--748
														4	CAD20319	putative potassium transporter	307.76	9.04E-82	31%	51%	28--621	21--649
														5	CAD20577	putative potassium transporter	307.38	1.18E-81	34%	53%	21--582	85--669
														6	XP_483290	putative high-affinity potassium transporter	306.22	2.63E-81	31%	51%	28--629	74--693
														7	CAB79319	putative potassium transport protein	305.83	3.44E-81	31%	51%	28--598	26--632
														8	BAE93236	potassium transporter	305.06	5.86E-81	32%	51%	27--598	20--616
														9	BAE93348	potassium transporter	303.91	1.31E-80	32%	52%	21--587	107--698
														10	AAX13997	putative high-affinity potassium transporter protein	303.52	1.71E-80	33%	52%	26--582	14--601
N115L	57840--57478	121	13,617	4.55		No Hit Found								0	No Hit Found	No Hit Found						
N116R	57916--58236	107	11,590	8.04	1	pfam03833	PolC_DP2, DNA polymerase II large subunit DP2..	27.98	2.76E-01	53%	60%	77--92	632--647	1	YP_619251	hypothetical protein Ldb1404	42.36	5.68E-03	39%	56%	54--99	58--103
					2	pfam05479	PsaN, Photosystem I reaction centre subunit N (PSAN or PSI-N). This family contains several Photosystem I reaction centre subunit N (PSI-N) proteins. The protein has no known function although it is localised in the thylakoid lumen. PSI-N is a small extrinsic subunit at the lumen side and is very likely involved in the docking of plastocyanin	27.00	5.64E-01	38%	58%	79--105	92--118	2	AAV85705	At2g33790	41.97	7.42E-03	40%	54%	59--102	47--90
N117R	58250--58678	143	16,313	11.34	1	COG1422	COG1422, Predicted membrane protein [Function unknown].	28.66	3.32E-01	25%	55%	46--93	62--106	1	AAC96504	a136R	57.00	2.21E-07	32%	50%	1--141	1--146
N119L	59913--58864	350	38,855	4.65	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	143.48	2.96E-35	36%	50%	88--343	1--236	1	AAC96462	PBCV-1 beta-1,3 glucanase	416.77	5.90E-115	56%	71%	9--345	12--361
					2	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..	93.52	3.36E-20	34%	47%	92--343	1--216	2	P23903	Glucan endo-1,3-beta-glucosidase A1 precursor ((1->3)-beta-glucan endohydrolase) ((1->3)-beta-glucanase A1)	118.24	4.34E-25	31%	45%	64--343	400--678
					3	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)	76.10	6.22E-15	30%	44%	85--343	1--256	3	YP_435911	Beta-glucanase/Beta-glucan synthetase	115.55	2.81E-24	32%	45%	77--343	322--571
					4	COG2273	SKN1, Beta-glucanase/Beta-glucan synthetase [Carbohydrate transport and metabolism].	61.61	1.56E-10	25%	39%	74--348	29--267	4	AAC60453	beta-1,3-glucanase	113.24	1.40E-23	31%	44%	84--343	424--681
					5	pfam00722	Glyco_hydro_16, Glycosyl hydrolases family 16..	59.47	5.68E-10	28%	45%	190--342	45--182	5	ZP_01113030	hypothetical protein MED297_09826	112.08	3.11E-23	31%	45%	84--343	33--297
					6	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	55.70	7.90E-09	24%	37%	89--329	2--283	6	ZP_01351075	Glucan endo-1,3-beta-D-glucosidase	112.08	3.11E-23	29%	43%	68--343	38--331

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	cd02183	GH16_GPI_glucanosyltransferase, GPI (glycosylphosphatidylinositol) - glucanosyltransferase is a GPI-anchored membrane protein present in the fungal cell wall that is thought to play an important role in cell wall biosynthesis. GPI-glucanosyltransferase belongs to a family of glycosyl hydrolases that 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	38.68	1.04E-03	28%	49%	190--283	58--144	7	YP_563606	glycoside hydrolase, family 16	111.69	4.06E-23	31%	44%	76--343	39--325
					8	cd02175	GH16_lichenase, Lichenase, also known as 1,3-1,4-beta-glucanase, is a member of glycosyl hydrolase family 16, that specifically cleaves 1,4-beta-D-glucosidic bonds in mixed-linked beta glucans that also contain 1,3-beta-D-glucosidic linkages. Natural substrates of beta-glucanase are beta-glucans from grain endosperm cell walls or lichenan from the Islandic moss, Cetraria islandica. This protein is found not only in bacteria but also in anaerobic fungi. This domain includes two seven-stranded antiparallel beta-sheets that are adjacent to one another forming a compact, jellyroll beta-sandwich structure	37.58	2.12E-03	29%	46%	126--281	30--157	8	ZP_00637494	Glucan endo-1,3-beta-D-glucosidase	109.00	2.63E-22	31%	43%	85--343	48--325
					9	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 bacteroidetes. This domain adopts a curved beta-sandwich conformation, with a tunnel-shaped active site cavity, referred to as a jellyroll fold	35.47	1.09E-02	30%	46%	81--212	3--117	9	NP_763201	Beta-glucanase/Beta-glucan synthetase	107.84	5.86E-22	28%	43%	49--343	12--333
					10	cd02176	GH16_XET, Xyloglucan endotransglycosylases (XETs) cleave and religate xyloglucan polymers in plant cell walls via a transglycosylation mechanism. Thus, XET is a key enzyme in all plant processes that require cell wall remodeling. Even though the overall structure of XET is a curved beta-sandwich similar to other enzymes in the glycosyl hydrolase family 16, parts of its substrate binding cleft are more reminiscent of the distantly related glycosyl hydrolase family 7	33.64	3.40E-02	34%	46%	190--283	50--142	10	BAD63242	endo-beta-1,3-glucanase	107.46	7.66E-22	31%	44%	83--343	31--277
N124R	60078--61757	560	63,740	8.49	1	COG1215	COG1215, Glycosyltransferases, probably involved in cell wall biogenesis [Cell envelope biogenesis, outer membrane].	90.76	4.57E-19	25%	44%	39--390	10--328	1	AAD26641	hyaluronan synthase	832.40	0.00E+00	69%	82%	1--550	8--557
					2	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)	54.92	2.35E-08	23%	44%	186--345	193--373	2	AAC96466	PBCV-1 hyaluronic acid synthetase	828.55	0.00E+00	69%	82%	1--550	8--557
					3	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..	47.14	5.09E-06	24%	43%	86--283	2--167	3	AAD26643	hyaluronan synthase	826.62	0.00E+00	68%	82%	1--550	8--557
														4	NP_037285	hyaluronan synthase 2	218.78	4.51E-55	29%	48%	7--530	12--532
														5	AAI09072	Hyaluronan synthase 2	218.39	5.89E-55	29%	48%	7--530	12--532
														6	AAC53309	hyaluronan synthase 2	218.39	5.89E-55	29%	48%	7--530	12--532
														7	BAC37733	unnamed protein product	218.39	5.89E-55	29%	48%	7--530	12--532
														8	XP_528222	PREDICTED: similar to hyaluronan synthase 2	218.39	5.89E-55	29%	48%	7--530	24--544
														9	BAB63264	hyaluronic acid synthase 2	216.85	1.71E-54	29%	48%	7--530	12--532
														10	NP_999218	hyaluronan synthase 2	216.85	1.71E-54	29%	48%	7--530	12--532
N128R	61944--62906	321	36,035	8.75	1	pfam01331	mRNA_cap_enzyme, mRNA capping enzyme, catalytic domain. This family represents the ATP binding catalytic domain of the mRNA capping enzyme..	97.66	1.64E-21	25%	43%	50--223	1--192	1	AAC96471	PBCV-1 mRNA guanylyltransferase	326.25	9.18E-88	52%	68%	2--314	12--324
					2	COG5226	CEG1, mRNA capping enzyme, guanylyltransferase (alpha) subunit [RNA processing and modification].	75.46	8.40E-15	21%	37%	20--317	17--362	2	1CKN_B	Chain B, Structure Of Guanylylated Mrna Capping Enzyme Complexed With Gtp	323.94	4.55E-87	52%	68%	2--314	12--324
					3	pfam03919	mRNA_cap_C, mRNA capping enzyme, C-terminal domain..	50.74	2.46E-07	26%	44%	228--315	3--110	3	AAL98788	putative RNA guanylyltransferase	71.25	5.32E-11	26%	42%	44--311	221--479
					4	COG1947	IspE, 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate synthase [Lipid metabolism].	29.47	5.92E-01	27%	38%	135--183	50--98	4	CAG09212	unnamed protein product	69.71	1.55E-10	24%	43%	9--317	234--560
														5	AAT68133	mRNA capping enzyme	68.55	3.45E-10	23%	44%	8--317	234--566
														6	AAX82373	mRNA capping enzyme	68.17	4.50E-10	24%	39%	44--311	220--478
														7	NP_974263	mRNA guanylyltransferase/ phosphoprotein phosphatase/ protein tyrosine/serine/threonine phosphatase	67.78	5.88E-10	22%	41%	33--318	337--650
														8	P78587	mRNA capping enzyme alpha subunit (mRNA guanylyltransferase) (GTP--RNA guanylyltransferase) (GTase)	65.47	2.92E-09	25%	39%	47--316	41--373
														9	1P16_B	Chain B, Structure Of An Mrna Capping Enzyme Bound To The Phosphorylated Carboxyl-Terminal Domain Of Rna Polymerase Ii	63.93	8.50E-09	25%	38%	47--316	41--373
														10	AAH67387	RNA guanylyltransferase and 5'-phosphatase	62.39	2.47E-08	23%	42%	8--313	231--563
N131L	63284--62922	121	13,791	7.42	1	COG5170	CDC55, Serine/threonine protein phosphatase 2A, regulatory subunit [Signal transduction mechanisms]	26.99	6.38E-01	35%	54%	58--84	59--85	1	AAC96969	A645R	80.11	2.41E-14	34%	57%	1--117	1--121
N132L	64178--63336	281	31,701	5.23	1	pfam00443	UCH, Ubiquitin carboxyl-terminal hydrolase..	109.71	3.76E-25	20%	40%	4--275	6--312	1	AAC96473	contains ubquitin carboxy-terminal hydrolase active sites; similar to human ubiquitin carboxy-terminal hydrolase, coresponds to Swiss-Prot Accession Number Q09879	261.92	1.71E-68	45%	67%	1--277	1--280
					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	81.90	8.61E-17	17%	33%	4--276	2--320	2	XP_654739	ubiquitin carboxyl-terminal hydrolase	60.46	7.57E-08	23%	41%	4--278	68--347
					3	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..	74.61	1.15E-14	21%	35%	6--276	4--335	3	EAR96319	Ubiquitin carboxyl-terminal hydrolase family protein	55.07	3.18E-06	20%	42%	6--277	425--736
					4	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..	63.74	2.55E-11	19%	37%	5--275	5--303	4	NP_990257	ubiquitin specific protease 2	52.76	1.58E-05	23%	39%	6--275	23--348
					5	cd02663	Peptidase_C19G, A subfamily of Peptidase C19. Peptidase C19 contains ubiquitinyl hydrolases. They are intracellular peptidases that remove ubiquitin molecules from polyubiquinated peptides by cleavage of isopeptide bonds. They hydrolyze bonds involving the carboxyl group of the C-terminal Gly residue of ubiquitin. The purpose of the de-ubiquitination is thought to be editing of the ubiquitin conjugates, which could rescue them from degradation, as well as recycling of the ubiquitin. The ubiquitin/proteasome system is responsible for most protein turnover in the mammalian cell, and with over 50 members, family C19 is one of the largest families of peptidases in the human genome..	44.15	2.15E-05	37%	63%	213--261	278--332	5	EAS01964	Ubiquitin carboxyl-terminal hydrolase family protein	51.60	3.52E-05	21%	38%	6--259	186--497

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	cd02662	Peptidase_C19F, 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..	44.00	2.20E-05	24%	43%	5--106	3--106	6	XP_394836	PREDICTED: similar to ENSANGP00000018711	49.29	1.75E-04	20%	38%	6--275	513--835
					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..	41.46	1.15E-04	26%	45%	4--105	3--102	7	AAC13729	ubiquitin specific protease 41	48.91	2.28E-04	21%	39%	6--276	23--349
					8	cd02673	Peptidase_C19Q, 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..	41.26	1.40E-04	24%	37%	6--109	4--124	8	XP_692507	PREDICTED: similar to ubiquitin specific protease 2 isoform b	47.37	6.64E-04	25%	41%	6--182	244--429
					9	cd02657	Peptidase_C19A, A subfamily of Peptidase C19. Peptidase C19 contains ubiquitinyl hydrolases. They are intracellular peptidases that remove ubiquitin molecules from polyubiquinated peptides by cleavage of isopeptide bonds. They hydrolyse bonds involving the carboxyl group of the C-terminal Gly residue of ubiquitin. The purpose of the de-ubiquitination is thought to be editing of the ubiquitin conjugates, which could rescue them from degradation, as well as recycling of the ubiquitin. The ubiquitin/proteasome system is responsible for most protein turnover in the mammalian cell, and with over 50 members, family C19 is one of the largest families of peptidases in the human genome..	39.53	4.87E-04	23%	40%	4--105	2--104	9	AAC68865	ubiquitin specific protease 66	46.60	1.13E-03	21%	39%	6--276	50--376
					10	COG5560	UBP12, Ubiquitin C-terminal hydrolase [Posttranslational modification, protein turnover, chaperones].	36.53	3.84E-03	22%	40%	6--170	270--455	10	NP_014213	Ubiquitin-specific protease that deubiquitinates ubiquitin-protein moieties; may regulate silencing by acting on Sir4p; involved in posttranscriptionally regulating Gap1p and possibly other transporters; primarily located in the nucleus; Ubp10p	44.67	4.30E-03	26%	48%	5--112	365--477
N135L	65168--64227	314	34,973	7.74	1	COG1405	SUA7, Transcription initiation factor TFIIB, Brf1 subunit/Transcription initiation factor TFIIB [Transcription].	66.09	5.92E-12	22%	43%	44--311	7--282	1	AAC96475	similar to Pyrococcus woesei factor TFIIB homolog, corresponds to GenBank Accession Number X70668	207.22	5.98E-52	36%	58%	30--312	1--290
					2	COG5635	COG5635, Predicted NTPase (NACHT family) [Signal transduction mechanisms]	29.29	5.77E-01	22%	44%	199--286	542--629	2	XP_635486	transcription initiation factor IIB	47.37	7.92E-04	22%	38%	42--304	26--301
														3	AAT43923	transcription initiation factor IIB	46.98	1.03E-03	21%	39%	44--310	17--306
														4	EAM94502	Transcription factor TFIIB	46.21	1.77E-03	20%	39%	44--310	18--307
														5	XP_760941	hypothetical protein UM04794.1	45.44	3.01E-03	20%	36%	42--306	50--351
														6	YP_142604	putative transcription initiation factor IIB	45.05	3.93E-03	19%	39%	48--305	153--438
														7	BAB60249	transcription initiation factor B	44.67	5.14E-03	21%	38%	44--306	19--304
														8	AAS51927	ADR007Cp	44.28	6.71E-03	22%	38%	44--304	31--322
														9	NP_187644	RNA polymerase II transcription factor/ transcription factor/ zinc ion binding	43.90	8.76E-03	21%	39%	49--304	15--288
														10	NP_594229	hypothetical protein SPAC16E8.16	43.90	8.76E-03	22%	38%	42--297	23--306
N137L	65675--65166	170	19,353	9.14	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	29.36	2.65E-01	29%	47%	37--89	2--51	1	AAC96477	A109L	128.64	7.47E-29	58%	73%	66--168	1--103
N140L	67579--65720	620	69,038	10.90	1	smart00220	S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-specific kinase subfamily. .	48.67	2.01E-06	33%	48%	63--257	4--151	1	AAC96646	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662	501.90	3.03E-140	45%	59%	1--618	1--609
					2	pfam00069	Pkinase, Protein kinase domain..	47.97	3.08E-06	30%	50%	63--257	4--151	2	AAC96650	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055	486.11	1.72E-135	44%	58%	1--618	15--568
					3	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	45.58	2.08E-05	33%	48%	63--257	5--152	3	AAC96947	RPQT-like (9x)	309.69	2.20E-82	46%	63%	11--344	7--336
					4	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	40.90	5.30E-04	44%	67%	225--251	171--198	4	AAC96461	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563	56.61	3.36E-06	34%	50%	457--541	30--132
					5	COG2334	COG2334, Putative homoserine kinase type II (protein kinase fold) [General function prediction only]	39.17	1.40E-03	40%	57%	225--255	200--230	5	AAC96984	similar to Chlorella virus PBCV-1 ORF A282L, corresponds to GenBank Accession Number U42580	58.54	8.85E-07	72%	89%	474--510	5--41
					6	COG0515	SPS1, Serine/threonine protein kinase [General function prediction only / Signal transduction mechanisms / Transcription / DNA replication, recombination, and repair]	38.99	1.92E-03	24%	45%	61--257	3--155	6	AAC48204	Hypothetical protein ZK84.1	45.44	7.75E-03	30%	44%	387--467	477--565
					7	COG0510	CotS, Predicted choline kinase involved in LPS biosynthesis [Cell envelope biogenesis, outer membrane]	36.92	7.52E-03	22%	50%	225--309	156--232	7	AAV63985	hydroxyproline-rich glycoprotein VSP-3	45.82	5.93E-03	35%	51%	386--457	384--459
					8	cd00892	PI3Kc_related, Proteins related to phosphoinositide 3-kinase (PI3K), catalytic domain; All of the members have been found to possess lipid kinase activity. Many show Ser/Thr protein kinase activity. Many PI3K-related proteins are involved in cell-cycle checkpoints. They share two additional domains FATC, at the very C-terminus and FAT N-terminal to the PI3K-like domain	36.69	8.05E-03	47%	57%	223--251	179--209	8	AAB54132	Uncoordinated protein 89, isoform a	54.68	1.28E-05	27%	45%	381--552	1387--1573
					9	COG0661	AarF, Predicted unusual protein kinase [General function prediction only]	36.46	9.58E-03	43%	57%	223--259	285--320	9	AAV34801	Uncoordinated protein 89, isoform g	54.68	1.28E-05	27%	45%	381--552	1387--1573
					10	COG5032	TEL1, Phosphatidylinositol kinase and protein kinases of the PI-3 kinase family [Signal transduction mechanisms / Cell division and chromosome partitioning / Chromatin structure and dynamics / DNA replication, recombination, and repair / Intracellular trafficking and secretion]	35.83	1.62E-02	61%	70%	227--250	1947--1970	10	AAH58068	Serine/arginine repetitive matrix 1	46.21	4.54E-03	36%	58%	393--470	622--699
N146R	67649--68707	353	39,850	8.80	1	pfam03673	UPF0128, Uncharacterized protein family (UPF0128). The members of this family are about 240 amino acids in length. The proteins are as yet uncharacterized..	31.48	1.54E-01	24%	42%	207--336	14--133	1	AAC96497	similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903	85.89	2.42E-15	44%	58%	9--111	81--188
					2	pfam04638	Pox_O1, Pox virus protein O1. The function of these viral proteins is not known..	29.49	7.12E-01	25%	40%	289--343	204--257	2	AAC96920	similar to Chlorella virus PBCV-1 ORF A154L, corresponds to GenBank Accession Number U42580	70.48	1.05E-10	35%	46%	5--110	112--218
														3	AAC96522	A154L	276.17	1.26E-72	40%	60%	9--349	6--347
														4	NP_077492	EsV-1-7	71.25	6.17E-11	30%	47%	4--151	79--232
														5	CAA64974	QI74 protein	47.75	7.30E-04	25%	35%	8--208	182--363
														6	XP_636514	hypothetical protein DDB0188101	52.37	2.96E-05	29%	37%	38--190	238--386
														7	EAR82303	Ubiquitin carboxyl-terminal hydrolase family protein	50.83	8.62E-05	24%	42%	26--199	661--818
														8	XP_384145	hypothetical protein FG03969.1	48.52	4.28E-04	22%	32%	5--192	87--268

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	EAR99846	Neurohypophysial hormones, N-terminal Domain containing protein	46.60	1.63E-03	22%	30%	40--198	353--530
														10	XP_644291	hypothetical protein DDB0230011	47.37	9.53E-04	26%	38%	4--170	201--350
N150L	68982--68698	95	10,247	6.50		No Hit Found								0	No Hit Found	No Hit Found						
N151R	69070--69546	159	17,621	5.03		No Hit Found								1	AAC96979	similar to Chlorella virus CVK2 unknown ORF, corresponds to GenBank Accession Number D16505	145.98	3.59E-34	51%	72%	16--153	1--139
														2	AAC96980	similar to Chlorella virus CVK2 unknown ORF, corresponds to GenBank Accession Number D16505	116.32	3.05E-25	39%	67%	3--152	22--170
N152R	69573--69890	106	12,326	10.98		No Hit Found								1	AAC96489	A121R	136.73	2.14E-31	63%	77%	8--104	1--96
														2	EAR81541	hypothetical protein TTHERM_01719410	72.40	4.96E-12	41%	58%	13--93	13--90
														3	AAL73473	Tlr 6Fp protein	62.00	6.71E-09	44%	66%	32--90	32--88
														4	AAL73467	Tlr 6Fp protein	61.23	1.14E-08	42%	66%	32--90	32--88
														5	EAR81120	Tlr 6Fp protein, putative	61.23	1.14E-08	42%	66%	32--90	32--88
														6	EAS05551	hypothetical protein TTHERM_01666140	60.08	2.55E-08	40%	66%	32--90	32--88
														7	EAR80857	Tlr 6Fp protein, putative	59.69	3.33E-08	40%	66%	32--90	32--88
														8	AAL73477	Tlr 6Fp protein	54.68	1.07E-06	38%	61%	32--90	32--88
														9	EAR81408	Tlr 6Fp protein-related	43.90	1.89E-03	53%	68%	32--63	7--38
N155R	69922--74763	1614	164,975	5.26		No Hit Found								1	AAC96490	PBCV-1 Vp260 protein	318.16	1.81E-84	35%	46%	409--1090	19--739
														2	AAA86307	glycoprotein Vp260	83.96	5.76E-14	35%	49%	1--190	587--766
														3	AAC96491	contains a leucine zipper motif; similar to Bacillus subtilis phage P2A preneck appendage protein, corresponds to Swiss-Prot Accession Number P07537	176.02	1.11E-41	36%	51%	1389--1610	41--304
														4	BAB83471	Vp260 like protein	66.24	1.24E-08	23%	36%	14--294	552--831
														5	BAB83468	Vp260 like protein	47.75	4.57E-03	24%	41%	14--185	95--257
														6	BAB83467	Vp260 like protein	70.48	6.59E-10	21%	36%	796--1428	34--579
														7	BAB83469	Vp260 like protein	48.14	3.50E-03	24%	40%	14--185	95--257
														8	BAB83470	Vp260 like protein	49.29	1.57E-03	25%	40%	14--185	95--257
														9	AAN05154	RB133	108.23	2.85E-21	39%	56%	1429--1570	309--453
														10	ZP_00739873	Cell surface protein	78.95	1.85E-12	33%	52%	1423--1559	277--422
N164R	74778--76070	431	48,976	10.91	1	pfam05804	KAP, Kinesin-associated protein (KAP). This family consists of several eukaryotic kinesin-associated (KAP) proteins. Kinesins are intracellular multimeric transport motor proteins that move cellular cargo on microtubule tracks. It has been shown that the sea urchin KRP85/95 holoenzyme associates with a KAP115 non-motor protein, forming a heterotrimeric complex in vitro, called the Kinesin-II	31.12	2.47E-01	24%	41%	265--345	83--169	1	AAC96984	similar to Chlorella virus PBCV-1 ORF A282L, corresponds to GenBank Accession Number U42580	209.53	1.93E-52	37%	53%	104--417	2--357
														2	AAC96650	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055	65.08	5.89E-09	46%	57%	69--141	369--451
														3	AAC96646	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662	48.14	7.45E-04	27%	42%	36--151	366--506
														4	AAC96461	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563	61.23	8.50E-08	44%	54%	7--73	49--132
N165L	76610--76065	182	20,841	8.76	1	pfam01096	TFIIS, Transcription factor S-II (TFIIS)..	66.54	1.64E-12	68%	74%	141--179	1--39	1	BAA04187	transcription elongation factor SII	281.95	6.32E-75	70%	82%	1--180	1--180
					2	smart00440	ZnF_C2C2, C2C2 Zinc finger; Nucleic-acid-binding motif in transcriptional elongation factor TFIIS and RNA polymerases. .	64.15	9.51E-12	56%	77%	141--180	1--40	2	AAC96492	contains a zinc ribbon domain; similar to Chlorella virus CVU1 TFIIS-like transcription factor, corresponds to GenBank Accession Number D29631	280.80	1.41E-74	70%	83%	1--180	1--180
					3	COG1594	RPB9, DNA-directed RNA polymerase, subunit M/Transcription elongation factor TFIIS [Transcription].	60.43	1.25E-10	54%	73%	142--179	74--111	3	S47662	transcription elongation factor TFIIS homolog - Chlorella virus CV-U1	279.26	4.10E-74	68%	82%	1--180	1--180
														4	BAA04186	transcription elongation factor SII	243.82	1.91E-63	66%	81%	1--164	1--164
														5	XP_458316	hypothetical protein DEHA0C15686g	83.57	3.30E-15	34%	49%	52--179	161--288
														6	XP_445674	unnamed protein product	82.03	9.60E-15	39%	56%	82--179	209--304
														7	AAS54500	AGR011Wp	82.03	9.60E-15	37%	52%	66--179	189--302
														8	XP_455506	unnamed protein product	81.65	1.25E-14	35%	50%	59--179	174--290
														9	XP_665338	transcription elongation factor TFIIS.h	80.88	2.14E-14	37%	54%	67--179	216--331
														10	XP_627328	transcription elongation factor TFIIS	80.88	2.14E-14	37%	54%	67--179	216--331
N168R	76715--77380	222	24,350	10.20		No Hit Found								1	AAC96495	A127R	291.20	1.68E-77	63%	76%	8--217	30--245
N169L	78142--77375	256	28,604	8.34		No Hit Found								1	AAC96377	A9R	204.91	2.10E-51	54%	77%	89--254	8--173
														2	AAC96972	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	94.74	3.06E-18	28%	45%	10--253	15--253
														3	AAC96643	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	89.35	1.29E-16	24%	45%	9--252	11--248
														4	AAC96818	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	85.50	1.86E-15	24%	46%	9--253	8--246
														5	AAC96447	A79R	76.64	8.63E-13	23%	47%	9--232	8--218
														6	AAC96545	A177R	73.94	5.59E-12	24%	44%	9--237	11--226
														7	AAU06304	hypothetical protein A275R	73.56	7.31E-12	25%	51%	89--252	4--167
														8	AAU06301	hypothetical protein A275R	72.40	1.63E-11	24%	51%	89--252	4--167
														9	AAU06302	hypothetical protein A275R	62.39	1.68E-08	26%	53%	121--253	1--132
N170R	78297--79217	307	33,545	5.48	1	pfam03747	ADP_ribosyl_GH, ADP-ribosylglycohydrolase. This family includes enzymes that ADP-ribosylations, for example ADP-ribosylarginine hydrolase EC:3.2.2.19 cleaves ADP-ribose-L-arginine. The family also includes dinitrogenase reductase activating glycohydrolase. Most surprisingly the family also includes jellyfish crystallins, these proteins appear to have lost the presumed active site residues	119.36	4.71E-28	28%	44%	10--272	2--286	1	EAR97823	ADP-ribosylglycohydrolase family protein	156.38	1.18E-36	33%	50%	9--283	28--344
														2	EAR82865	ADP-ribosylglycohydrolase family protein	138.27	3.32E-31	30%	49%	9--269	38--338
														3	EAS04429	ADP-ribosylglycohydrolase family protein	132.11	2.38E-29	30%	49%	9--272	41--348
														4	ZP_00531396	ADP-ribosylglycohydrolase	130.18	9.03E-29	31%	48%	5--285	18--295
														5	ABB23694	putative ribosylglycohydrolase	128.26	3.43E-28	31%	46%	11--285	18--290
														6	ZP_00533899	ADP-ribosylglycohydrolase	126.72	9.99E-28	31%	48%	11--285	12--283
														7	EAR83125	ADP-ribosylglycohydrolase family protein	121.71	3.21E-26	30%	45%	11--267	179--485
														8	YP_463645	Inositol monophosphatase/ADP-ribosylglycohydrolase	120.94	5.48E-26	30%	45%	11--291	335--639
														9	ZP_00661442	ADP-ribosylglycohydrolase	120.55	7.16E-26	30%	47%	11--285	12--284
														10	ZP_00589296	ADP-ribosylglycohydrolase	117.09	7.91E-25	31%	46%	6--286	9--257
N173R	79199--79600	134	15,106	9.33		No Hit Found								1	AAC96416	A48R	95.90	4.31E-19	37%	62%	18--132	5--123
N174L	79957--79595	121	12,049	3.55		No Hit Found								0	No Hit Found	No Hit Found						
N176R	80040--80495	152	17,368	4.90	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	28.84	2.76E-01	37%	49%	6--50	1--44	1	AAZ21386	hypothetical protein SAR11_0565	94.74	9.50E-19	39%	62%	1--117	1--115
N177L	81844--80465	460	51,947	5.86	1	COG0637	COG0637, Predicted phosphatase/phosphohexomutase [General function prediction only].	100.95	2.51E-22	31%	47%	9--198	2--190	1	ZP_01245443	Nucleotidyl transferase	211.46	5.54E-53	50%	67%	219--450	6--242
					2	COG0546	Gph, Predicted phosphatases [General function prediction only].	78.65	1.45E-15	29%	46%	7--196	2--191	2	CAH09375	putative nucleotidyltransferase	206.84	1.37E-51	48%	64%	220--450	2--236
					3	pfam00702	Hydrolase, haloacid dehalogenase-like hydrolase. This family are structurally different from the alpha/ beta hydrolase family (pfam00561). This family includes L-2-haloacid dehalogenase, epoxide hydrolases and phosphatases. The structure of the family consists of two domains. One is an inserted four helix bundle, which is the least well conserved region of the alignment. The rest of the fold is composed of the core alpha/beta domain	68.09	2.34E-12	25%	39%	10--190	1--191	3	ZP_00368786	lipopolysaccharide biosynthesis protein, putative	200.68	9.78E-50	48%	65%	220--450	2--238
					4	COG1209	RfbA, dTDP-glucose pyrophosphorylase [Cell envelope biogenesis, outer membrane].	63.27	6.29E-11	25%	42%	223--446	4--228	4	ZP_00371454	lipopolysaccharide biosynthesis protein, putative	196.44	1.85E-48	49%	63%	220--450	1--236
					5	COG1011	COG1011, Predicted hydrolase (HAD superfamily) [General function prediction only].	60.30	4.54E-10	22%	35%	28--203	30--208	5	AAT91796	putative nucleotidyl transferase family protein	195.67	3.15E-48	43%	63%	220--450	2--240

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					6	COG1208	GCD1, Nucleoside-diphosphate-sugar pyrophosphorylase involved in lipopolysaccharide biosynthesis/translation initiation factor 2B, gamma/epsilon subunits (eIF-2Bgamma/eIF-2Bepsilon) [Cell envelope biogenesis, outer membrane / Translation, ribosomal structure and biogenesis]	58.43	1.77E-09	25%	42%	222--447	4--228	6	YP_419519	dTDP-glucose pyrophosphorylase	181.80	4.70E-44	43%	59%	217--450	2--240
					7	pfam00483	NTP_transferase, Nucleotidyl transferase. This family includes a wide range of enzymes which transfer nucleotides onto phosphosugars..	55.98	9.93E-09	19%	38%	223--392	3--184	7	AAO54248	lipopolysaccharide biosynthesis protein, putative	173.71	1.28E-41	40%	57%	220--450	2--239
					8	COG4750	LicC, CTP:phosphocholine cytidylyltransferase involved in choline phosphorylation for cell surface LPS epitopes [Cell envelope biogenesis, outer membrane]	41.05	3.27E-04	45%	68%	220--258	1--39	8	ZP_00207801	COG1209: dTDP-glucose pyrophosphorylase	171.79	4.87E-41	42%	59%	226--450	1--229
					9	pfam02348	CTP_transf_3, Cytidylyltransferase. This family consists of two main Cytidylyltransferase activities: 1) 3-deoxy-manno-octulosonate cytidylyltransferase, EC:2.7.7.38 catalysing the reaction:- CTP + 3-deoxy-D-manno-octulosonate <=> diphosphate + CMP-3-deoxy-D-manno-octulosonate, 2) acylneuraminate cytidylyltransferase EC:2.7.7.43, catalysing the reaction:- CTP + N-acylneuraminate <=> diphosphate + CMP-N-acylneuraminate. NeuAc cytydilyltransferase of Mannheimia haemolytica has been characterised describing kinetics and regulation by substrate charge, energetic charge and amino-sugar demand..	39.54	8.31E-04	19%	36%	222--392	2--175	9	AAY35682	Nucleotidyl transferase	168.70	4.12E-40	38%	58%	220--450	2--239
					10	COG1210	GalU, UDP-glucose pyrophosphorylase [Cell envelope biogenesis, outer membrane]	37.41	3.72E-03	24%	35%	223--389	8--210	10	AAZ34091	lipopolysaccharide biosynthesis protein, putative	164.85	5.95E-39	38%	55%	220--450	2--239
N180L	83389--81878	504	57,895	5.18		No Hit Found								1	ZP_00520036	Methyltransferase FkbM	66.24	3.27E-09	29%	45%	34--232	402--595
														2	ZP_01305260	hypothetical protein SKA58 11615	54.30	1.29E-05	27%	45%	37--185	45--196
														3	AAV90092	hypothetical protein ZMO1468	53.53	2.19E-05	31%	52%	282--408	398--533
														4	YP_439595	methyltransferase, FkbM family domain protein	48.14	9.22E-04	26%	40%	32--237	13--223
														5	ZP_01090243	methyltransferase FkbM	45.82	4.57E-03	27%	41%	34--167	38--166
N183L	83996--83331	222	25,840	7.70		No Hit Found								1	AAV90092	hypothetical protein ZMO1468	50.45	5.00E-05	33%	54%	69--168	428--533
N185L	84514--84026	163	18,262	11.11	1	pfam04856	Securin, Securin sister-chromatid separation inhibitor. Securin is also known as pituitary tumour-transforming gene product. Over-expression of securin is associated with a number of tumours, and it has been proposed that this may be due to erroneous chromatid separation leading to chromosome gain or loss..	28.92	3.34E-01	27%	44%	98--157	90--153	1	AAC96978	A662L	210.69	1.31E-53	64%	79%	6--161	16--171
														2	ABE89007	Mpv17/PMP22	71.25	1.24E-11	28%	51%	34--160	92--219
														3	NP_565983	unknown protein	64.31	1.52E-09	24%	49%	27--160	98--232
														4	BAD87006	unknown protein	63.16	3.39E-09	25%	50%	34--160	112--239
														5	XP_962431	hypothetical protein	50.45	2.27E-05	23%	44%	34--156	129--255
														6	AAM62733	contains similarity to 22 kDa peroxisomal membrane protein	49.68	3.87E-05	25%	51%	34--135	122--225
														7	BAB08278	unnamed protein product	49.68	3.87E-05	25%	51%	34--135	121--224
														8	NP_568621	unknown protein	49.68	3.87E-05	25%	51%	34--135	121--224
														9	AAN46791	At5g19750/T29J13 170	47.37	1.92E-04	19%	43%	13--161	139--284
														10	AAM65990	unknown	47.37	1.92E-04	19%	43%	13--161	140--285
N186L	85183--84707	159	18,620	10.37	1	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	33.51	1.20E-02	24%	39%	8--94	1--83	1	AAC96502	similar to bacteriophage T4 intron-associated endonuclease, corresponds to Swiss-Prot Accession Number P13299	140.97	1.16E-32	52%	64%	1--147	1--148
					2	COG2989	COG2989, Uncharacterized protein conserved in bacteria [Function unknown]	29.19	2.78E-01	28%	55%	12--103	442--518	2	AAC96973	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	48.91	5.98E-05	38%	50%	9--93	2--85
					3	pfam04180	LTV, Low temperature viability protein..	27.39	8.88E-01	32%	54%	82--149	333--396	3	YP_293795	putative endonuclease	48.52	7.81E-05	34%	49%	9--92	3--88
														4	AAY63899	reticulocyte binding protein 1	43.51	2.51E-03	26%	45%	13--135	465--583
														5	AAC96655	PBCV-1 33kd peptide	41.59	9.55E-03	30%	54%	14--94	11--91
														6	CAA31562	unidentified protein	41.59	9.55E-03	34%	51%	2--89	61--155
														7	NP_074961	orf261	41.59	9.55E-03	34%	51%	2--89	60--154
N187L	86288--85218	357	41,118	8.86	1	smart00507	HNHc, HNH nucleases; .	30.44	3.67E-01	31%	46%	56--102	2--50	1	AAC96722	A354R	200.68	6.86E-50	42%	61%	110--355	1--235
					2	pfam01844	HNH, HNH endonuclease..	29.91	5.47E-01	29%	34%	61--102	2--43	2	AAC96790	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	77.41	8.76E-13	29%	49%	54--322	51--319
														3	AAL98037	hypothetical phage protein	45.82	2.82E-03	26%	48%	20--155	11--147
N189L	86927--86580	116	13,059	10.04		No Hit Found								1	AAC96507	A139L	117.47	1.39E-25	56%	78%	26--108	25--107
N191R	87022--88587	522	58,869	8.65	1	COG1215	COG1215, Glycosyltransferases, probably involved in cell wall biogenesis [Cell envelope biogenesis, outer membrane]	131.59	2.17E-31	20%	35%	66--460	9--412	1	AAC96840	similar to Acetobacter cellulose synthase, corresponds to Swiss-Prot Accession Number P21877	800.82	0.00E+00	72%	84%	1--520	1--517
					2	pfam00535	Glycos_transf_2, Glycosyl transferase. Diverse family, transferring sugar from UDP-glucose, UDP-N-acetyl- galactosamine, GDP-mannose or CDP-abequose, to a range of substrates including cellulose, dolichol phosphate and teichoic acids.	64.09	4.39E-11	18%	37%	118--280	1--164	2	XP_959316	hypothetical protein	434.88	3.64E-120	44%	60%	21--516	168--670
					3	pfam03552	Cellulose_synt, Cellulose synthase. Cellulose, an aggregate of unbranched polymers of beta-1,4-linked glucose residues, is the major component of wood and thus paper, and is synthesised by plants, most algae, some bacteria and fungi, and even some animals. The genes that synthesise cellulose in higher plants differ greatly from the well-characterized genes found in Acetobacter and Agrobacterium sp. More correctly designated as 'cellulose synthase catalytic subunits', plant cellulose synthase (CesA) proteins are integral membrane proteins, approximately 1,000 amino acids in length. There are a number of highly conserved residues, including several motifs shown to be necessary for processive glycosyltransferase activity.	43.36	6.56E-05	30%	46%	292--387	517--618	3	EAQ91514	conserved hypothetical protein	417.93	4.61E-115	42%	59%	19--500	391--894
					4	COG1216	COG1216, Predicted glycosyltransferases [General function prediction only]	42.81	1.10E-04	21%	34%	190--312	82--213	4	YP_482343	Cellulose synthase (UDP-forming)	273.09	1.83E-71	35%	54%	23--453	60--492
					5	pfam05601	DUF774, Protein of unknown function (DUF774). This family consists of several uncharacterised Actinomycete proteins of unknown function..	31.87	1.99E-01	32%	46%	136--164	62--90	5	AAO54558	cellulose synthase, catalytic subunit	158.30	6.58E-37	34%	51%	111--367	155--417
					6	COG0463	WcaA, Glycosyltransferases involved in cell wall biogenesis [Cell envelope biogenesis, outer membrane]	30.88	4.13E-01	12%	26%	113--395	1--278	6	NP_902348	cellulose synthase, subunit A	157.53	1.12E-36	34%	51%	115--385	264--540
					7	COG2943	MdoH, Membrane glycosyltransferase [Cell envelope biogenesis, outer membrane]	29.95	7.60E-01	20%	34%	98--457	122--500	7	AAL71842	WssB	154.84	7.28E-36	34%	50%	113--367	157--417
														8	AAM38361	celullose synthase	154.45	9.51E-36	32%	49%	115--405	153--451
														9	NP_643825	celullose synthase	154.45	9.51E-36	32%	49%	115--405	155--453
														10	YP_365375	cellulose synthase catalytic subunit	154.07	1.24E-35	32%	49%	115--405	155--453
N195R	88652--91846	1065	112,990	10.84		No Hit Found								1	BAE02830	surface protein	107.84	2.37E-21	83%	90%	1002--1063	1012--1073
														2	AAC96508	PBCV-1 surface protein	46.98	4.97E-03	37%	48%	485--566	949--1031
														3	BAD22850	surface protein	46.98	4.97E-03	37%	48%	485--566	949--1031
														4	BAD12236	surface protein	46.60	6.49E-03	37%	48%	485--566	949--1031
														5	T17636	proline-rich protein A145R - Chlorella virus PBCV-1	46.98	4.97E-03	37%	48%	485--566	158--240
														6	AAC96752	Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	51.22	2.64E-04	36%	50%	495--563	195--263
														7	AAC96773	Pro-, Lys-rich, PAPK (30x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	48.91	1.31E-03	32%	58%	638--699	4--65
														8	AAC96654	A286R	50.06	5.87E-04	20%	40%	139--416	68--346
														9	YP_535207	Mucus binding protein	46.98	4.97E-03	28%	39%	734--865	723--853
														10	AAT30727	conserved repeat domain protein	46.60	6.49E-03	24%	40%	173--373	741--954
N205L	92196--91882	105	11,647	6.50	1	COG0731	COG0731, Fe-S oxidoreductases [Energy production and conversion]	28.34	1.93E-01	21%	43%	39--80	21--68	1	AAC96518	A150L	78.18	9.08E-14	41%	59%	1--100	5--105
N206R	92301--93668	456	50,729	8.78	1	COG1061	SSL2, DNA or RNA helicases of superfamily II [Transcription / DNA replication, recombination, and repair]	137.92	2.16E-33	26%	40%	54--397	7--379	1	AAC96521	similar to phage T5 helicase, corresponds to Swiss-Prot Accession Number P11107	613.99	3.69E-174	66%	81%	4--453	7--459
					2	smart00487	DEXDc, DEAD-like helicases superfamily; .	63.32	5.82E-11	23%	37%	94--233	13--173	2	NP_077551	EsV-1-66	179.87	1.78E-43	32%	50%	21--426	17--443
					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	55.86	9.68E-09	23%	34%	110--230	2--143	3	YP_142750	VV A18 helicase	168.70	4.10E-40	28%	48%	3--430	8--453
					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..	54.32	2.71E-08	25%	37%	110--230	2--144	4	ZP_00754006	COG1061: DNA or RNA helicases of superfamily II	146.36	2.18E-33	29%	45%	85--430	419--776

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	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	44.64	2.48E-05	18%	32%	102--233	19--175	5	YP_437404	DNA or RNA helicase of superfamily II	140.58	1.20E-31	29%	46%	87--430	421--776
					6	COG4096	HsdR, Type I site-specific restriction-modification system, R (restriction) subunit and related helicases [Defense mechanisms].	43.81	4.13E-05	26%	40%	113--231	190--321	6	ZP_01111109	DNA or RNA helicase of superfamily II	139.81	2.04E-31	29%	46%	85--443	419--793
					7	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, FTI 1)	43.73	5.08E-05	22%	38%	110--230	18--156	7	NP_309332	hypothetical protein ECs1305	135.58	3.85E-30	30%	49%	81--430	417--777
					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	42.22	1.21E-04	23%	40%	302--395	17--123	8	NP_287072	putative helicase	135.58	3.85E-30	30%	49%	81--430	425--785
					9	COG1197	Mfd, Transcription-repair coupling factor (superfamily II helicase) [DNA replication, recombination, and repair / Transcription].	42.17	1.34E-04	26%	43%	89--230	607--752	9	ZP_00532208	Type III restriction enzyme, res subunit:DEAD/DEAH box helicase, N-terminal	128.26	6.14E-28	29%	46%	51--430	394--790
					10	COG0610	COG0610, Type I site-specific restriction-modification system, R (restriction) subunit and related helicases [Defense mechanisms].	38.12	2.05E-03	22%	36%	95--338	258--517	10	ZP_00665630	Type III restriction enzyme, res subunit:DEAD/DEAH box helicase, N-terminal	127.10	1.37E-27	28%	45%	81--430	481--849
N209L	94072--93716	119	12,775	3.90	1	pfam03115	Astro_capsid, Astrovirus capsid protein precursor. This product is encoded by astrovirus ORF2, one of the three astrovirus ORFs (1a, 1b, 2). The 87kD precursor protein undergoes an intracellular cleavage to form a 79kD protein. Subsequently, extracellular trypsin cleavage yields the three proteins forming the infectious virion..	26.55	7.67E-01	26%	43%	71--117	640--686	1	AAC96525	A157L	93.59	2.12E-18	45%	56%	1--117	1--109
N210L	94496--94128	123	14,411	4.26		No Hit Found								0	No Hit Found	No Hit Found						
N211L	94977--94603	125	13,445	8.66		No Hit Found								0	No Hit Found	No Hit Found						
N214L	95457--95167	97	10,869	9.32		No Hit Found								1	AAC96533	A165L	72.79	3.93E-12	39%	53%	5--95	171--279
N216L	95974--95516	153	17,584	10.22	1	cd01973	Nitrogenase_VFe_beta_like, Nitrogenase_VFe_beta -like: Nitrogenase VFe protein, beta subunit like. This group contains proteins similar to the beta subunits of the VFe protein of the vanadium-dependent (V-) nitrogenase. Nitrogenase catalyzes the ATP-dependent reduction of dinitrogen (N2) to ammonia. In addition to V-nitrogenase there is a molybdenum (Mo)-dependent nitrogenase and an iron only (Fe-) nitrogenase. The Mo-nitrogenase is the most widespread and best characterized of these systems. These systems consist of component 1 (VFe protein, FeFe protein or, MoFe protein respectively) and, component 2 (Fe protein). MoFe is an alpha2beta2 tetramer, V-and Fe-nitrogenases are alpha2beta2delta2 hexamers. The alpha and beta subunits of VFe and FeFe are similar to the alpha and beta subunits of MoFe. For MoFe each alphabeta pair contains one P-cluster (at the alphabeta interface) and, one molecule of iron molybdenum cofactor (FeMoco) contained within the alpha subunit. The Fe protein which has a practically identical structure in all three systems, it contains a single [4Fe	29.52	1.89E-01	41%	49%	47--84	342--383	1	AAC96533	A165L	85.50	5.74E-16	37%	58%	5--118	20--134
														2	ZP_00133232	COG1219: ATP-dependent protease Clp, ATPase subunit	41.97	7.28E-03	29%	50%	31--149	16--124
														3	ZP_00123014	COG1219: ATP-dependent protease Clp, ATPase subunit	41.97	7.28E-03	29%	50%	31--149	20--128
N218R	96021--96887	289	33,100	5.82	1	COG5377	COG5377, Phage-related protein, predicted endonuclease [DNA replication, recombination, and repair].	30.01	3.41E-01	23%	32%	30--169	2--148	1	AAC96534	PBCV-1 exonuclease	298.52	1.72E-79	58%	73%	9--253	1--246
														2	NP_077549	EsV-1-64	88.58	2.71E-16	27%	50%	21--221	1--191
														3	AAG28903	F12A21.19	78.57	2.80E-13	30%	45%	36--237	95--294
														4	NP_176934	unknown protein	78.57	2.80E-13	30%	45%	36--237	116--315
														5	AAR26918	FirrV-1-B43 precursor	76.26	1.39E-12	30%	46%	19--223	3--197
														6	CAF27488	Exonuclease	73.94	6.90E-12	29%	44%	36--224	3--200
														7	CAF27185	Exonuclease	73.94	6.90E-12	29%	44%	36--224	3--200
														8	YP_612778	exonuclease	72.02	2.62E-11	32%	48%	36--183	3--155
														9	XP_480865	unknown protein	71.63	3.42E-11	29%	45%	35--216	134--317
														10	XP_472944	OSJNBa0081L15.20	70.86	5.84E-11	29%	45%	35--216	141--324
N221R	97092--97394	101	10,665	4.11		No Hit Found								1	AAC96536	A168R	93.97	1.62E-18	53%	66%	9--95	70--161
N222L	98212--97385	276	30,232	9.99	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	116.20	3.75E-27	31%	49%	16--189	1--179	1	AAC96541	similar to E. coli hypothetical protein, corresponds to Swiss-Prot Accession Number P39407	322.40	1.03E-86	57%	77%	5--274	14--288
					2	COG1752	RssA, Predicted esterase of the alpha-beta hydrolase superfamily [General function prediction only].	102.08	6.31E-23	31%	48%	16--228	14--229	2	ZP_00240206	Patatin-like phospholipase family	106.69	8.91E-22	34%	53%	15--185	7--190
					3	COG4667	COG4667, Predicted esterase of the alpha-beta hydrolase superfamily [General function prediction only].	51.42	1.12E-07	24%	40%	16--208	14--202	3	XP_800775	PREDICTED: hypothetical protein XP_795682, partial	99.37	1.42E-19	35%	50%	14--186	76--265
					4	COG3621	COG3621, Patatin [General function prediction only].	31.14	1.52E-01	41%	58%	5--66	6--65	4	YP_133138	hypothetical protein PBPRB1472	90.12	8.63E-17	28%	48%	14--256	6--261
														5	XP_789091	PREDICTED: hypothetical protein XP_783998	89.74	1.13E-16	32%	51%	14--186	86--275
														6	NP_905966	hypothetical protein PG1879	89.35	1.47E-16	32%	52%	14--185	6--194
														7	NP_149926	463L	86.66	9.54E-16	29%	54%	14--206	25--215
														8	AAS94389	phospholipase, patatin family	85.89	1.63E-15	30%	50%	15--186	4--194
														9	ABB37620	esterase of the alpha-beta hydrolase superfamily-like	84.73	3.63E-15	30%	50%	15--185	4--193
														10	YP_142800	patatin-like phospholipase (463L)	83.19	1.06E-14	25%	49%	15--238	57--285
N223L	99133--98228	302	34,030	4.75	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	175.01	8.41E-45	31%	55%	36--295	1--256	1	AAU06282	protein kinase A248R	187.96	3.55E-46	38%	58%	21--295	32--305
					2	smart00220	S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosohotransferases. Serine or threonine-soecific kinase subfamily. .	175.02	8.98E-45	31%	52%	37--295	1--256	2	AAU06280	protein kinase A248R	187.19	6.05E-46	37%	58%	21--295	32--305
					3	pfam00069	Pkinase, Protein kinase domain..	164.30	1.48E-41	28%	52%	37--295	1--258	3	AAU06275	protein kinase A248R	187.19	6.05E-46	37%	58%	21--295	11--284
					4	COG0515	SPS1, Serine/threonine protein kinase [General function prediction only / Signal transduction mechanisms / Transcription / DNA replication, recombination, and repair].	117.18	2.27E-27	28%	47%	36--296	1--279	4	AAC96616	PBCV-1 protein kinase	182.19	1.95E-44	37%	58%	24--295	32--304
					5	smart00219	TyrKc, Tyrosine kinase, catalytic domain; Phosphotransferases. Tyrosine-specific kinase subfamily. .	91.39	1.42E-19	28%	48%	39--227	3--192	5	AAA87065	serine/threonine protein kinase	180.26	7.39E-44	36%	58%	24--295	28--300
					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..	88.71	7.38E-19	27%	46%	36--227	8--203	6	AAU06274	protein kinase A248R	156.76	8.75E-37	38%	59%	79--295	22--238
					7	COG3642	COG3642, Mn2+-dependent serine/threonine protein kinase [Signal transduction mechanisms].	52.93	4.66E-08	28%	43%	112--211	66--180	7	AAU06285	protein kinase A248R	150.60	6.27E-35	37%	58%	81--295	2--217

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	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..	37.17	2.49E-03	21%	37%	92--167	74--155	8	AAC96657	similar to PBCV-1 serine/threonine protein kinase, corresponds to GenBank Accession Number U14660	129.03	1.95E-28	31%	52%	41--295	22--279
					9	COG4248	COG4248, Uncharacterized protein with protein kinase and helix-hairpin-helix DNA-binding domains [General function prediction only].	33.14	4.19E-02	20%	40%	127--201	114--189	9	AAW40743	serine/threonine-protein kinase, putative	120.55	6.95E-26	31%	51%	36--298	114--404
					10	COG0478	COG0478, RIO-like serine/threonine protein kinase fused to N-terminal HTH domain [Signal transduction mechanisms]	30.59	2.57E-01	25%	47%	86--174	158--246	10	AAU06286	protein kinase A248R	119.40	1.55E-25	45%	63%	146--295	2--152
N227L	101362--99221	714	79,825	6.96	1	COG4581	COG4581, Superfamily II RNA helicase [DNA replication, recombination, and repair].	368.92	9.77E-103	40%	55%	12--409	119--568	1	AAC96609	contains ATP-GTP binding motif; similar to Saccharomyces cervisiae antiviral protein SKI2, corresponds to Swiss-Prot Accession Number P35207	699.12	0.00E+00	49%	66%	4--712	2--725
					2	COG1204	COG1204, Superfamily II helicase [General function prediction only].	196.76	6.64E-51	31%	50%	30--400	50--430	2	XP_652111	DEAD/DEAH box helicase	306.22	2.89E-81	37%	53%	13--527	96--654
					3	COG1202	COG1202, Superfamily II helicase, archaea-specific [General function prediction only].	126.23	1.03E-29	33%	50%	27--378	232--552	3	XP_970408	PREDICTED: similar to CG4152-PA	303.52	1.87E-80	38%	56%	13--428	138--594
					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	106.66	8.20E-24	23%	37%	5--190	3--206	4	EAL27780	GA10159-PA	303.52	1.87E-80	35%	50%	3--542	244--850
					5	smart00487	DEXDc, DEAD-like helicases superfamily; .	94.52	3.52E-20	26%	42%	5--175	1--190	5	XP_970674	PREDICTED: similar to CG10210-PA	302.75	3.20E-80	37%	53%	13--502	239--765
					6	COG1205	COG1205, Distinct helicase family with a unique C-terminal domain including a metal-binding cysteine cluster [General function prediction only]	83.14	1.06E-16	24%	41%	3--365	61--410	6	EAS01065	DEAD/DEAH box helicase family protein	300.06	2.07E-79	38%	54%	13--447	133--599
					7	COG1201	Lhr, Lhr-like helicases [General function prediction only].	79.14	1.81E-15	26%	43%	17--374	27--358	7	NP_524465	twister CG10210-PA	299.67	2.71E-79	37%	52%	3--464	246--760
					8	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	78.59	2.53E-15	26%	43%	28--154	1--143	8	CAC35050	putative helicase Ski2	299.67	2.71E-79	37%	52%	3--464	246--760
					9	COG0513	SmB, Superfamily II DNA and RNA helicases [DNA replication, recombination, and repair / Transcription / Translation, ribosomal structure and biogenesis]	67.88	4.75E-12	21%	40%	17--424	56--425	9	NP_001012962	superkiller viralicidic activity 2-like 2	298.13	7.88E-79	38%	56%	13--447	127--598
					10	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..	67.41	5.63E-12	27%	45%	28--154	1--142	10	EAA08476	ENSANGP00000014710	295.82	3.91E-78	39%	54%	13--443	269--745
N231R	101470--101790	107	12,564	10.56		No Hit Found								0	No Hit Found	No Hit Found						
N232L	103319--101802	506	56,634	9.80	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	524.14	1.42E-149	38%	53%	29--498	1--470	1	AAC96605	PBCV-1 homospermidine synthase	711.45	0.00E+00	69%	81%	23--502	36--516
					2	COG1748	LYS9, Saccharopine dehydrogenase and related proteins [Amino acid transport and metabolism].	33.72	5.80E-02	24%	49%	30--271	2--184	2	ZP_00589757	Homospermidine synthase	333.18	1.44E-89	38%	57%	29--476	5--452
					3	pfam01967	MoaC, MoaC family. Members of this family are involved in molybdenum cofactor biosynthesis. However their molecular function is not known..	33.29	7.35E-02	21%	42%	86--154	22--94	3	AAM05046	homospermidine synthase	300.06	1.35E-79	36%	55%	29--476	13--459
					4	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	32.53	1.13E-01	22%	46%	32--187	1--137	4	NP_632190	homospermidine synthase	295.82	2.55E-78	35%	55%	29--476	6--452
					5	cd00528	MoaC, MoaC family. Members of this family are involved in molybdenum cofactor (Moco) biosynthesis, an essential cofactor of a diverse group of redox enzymes. MoaC, a small hexameric protein, converts, together with MoaA, a guanosine derivative to the precursor Z by inserting the carbon-8 of the purine between the 2' and 3' ribose carbon atoms, which is the first of three phases of Moco biosynthesis..	30.11	6.14E-01	22%	38%	86--154	22--94	5	AAZ71589	homospermidine synthase	283.49	1.31E-74	34%	53%	29--476	6--452
														6	ZP_00584810	Homospermidine synthase	246.51	1.78E-63	33%	51%	32--476	14--446
														7	AAU28555	homospermidine synthase	241.89	4.37E-62	34%	53%	28--476	13--446
														8	CAH16656	hypothetical protein	240.74	9.74E-62	34%	53%	28--476	13--446
														9	YP_421629	Homospermidine synthase	234.19	9.12E-60	32%	51%	35--476	34--462
														10	AAO55386	homospermidine synthase	228.02	6.53E-58	34%	53%	79--480	56--448
N237L	103707--103336	124	14,303	5.61		No Hit Found								0	No Hit Found	No Hit Found						
N239R	103911--104258	116	12,766	10.69	1	cd01235	PH_SETbf, Set binding factor Pleckstrin Homology (PH) domain. Set binding factor is a myotubularin-related pseudo-phosphatase consisting of a Denn domain, a Gram domain, an inactive phosphatase domain, a SID motif and a C-terminal PH domain. Its PH domain is predicted to bind lipids based upon its ability to respond to phosphatidylinositol 3-kinase ..	28.87	1.74E-01	28%	50%	20--56	16--52	1	AAC96602	A234L	93.59	2.15E-18	44%	60%	1--113	1--107
N240L	104517--104242	92	11,364	10.47		No Hit Found								1	AAC96601	A233R	85.89	4.40E-16	48%	70%	5--86	11--91
N241L	104802--104536	89	10,140	4.86	1	COG0695	GrxC, Glutaredoxin and related proteins [Posttranslational modification, protein turnover, chaperones].	43.41	7.16E-06	34%	53%	3--72	4--72	1	CAE58431	Hypothetical protein CBG01566	50.45	2.07E-05	33%	51%	4--83	31--114
					2	pfam00462	Glutaredoxin, Glutaredoxin..	35.23	2.08E-03	26%	41%	3--70	3--71	2	AAZ82854	Hypothetical protein C07G1.8	49.29	4.61E-05	33%	51%	4--83	31--114
														3	NP_077613	EsV-1-128	48.91	6.02E-05	36%	57%	3--72	8--74
														4	ZP_01076784	Glutaredoxin	47.75	1.34E-04	34%	52%	2--77	4--78
														5	ZP_01301447	hypothetical protein RqrvI_01000777	47.37	1.75E-04	31%	53%	4--84	6--85
														6	AAZ27138	glutaredoxin	46.21	3.90E-04	32%	61%	4--70	6--69
														7	ZP_01166994	putative glutaredoxin 1	46.21	3.90E-04	33%	54%	3--74	5--75
														8	AAV29888	NT02FT1163	44.67	1.14E-03	36%	48%	4--77	5--86
														9	YP_513693	Glutaredoxin 1	43.51	2.53E-03	36%	47%	4--77	5--86
														10	YP_563095	Glutaredoxin, GrxA	43.13	3.31E-03	33%	50%	4--76	5--81
N242L	105406--104831	192	20,946	10.14		No Hit Found								1	AAC96598	A230R	155.99	6.01E-37	39%	65%	2--190	3--194
N243R	105431--105670	80	8,656	5.75	1	COG1647	COG1647, Esterase/lipase [General function prediction only].	28.30	2.46E-01	28%	43%	29--75	70--116	0	No Hit Found	No Hit Found						
N244R	105688--106083	132	14,773	10.00	1	cd01043	DPS, DPS (DNA Protecting protein under Starved conditions) domain is a member of a broad superfamily of ferritin-like diiron-carboxylate proteins. Some DPS proteins nonspecifically bind DNA, protecting it from cleavage caused by reactive oxygen species such as the hydroxyl radicals produced during oxidation of Fe(II) by hydrogen peroxide. These proteins assemble into dodecameric structures, some form DPS-DNA co-crystalline complexes, and possess iron and H2O2 detoxification capabilities. Expression of DPS is induced by oxidative or nutritional stress, including metal ion starvation. Members of the DPS family are homopolymers formed by 12 four-helix bundle subunits that assemble with 23 symmetry into a hollow shell. The DPS ferroxidase site is unusual in that it is not located in a four-helix bundle as is in ferritin, but is shared by 2-fold symmetry-related subunits providing the iron ligands. Many DPS sequences (e.g., E. coli) display an N-terminal extension of variable length that contains two or three positively charged lysine residues that	32.92	1.38E-02	21%	41%	18--128	10--136	1	AAC96595	A227L	163.70	1.64E-39	64%	79%	10--129	18--136
					2	COG0783	Dps, DNA-binding ferritin-like protein (oxidative damage protectant) [Inorganic ion transport and metabolism].	29.46	1.26E-01	35%	52%	18--67	27--79	2	NP_049716	MobD.6 hypothetical protein	48.52	7.70E-05	27%	50%	29--129	33--127
N246L	109032--106141	964	97,310	3.61		No Hit Found								1	BAB83469	Vp260 like protein	112.85	6.61E-23	25%	41%	487--830	39--493
														2	BAB83468	Vp260 like protein	115.93	7.81E-24	28%	45%	602--915	11--348
														3	BAB83470	Vp260 like protein	114.01	2.97E-23	27%	41%	590--915	39--348
														4	AAC96490	PBCV-1 Vp260 protein	52.76	8.12E-05	28%	44%	665--847	8--199
														5	BAB83471	Vp260 like protein	93.20	5.42E-17	23%	41%	457--955	30--527
														6	BAB83467	Vp260 like protein	47.75	2.61E-03	22%	40%	16--274	792--1063

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	AAA86307	glycoprotein Vp260	51.60	1.81E-04	28%	44%	665–847	8–199
														8	AAC96397	Asn/Thr/Ser/Val rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	68.55	1.43E-09	26%	40%	491–833	7–349
														9	AAC96382	Asn/Thr/Ser/Val rich protein	103.61	4.01E-20	22%	39%	16–726	491–1250
														10	AAC96386	Asn/Thr/Ser/Ile rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	81.65	1.63E-13	21%	39%	16–760	555–1304
N253L	110377–109085	431	44,963	4.18		No Hit Found								1	BAB83467	Vp260 like protein	53.53	1.77E-05	22%	41%	19–284	19–281
														2	BAB83471	Vp260 like protein	57.00	1.60E-06	24%	37%	22–383	22–405
														3	BAB83468	Vp260 like protein	70.09	1.83E-10	27%	44%	11–209	571–812
														4	BAB83469	Vp260 like protein	68.55	5.33E-10	27%	43%	31–239	501–743
														5	BAB83470	Vp260 like protein	63.16	2.24E-08	26%	42%	11–236	571–849
														6	AAA86307	glycoprotein Vp260	44.67	8.24E-03	24%	41%	4–261	220–490
														7	AAC96386	Asn/Thr/Ser/Ile rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	50.45	1.50E-04	21%	37%	5–372	474–867
														8	AAC96393	Asn/Thr/Ser/Gly rich protein; similar to E. coli adhesin AIDA-I precursor, corresponds to Swiss-Prot Accession Number Q03155	51.22	8.80E-05	24%	42%	25–233	94–312
														9	ZP_01250137	hypothetical protein Bpse110 02005980	55.07	6.09E-06	25%	38%	17–378	267–653
														10	AAC96382	Asn/Thr/Ser/Val rich protein	54.68	7.96E-06	22%	40%	31–377	154–548
N254L	111615–110404	404	45,168	6.32	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 PBCV-1 the major capsid protein is a glycoprotein	251.38	1.27E-67	31%	50%	3–398	5–443	1	AAC96917	similar to Chlorella virus PBCV-1 ORF A11L, corresponds to GenBank Accession Number U42580	244.20	6.49E-63	35%	55%	3–402	4–400
														2	AAC96379	contains aminoacyl-tRNA synthetase class-II signature	234.19	6.72E-60	35%	54%	2–402	4–403
														3	AAC96378	similar to PBCV-1 major capsid protein, corresponds to Swiss-Prot Accession Number P30328	220.71	7.69E-56	33%	56%	3–400	6–399
														4	AAC96798	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	214.16	7.19E-54	33%	50%	3–402	5–437
														5	BAA76601	major capsid protein MCP1	213.00	1.60E-53	32%	50%	3–402	5–437
														6	BAA76600	major capsid protein	210.69	7.95E-53	32%	50%	3–402	5–436
														7	BAA22198	major capsid protein Vp54	206.45	1.50E-51	32%	49%	3–402	5–437
														8	AAC27492	major capsid protein Vp49	204.53	5.70E-51	31%	50%	3–402	5–432
														9	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	202.99	1.66E-50	34%	50%	52–402	31–413
														10	1M3Y_D	Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid Containing, Dna Virus	202.99	1.66E-50	34%	50%	52–402	31–413
N255R	111871–112290	140	15,361	9.27		No Hit Found								1	AAC96922	A567L	77.03	2.02E-13	37%	52%	12–127	10–123
N259L	114151–112286	622	67,016	7.97	1	cd03272	ABC_SMC3_euk, Eukaryotic SMC3 proteins; SMC proteins are large (approximately 110 to 170 kDa), and each is arranged into five recognizable domains. Amino-acid sequence homology of SMC proteins between species is largely confined to the amino- and carboxy-terminal globular domains. The amino-terminal domain contains a 'Walker A' nucleotide-binding domain (GxxGxGKS/T, in the single-letter amino-acid code), which by mutational studies has been shown to be essential in several proteins. The carboxy-terminal domain contains a sequence (the DA-box) that resembles a 'Walker B' motif, and a motif with homology to the signature sequence of the ATP-binding cassette (ABC) family of ATPases. The sequence homology within the carboxy-terminal domain is relatively high within the SMC1-SMC4 group, whereas SMC5 and SMC6 show some divergence in both of these sequences. In eukaryotic cells, the proteins are found as heterodimers of SMC1 paired with SMC3, SMC2 with SMC4, and SMC5 with SMC6 (formerly known as Rad18)..	39.63	1.03E-03	20%	46%	29–176	212–352	1	AAC96921	A565R	45.44	7.78E-03	47%	54%	365–406	253–293
					2	COG4372	COG4372, Uncharacterized protein conserved in bacteria with the myosin-like domain [Function unknown].	36.64	8.29E-03	21%	44%	38–182	77–214	2	BAA11343	DNA binding protein	45.44	7.78E-03	47%	54%	365–406	253–293
					3	pfam01559	Zein, Zein seed storage protein. Zeins are seed storage proteins. They are unusually rich in glutamine, proline, alanine, and leucine residues and their sequences show a series of tandem repeats..	33.98	6.21E-02	29%	44%	97–248	75–226	3	AAC96752	Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	63.93	2.11E-08	45%	53%	251–312	29–99
					4	COG5064	SRP1, Karyopherin (importin) alpha [Intracellular trafficking and secretion].	33.47	8.80E-02	24%	53%	112–191	37–117	4	NP_620240	golgi-associated protein GCP360	57.38	1.98E-06	19%	42%	29–268	218–480
					5	cd03277	ABC_SMC5_euk, Eukaryotic SMC5 proteins; SMC proteins are large (approximately 110 to 170 kDa), and each is arranged into five recognizable domains. Amino-acid sequence homology of SMC proteins between species is largely confined to the amino- and carboxy-terminal globular domains. The amino-terminal domain contains a 'Walker A' nucleotide-binding domain (GxxGxGKS/T, in the single-letter amino-acid code), which by mutational studies has been shown to be essential in several proteins. The carboxy-terminal domain contains a sequence (the DA-box) that resembles a 'Walker B' motif, and a motif with homology to the signature sequence of the ATP-binding cassette (ABC) family of ATPases. The sequence homology within the carboxy-terminal domain is relatively high within the SMC1-SMC4 group, whereas SMC5 and SMC6 show some divergence in both of these sequences. In eukaryotic cells, the proteins are found as heterodimers of SMC1 paired with SMC3, SMC2 with SMC4, and SMC5 with SMC6 (formerly known as Rad18)..	33.10	9.29E-02	14%	43%	45–183	171–299	5	EAS36015	hypothetical protein CIMG_01369	51.22	1.42E-04	23%	45%	32–227	556–762
					6	COG0497	RecN, ATPase involved in DNA repair [DNA replication, recombination, and repair].	33.26	9.93E-02	19%	38%	31–155	279–413	6	NP_186943	unknown protein	50.45	2.42E-04	21%	50%	32–182	295–440
					7	COG5280	COG5280, Phage-related minor tail protein [Function unknown].	32.70	1.27E-01	19%	32%	47–250	120–325	7	EAR81593	Protein kinase domain containing protein	49.29	5.39E-04	25%	46%	54–185	450–592
					8	cd00584	Prefoldin_alpha, Prefoldin alpha subunit; Prefoldin is a hexameric molecular chaperone complex, found in both eukaryotes and archaea, that binds and stabilizes newly synthesized polypeptides allowing them to fold correctly. The complex contains two alpha and four beta subunits, the two subunits being evolutionarily related. In archaea, there is usually only one gene for each subunit while in eukaryotes there two or more paralogous genes encoding each subunit adding heterogeneity to the structure of the hexamer. The structure of the complex consists of a double beta barrel assembly with six protruding coiled-coils..	32.91	1.34E-01	25%	42%	90–185	12–118	8	XP_730668	hypothetical protein PY02748	48.52	9.19E-04	21%	45%	27–262	266–489
					9	COG0712	AtpH, F0F1-type ATP synthase, delta subunit (mitochondrial oligomycin sensitivity protein) [Energy production and conversion].	32.13	2.17E-01	18%	41%	37–190	8–150	9	AAS10477	merozoite surface protein 3b	48.14	1.20E-03	24%	41%	30–245	387–609
					10	COG1842	PspA, Phage shock protein A (IM30), suppresses sigma54-dependent transcription [Transcription / Signal transduction mechanisms].	31.79	2.42E-01	22%	43%	66–189	13–156	10	BAB66120	337aa long hypothetical protein	47.37	2.05E-03	26%	44%	29–180	35–190
N262R	114211–114780	190	20,299	4.71		No Hit Found								1	AAC96919	similar to Chlorella virus CVK2 DNA binding protein, corresponds to GenBank Accession Number D78305	91.66	1.38E-17	43%	71%	65–167	61–163
														2	BAA11342	DNA binding protein	91.28	1.80E-17	43%	71%	65–167	61–163
N263R	114795–115994	400	42,800	10.56	1	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. .	34.14	2.74E-02	29%	47%	1–79	477–550	1	AAC96919	similar to Chlorella virus CVK2 DNA binding protein, corresponds to GenBank Accession Number D78305	53.91	1.23E-05	36%	57%	1–82	191–272
					2	COG2811	NtpF, Archaeal/vacuolar-type H+-ATPase subunit H [Energy production and conversion].	32.63	8.67E-02	23%	46%	1–79	2–73	2	BAA11342	DNA binding protein	51.22	7.95E-05	35%	56%	1–82	191–272
					3	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..	31.64	1.73E-01	28%	42%	5–82	96–177	3	BAA83789	alginate lyase	97.44	9.67E-19	35%	51%	212–377	153–317
					4	COG3064	TolA, Membrane protein involved in colicin uptake [Cell envelope biogenesis, outer membrane].	31.30	2.16E-01	27%	45%	13–80	138–205	4	AAC96583	PBCV-1 alginate lyase	95.90	2.81E-18	35%	50%	212–377	141–305

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	COG0711	AtpF, F0F1-type ATP synthase, subunit b [Energy production and conversion]	29.07	9.59E-01	21%	42%	31--79	58--106	5	BAB19127	vAL-1	94.74	6.27E-18	34%	51%	212--377	169--333
														6	EAL19679	hypothetical protein CNBG3070	48.14	6.73E-04	26%	44%	189--372	157--344
														7	AAW44587	hypothetical protein CNG01710	48.14	6.73E-04	26%	44%	189--372	147--334
														8	AAW45360	expressed protein	46.98	1.50E-03	26%	44%	218--388	324--489
														9	EAL19065	hypothetical protein CNBH1670	46.98	1.50E-03	26%	44%	218--388	324--489
														10	BAE48156	hypothetical methionyl-tRNA synthetase	45.82	3.34E-03	28%	41%	228--372	82--231
N266R	116058--116723	222	24,410	10.29	1	pfam04443	LuxE, Acyl-protein synthetase, LuxE. LuxE is an acyl-protein synthetase found in bioluminescent bacteria. LuxE catalyses the formation of an acyl-protein thioester from a fatty acid and a protein. This is the second step in the bioluminescent fatty acid reduction system, which converts tetradecanoic acid to the aldehyde substrate of the luciferase-catalysed bioluminescence reaction A conserved cysteine found at position 364 in Photobacterium phosphoreum LuxE is thought to be acylated during the transfer of the acyl group from the synthetase subunit to the reductase. The carboxyl terminal of the synthetase is though to act as a flexible arm to transfer acyl groups between the sites of activation and reduction. This family also includes Vibrio cholerae RBFN protein, which is involved in the biosynthesis of the O-antigen component 3-deoxy-L-glycero-tetronic acid	31.07	1.29E-01	25%	36%	175--219	158--202	1	AAC96918	A559L	80.49	4.51E-14	29%	45%	1--178	1--185
N267L	117565--116714	284	32,264	8.54	1	cd01170	THZ_kinase, 4-methyl-5-beta-hydroxyethylthiazole (Thz) kinase catalyzes the phosphorylation of the hydroxylgroup of Thz. A reaction that allows cells to recycle Thz into the thiamine biosynthesis pathway, as an alternative to its synthesis from cysteine, tyrosine and 1-deoxy-D-xylulose-5-phosohate..	29.76	3.85E-01	32%	45%	224--260	42--80	1	AAT60033	intein-containing protein	44.67	4.39E-03	23%	43%	94--272	68--244
N269L	118036--117614	141	14,826	4.53	1	pfam00692	dUTPase, dUTPase. dUTPase hydrolyses dUTP to dUMP and pyrophosphate..	112.64	1.47E-26	48%	67%	11--126	2--116	1	AAW51452	deoxyuridine triphosphatase	185.65	4.02E-46	71%	85%	2--127	3--128
					2	COG0717	Dcd, Deoxycytidine deaminase [Nucleotide transport and metabolism].	48.73	2.88E-07	27%	46%	32--117	74--158	2	AAC96912	similar to tomato dUTP pyrophosphatase, corresponds to GenBank Accesssion Number S40549	184.50	8.96E-46	69%	87%	2--127	3--128
					3	cd01458	vWA_ku, Ku70/Ku80 N-terminal domain. The Ku78 heterodimer (composed of Ku70 and Ku80) contributes to genomic integrity through its ability to bind DNA double-strand breaks (DSB) in a preferred orientation. DSB's are repaired by either homologues recombination or non-homologues end joining and facilitate repair by the non-homologous end-joining pathway (NHEJ). The Ku heterodimer is required for accurate process that tends to preserve the sequence at the junction. Ku78 is found in all three kingdoms of life. However, only the eukaryotic proteins have a vWA domain fused to them at their N-termini. The vWA domain is not involved in DNA binding but may very likey mediate Ku78's interactions with other proteins. Members of this subgroup lack the conserved MIDAS motif	27.97	5.19E-01	29%	44%	65--127	28--87	3	AAW51453	deoxyuridine triphosphatase	182.96	2.61E-45	69%	84%	2--127	3--128
					4	pfam03731	Ku_N, Ku70/Ku80 N-terminal alpha/beta domain. The Ku heterodimer (composed of Ku70 and Ku80) contributes to genomic integrity through its ability to bind DNA double-strand breaks and facilitate repair by the non-homologous end-joining pathway. This is the amino terminal alpha/beta domain. This domain only makes a small contribution to the dimer interface. The domain comprises a six stranded beta sheet of the Rossman fold	27.22	7.28E-01	28%	54%	73--127	35--85	4	NP_190278	dUTP diphosphatase/ hydrolase	148.29	7.12E-35	60%	74%	3--127	29--153
					5	COG2433	COG2433, Uncharacterized conserved protein [Function unknown].	27.18	8.95E-01	23%	46%	69--127	5--61	5	EAS33768	hypothetical protein CIMG_04792	148.29	7.12E-35	61%	76%	4--127	27--150
														6	XP_750039	dUTPase	147.90	9.29E-35	59%	75%	4--127	114--237
														7	XP_657875	hypothetical protein AN0271.2	147.52	1.21E-34	59%	76%	4--127	71--194
														8	AAB22611	deoxyuridine triphosphatase; dUTPase; P18	147.52	1.21E-34	60%	72%	3--126	32--155
														9	XP_469212	putative deoxyuridine triphosphatase	143.67	1.75E-33	58%	74%	5--126	90--211
														10	ABF98101	Deoxyuridine 5'-triphosphate nucleotidohydrolase, putative, expressed	143.67	1.75E-33	58%	74%	5--126	36--157
N271R	118157--118981	275	30,440	8.75	1	pfam00352	TBP, Transcription factor TFIID (or TATA-binding protein, TBP)..	37.41	1.89E-03	24%	48%	142--237	3--85	1	AAC96913	similar to Sulfolobus TATA-binding protein, corresponds to GenBank Accession Number S55311'	180.64	4.82E-44	41%	60%	25--268	13--268
														2	YP_142807	TATA-box binding protein (TBP)	49.68	1.28E-04	25%	40%	75--252	138--346
N274L	120289--118979	437	48,753	9.11	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 PBCV-1 the major capsid protein is a glycoprotein	362.71	4.67E-101	40%	57%	36--431	2--443	1	AAC96917	similar to Chlorella virus PBCV-1 ORF A11L, corresponds to GenBank Accession Number U42580	509.22	1.20E-142	60%	76%	36--435	1--400
														2	AAC96379	contains aminoacyl-tRNA synthetase class-II signature	396.74	8.71E-109	49%	67%	37--435	3--403
														3	AAC96378	similar to PBCV-1 major capsid protein, corresponds to Swiss-Prot Accession Number P30328	320.86	6.07E-86	40%	62%	37--433	4--399
														4	AAC27492	major capsid protein Vp49	274.63	4.99E-72	38%	53%	37--435	3--432
														5	BAA76601	major capsid protein MCP1	269.24	2.10E-70	39%	54%	37--435	3--437
														6	AAC96798	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	268.86	2.74E-70	38%	54%	37--435	3--437
														7	BAA76600	major capsid protein	265.77	2.32E-69	37%	54%	37--435	3--436
														8	BAA22198	major capsid protein Vp54	260.00	1.27E-67	37%	53%	37--435	3--437
														9	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	238.81	3.04E-61	37%	53%	67--435	9--413
														10	1M3Y_D	Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid Containing Dna Virus	238.81	3.04E-61	37%	53%	67--435	9--413
N277L	121892--120225	556	62,182	9.61	1	COG0553	HepA, Superfamily II DNA/RNA helicases, SNF2 family [Transcription / DNA replication, recombination, and repair]	168.40	1.95E-42	26%	42%	81--533	314--851	1	AAC96911	similar to Caenorhabditis transcription activator, corresponds to Swiss-Prot Accession Number P41877	450.28	9.12E-125	50%	71%	103--552	7--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, FTI 1)	127.70	3.45E-30	26%	44%	106--357	1--287	2	BAC15031	helicase (swi/snf family)	167.55	1.18E-39	30%	51%	103--517	606--1027
					3	COG1061	SSL2, DNA or RNA helicases of superfamily II [Transcription / DNA replication, recombination, and repair]	82.84	9.09E-17	22%	37%	102--361	36--286	3	CAE79539	putative helicase/SNF2 family domain protein	164.47	1.00E-38	29%	49%	98--517	856--1294
					4	smart00487	DEXDc, DEAD-like helicases superfamily; .	79.50	1.08E-15	17%	36%	103--288	9--199	4	NP_225044	SWI/SNF family helicase 2	163.31	2.23E-38	29%	47%	99--532	698--1154
					5	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	78.04	2.56E-15	27%	44%	370--497	3--130	5	CAC96916	lin1685	162.16	4.96E-38	29%	46%	99--548	612--1064
					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..	73.96	4.75E-14	25%	49%	124--259	1--143	6	BAA99057	SWI/SNF family helicase_2	161.00	1.11E-37	29%	46%	99--532	698--1154
					7	smart00490	HELICc, helicase superfamily c-terminal domain; .	63.71	5.64E-11	25%	46%	409--490	3--82	7	AAF73530	helicase, Snf2 family	159.84	2.46E-37	28%	47%	97--532	711--1169
					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..	62.54	1.44E-10	27%	47%	411--490	1--78	8	YP_014262	helicase, Snf2 family	159.46	3.22E-37	29%	46%	99--532	612--1053
					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	60.10	6.98E-10	18%	36%	125--259	2--144	9	ZP_00231434	helicase, Snf2 family	159.46	3.22E-37	29%	46%	99--532	612--1053
					10	COG0513	SrmB, Superfamily II DNA and RNA helicases [DNA replication, recombination, and repair / Transcription / Translation, ribosomal structure and biogenesis]	45.92	1.14E-05	20%	40%	348--492	227--364	10	AAC68303	SWF/SNF family helicase	158.69	5.49E-37	28%	47%	97--532	693--1151
N284R	121783--122544	254	27,361	4.17	1	COG4934	COG4934, Predicted protease [Posttranslational modification, protein turnover, chaperones]	28.42	9.66E-01	21%	36%	164--217	80--133	0	No Hit Found	No Hit Found						
N287L	122844--122626	73	6,960	12.35		No Hit Found								0	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
N288R	122936--123199	88	9,777	10.68	1	COG2339	COG2339, Predicted membrane protein [Function unknown].	26.43	9.34E-01	31%	43%	28--69	40--82	1	AAC96886	A519L	103.99	1.58E-21	64%	75%	7--80	4--77
N289R	123228--124391	388	44,114	9.57	1	COG1231	COG1231, Monoamine oxidase [Amino acid transport and metabolism].	52.68	7.73E-08	18%	30%	1--384	1--439	1	BAA83788	orf1	560.45	3.85E-158	71%	82%	8--383	19--394
					2	pfam03738	GSP_synth, Glutathionylspermidine synthase. This region contains the Glutathionylspermidine synthase enzymatic activity EC:6.3.1.8. This is the C-terminal region in bienzymes. Glutathionylspermidine (GSP) synthetases of Trypanosomatidae and Escherichia coli couple hydrolysis of ATP (to ADP and Pi) with formation of an amide bond between spermidine and the glycine carboxylate of glutathione (gamma-Glu-Cys-Gly). In the pathogenic trypanosomatids, this reaction is the penultimate step in the biosynthesis of the antioxidant metabolite, trypanothione (N1,N8-bis-(glutathionyl)spermidine), and is a target for drug design..	30.65	3.66E-01	28%	42%	117--184	192--257	2	AAC96585	similar to bovine monoamine oxidase, corresponds to Swiss-Prot Accession Number P21398	549.67	6.79E-155	69%	81%	8--383	16--391
N293R	124417--125514	366	40,178	10.11		No Hit Found								1	AAC96583	PBCV-1 alginate lyase	416.39	8.27E-115	64%	73%	61--362	5--320
														2	BAA83789	alginate lyase	406.76	6.56E-112	60%	70%	61--362	5--332
														3	BAB19127	vAL-1	404.06	4.25E-111	55%	65%	1--362	1--348
														4	BAE48156	hypothetical methionyl-tRNA synthetase	80.11	1.41E-13	28%	44%	125--357	5--247
														5	AAC96919	similar to Chlorella virus CVK2 DNA binding protein, corresponds to GenBank Accession Number D78305	65.47	3.58E-09	30%	50%	230--363	501--648
														6	BAA11342	DNA binding protein	64.31	7.98E-09	31%	48%	230--363	500--647
														7	AAC96539	similar to PBCV-1 ORF A41R, corresponds to GenBank Accession Number U17055	48.91	3.47E-04	40%	56%	3--59	4--60
														8	AAC96508	PBCV-1 surface protein	44.67	6.54E-03	42%	62%	3--52	906--955
														9	BAD22850	surface protein	44.67	6.54E-03	42%	62%	3--52	906--955
														10	T17636	proline-rich protein A145R - Chlorella virus PBCV-1	44.67	6.54E-03	42%	62%	3--52	115--164
N299R	125532--126188	219	24,231	10.35		No Hit Found								0	No Hit Found	No Hit Found						
N303R	126293--126754	154	17,595	6.55		No Hit Found								1	AAC96582	A214L	130.57	1.55E-29	46%	65%	17--149	3--135
														2	XP_767961	hypothetical protein GLP_228_21235_22845	41.97	7.24E-03	23%	49%	7--113	145--255
N304R	126830--127282	151	16,662	4.56		No Hit Found								1	AAC96581	A213L	139.04	4.41E-32	53%	79%	1--126	1--125
N307L	128331--127321	337	37,918	4.05		No Hit Found								0	No Hit Found	No Hit Found						
N312L	129519--128398	374	41,982	5.21	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..	198.56	8.96E-52	39%	57%	25--260	4--246	1	AAC96575	PBCV-1 arginine decarboxylase	497.28	3.82E-139	63%	79%	1--372	1--372
					2	COG0019	LysA, Diaminopimelate decarboxylase [Amino acid transport and metabolism].	195.50	7.90E-51	27%	44%	19--372	29--394	2	CAE02644	ornithine decarboxylase	269.63	1.29E-70	40%	59%	1--364	43--407
					3	pfam00278	Orn_DAP_Arg_deC, Pyridoxal-dependent decarboxylase, C-terminal sheet domain. These pyridoxal-dependent decarboxylases act on ornithine. lysine. arginine and related substrates. .	89.49	5.74E-19	30%	46%	263--365	1--103	3	CAA71498	ornithine decarboxylase	266.93	8.39E-70	38%	57%	1--364	48--418
					4	COG1166	SpeA, Arginine decarboxylase (spermidine biosynthesis) [Amino acid transport and metabolism].	53.33	4.42E-08	23%	41%	45--270	118--374	4	XP_968571	PREDICTED: similar to Ornithine decarboxylase (ODC)	266.54	1.10E-69	39%	57%	9--364	26--385
					5	COG1387	HIS2, Histidinol phosphatase and related hydrolases of the PHP family [Amino acid transport and metabolism / General function prediction only].	29.54	6.66E-01	24%	44%	152--219	15--87	5	AAL87201	ornithine decarboxylase	266.16	1.43E-69	40%	59%	1--364	50--420
														6	BAA83427	ornithine decarboxylase	265.77	1.87E-69	39%	58%	1--364	47--417
														7	AAL83709	putative ornithine decarboxylase 1	264.62	4.16E-69	40%	59%	1--364	50--420
														8	NP_571876	ornithine decarboxylase 1	263.85	7.10E-69	37%	56%	2--372	21--409
														9	AAK13622	ornithine decarboxylase	263.46	9.27E-69	38%	57%	1--364	48--418
														10	CAB61758	ornithine decarboxylase	263.08	1.21E-68	38%	60%	13--371	54--428
N313L	130282--129647	212	23,020	11.49	1	cd01839	SGNH_arylesterase_like, SGNH_hydrolase subfamily, similar to arylesterase (7-aminocephalosporanic acid-deacetylating enzyme) of A. tumefaciens. SGNH hydrolases are a diverse family of lipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alpha/beta hydrolase family and unique among all known hydrolases; its active site closely resembles the Ser-His-Asp(Glu) triad found in other serine hydrolases	28.33	7.78E-01	32%	48%	67--111	89--133	1	AAC96573	A205R	176.79	4.21E-43	52%	62%	13--203	9--205
														2	AAC96461	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563	44.67	2.50E-03	35%	53%	126--191	3--69
N315L	130984--130319	222	23,976	4.83		No Hit Found								1	AAC96571	A203R	269.63	5.24E-71	63%	74%	1--220	1--213
N317R	131031--131372	114	12,026	4.52	1	COG2217	ZntA, Cation transport ATPase [Inorganic ion transport and metabolism].	26.73	7.52E-01	28%	52%	13--38	13--38	1	AAC96570	A202L	117.09	1.77E-25	52%	71%	3--112	2--112
N318L	131927--131367	187	20,852	10.87	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..	131.90	4.20E-32	35%	53%	26--159	4--139	1	AAC96673	similar to human protein Tyr-phosphatase, corresponds to GenBank Accession Number U27193	242.66	4.62E-63	66%	85%	6--168	7--169
					2	pfam00782	DSPc, Dual specificity phosphatase, catalytic domain. Ser/Thr and Tyr protein phosphatases. The enzymes have a tertiary fold is highly similar to that of tyrosine-specific phosphatases, except for a "recognition" region..	124.98	4.44E-30	35%	55%	24--161	2--139	2	NP_998405	dual specificity phosphatase 16	87.43	2.48E-16	34%	52%	24--164	159--300
					3	smart00195	DSPc, Dual specificity phosphatase, catalytic domain; .	122.69	2.52E-29	34%	56%	24--161	2--139	3	XP_543810	PREDICTED: similar to Dual specificity protein phosphatase 16 (Mitogen-activated protein kinase phosphatase 7) (MAP kinase phosphatase 7) (MKP-7)	84.73	1.61E-15	34%	53%	24--163	159--299
					4	COG2453	CDC14, Predicted protein-tyrosine phosphatase [Signal transduction mechanisms].	51.24	7.52E-08	27%	50%	58--143	65--147	4	AAH42101	DUSP16 protein	83.57	3.58E-15	34%	53%	24--163	159--299
					5	smart00404	PTPc motif, Protein tyrosine phosphatase, catalytic domain motif; .	35.01	5.22E-03	22%	48%	99--147	36--90	5	AAH31643	Unknown (protein for IMAGE:5176724)	83.57	3.58E-15	34%	53%	24--163	110--250
					6	pfam00102	Y_phosphatase, Protein-tyrosine phosphatase..	33.42	1.68E-02	22%	50%	98--143	166--216	6	AAI09236	Dual specificity phosphatase 16	83.57	3.58E-15	34%	53%	24--163	159--299
					7	smart00194	PTPc, Protein tyrosine phosphatase, catalytic domain; .	31.45	7.33E-02	23%	47%	72--143	168--243	7	AAI09235	Dual specificity phosphatase 16	83.57	3.58E-15	34%	53%	24--163	159--299
					8	COG2365	COG2365, Protein tyrosine/serine phosphatase [Signal transduction mechanisms].	31.19	8.72E-02	37%	50%	101--139	136--174	8	XP_520751	PREDICTED: similar to KIAA1700 protein	83.57	3.58E-15	34%	53%	24--163	268--408
					9	pfam05693	Glycogen_syn, Glycogen synthase. This family consists of the eukaryotic glycogen synthase proteins GYS1, GYS2 and GYS3. Glycogen synthase (GS) is the enzyme responsible for the synthesis of -1,4-linked glucose chains in glycogen. It is the rate limiting enzyme in the synthesis of the polysaccharide, and its activity is highly regulated through phosphorylation at multiple sites and also by allosteric effectors, mainly glucose 6-ohosohate (G6P)\..	29.88	2.28E-01	31%	43%	78--148	174--249	9	BAB21791	KIAA1700 protein	83.57	3.58E-15	34%	53%	24--163	184--324
					10	cd00047	PTPc, Protein tyrosine phosphatases (PTP) catalyze the dephosphorylation of phosphotyrosine peptides; they regulate phosphotyrosine levels in signal transduction pathways. The depth of the active site cleft renders the enzyme specific for phosphorylated Tyr (pTyr) residues, instead of pSer or pThr. This family has a distinctive active site signature motif, HCSAGxGRxG. Characterized as either transmembrane, receptor-like or non-transmembrane (soluble) PTPs. Receptor-like PTP domains tend to occur in two copies in the cytoplasmic region of the transmembrane proteins, only one copy may be active..	29.06	3.51E-01	42%	74%	101--120	166--185	10	AAH59232	Dusp16 protein	82.42	7.98E-15	35%	52%	24--162	159--298
N320L	132255--131986	90	10,333	9.82		No Hit Found								1	AAC96989	A687R	43.13	3.29E-03	36%	52%	16--83	6--74
N321L	132624--132364	87	10,211	3.83	1	cd03317	NAAAR, N-acylamino acid racemase (NAAAR), an octameric enzyme that catalyzes the racemization of N-acylamino acids. NAAARs act on a broad range of N-acylamino acids rather than amino acids. Enantiopure amino acids are of industrial interest as chiral building blocks for antibiotics, herbicides, and drugs. NAAAR is a member of the enolase superfamily, characterized by the presence of an enolate anion intermediate which is generated by abstraction of the alpha-proton of the carboxylate substrate by an active site residue and is stabilized by coordination to the essential Mg2+ ion	27.84	2.86E-01	41%	66%	36--69	49--78	1	AAC96674	A306L	102.45	4.62E-21	82%	92%	34--85	35--86
N323L	133084--132743	114	12,977	5.02		No Hit Found								1	AAC96676	A308L	46.21	3.83E-04	38%	66%	43--98	20--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
N325L	133669--133160	170	18,092	8.10	1	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..	30.09	1.42E-01	42%	55%	91--123	7--40	1	AAC96678	A310L	184.50	1.15E-45	53%	70%	1--168	1--170
N327L	134603--133710	298	33,061	9.03	1	pfam03658	UPF0125, Uncharacterized protein family (UPF0125)..	28.62	9.59E-01	25%	40%	220--292	19--84	1	AAA88833	33 kDa in vitro translation peptide	340.12	5.45E-92	71%	82%	56--290	1--230
														2	NP_077561	EsV-1-76	83.57	9.20E-15	28%	43%	60--291	10--259
														3	AAR26966	FirrV-1-I1	58.15	4.14E-07	24%	38%	64--268	10--222
														4	AAR26885	FirrV-1-B10	55.84	2.05E-06	25%	39%	56--277	13--247
N329L	134854--134597	86	9,357	8.92		No Hit Found								1	AAC96681	A313L	46.98	2.31E-04	57%	69%	11--43	1--33
N330R	134941--135177	79	8,884	5.01		No Hit Found								1	AAC96682	A314R	117.86	1.05E-25	75%	80%	2--77	5--80
N331R	135231--136355	375	41,182	4.87		No Hit Found								1	AAW39962	membrane protein, putative	88.58	4.07E-16	26%	46%	130--357	18--248
														2	ZP_01139209	membrane protein, putative	87.43	9.06E-16	26%	46%	130--357	18--248
														3	CAI82895	hypothetical membrane protein	86.27	2.02E-15	25%	46%	130--357	18--248
														4	AAL80842	hypothetical protein	81.26	6.50E-14	22%	52%	126--358	2--231
														5	AAB85442	unknown	75.49	3.56E-12	26%	47%	119--358	14--262
N333R	136377--136733	119	12,895	10.56	1	COG5169	HSF1, Heat shock transcription factor [Transcription].	27.68	3.71E-01	27%	45%	51--106	112--167	1	AAC96688	A320R	68.17	9.56E-11	31%	47%	1--115	24--138
N335R	136751--137095	115	11,866	4.87		No Hit Found								1	AAC96689	A321R	75.49	5.87E-13	33%	61%	4--113	11--119
N336L	137697--137194	168	19,111	5.50		No Hit Found								1	AAC96690	A322L	121.32	1.14E-26	38%	56%	1--159	1--170
N339L	138996--137725	424	46,766	5.00	1	pfam05976	DUF893, Bacterial membrane protein of unknown function (DUF893). This family consists of several putative bacterial membrane proteins of unknown function..	30.23	4.37E-01	29%	46%	104--196	239--323	1	AAC96692	A324L	367.08	7.07E-100	49%	60%	11--422	13--453
					2	COG5602	SIN3, Histone deacetylase complex, SIN3 component [Chromatin structure and dynamics].	29.64	7.46E-01	26%	33%	319--394	334--407	2	AAR26897	FirrV-1-B22	79.72	2.26E-13	32%	55%	118--254	136--270
					3	pfam06459	RR_TM4-6, Ryanodine Receptor TM 4-6. This region covers TM regions 4-6 of the ryanodine receptor 1 family	29.22	9.89E-01	20%	32%	318--419	11--114	3	NP_077588	EsV-1-103	75.87	3.26E-12	31%	54%	121--256	232--365
														4	YP_142803	unknown	56.23	2.67E-06	30%	52%	140--250	225--333
														5	YP_294142	hypothetical protein EhV_384	48.14	7.27E-04	26%	48%	122--258	183--315
														6	AAL98780	ORF056L	47.75	9.50E-04	26%	50%	117--253	62--196
														7	AAT71869	ORF054L	44.67	8.04E-03	25%	50%	117--253	62--196
														8	AAX82366	ORF57L	44.67	8.04E-03	25%	50%	117--253	62--196
														9	NP_149530	067R	44.67	8.04E-03	26%	48%	120--261	232--372
N342L	139681--139040	214	24,766	8.76	1	COG0494	MutT, NTP pyrophosphohydrolases including oxidative damage repair enzymes [DNA replication, recombination, and repair / General function prediction only]	34.96	8.28E-03	21%	34%	10--180	5--161	1	AAC96694	A326L	234.96	1.31E-60	61%	81%	14--188	1--175
					2	pfam00293	NUDIX, NUDIX domain..	32.13	5.60E-02	23%	41%	17--155	4--119	2	CAH13320	hypothetical protein	43.90	4.29E-03	29%	53%	17--109	423--513
					3	cd00778	ProRS_core_arch_euk, Prolyl-tRNA synthetase (ProRS) class II core catalytic domain. ProRS is a homodimer. It is responsible for the attachment of proline to the 3' OH group of ribose of the appropriate tRNA. This domain is primarily responsible for ATP-dependent formation of the enzyme bound aminoacyl-adenylate. Class II assignment is based upon its structure and the presence of three characteristic sequence motifs in the core domain. This subfamily contains the core domain of ProRS from archaea, the cytoplasm of eukaryotes and some bacteria..	28.20	8.67E-01	23%	47%	102--144	104--147	3	YP_294156	hypothetical protein EhV_398	43.90	4.29E-03	27%	48%	15--134	14--141
														4	XP_647200	hypothetical protein DDB0216539	43.13	7.31E-03	27%	51%	17--147	250--390
N346L	140751--139690	354	40,496	9.78		No Hit Found								1	AAC96696	A328L	248.44	2.83E-64	37%	58%	1--352	1--350
N351R	141062--141790	243	26,762	5.32		No Hit Found								1	AAA88833	33 kDa in vitro translation peptide	340.12	3.81E-92	71%	82%	1--235	1--230
														2	NP_077561	EsV-1-76	84.73	2.88E-15	28%	43%	5--236	10--259
														3	AAR26966	FirrV-1-I1	58.54	2.21E-07	26%	40%	9--213	10--222
														4	AAR26885	FirrV-1-B10	57.38	4.93E-07	26%	40%	1--222	13--247
tRNAs																						
Ile	141848--141920	72 bs					anticodon: TAT															
Leu	141945--142028	83 bs					anticodon: TAA															
Arg	142053--142125	72 bs					anticodon: TCT															
Gly	142129--142199	70 bs					anticodon: TCC															
Asn	142223--142296	73 bs					anticodon: GTT															
Asn	142320--142393	73 bs					anticodon: GTT															
Tyr	142416--142501	85 bs					anticodon: GTA															
Intron	142453--142465	13 bs					Intron (142453-142465)															
Lys	142504--142576	72 bs					anticodon: CTT															
Thr	142738--142809	71 bs					anticodon: CGT															
N354L	143383--142856	176	20,240	6.76		No Hit Found								1	AAC96705	A337L	66.24	5.00E-10	48%	59%	100--173	5--78
N357R	143563--143880	106	11,635	8.82		No Hit Found								0	No Hit Found	No Hit Found						
N359L	145663--143981	561	60,560	9.73		No Hit Found								1	AAC96710	A342L	736.49	0.00E+00	68%	78%	16--550	37--561
N364R	145777--147096	440	48,944	4.51	1	COG3534	AbfA, Alpha-L-arabinofuranosidase [Carbohydrate transport and metabolism]	41.82	1.68E-04	25%	41%	57--247	66--256	1	CAD86595	cellulase precursor	55.84	3.68E-06	23%	35%	54--367	282--574
					2	pfam03662	Glyco_hydro_79n, Glycosyl hydrolase family 79, N-terminal domain. Family of endo-beta-N-glucuronidase, or heparanase. Heparan sulfate proteoglycans (HSPGs) play a key role in the self- assembly, insolubility and barrier properties of basement membranes and extracellular matrices. Hence, cleavage of heparan sulfate (HS) affects the integrity and functional state of tissues and thereby fundamental normal and pathological phenomena involving cell migration and response to changes in the extracellular micro-environment. Heparanase degrades HS at specific intra-chain sites. The enzyme is synthesised as a latent approximately 65 kDa protein that is processed at the N-terminus into a highly active approximately 50 kDa form. Experimental evidence suggests that heparanase may facilitate both tumour cell invasion and neovascularization, both critical steps in cancer progression. The enzyme is also involved in cell migration associated with inflammation and autoimmunity	40.37	4.12E-04	30%	48%	146--237	149--243	2	CAF22222	putative cellulase	49.68	2.64E-04	27%	39%	57--236	273--416
					3	COG1554	ATH1, Trehalose and maltose hydrolases (possible phosphorylases) [Carbohydrate transport and metabolism].	32.96	7.22E-02	36%	43%	62--162	394--484	3	ZP_00766677	similar to Alpha-L-arabinofuranosidase	49.29	3.45E-04	24%	38%	57--396	105--466
					4	COG1809	COG1809, Uncharacterized conserved protein [Function unknown].	31.43	2.03E-01	48%	76%	205--230	22--47	4	BAA10965	cellulose-binding protein	48.14	7.68E-04	23%	37%	57--290	677--879
					5	pfam03632	Glyco_hydro_65m, Glycosyl hydrolase family 65 central catalytic domain. This family of glycosyl hydrolases contains vacuolar acid trehalase and maltose phosphorylase.Maltose phosphorylase (MP) is a dimeric enzyme that catalyses the conversion of maltose and inorganic phosphate into beta-D-glucose-1-phosphate and glucose. The central domain is the catalytic domain, which binds a phosphate ion that is proximal the the highly conserved Glu. The arrangement of the phosphate and the glutamate is thought to cause nucleophilic attack on the anomeric carbon atom. The catalytic domain also forms the majority of the dimerisation interface	31.03	2.61E-01	29%	46%	62--167	76--159	5	AAC45377	endoglucanase F precursor	47.37	1.31E-03	23%	41%	143--290	728--879
N367L	147566--147138	143	16,043	10.44		No Hit Found								1	BAA22200	URF14.2	158.30	7.04E-38	68%	81%	14--122	1--109

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	BAA22202	URF14.2	157.92	9.20E-38	68%	82%	14--122	1--109
														3	BAA22201	URF14.2	157.53	1.20E-37	67%	81%	14--122	1--109
														4	BAA22199	URF14.2	156.38	2.68E-37	66%	81%	14--122	1--109
														5	AAC96717	A349L	111.69	7.56E-24	73%	86%	16--83	74--141
N369L	149048--148017	344	38,430	8.32	1	cd00315	Cyt_C5_DNA_methylase, Cytosine-C5 specific DNA methylases; Methyl transfer reactions play an important role in many aspects of biology. Cytosine-specific DNA methylases are found both in prokaryotes and eukaryotes. DNA methylation, or the covalent addition of a methyl group to cytosine within the context of the CpG dinucleotide, has profound effects on the mammalian genome. These effects include transcriptional repression via inhibition of transcription factor binding or the recruitment of methyl-binding proteins and their associated chromatin remodeling factors, X chromosome inactivation, imprinting and the suppression of parasitic DNA sequences. DNA methylation is also essential for proper embryonic development and is an important player in both DNA repair and genome stability	162.40	6.83E-41	30%	43%	5--259	4--271	1	AAC96884	M.CviAII cytosine DNA methyltransferase	447.20	3.96E-124	61%	74%	1--340	1--342
					2	pfam00145	DNA methylase, C-5 cytosine-specific DNA methylase..	156.24	4.66E-39	29%	44%	5--264	4--267	2	AAC96897	M.CviAIV cytosine DNA methyltransferase	437.96	2.40E-121	61%	76%	1--328	2--332
					3	COG0270	Dcm, Site-specific DNA methylase [DNA replication, recombination, and repair]	108.24	1.21E-24	28%	44%	1--191	3--197	3	AAV84097	CviPII m5C DNA methyltransferase	309.30	1.29E-82	47%	61%	2--340	16--357
					4	pfam05478	Prominin, Prominin. The prominins are an emerging family of proteins that among the multispan membrane proteins display a novel topology. Mouse prominin and human prominin (mouse)-like 1 (PROML1) are predicted to contain five membrane spanning domains, with an N-terminal domain exposed to the extracellular space followed by four, alternating small cytoplasmic and large extracellular, loops and a cytoplasmic C-terminal domain. The exact function of prominin is unknown although in humans defects in PROM1, the gene coding for prominin, cause retinal degeneration	30.65	3.11E-01	37%	57%	205--240	200--235	4	AAC64006	cytosine methyltransferase	286.19	1.17E-75	44%	58%	2--342	3--358
														5	AAC55063	cytosine methyltransferase	263.08	1.06E-68	40%	56%	5--342	6--362
														6	AAC96987	nonfunctional M.CviAV cytosine DNA methyltransferase	261.92	2.36E-68	40%	56%	5--342	6--362
														7	AAR23218	gp67	101.29	5.34E-20	33%	49%	2--162	6--192
														8	NP_818425	gp127	92.43	2.48E-17	35%	51%	5--155	6--165
														9	NP_150145	putative DNA methylase	92.05	3.24E-17	32%	44%	5--222	4--226
														10	AAC98421	methyl transferase	92.05	3.24E-17	33%	47%	5--170	4--178
N372L	149874--149227	216	23,956	3.74	1	COG1966	CstA, Carbon starvation protein, predicted membrane protein [Signal transduction mechanisms]	28.31	8.02E-01	28%	53%	5--37	2--34	1	AAC96720	Asp/Glu rich; DAEDDDIYxxET (2X) negative charge cluster	256.91	3.30E-67	63%	71%	1--214	1--207
														2	YP_142843	unknown	51.22	2.75E-05	32%	58%	135--214	140--221
N373L	150935--149967	323	36,800	4.01		No Hit Found								1	AAC96725	A357L	164.08	6.10E-39	50%	72%	140--300	86--249
N377L	151591--150974	206	23,714	5.64		No Hit Found								0	No Hit Found	No Hit Found						
N380R	151657--152352	232	26,010	9.70	1	smart00497	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished)	47.43	1.58E-06	37%	54%	175--228	1--53	1	AAC96973	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	215.31	1.28E-54	50%	63%	1--227	1--224
					2	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	44.30	1.26E-05	34%	54%	1--91	1--83	2	AAA88832	unknown	201.06	2.50E-50	44%	60%	1--228	1--242
					3	cd00283	GIY-YIG_Clerm, GIY(10-11)YIG family of class I homing endonucleases C-terminus (GIY-YIG_Clerm). 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	41.14	1.33E-04	46%	61%	111--226	19--113	3	AAC96655	PBCV-1 33kd peptide	172.56	9.51E-42	41%	54%	5--227	7--248
					4	pfam07453	NUMOD1, NUMOD1 domain..	37.33	1.52E-03	43%	63%	175--205	1--31	4	AAC96862	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	141.35	2.35E-32	41%	55%	1--201	1--196
					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	36.68	2.71E-03	28%	47%	1--88	1--89	5	YP_293795	putative endonuclease	72.40	1.34E-11	39%	55%	2--108	3--110
					6	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	29.44	3.90E-01	34%	53%	1--31	16--48	6	AAC96906	A539R	53.14	8.41E-06	34%	53%	4--104	34--133
														7	CAA25939	unnamed protein product	50.06	7.12E-05	32%	54%	155--227	75--143
														8	NP_049674	MobB homing endonuclease	50.06	7.12E-05	32%	54%	155--227	112--180
														9	AAC96502	similar to bacteriophage T4 intron-associated endonuclease, corresponds to Swiss-Prot Accession Number P13299	48.91	1.59E-04	34%	52%	2--89	9--95
														10	NP_899393	SeqD	48.91	1.59E-04	31%	43%	1--173	1--157
N383R	152377--155754	1126	123,633	10.87	1	smart00490	HELICc, helicase superfamily c-terminal domain; .	37.52	1.02E-02	36%	60%	848--890	38--80	1	AAC96731	similar to chicken vitellogenin II, corresponds to Swiss-Prot Accession Number P02845	835.48	0.00E+00	61%	75%	352--1028	1--651
					2	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..	36.35	1.96E-02	24%	51%	841--890	27--76	2	AAC96728	A360R	240.74	2.50E-61	55%	68%	4--225	14--239
					3	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	31.82	4.33E-01	26%	57%	848--890	79--121	3	AAC96729	A361R	130.18	4.76E-28	70%	84%	248--331	2--85
														4	ABF82117	hypothetical protein MIV087L	58.92	1.35E-06	25%	41%	785--938	414--556
														5	YP_142731	putative NTPase I	49.68	8.18E-04	24%	46%	833--941	608--721
														6	YP_142917	helicase conserved C-terminal domain protein	46.21	9.04E-03	30%	51%	832--912	437--519
														7	CAJ57278	putative helicase	46.21	9.04E-03	26%	44%	842--965	318--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
N387R	155817--156737	307	35,901	10.26	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	56.55	4.35E-09	46%	66%	126--187	1--60	1	AAA88832	unknown	84.34	5.68E-15	34%	51%	29--198	1--175
					2	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	31.20	1.69E-01	32%	57%	33--117	6--83	2	AAC96862	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	79.72	1.40E-13	28%	45%	29--276	1--218
					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	30.90	1.99E-01	25%	43%	30--112	2--85	3	AAC96973	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	74.33	5.88E-12	28%	41%	29--276	1--220
														4	AAC96655	PBCV-1 33kd peptide	73.17	1.31E-11	28%	44%	39--276	15--244
														5	AAC49244	ORF301	66.24	1.60E-09	30%	50%	38--186	83--240
														6	AAU16837	GIY-YIG catalytic domain containing protein; possible intron encoded endonuclease	63.54	1.04E-08	29%	46%	28--180	2--155
														7	NP_074951	orf305	57.77	5.70E-07	28%	52%	42--178	85--227
														8	CAC51107	putative GIY-YIG endonuclease	57.00	9.72E-07	25%	40%	27--274	5--225
														9	AAC49248	ORF211	55.84	2.16E-06	44%	68%	104--161	86--139
														10	AAK09365	intron encoded Bmol	53.91	8.23E-06	23%	41%	40--249	14--242
N389L	157485--156868	206	23,558	10.99	1	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	34.28	1.14E-02	21%	38%	76--172	1--83	1	AAC96747	A379L	164.85	1.54E-39	44%	60%	12--203	5--203
					2	COG0322	UvrC, Nuclease subunit of the excinuclease complex [DNA replication, recombination. and repair].	28.33	7.66E-01	41%	47%	72--111	16--50	2	NP_911913	putative probable transcription repressor HOTR	46.60	6.11E-04	25%	40%	42--182	29--193
														3	CAA04677	putative transcription repressor HOTR	45.82	1.04E-03	25%	40%	42--198	26--206
N391R	157646--158395	250	28,257	8.78	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	66.95	2.30E-12	46%	63%	101--172	1--72	1	AAA88832	unknown	246.90	4.59E-64	51%	70%	1--244	1--240
					2	smart00497	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished). .	42.04	6.10E-05	35%	52%	193--246	1--53	2	AAC96973	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	213.77	4.31E-54	49%	61%	1--247	1--226
					3	pfam07453	NUMOD1, NUMOD1 domain..	37.33	1.74E-03	44%	63%	193--225	1--33	3	AAC96655	PBCV-1 33kd peptide	198.75	1.43E-49	46%	61%	5--244	7--247
					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. .	36.29	3.93E-03	26%	46%	1--88	1--88	4	AAC96862	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	171.79	1.88E-41	45%	62%	1--219	1--196
					5	pfam05118	Asp_Arg_Hydrox, Aspartyl/Asparaginyl beta-hydroxylase. Iron (II)/2-oxoglutarate (2-OG)-dependent oxygenases catalyse oxidative reactions in a range of metabolic processes. Proline 3-hydroxylase hydroxylates proline at position 3, the first of a 2-OG oxygenase catalysing oxidation of a free alpha-amino acid. The structure of proline 3-hydroxylase contains the conserved motifs present in other 2-OG oxygenases including a jelly roll strand core and residues binding iron and 2-oxoglutarate, consistent with divergent evolution within the extended family. This family represent the arginine, asparagine and proline hydroxylases. The aspartyl/asparaginyl beta-hydroxylase (EC:1.14.11.16) specifically hydroxylates one aspartic or asparagine residue in certain epidermal growth factor-like domains of a number of proteins..	30.27	2.43E-01	26%	39%	20--91	7--69	5	AAK09365	intron encoded Bmol	82.42	1.50E-14	34%	48%	4--219	5--242
					6	COG4678	COG4678, Muramidase (phage lambda lysozyme) [Carbohydrate transport and metabolism].	28.40	9.87E-01	23%	42%	61--116	24--84	6	NP_899393	SegD	81.65	2.56E-14	33%	46%	1--196	1--205
														7	CAA73995	unnamed protein product	76.64	8.22E-13	31%	49%	14--204	124--311
														8	AAC49244	ORF301	75.49	1.83E-12	32%	52%	9--159	80--240
														9	NP_074951	orf305	71.63	2.65E-11	34%	53%	4--162	74--236
														10	YP_293795	putative endonuclease	67.40	4.99E-10	37%	55%	2--109	3--110
N395R	158437--159858	474	53,895	6.11	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 PBCV-1 the major capsid protein is a glycoprotein	182.43	8.07E-47	29%	45%	81--470	3--422	1	AAC96751	similar to PBCV-1 major capsid protein, corresponds to Swiss-Prot Accession Number P30328	288.50	3.72E-76	52%	72%	148--405	1--257
					2	pfam03635	Vps35, Vacuolar protein sorting-associated protein 35. Vacuolar protein sorting-associated protein (Vps) 35 is one of around 50 proteins involved in protein trafficking. In particular, Vps35 assembles into a retromer complex with at least four other proteins Vps5, Vps17, Vps26 and Vps29. Vps35 contains a central region of weaker sequence similarity, thought to indicate the presence of at least three domains..	29.51	9.05E-01	21%	39%	86--144	484--546	2	BAE06835	hypothetical major capsid protein	169.09	3.29E-40	31%	49%	81--446	3--401
														3	BAA76601	major capsid protein MCP1	142.12	4.31E-32	30%	45%	81--438	3--380
														4	BAA22198	major capsid protein Vp54	139.81	2.14E-31	29%	42%	81--469	3--410
														5	AAC27492	major capsid protein Vp49	139.43	2.79E-31	30%	44%	81--438	3--371
														6	AAC96798	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	139.43	2.79E-31	30%	42%	81--469	3--410
														7	BAA76600	major capsid protein	139.04	3.65E-31	29%	42%	81--469	3--409
														8	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	115.93	3.31E-24	28%	40%	103--469	1--386
														9	1M3Y_D	Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid Containing, Dna Virus	115.55	4.32E-24	28%	40%	103--469	1--386
														10	YP_142795	capsid protein	67.40	1.35E-09	31%	46%	284--430	310--465

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
N400L	161546--159861	562	62,960	11.03	1	cd00233	VIP2, VIP2; A family of actin-ADP-ribosylating toxin. A member of the Bacillus-prodiced vegetative insecticidal proteins (VIPs) possesses high specificity against the major insect pest, corn rootworms, and belongs to a classs of binary toxins and regulators of biological pathways distinct from classical A-B toxins. A novel family of insecticidal ADP-ribosyltransferses were isolated from Bacillus cereus during vegetative growth, where VIP1 likely targets insect cells and VIP2 ribosylates actin. VIP2 shares significant sequence similarity with enzymatic components of other binary toxins, Clostridium botulinum C2 toxin, C. perfringens iota toxin, C. piroforme toxin, C. piroforme toxin and C. difficile toxin..	33.38	8.72E-02	36%	55%	479--540	128--183	1	AAC96461	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563	57.77	1.34E-06	40%	54%	61--150	3--85
					2	pfam03148	Tektin, Tektin family. Tektins are cytoskeletal proteins. They have been demonstrated in such cellular sites as centrioles, basal bodies, and along ciliary and flagellar doublet microtubules. Tektins form unique protofilaments, organised as longitudinal polymers of tektin heterodimers with axial periodicity matching tubulin. Tektin polypeptides consist of several alpha-helical regions that are predicted to form coiled coils. Indeed, tektins share considerable structural similarities with intermediate filament proteins. Possible functional roles for tektins are: stabilisation of tubulin protofilaments; attachment of A and B-tubules in ciliary/flagellar microtubule doublets and C-tubules in centrioles; binding of axonemal components	31.40	2.94E-01	33%	48%	218--300	27--106	2	AAC96650	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055	73.94	1.80E-11	37%	46%	94--238	396--556
														3	AAC96459	a91L	115.16	7.06E-24	48%	66%	436--560	1--126
														4	AAC96646	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662	48.52	8.11E-04	31%	43%	115--238	465--597
														5	AAC96984	similar to Chlorella virus PBCV-1 ORF A282L, corresponds to GenBank Accession Number U42580	63.93	1.87E-08	61%	71%	58--114	2--55
N403R	161664--162374	237	27,164	4.53	1	COG1623	COG1623, Predicted nucleic-acid-binding protein (contains the HHH domain) [General function prediction only]	29.47	4.27E-01	22%	48%	2--48	301--347	1	YP_214596	hypothetical protein PSSM4_035	56.23	1.04E-06	27%	44%	6--205	306--511
					2	pfam06329	Fungal_ODC_AZ, Fungal ornithine decarboxylase antizyme. This family consists of several fungal ornithine decarboxylase antizyme proteins. The polyamine biosynthetic enzyme ornithine decarboxylase (ODC) is degraded by the 26 S proteasome via a ubiquitin-independent pathway. Its degradation is greatly accelerated by association with the polyamine-induced regulatory protein antizyme 1 (AZ1). This family is specific to fungal species but is related to the pfam02100 family	29.24	4.63E-01	37%	50%	195--227	110--140	2	CAJ74836	Conserved Hypothetical Protein	45.05	2.40E-03	27%	41%	33--146	48--187
N404R	162413--163957	515	59,192	8.27	1	pfam01171	ATP_bind_3, PP-loop family. This family of proteins belongs to the PP-loop superfamily..	125.36	1.62E-29	28%	48%	210--413	1--204	1	AAC96914	similar to MesJ cell cycle protein	523.86	5.86E-147	52%	67%	15--508	1--497
					2	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 stronaly conserved motif SGGXD at the N terminus.	123.41	6.24E-29	29%	48%	210--395	1--185	2	T18059	hypothetical protein A557L - Chlorella virus PBCV-1	102.45	4.21E-20	52%	66%	40--132	14--108
					3	COG0037	MesJ, Predicted ATPase of the PP-loop superfamily implicated in cell cycle control [Cell division and chromosome partitioning]	104.93	2.32E-23	24%	44%	195--403	6--218	3	YP_039960	hypothetical protein SAR0510	87.04	1.83E-15	23%	46%	211--509	14--320
					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 stronaly conserved motif SGGKD at the N terminus..	65.65	1.28E-11	23%	41%	210--381	1--180	4	AAV42170	putative cell cycle	86.66	2.39E-15	24%	45%	210--472	23--293
					5	pfam06041	DUF924, Bacterial protein of unknown function (DUF924). This family consists of several hypothetical bacterial proteins of unknown function..	46.83	6.42E-06	24%	39%	16--119	4--117	5	YP_415957	hypothetical protein SAB0458	86.27	3.12E-15	23%	46%	211--509	14--320
					6	cd01712	ThiI, ThiI is required for thiazole synthesis in the thiamine biosynthesis pathway. It belongs to the Adenosine Nucleotide Hydrolysis superfamily and predicted to bind to Adenosine nucleotide..	38.65	1.66E-03	23%	41%	215--370	6--153	6	BAB56671	conserved hypothetical protein	85.89	4.08E-15	23%	46%	211--509	14--320
					7	cd01713	PAPS_reductase, This domain is found in phosphoadenosine phosphosulphate (PAPS) reductase enzymes or PAPS sulphotransferase. PAPS reductase is part of the adenine nucleotide alpha hydrolases superfamily also including N type ATP PPases and ATP sulphurylases. A highly modified version of the P loop, the fingerprint peptide of mononucleotide-binding proteins, is present in the active site of the protein, which appears to be a positively charged cleft containing a number of conserved arginine and lysine residues. Although PAPS reductase has no ATPase activity, it shows a striking similarity to the structure of the ATP pyrophosphatase (ATP PPase) domain of GMP synthetase, indicating that both enzyme families have evolved from a common ancestral nucleotide-binding fold. The enzyme uses thioredoxin as an electron donor for the reduction of PAPS to phospho-adenosine-phosphate (PAP) . It is also found in NodP nodulation protein P from Rhizobium meliloti which has ATP sulphurylase activity (sulphate adenylate transferase)	37.34	4.98E-03	22%	39%	210--377	1--163	7	YP_499061	hypothetical protein SAOUHSC_00484	85.89	4.08E-15	23%	46%	211--509	14--320
					8	COG0175	CysH, 3'-phosphoadenosine 5'-phosphosulfate sulfotransferase (PAPS reductase)/FAD synthetase and related enzymes [Amino acid transport and metabolism / Coenzyme metabolism].	35.41	1.85E-02	19%	35%	192--378	23--199	8	EAN10266	PP-loop	84.34	1.19E-14	24%	47%	206--433	28--260
					9	COG0301	ThiI, Thiamine biosynthesis ATP pyrophosphatase [Coenzyme metabolism]	34.87	2.22E-02	24%	45%	210--370	177--329	9	CAG42241	conserved hypothetical protein	83.96	1.55E-14	23%	46%	211--509	14--320
					10	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..	34.82	2.69E-02	27%	38%	211--384	1--154	10	YP_428986	tRNA(Ile)-lysidine synthetase-like	81.26	1.00E-13	27%	51%	210--408	20--217
N410L	164225--163956	90	9,651	10.15		No Hit Found								0	No Hit Found	No Hit Found						
N412R	164354--164968	205	23,554	7.25	1	COG4942	COG4942, Membrane-bound metallopeptidase [Cell division and chromosome partitioning]	29.16	4.07E-01	18%	42%	65--115	46--96	1	AAC96971	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	118.63	1.25E-25	34%	53%	8--201	7--185
														2	NP_064894	Hypothetical protein	42.74	8.71E-03	36%	54%	32--110	114--201
N413R	165118--165882	255	28,764	8.33		No Hit Found								1	AAC96377	A9R	213.00	7.66E-54	57%	78%	88--253	8--173
														2	AAC96643	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	106.69	7.73E-22	28%	46%	3--251	5--248
														3	AAC96972	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	101.68	2.49E-20	28%	46%	3--252	8--253
														4	AAC96818	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	99.75	9.45E-20	27%	46%	3--252	2--246
														5	AAC96545	A177R	99.75	9.45E-20	28%	48%	3--240	5--230
														6	AAC96447	A79R	89.35	1.28E-16	28%	46%	3--231	2--218
														7	AAU06304	hypothetical protein A275R	81.26	3.48E-14	31%	50%	88--251	4--167
														8	AAU06301	hypothetical protein A275R	80.11	7.74E-14	30%	50%	88--251	4--167
														9	AAU06302	hypothetical protein A275R	68.17	3.05E-10	32%	52%	120--252	1--132
														10	AAU06299	hypothetical protein A275R	43.90	6.15E-03	41%	52%	189--251	1--59
N416L	166454--165900	185	20,724	7.72	1	pfam02940	mRNA_triPase, mRNA capping enzyme, beta chain. The beta chain of mRNA capping enzyme has triphosphatase activity. The function of the capping enzyme also depends on the guanylyltransferase activity conferred by the alpha chain (see pfam01331).	42.31	4.00E-05	34%	55%	94--156	191--256	1	AAC96817	PBCV-1 RNA triphosphatase	185.65	6.48E-46	55%	69%	5--183	11--193
														2	XP_636333	hypothetical protein DDB0188303	42.74	6.79E-03	39%	56%	94--156	155--219
N418R	166478--167233	252	28,265	7.95	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	113.63	2.26E-26	29%	44%	60--248	1--178	1	AAC96453	PBCV-1 prolyl 4-hydroxylase	224.94	1.90E-57	48%	63%	19--250	3--241

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
							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														
					2	pfam03171	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	41.64	1.00E-04	27%	39%	145--249	2--96	2	AAZ62310	Procollagen-proline,2-oxoglutarate-4-dioxygenase	93.59	6.61E-18	33%	50%	60--248	90--274
														3	CAD16521	putative prolyl 4-hydroxylase alpha subunit homologue oxidoreductase protein	90.89	4.28E-17	32%	47%	60--248	97--281
														4	EAP73718	Prolyl 4-hydroxylase alpha subunit	90.12	7.31E-17	32%	47%	60--248	97--281
														5	YP_585239	2OG-Fe(II) oxygenase	87.81	3.63E-16	32%	46%	58--248	99--285
														6	YP_550510	Procollagen-proline,2-oxoglutarate-4-dioxygenase	85.89	1.38E-15	31%	47%	60--248	90--272
														7	ZP_00984285	hypothetical protein BdolA_01003928	84.73	3.07E-15	30%	46%	60--248	73--257
														8	ZP_01271629	2OG-Fe(II) oxygenase	84.34	4.01E-15	31%	52%	60--248	79--263
														9	YP_142947	prolyl 4-hydroxylase	81.65	2.60E-14	28%	44%	56--248	56--237
														10	AAF08583	unknown protein	78.57	2.20E-13	32%	43%	60--247	45--223
							Thioredoxin, Thioredoxin. Thioredoxins are small enzymes that participate in redox reactions, via the reversible oxidation of an active														
N420R	167270--167593	108	11,872	10.26	1	pfam00085	centre disulfide bond. Some members with only the active site are not separated from the noise..	61.78	1.78E-11	32%	49%	29--105	33--108	1	AAC96816	contains cytochrome C family heme-binding site signature; similar to maize protein disulphide isomerase, correspond to Swiss-Prot Accession Number P52588	127.87	1.02E-28	55%	73%	4--103	6--105
					2	COG3118	COG3118, Thioredoxin domain-containing protein [Posttranslational modification, protein turnover, chaperones].	39.89	6.93E-05	20%	39%	13--103	39--129	2	AAQ23042	transglutaminase	67.01	2.14E-10	35%	58%	10--104	6--105
					3	COG0526	TrxA, Thiol-disulfide isomerase and thioredoxins [Posttranslational modification, protein turnover, chaperones / Energy production and conversion].	33.34	6.49E-03	14%	33%	27--106	42--126	3	CAH65062	hypothetical protein	67.01	2.14E-10	46%	62%	29--102	60--134
					4	pfam01216	Calsequestrin, Calsequestrin..	28.43	2.02E-01	30%	59%	54--99	95--139	4	XP_419952	PREDICTED: similar to Protein disulfide isomerase A6 precursor (Protein disulfide isomerase P5) (Thioredoxin domain containing protein 7)	45.44	6.68E-04	32%	48%	29--100	196--271
														5	AAF78087	protein disulfide isomerase ER-60	66.24	3.66E-10	39%	59%	23--101	43--119
														6	CAG03659	unnamed protein product	50.45	2.08E-05	35%	49%	23--97	185--261
														7	ABC75564	protein disulfide isomerase-related protein P5 precursor	48.52	7.89E-05	31%	48%	23--100	167--246
														8	AAH67155	Hypothetical protein zgc:77086	63.93	1.82E-09	36%	56%	23--104	45--124
														9	AAC24752	transglutaminase precursor	63.54	2.37E-09	35%	56%	10--104	35--134
														10	AAH44507	Protein disulfide isomerase-related protein (provisional)	50.06	2.71E-05	35%	47%	23--100	186--265
N421R	167673--169067	465	52,190	5.95	1	COG0661	AarF, Predicted unusual protein kinase [General function prediction only]	176.29	5.77E-45	27%	42%	24--385	37--423	1	AAC96813	similar to Clostridium pasteurianum ORF, corresponds to GenBank Accession Number 728353	615.92	9.91E-175	64%	84%	13--463	11--462
							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														
					2	pfam03109	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..	113.77	4.19E-26	41%	57%	96--209	7--119	2	YP_477188	ABC1 domain protein	139.81	2.08E-31	26%	48%	49--422	68--475
					3	COG4670	COG4670, Acyl CoA:acetate/3-ketoacid CoA transferase [Lipid metabolism].	30.19	5.76E-01	31%	47%	250--366	368--476	3	YP_473700	ABC1 domain protein	134.81	6.69E-30	26%	48%	49--392	79--453
														4	YP_399885	hypothetical protein Synpcc7942_0868	134.04	1.14E-29	26%	47%	49--392	83--452
														5	BAB66733	488aa long conserved hypothetical protein	132.88	2.54E-29	24%	46%	12--457	9--459
														6	ZP_00673531	ABC-1	132.11	4.34E-29	26%	47%	49--384	70--431
														7	YP_256267	universally conserved protein	131.34	7.40E-29	24%	43%	12--460	9--467
														8	ABA22988	ABC-1	130.57	1.26E-28	25%	46%	37--392	71--452
														9	BAB76214	alr4515	129.80	2.15E-28	25%	46%	37--392	82--463
														10	BAC92119	glr4178	129.41	2.81E-28	25%	46%	54--394	102--470
							V-ATPase_G, Vacuolar (H+)-ATPase G subunit. This family represents the eukaryotic vacuolar (H+)-ATPase (V-ATPase) G subunit. V-ATPases generate an acidic environment in several intracellular compartments. Correspondingly, they are found as membrane-attached proteins in														
N426R	169131--169475	115	13,217	5.50	1	pfam03179	several organelles. They are also found in the plasma membranes of some specialised cells. V-ATPases consist of peripheral (V1) and membrane integral (V0) heteromultimeric complexes. The G subunit is part of the V1 subunit, but is also thought to be strongly attached to the V0 complex. It may be involved in the coupling of ATP degradation to H+ translocation.	28.73	1.72E-01	21%	49%	27--98	26--97	1	AAC96812	A444L	71.25	1.11E-11	38%	55%	4--112	2--99
					2	COG1322	COG1322, Uncharacterized protein conserved in bacteria [Function unknown].	28.42	1.89E-01	14%	46%	22--100	107--185	2	AAQ96243	LRRGT00030	43.90	1.89E-03	30%	53%	13--107	144--234
					3	COG0484	DnaJ, DnaJ-class molecular chaperone with C-terminal Zn finger domain [Posttranslational modification, protein turnover, chaperones].	28.27	2.43E-01	38%	65%	29--63	17--51	3	XP_215260	PREDICTED: similar to LRRGT00030	43.90	1.89E-03	30%	53%	13--107	149--239
N428R	169736--171154	473	51,078	10.86		No Hit Found								0	No Hit Found	No Hit Found						
N430L	171587--171159	143	15,847	4.41		No Hit Found								1	AAC96809	A441L	140.97	1.16E-32	51%	67%	6--141	2--137
N433R	171678--172577	300	33,203	5.57		No Hit Found								1	AAC96811	A443R	182.57	1.47E-44	34%	55%	6--297	8--306
N434L	172956--172621	112	12,602	6.50		No Hit Found								1	AAC96807	A439R	121.71	7.23E-27	58%	81%	19--103	21--105
N436R	172977--173240	88	10,083	11.02	1	COG0695	GrxC, Glutaredoxin and related proteins [Posttranslational modification, protein turnover, chaperones].	52.66	1.13E-08	34%	55%	3--72	5--72	1	AAC96806	similar to E. coli glutaredoxin, corresponds to Swiss-Prot Accession Number P37687	95.13	7.35E-19	58%	84%	3--72	5--74
					2	pfam00462	Glutaredoxin, Glutaredoxin..	39.85	7.56E-05	25%	49%	1--72	2--73	2	ABE05174	Glutaredoxin GrxC	54.30	1.44E-06	31%	57%	1--70	9--78
					3	COG0278	COG0278, Glutaredoxin-related protein [Posttranslational modification, protein turnover, chaperones]	27.10	4.98E-01	27%	48%	10--74	28--92	3	NP_901706	probable peroxiredoxin/glutaredoxin family protein	50.83	1.59E-05	38%	56%	1--74	173--241
														4	AAL51366	GLUTAREDOXIN	50.83	1.59E-05	33%	55%	1--78	8--78
														5	AAN30771	glutaredoxin 3	50.83	1.59E-05	33%	55%	1--78	4--74
														6	YP_465226	Glutaredoxin, GrxC	49.29	4.63E-05	31%	59%	3--71	8--72
														7	YP_154197	glutaredoxin 3	45.44	6.68E-04	33%	51%	1--74	4--73
														8	ZP_00374087	glutaredoxin family protein	45.44	6.68E-04	33%	50%	1--78	4--80
														9	AAZ41226	glutaredoxin 3	44.67	1.14E-03	32%	55%	3--70	6--73
														10	ZP_00971854	COG0695: Glutaredoxin and related proteins	44.67	1.14E-03	28%	54%	1--70	4--69
							dNK, Deoxyribonucleoside kinase (dNK) catalyzes the phosphorylation of deoxyribonucleosides to yield corresponding monophosphates (dNMPs). This family consists of various deoxynucleoside kinases														
N438R	173248--173814	189	21,593	5.53	1	cd01673	including deoxyribo- cytidine (EC 2.7.1.74), guanosine (EC 2.7.1.113), adenosine (EC 2.7.1.76), and thymidine (EC 2.7.1.21) kinases. They are key enzymes in the salvage of deoxyribonucleosides originating from extra- or intracellular breakdown of DNA	102.24	3.21E-23	31%	52%	3--165	1--178	1	AAC96784	contains ATP/GTP-binding site motif A; similar to Bacillus subtilus 24.1 kDa protein, corresponds to Swiss-Prot Accession Number P37530	191.05	1.64E-47	50%	68%	3--181	2--187
					2	COG1428	COG1428, Deoxynucleoside kinases [Nucleotide transport and metabolism].	91.47	6.18E-20	34%	54%	1--160	4--178	2	ABF82059	hypothetical protein MIV029R	91.28	1.77E-17	31%	56%	3--183	4--189
							dNK, Deoxynucleoside kinase. This family consists of various deoxynucleoside kinases cytidine EC:2.7.1.74, guanosine EC:2.7.1.113,														
					3	pfam01712	adenosine EC:2.7.1.76 and thymidine kinase EC:2.7.1.21 (which also phosphorylates deoxyuridine and deoxycytosine.) These enzymes catalyse the production of deoxynucleotide 5'-monophosphate from a deoxynucleoside. Using ATP and yielding ADP in the process..	80.31	1.44E-16	31%	49%	53--183	1--141	3	NP_149606	143R	88.20	1.50E-16	31%	52%	1--152	1--157
					4	COG0125	Tmk, Thymidylate kinase [Nucleotide transport and metabolism].	53.35	1.90E-08	23%	45%	4--184	6--204	4	AAU10875	deoxynucleoside kinases	76.64	4.52E-13	31%	55%	3--152	4--161
							TMPK, Thymidine monophosphate kinase (TMPK), also known as thymidylate kinase, catalyzes the phosphorylation of thymidine														
					5	cd01672	monophosphate (TMP) to thymidine diphosphate (TDP) utilizing ATP as its preferred phophoryl donor. TMPK represents the rate-limiting step in either de novo or salvage biosynthesis of thymidine triphosphate (TTP)..	51.43	6.22E-08	19%	42%	3--181	2--198	5	CAC84481	thymidine kinase	75.49	1.01E-12	26%	46%	4--182	13--208

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	cd02030	NDUO42, NADH:Ubiquinone oxioreductase, 42 kDa (NDUO42) is a family of proteins that are highly similar to deoxyribonucleoside kinases (dNK). Members of this family have been identified as one of the subunits of NADH:Ubiquinone oxioreductase (complex I), a multi-protein complex located in the inner mitochondrial membrane. The main function of the complex is to transport electrons from NADH to ubiquinone, which is accompanied by the translocation of protons from the mitochondrial matrix to the inter membrane space..	43.31	2.00E-05	27%	45%	3--153	1--184	6	YP_040010	putative deoxyadenosine kinase protein	72.79	6.53E-12	31%	47%	3--181	11--211
					7	cd02019	NK, Nucleoside/nucleotide kinase (NK) is a protein superfamily consisting of multiple families of enzymes that share structural similarity and are functionally related to the catalysis of the reversible phosphate group transfer from nucleoside triphosphates to nucleosides/nucleotides, nucleoside monophosphates, or sugars. Members of this family play a wide variety of essential roles in nucleotide metabolism, the biosynthesis of coenzymes and aromatic compounds, as well as the metabolism of sugar and sulfate.	42.11	3.79E-05	21%	38%	3--159	1--165	7	YP_416005	deoxypurine kinase subunit	72.79	6.53E-12	31%	48%	3--181	11--211
					8	pfam02223	Thymidylate kin, Thymidylate kinase..	40.65	1.12E-04	19%	37%	6--177	1--186	8	YP_499111	hypothetical protein SAOUHSC_00539	71.25	1.90E-11	34%	51%	3--153	11--172
					9	COG0572	Udk, Uridine kinase [Nucleotide transport and metabolism].	35.22	5.22E-03	22%	44%	3--159	10--179	9	CAC84464	thymidine kinase	70.48	3.24E-11	26%	51%	4--174	9--196
					10	cd02022	DPCK, Dephospho-coenzyme A kinase (DPCK, EC 2.7.1.24) catalyzes the phosphorylation of dephosphocoenzyme A (dCoA) to yield CoA, which is the final step in CoA biosynthesis..	35.14	6.14E-03	25%	40%	3--134	1--142	10	AAW53624	deoxynucleoside kinase family protein	70.48	3.24E-11	30%	48%	3--181	11--211
N441L	174871--173786	362	40,197	8.53	1	COG1910	COG1910, Periplasmic molybdate-binding protein/domain [Inorganic ion transport and metabolism].	35.64	9.41E-03	42%	58%	50--83	97--130	1	AAC96785	A417L	211.46	3.97E-53	36%	56%	1--348	1--372
					2	pfam03215	Rad17, Rad17 cell cycle checkpoint protein..	34.61	1.78E-02	33%	53%	25--77	15--72	2	YP_142765	unknown	47.75	7.60E-04	26%	42%	115--346	210--468
N445L	175247--174900	116	13,136	10.69	1	pfam02404	SCF, Stem cell factor. Stem cell factor (SCF) is a homodimer involved in hematopoiesis. SCF binds to and activates the SCF receptor (SCFR), a receptor tyrosine kinase. The crystal structure of human SCF has been resolved and a potential receptor-binding site identified..	26.62	8.64E-01	22%	49%	5--68	66--129	1	AAC96788	A420L	69.71	3.32E-11	48%	65%	45--114	1--70
N446R	175027--175434	136	15,397	10.41	1	COG1177	PotC, ABC-type spermidine/putrescine transport system, permease component II [Amino acid transport and metabolism].	27.44	5.85E-01	31%	53%	3--68	195--265	1	AAC96789	A421R	96.67	2.51E-19	52%	71%	47--128	8--91
N447L	175712--175458	85	9,585	7.92		No Hit Found								0	No Hit Found	No Hit Found						
N449R	175739--176215	159	17,816	4.79	1	COG1048	AcnA, Aconitase A [Energy production and conversion].	28.74	3.61E-01	33%	54%	99--147	519--567	1	AAC96791	A423R	76.64	2.68E-13	34%	62%	39--154	36--155
N451R	176270--176593	108	12,209	11.33		No Hit Found								0	No Hit Found	No Hit Found						
N452R	176630--176983	118	13,306	4.82		No Hit Found								1	AAC96794	A426R	97.06	1.93E-19	43%	66%	6--116	4--114
N453L	177331--176978	118	13,316	5.26	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..	43.67	5.19E-06	24%	48%	25--105	25--107	1	AAC96795	contains thioredoxin active site-like sequence; similar to Synechocystis thioredoxin-like protein, corresponds to Swiss-Prot Accession Number P52232	103.61	2.06E-21	43%	66%	2--115	7--119
					2	COG3118	COG3118, Thioredoxin domain-containing protein [Posttranslational modification, protein turnover, chaperones].	32.19	1.52E-02	26%	46%	33--101	55--129	2	NP_072786	thioredoxin (trxA)	56.23	3.77E-07	30%	53%	4--103	2--102
					3	COG0526	TrxA, Thiol-disulfide isomerase and thioredoxins [Posttranslational modification, protein turnover, chaperones / Energy production and conversion].	31.79	2.28E-02	26%	49%	25--105	36--124	3	NP_857580	THIOREDOXIN TRXC (TRX) (MPT46)	53.53	2.45E-06	31%	54%	17--105	21--112
					4	COG0259	PdxH, Pyridoxamine-phosphate oxidase [Coenzyme metabolism].	28.30	2.43E-01	56%	75%	80--96	56--72	4	ZP_01043911	Thioredoxin domain-containing protein	51.99	7.12E-06	31%	50%	1--101	6--107
					5	cd02661	Peptidase_C19E, A subfamily of Peptidase C19. Peptidase C19 contains ubiquitinyl hydrolases. They are intracellular peptidases that remove ubiquitin molecules from polyubiquitinated 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..	28.30	2.57E-01	28%	49%	29--73	134--177	5	YP_521223	hypothetical protein DSY4990	51.60	9.29E-06	28%	52%	8--107	9--108
														6	ZP_01192583	Thioredoxin	51.60	9.29E-06	32%	53%	17--105	18--109
														7	AAT27628	thioredoxin	51.22	1.21E-05	23%	56%	25--104	21--100
														8	ZP_01204863	Thioredoxin	51.22	1.21E-05	31%	54%	17--105	19--110
														9	AAG21048	thioredoxin; TrxA1 2	50.83	1.59E-05	30%	51%	11--101	23--114
														10	BAE56042	unnamed protein product	50.83	1.59E-05	29%	51%	2--90	1--92
N457L	177735--177394	114	12,710	5.76	1	COG3118	COG3118, Thioredoxin domain-containing protein [Posttranslational modification, protein turnover, chaperones].	40.66	4.01E-05	38%	56%	55--89	81--115	1	AAC96795	contains thioredoxin active site-like sequence; similar to Synechocystis thioredoxin-like protein, corresponds to Swiss-Prot Accession Number P52232	52.76	4.09E-06	28%	52%	1--105	7--110
					2	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..	39.05	1.33E-04	31%	53%	40--100	44--103	2	ZP_00673542	Thioredoxin	42.74	4.24E-03	26%	61%	40--106	41--107
					3	COG3640	CooC, CO dehydrogenase maturation factor [Cell division and chromosome partitioning]	28.64	1.56E-01	23%	47%	30--76	13--56	3	ABE04083	Thioredoxin	41.59	9.43E-03	42%	64%	40--81	55--96
														4	ZP_01228494	thioredoxin	41.59	9.43E-03	34%	61%	49--100	50--101
														5	ZP_00857352	Thioredoxin	41.59	9.43E-03	39%	60%	47--97	87--137
N458L	178083--177766	106	11,770	7.34		No Hit Found								1	AAC96795	contains thioredoxin active site-like sequence; similar to Synechocystis thioredoxin-like protein, corresponds to Swiss-Prot Accession Number P52232	53.14	3.12E-06	32%	53%	2--95	9--106
N460L	179433--178102	444	51,180	5.46	1	cd00204	ANK, ankyrin repeats; ankyrin repeats mediate protein-protein interactions in very diverse families of proteins. The number of ANK repeats in a protein can range from 2 to over 20 (ankyrins, for example). ANK repeats may occur in combinations with other types of domains. The structural repeat unit contains two antiparallel helices and a beta-hairpin, repeats are stacked in a superhelical arrangement; this alignment contains 4 consecutive repeats.	43.53	5.90E-05	27%	55%	26--121	3--95	1	AAC96797	A429L	210.69	9.06E-53	33%	52%	7--437	34--466
														2	NP_200815	protein binding	54.68	8.31E-06	25%	43%	51--237	61--250
														3	ABF95306	ankyrin repeat family protein, putative, expressed	51.22	9.19E-05	23%	43%	49--237	205--396
														4	XP_780211	PREDICTED: similar to Ankyrin-2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid)	51.22	9.19E-05	23%	41%	11--271	337--595
														5	BAB03143	ankyrin-like protein	50.83	1.20E-04	22%	43%	49--253	599--805
														6	NP_187842	protein binding	49.68	2.67E-04	22%	43%	49--237	114--305
														7	ABE93792	Ankyrin	48.91	4.56E-04	24%	43%	30--237	27--242
														8	AAM62711	ankyrin-like protein	47.37	1.33E-03	21%	43%	49--237	58--249
														9	AAY61232	Guanosine polyphosphate pyrophosphohydrolases/synthetases homolog	44.67	8.60E-03	23%	43%	26--217	730--914
														10	NP_192256	protein binding	46.98	1.73E-03	22%	41%	27--281	198--454
N462R	179472--179855	128	14,974	9.10		No Hit Found								0	No Hit Found	No Hit Found						
N464R	179846--180130	95	10,649	10.71		No Hit Found								1	AAC96938	A601R	93.97	1.60E-18	53%	64%	11--91	12--101
														2	AAC96975	A656L	46.98	2.24E-04	40%	64%	45--89	51--100
N465R	180130--181275	382	42,891	10.09		No Hit Found								1	AAC96599	contains ATP/GTP-binding motif A	489.57	8.15E-137	61%	76%	6--380	3--383
														2	AAC96957	similar to Chlorella virus PBCV-1 ORF A231L, corresponds to GenBank Accession Number U42580	305.45	2.18E-81	44%	62%	14--340	104--428
N469R	181355--181765	137	15,763	6.80		No Hit Found								0	No Hit Found	No Hit Found						
N470L	183182--181881	434	47,713	6.94	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 PBCV-1 the major capsid protein is a glycoprotein	428.58	5.73E-121	53%	64%	1--428	1--443	1	AAC27492	major capsid protein Vp49	738.03	0.00E+00	86%	87%	1--432	1--432

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	AAC96798	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	618.23	1.82E-175	72%	78%	1--432	1--437
														3	BAA76601	major capsid protein MCP1	613.22	5.86E-174	72%	78%	1--432	1--437
														4	BAA22198	major capsid protein Vp54	611.30	2.23E-173	72%	77%	1--432	1--437
														5	BAA76600	major capsid protein	606.29	7.17E-172	71%	77%	1--432	1--436
														6	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	571.24	2.56E-161	71%	77%	25--432	1--413
														7	1M3Y_D	Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid Containing, Dna Virus	571.24	2.56E-161	71%	77%	25--432	1--413
														8	BAE06835	hypothetical major capsid protein	305.83	2.00E-81	40%	53%	1--432	1--440
														9	AAC96917	similar to Chlorella virus PBCV-1 ORF A11L, corresponds to GenBank Accession Number U42580	236.11	1.95E-60	35%	50%	3--432	2--400
														10	AAC96379	contains aminoacyl-tRNA synthetase class-II signature	233.80	9.67E-60	36%	50%	1--432	1--403
N471L	184168--183284	295	35,080	8.32	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. .	35.14	1.03E-02	20%	36%	28--112	2--88	1	AAC96862	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	72.02	2.72E-11	30%	49%	30--185	4--166
					2	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	33.89	2.28E-02	34%	54%	30--116	4--83	2	AAA88832	unknown	64.31	5.67E-09	29%	45%	30--212	4--180
					3	cd03274	ABC_SMC4_euk, Eukaryotic SMC4 proteins; SMC proteins are large (approximately 110 to 170 kDa), and each is arranged into five recognizable domains. Amino-acid sequence homology of SMC proteins between species is largely confined to the amino- and carboxy-terminal globular domains. The amino-terminal domain contains a 'Walker A' nucleotide-binding domain (GxxGxGKS/T, in the single-letter amino-acid code), which by mutational studies has been shown to be essential in several proteins. The carboxy-terminal domain contains a sequence (the DA-box) that resembles a 'Walker B' motif, and a motif with homology to the signature sequence of the ATP-binding cassette (ABC) family of ATPases. The sequence homology within the carboxy-terminal domain is relatively high within the SMC1-SMC4 group, whereas SMC5 and SMC6 show some divergence in both of these sequences. In eukaryotic cells, the proteins are found as heterodimers of SMC1 paired with SMC3, SMC2 with SMC4, and SMC5 with SMC6 (formerly known as Rad18)..	32.27	8.15E-02	18%	38%	123--293	885--1060	3	AAC96973	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	57.38	6.93E-07	39%	51%	30--132	4--105
					4	pfam04189	eIF3_gamma, Eukaryotic initiation factor 3, gamma subunit. eIF-3 is a multi-subunit complex that stimulates translation initiation in vitro at several different steps. This family corresponds to the gamma subunit if eIF3..	31.52	1.16E-01	23%	47%	150--242	75--165	4	AAC96655	PBCV-1 33kd peptide	56.61	1.18E-06	28%	45%	40--199	19--192
					5	cd03277	ABC_SMC5_euk, Eukaryotic SMC5 proteins; SMC proteins are large (approximately 110 to 170 kDa), and each is arranged into five recognizable domains. Amino-acid sequence homology of SMC proteins between species is largely confined to the amino- and carboxy-terminal globular domains. The amino-terminal domain contains a 'Walker A' nucleotide-binding domain (GxxGxGKS/T, in the single-letter amino-acid code), which by mutational studies has been shown to be essential in several proteins. The carboxy-terminal domain contains a sequence (the DA-box) that resembles a 'Walker B' motif, and a motif with homology to the signature sequence of the ATP-binding cassette (ABC) family of ATPases. The sequence homology within the carboxy-terminal domain is relatively high within the SMC1-SMC4 group, whereas SMC5 and SMC6 show some divergence in both of these sequences. In eukaryotic cells, the proteins are found as heterodimers of SMC1 paired with SMC3, SMC2 with SMC4, and SMC5 with SMC6 (formerly known as Rad18)..	31.17	1.70E-01	22%	48%	130--240	194--308	5	YP_293795	putative endonuclease	56.61	1.18E-06	39%	55%	38--118	14--95
														6	XP_654864	hypothetical protein 42.t00003	43.90	7.93E-03	24%	51%	121--245	1073--1200
														7	EAR95972	Protein kinase domain containing protein	44.28	6.07E-03	25%	45%	123--292	972--1136
														8	CAG08671	unnamed protein product	44.67	4.65E-03	33%	52%	123--238	338--441
														9	AAC26930	Hypothetical protein F36H12.3	51.22	4.97E-05	29%	51%	123--248	95--217
														10	XP_654034	hypothetical protein 64.t00001	50.06	1.11E-04	25%	54%	123--233	151--261
N472R	184277--186826	850	95,405	6.12	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	49.83	1.32E-06	30%	43%	256--383	2--124	1	AAC96482	A114R	639.80	0.00E+00	64%	77%	373--847	5--477
					2	COG0463	WcaA, Glycosyltransferases involved in cell wall biogenesis [Cell envelope biogenesis, outer membrane]	47.44	7.61E-06	24%	45%	251--380	2--121	2	AAC96479	A111R	489.96	1.72E-136	58%	75%	1--372	1--379
					3	COG1215	COG1215, Glycosyltransferases, probably involved in cell wall biogenesis [Cell envelope biogenesis, outer membrane]	44.15	7.72E-05	24%	38%	251--367	53--167	3	NP_872956	possible glycosyltransferase	90.51	3.05E-16	32%	46%	254--426	6--189
					4	COG1216	COG1216, Predicted glycosyltransferases [General function prediction only]	38.19	4.16E-03	25%	46%	252--369	3--116	4	CAG34747	hypothetical protein	87.81	1.97E-15	29%	45%	1--225	1--225
					5	pfam03016	Exostosin, Exostosin family. The EXT family is a family of tumour suppressor genes. Mutations of EXT1 on 8q24.1, EXT2 on 11p11-13, and EXT3 on 19p have been associated with the autosomal dominant disorder known as hereditary multiple exostoses (HME). This is the most common known skeletal dysplasia. The chromosomal locations of other EXT genes suggest association with other forms of neoplasia. EXT1 and EXT2 have both been shown to encode a heparan sulphate polymerase with both D-glucuronyl (GlcA) and N-acetyl-D-glucosaminoglycan (GlcNAC) transferase activities. The nature of the defect in heparan sulphate biosynthesis in HME is unclear	35.03	3.45E-02	23%	43%	657--818	177--325	5	ZP_00202013	COG0463: Glycosyltransferases involved in cell wall biogenesis	86.27	5.74E-15	32%	49%	254--426	48--231
														6	AAU37201	unknown	81.26	1.85E-13	26%	40%	254--495	6--259
														7	ZP_00154867	COG1216: Predicted glycosyltransferases	78.95	9.17E-13	29%	44%	254--427	29--215
														8	AAU03775	conserved hypothetical protein	59.31	7.52E-07	25%	50%	58--206	103--250
														9	ZP_00340129	hypothetical protein RakaH01000503	58.54	1.28E-06	28%	51%	58--206	107--254
														10	EAN09672	hypothetical protein EfaeDRAFT_1144	57.77	2.19E-06	23%	43%	3--201	2--204
N477R	186874--187335	154	16,607	8.18		No Hit Found								1	AAC96800	Lys-rich	114.01	1.50E-24	46%	67%	1--109	1--109
N480R	187361--187714	118	12,745	11.93		No Hit Found								1	AAC96925	A571R	156.76	2.05E-37	79%	89%	1--101	1--101
														2	AAC96803	Arg-rich	56.23	3.77E-07	61%	70%	1--44	4--47
														3	AAC96573	A205R	45.05	8.70E-04	57%	75%	7--39	174--206
N482R	187960--188919	320	36,440	8.67	1	smart00507	HNHc, HNH nucleases; .	29.67	5.70E-01	36%	45%	127--168	10--52	1	AAC96845	Lys-, Arg-rich; contains eukaryotic putative RNA-binding region RNP-1 signature; similar to PBCV-1 ORF A267L, corresponds to GenBank Accession Number U42580	188.73	2.27E-46	35%	47%	9--312	3--305
					2	cd00085	HNHc, HNH nucleases; HNH endonuclease signature which is found in viral, prokaryotic, and eukaryotic proteins. The alignment includes members of the large group of homing endonucleases, yeast intron 1 protein, MutS, as well as bacterial colicins, pyocins, and anaredoxins..	28.89	8.81E-01	30%	43%	127--172	11--57	2	AAC96857	Lys-, Glu-rich	177.56	5.24E-43	34%	47%	9--312	3--305
					3	COG1110	COG1110, Reverse gyrase [DNA replication, recombination, and repair].	28.74	9.54E-01	30%	52%	4--27	692--715	3	YP_142777	unknown	73.17	1.39E-11	38%	56%	194--295	251--349
														4	AAC96635	A267L	85.11	3.54E-15	34%	47%	114--293	73--238
N484L	189482--189174	103	10,874	10.74	1	pfam05854	MC1, Non-histone chromosomal protein MC1. This family consists of archaeal chromosomal protein MC1 sequences which protect DNA against thermal denaturation..	42.35	1.23E-05	37%	53%	5--68	2--70	1	AAC96805	similar to Methanothrix chromosomal protein MC1A, corresponds to Swiss-Prot Accession Number P15251	134.81	8.23E-31	69%	81%	5--99	6--100
														2	P15249	Chromosomal protein MC1a	54.30	1.41E-06	42%	50%	7--90	4--84
														3	P15250	Chromosomal protein MC1b	50.83	1.56E-05	45%	60%	7--66	4--62
														4	ZP_01153404	Non-histone chromosomal MC1	49.68	3.48E-05	42%	57%	4--66	6--67
														5	P15251	Chromosomal protein MC1c	48.14	1.01E-04	45%	58%	6--67	4--62
														6	P12770	Chromosomal protein MC1	45.82	5.03E-04	35%	50%	4--93	1--90
														7	A25343	nonhistone chromosomal protein HMb - Methanosarcina barkeri	43.51	2.50E-03	40%	54%	4--67	1--69
														8	AAZ70150	chromosomal protein	43.13	3.26E-03	40%	54%	4--67	2--70
N485R	189557--189775	73	7,645	10.67		No Hit Found								0	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
N487L	190133--189810	108	12,183	4.88	1	COG3164	COG3164, Predicted membrane protein [Function unknown].	27.59	3.50E-01	33%	54%	3--27	1135--1159	1	AAC96778	A410L	113.62	2.00E-24	48%	74%	4--105	8--110
														2	AAC96776	A408L	78.18	9.30E-14	33%	61%	3--102	40--139
														3	NP_077527	EsV-1-42	50.83	1.59E-05	30%	50%	3--101	6--110
														4	AAR26867	FirrV-1-A43	49.68	3.54E-05	33%	51%	3--103	5--110
N488R	190231--190728	166	18,734	5.30		No Hit Found								1	AAC96779	A411R	79.34	4.76E-14	37%	56%	1--131	1--133
N491R	190757--191299	181	20,356	6.54	1	COG5221	DOP1, Dopey and related predicted leucine zipper transcription factors [Transcription].	28.45	5.72E-01	22%	39%	65--174	575--682	1	AAC96780	A412R	187.19	2.08E-46	51%	67%	1--178	1--179
N493L	192037--191294	248	26,968	8.99		No Hit Found								1	AAC96781	Gln-rich, QQQQM(4x); similar to human transcription factor TFIID, corresponds to Swiss-Prot Accession Number P20226	146.75	6.36E-34	37%	50%	2--245	3--244
N494R	192130--192828	233	25,752	5.00	1	pfam03254	XG_FTase, Xyloglucan fucosyltransferase. Plant cell walls are crucial for development, signal transduction, and disease resistance in plants. Cell walls are made of cellulose, hemicelluloses, and pectins. Xyloglucan (XG), the principal load-bearing hemicellulose of dicotyledonous plants, has a terminal fucosyl residue. This fucosyltransferase adds this residue..	28.81	6.83E-01	21%	44%	12--144	250--393	1	AAC96776	A408L	209.53	7.09E-53	51%	75%	1--187	37--233
														2	NP_077527	EsV-1-42	62.39	1.40E-08	26%	49%	5--127	7--142
														3	AAC96778	A410L	58.54	2.02E-07	29%	53%	4--97	7--108
														4	AAR26867	FirrV-1-A43	56.23	1.00E-06	24%	51%	4--130	5--143
N496L	193660--192842	273	28,323	12.94	1	pfam04094	DUF390, Protein of unknown function (DUF390). This family of long proteins are currently only found in the rice genome. They have no known function. However they may be some kind of transposable element..	30.18	2.58E-01	35%	44%	149--237	72--157	1	YP_559513	Hypothetical proline rich protein	46.21	1.40E-03	36%	48%	121--194	246--311
					2	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..	29.88	3.22E-01	33%	42%	172--239	503--570	2	ZP_01147863	Initiation factor 2:Small GTP-binding protein domain	80.88	5.13E-14	37%	44%	123--252	198--336
														3	ABB45731	collagen type III alpha 1	44.28	5.32E-03	38%	39%	123--185	110--172
														4	AAB96638	precollagen D	55.45	2.31E-06	36%	42%	129--258	104--228
														5	P04258	Collagen alpha-1(III) chain	58.15	3.56E-07	34%	38%	153--257	7--102
														6	XP_588040	PREDICTED: similar to Collagen alpha 1(III) chain precursor isoform 1	48.91	2.16E-04	32%	42%	124--213	1115--1200
														7	XP_883522	PREDICTED: similar to Collagen alpha 1(III) chain precursor isoform 8	56.23	1.35E-06	31%	36%	131--257	94--231
														8	XP_883483	PREDICTED: similar to Collagen alpha 1(III) chain precursor isoform 7	48.91	2.16E-04	32%	42%	124--213	1133--1218
														9	XP_883439	PREDICTED: similar to Collagen alpha 1(III) chain precursor isoform 6	47.37	6.28E-04	34%	38%	153--257	1--111
														10	XP_871448	PREDICTED: similar to Collagen alpha 1(III) chain precursor isoform 2	48.91	2.16E-04	32%	42%	124--213	1133--1218
N500R	193784--194935	384	43,924	8.00	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..	32.12	1.25E-01	27%	39%	116--253	51--167	0	No Hit Found	No Hit Found						
N504R	194966--195598	211	23,349	8.61	1	COG1290	QcrB, Cytochrome b subunit of the bc complex [Energy production and conversion].	28.34	7.78E-01	17%	35%	10--85	299--377	1	AAC96775	A407L	309.30	5.37E-83	70%	82%	1--209	1--209
														2	YP_142822	unknown	47.37	3.80E-04	23%	42%	12--208	41--225
N506L	196264--195644	207	23,029	7.81	1	smart00731	SprT, SprT homologues. Predicted to have roles in transcription elongation. Contains a conserved HExxH motif, indicating a metalloprotease function	35.66	4.74E-03	28%	40%	122--201	56--146	1	AAC96888	contains Gln-rich, neutral zinc metallopeptidase, zinc binding region signature	216.85	3.46E-55	62%	79%	1--167	1--165
					2	COG0501	HtpX, Zn-dependent protease with chaperone function [Posttranslational modification, protein turnover, chaperones].	33.36	2.26E-02	38%	53%	100--138	130--170	2	YP_142679	metal-dependent hydrolase	56.23	7.81E-07	27%	46%	4--181	3--191
					3	COG0503	Apt, Adenine/guanine phosphoribosyltransferases and related PRPP-binding proteins [Nucleotide transport and metabolism].	31.42	8.08E-02	28%	43%	52--108	54--108	3	NP_149599	136R	48.52	1.63E-04	34%	59%	100--162	98--163
					4	pfam01435	Peptidase_M48, Peptidase family M48..	30.83	1.39E-01	40%	53%	122--166	51--91	4	ABF82072	hypothetical protein MIV042R	45.44	1.38E-03	33%	60%	100--163	81--145
N507R	196327--196899	191	20,846	9.33		No Hit Found								1	AAC96890	A523R	230.34	2.47E-59	70%	80%	30--186	7--168
N512R	196920--197366	149	16,359	8.39		No Hit Found								1	AAC96893	A526R	129.03	4.60E-29	50%	62%	18--147	1--129
N513R	197365--197649	95	10,969	10.78		No Hit Found								1	AAC96894	A527R	76.64	2.64E-13	60%	74%	6--69	5--69
N514L	197904--197689	72	8,035	10.35		No Hit Found								1	AAC96898	A531L	48.91	5.92E-05	54%	80%	22--62	24--65
N515L	198165--197926	80	8,518	9.37		No Hit Found								1	AAC96899	A532L	96.67	2.50E-19	60%	77%	1--72	1--74
N516R	198415--200001	529	57,224	4.98	1	pfam06013	DUF909, Bacterial protein of unknown function (DUF909). This family consists of several short bacterial proteins of unknown function	34.47	3.38E-02	30%	50%	94--124	56--86	1	AAC96900	A533R	436.42	1.28E-120	58%	73%	1--362	1--365
														2	AAA66400	unknown protein	434.11	6.33E-120	58%	73%	1--362	1--365
														3	AAC96901	a534R	154.45	9.68E-36	77%	88%	438--526	16--104
														4	NP_077576	EsV-1-91	52.37	5.18E-05	27%	45%	3--161	8--158
														5	AAX13509	EtpA	47.37	1.67E-03	24%	40%	79--377	1175--1447
														6	XP_654515	hypothetical protein 51.t00023	46.60	2.84E-03	25%	35%	112--299	39--227
N518L	200210--199992	73	8,156	4.08		No Hit Found								1	AAC96902	A535L	69.71	3.23E-11	50%	68%	1--71	1--71
N520L	200500--200258	81	9,159	10.91		No Hit Found								1	AAC96903	A536L	60.85	1.52E-08	49%	66%	17--79	17--73
N521L	201271--200522	250	28,688	5.31	1	cd01293	Bact_CD, Bacterial cytosine deaminase and related metal-dependent hydrolases. Cytosine deaminases (CDs) catalyze the deamination of cytosine, producing uracil and ammonia. They play an important role in pyrimidine salvage. CDs are present in prokaryotes and fungi, but not mammalian cells. The bacterial enzymes, but not the fungal enzymes, are related to the adenosine deaminases (ADA). The bacterial enzymes are iron dependent and hexameric	29.43	4.18E-01	30%	43%	73--127	92--148	1	AAC96904	A537L	103.22	8.20E-21	30%	46%	4--248	10--263
N524R	201389--201898	170	18,987	10.49	1	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	32.35	3.52E-02	38%	54%	29--68	5--44	1	AAC96906	A539R	179.87	2.82E-44	51%	70%	1--168	1--173
														2	AAC96502	similar to bacteriophage T4 intron-associated endonuclease, corresponds to Swiss-Prot Accession Number P13299	50.45	2.58E-05	53%	68%	28--74	11--55
														3	AAC96862	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	42.74	5.39E-03	29%	44%	37--145	15--123
N525L	205835--201867	1323	133,329	4.77		No Hit Found								1	AAC96907	A540L	227.64	2.62E-57	26%	40%	28--897	138--1055
														2	ABE69167	A540L	90.89	3.82E-16	24%	36%	28--452	172--543
														3	YP_589431	hypothetical protein Acid345 0352	50.45	5.72E-04	26%	41%	28--194	701--864
														4	ZP_00950302	outer membrane protein	48.14	2.84E-03	22%	35%	837--1205	58--446
														5	AAA66404	unknown protein	150.21	5.31E-34	37%	53%	1085--1312	9--273
														6	YP_214529	possible T4-like proximal tail fiber	94.74	2.65E-17	24%	36%	569--1223	29--598
														7	ZP_00533161	Hep_Hag	109.00	1.36E-21	26%	42%	582--959	71--447
														8	ZP_01257209	haemagluttinin family protein	50.06	7.47E-04	26%	39%	32--210	156--334
														9	ZP_01059081	hypothetical protein MED217 15260	49.29	1.27E-03	24%	38%	30--208	221--394
														10	ZP_00532602	Hep_Hag	47.75	3.71E-03	24%	37%	9--210	655--868

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
N531L	207897--205900	666	73,134	8.43		No Hit Found								1	AAC96403	contains Pro-rich Px motif, PAPK (19X); similar to Arabidopsis anter-specific Pro-rich protein, corresponds to Swiss-Prot Accession Number P40602	50.06	3.42E-04	28%	45%	174--291	30--134
														2	BAA11344	DNA binding protein	51.22	1.54E-04	26%	36%	133--289	95--261
N533L	208755--207985	257	29,166	8.44		No Hit Found								1	AAC96377	A9R	222.25	1.28E-56	63%	79%	85--250	8--173
														2	AAC96972	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	129.03	1.48E-28	31%	51%	1--251	4--255
														3	AAC96818	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	118.24	2.61E-25	29%	50%	5--249	2--246
														4	AAC96643	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	113.24	8.39E-24	28%	49%	5--248	5--248
														5	AAC96545	A177R	91.66	2.61E-17	25%	44%	5--252	5--245
														6	AAC96447	A79R	91.28	3.41E-17	27%	45%	5--228	2--218
														7	AAC96971	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	85.89	1.43E-15	64%	90%	5--55	2--52
														8	AAU06304	hypothetical protein A275R	58.92	1.88E-07	27%	46%	85--248	4--167
														9	AAU06301	hypothetical protein A275R	57.77	4.18E-07	26%	46%	85--248	4--167
														10	AAU06302	hypothetical protein A275R	57.00	7.13E-07	28%	47%	117--251	1--134
N535L	213044--209160	1295	130,191	6.38	1	COG0196	RibF, FAD synthase [Coenzyme metabolism].	35.26	5.65E-02	25%	39%	1026--1143	164--274	1	AAC96907	A540L	101.29	2.76E-19	24%	36%	7--562	523--1052
														2	ABE69167	A540L	50.45	5.59E-04	26%	41%	9--170	389--544
														3	YP_589431	hypothetical protein Acid345_0352	52.37	1.47E-04	23%	46%	28--198	438--599
														4	AAA66404	unknown protein	137.50	3.48E-30	37%	52%	1080--1292	50--280
														5	ZP_01257209	haemaqluttinin family protein	80.49	5.04E-13	23%	35%	30--465	24--412
														6	ZP_00533161	Hep_Hag	60.46	5.40E-07	22%	45%	12--209	54--249
														7	YP_214529	possible T4-like proximal tail fiber	85.11	2.05E-14	21%	35%	12--652	362--960
														8	ZP_00945173	Collagen alpha 2(I) chain precursor	72.40	1.37E-10	25%	33%	335--730	736--1083
														9	ZP_00532602	Hep_Hag	49.68	9.53E-04	24%	38%	40--209	684--867
														10	ZP_00950302	outer membrane protein	51.22	3.28E-04	25%	45%	28--178	640--784
N540L	214063--213113	317	36,864	10.55	1	COG0436	COG0436, Aspartate/tyrosine/aromatic aminotransferase [Amino acid transoort and metabolism].	29.12	7.54E-01	28%	44%	33--87	124--174	0	No Hit Found	No Hit Found						
N542L	215171--214032	380	41,731	10.39	1	COG3007	COG3007, Uncharacterized paraquat-inducible protein B [Function unknown].	29.53	7.19E-01	37%	47%	282--348	74--133	1	AAC96985	STKPP (11x); similar to Gossypium Pro-rich wall protein, corresponds to GenBank Accession Number U04267	181.80	3.60E-44	46%	60%	159--378	163--384
N546L	215578--215141	146	16,059	7.00	1	cd01286	deoxycytidylate_deaminase, Deoxycytidylate deaminase domain. Deoxycytidylate deaminase catalyzes the deamination of dCMP to dUMP, providing the nucleotide substrate for thymidylate synthase. The enzyme binds Zn++, which is required for catalytic activity. The activity of the enzyme is allosterically regulated by the ratio of dCTP to dTTP not only in eukaryotic cells but also in T-even phage-infected Escherichia coli, with dCTP acting as an activator and dTTP as an inhibitor	128.08	3.69E-31	39%	55%	9--135	4--131	1	AAC96936	similar to Vibrio fischeri dCMP deaminase, corresponds to Swiss-Prot Accession Number P33968	201.45	7.16E-51	65%	81%	4--144	2--142
					2	COG2131	ComEB, Deoxycytidylate deaminase [Nucleotide transport and metabolism].	104.27	5.66E-24	36%	54%	9--141	12--148	2	YP_223954	deoxycytidylate deaminase	124.02	1.45E-27	52%	65%	9--139	110--237
					3	pfam00383	dCMP_cyt_deam, Cytidine and deoxycytidylate deaminase zinc-binding region..	86.57	9.86E-19	41%	60%	8--113	5--100	3	YP_422060	Deoxycytidylate deaminase	109.38	3.70E-23	44%	62%	21--141	20--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..	58.87	2.71E-10	29%	50%	14--113	6--92	4	ZP_00052863	COG2131: Deoxycytidylate deaminase	109.38	3.70E-23	44%	62%	21--141	20--139
					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	56.43	1.36E-09	32%	55%	10--120	1--100	5	NP_975562	dCMP deaminase	99.37	3.83E-20	40%	56%	9--142	13--147
					6	cd01284	Riboflavin_deaminase-reductase, Riboflavin-specific deaminase. Riboflavin biosynthesis protein RibD (Diaminohydroxyphosphoribosylaminopyrimidine deaminase) catalyzes the deamination of 2,5-diamino-6-ribosylamino-4(3H)-pyrimidinone 5'-phosphate, which is an intermediate step in the biosynthesis of riboflavin.The ribG gene of Bacillus subtilis and the ribD gene of E. coli are bifunctional and contain this deaminase domain and a reductase domain which catalyzes the subsequent reduction of the ribosyl side chain	55.25	2.72E-09	38%	56%	28--113	21--93	6	YP_424377	probable deoxycytidylate deaminase	99.37	3.83E-20	40%	56%	9--142	6--140
					7	COG0590	CumB, Cytosine/adenosine deaminases [Nucleotide transport and metabolism / Translation. ribosomal structure and biogenesis].	54.57	5.04E-09	34%	50%	1--113	3--104	7	ZP_00527356	Cytidine/deoxycytidylate deaminase, zinc-binding region	98.98	5.01E-20	45%	68%	21--113	22--115
					8	COG0117	RibD, Pyrimidine deaminase [Coenzyme metabolism].	54.10	7.39E-09	28%	51%	1--113	1--100	8	NP_001016265	hypothetical protein LOC549019	98.60	6.54E-20	37%	56%	8--140	28--162
														9	AAH68731	MGC81193 protein	98.21	8.54E-20	38%	56%	8--140	28--162
														10	YP_293777	putative deoxycytidylate deaminase	97.44	1.46E-19	50%	67%	21--113	35--134
N548R	215695--216513	273	30,481	8.08	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	150.74	1.47E-37	29%	50%	15--265	1--256	1	AAC96657	similar to PBCV-1 serine/threonine protein kinase, corresponds to GenBank Accession Number U14660	136.73	7.88E-31	33%	49%	4--265	6--279
					2	smart00220	S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-specific kinase subfamily	144.97	8.97E-36	28%	48%	16--265	1--256	2	XP_387297	hypothetical protein FG07121.1	98.21	3.10E-19	27%	46%	14--269	285--556
					3	pfam00069	Pkinase, Protein kinase domain..	143.89	1.70E-35	30%	49%	16--265	1--258	3	CAJ02527	mitogen-activated protein kinase kinase 2	96.67	9.03E-19	29%	47%	19--268	11--265
					4	COG0515	SPS1, Serine/threonine protein kinase [General function prediction only / Signal transduction mechanisms / Transcription / DNA replication, recombination, and repair]	106.78	2.54E-24	26%	44%	16--269	2--282	4	CAC07966	putative mitogen-activated protein kinase kinase 2	95.52	2.01E-18	29%	47%	19--268	11--265
					5	smart00219	TyrKc, Tyrosine kinase, catalytic domain; Phosphotransferases. Tyrosine-specific kinase subfamily. .	74.44	1.48E-14	24%	44%	17--258	2--250	5	AAU06280	protein kinase A248R	93.59	7.65E-18	25%	46%	14--268	46--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	68.68	7.32E-13	24%	43%	13--261	6--264	6	AAU06275	protein kinase A248R	93.59	7.65E-18	25%	46%	14--268	25--287
					7	COG3894	COG3894, Uncharacterized metal-binding protein [General function prediction only].	35.69	7.13E-03	21%	35%	130--247	105--227	7	XP_578647	PREDICTED: similar to hypothetical protein 4930509O22	92.43	1.70E-17	30%	46%	22--265	734--976
					8	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	31.66	1.10E-01	36%	52%	116--157	156--198	8	NP_180595	CIPK11; ATP binding / kinase/ protein kinase/ protein serine/threonine kinase/ protein-tyrosine kinase	91.66	2.91E-17	28%	47%	22--264	27--276
					9	COG2334	COG2334, Putative homoserine kinase type II (protein kinase fold) [General function prediction only]	30.70	1.92E-01	27%	43%	114--158	183--227	9	AAL16166	At2g30360/T9D9.17	91.66	2.91E-17	28%	47%	22--264	27--276

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	COG3173	COG3173, Predicted aminoglycoside phosphotransferase [General function prediction only]	29.61	3.86E-01	42%	52%	125--157	195--226	10	AAU06282	protein kinase A248R	91.28	3.79E-17	25%	45%	14--268	46--308
N549R	216631--219810	1060	117,600	9.53	1	smart00433	TOP2c, Topoisomerase II; Eukaryotic DNA topoisomerase II, GyrB, ParE .	453.09	7.69E-128	32%	49%	50--610	1--589	1	AAU95770	topoisomerase II	1889.39	0.00E+00	90%	92%	1--1058	1--1058
					2	smart00434	TOP4c, DNA Topoisomerase IV; Bacterial DNA topoisomerase IV, GyrA, ParC.	392.63	1.11E-109	33%	50%	626--1017	1--423	2	AAC96932	PBCV-1 DNA topoisomerase II	1338.55	0.00E+00	63%	75%	3--1056	2--1061
					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.	362.96	8.47E-101	33%	51%	644--1017	1--401	3	NP_584718	DNA TOPOISOMERASE II	877.09	0.00E+00	47%	64%	5--1017	8--1030
					4	COG0187	GyrB, Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), B subunit [DNA replication, recombination, and repair].	319.39	1.17E-87	27%	43%	2--626	8--635	4	NP_189031	TOPII (TOPOISOMERASE II); ATP binding / DNA binding / DNA topoisomerase (ATP-hydrolyzing)	846.27	0.00E+00	46%	63%	5--1000	35--1072
					5	pfam00521	DNA_topoisoIV, DNA gyrase/topoisomerase IV, subunit A..	271.27	3.49E-73	30%	45%	646--1017	1--397	5	AAW40881	DNA topoisomerase II, putative	844.34	0.00E+00	45%	61%	5--1000	112--1153
					6	COG0188	GyrA, Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), A subunit [DNA replication, recombination, and repair].	251.33	3.96E-67	23%	44%	617--1015	2--428	6	XP_467311	putative DNA topoisomerase II	834.71	0.00E+00	45%	60%	5--1000	32--1097
					7	cd03365	TOPRIM_TopoIIA_TopoIIA, TOPRIM_TopoIIA: topoisomerase-primase (TOPRIM) nucleotidyl transferase/hydrolase domain of the type found in proteins of the type IIA family of DNA topoisomerases similar to Saccharomyces cerevisiae Topoisomerase II. TopoIIA enzymes cut both strands of the duplex DNA to remove (relax) both positive and negative supercoils in DNA. These enzymes covalently attach to the 5' ends of the cut DNA, separate the free ends of the cleaved strands, pass another region of the duplex through this gap, then rejoin the ends. These proteins also catenate/ decatenate duplex rings. The TOPRIM domain has two conserved motifs, one of which centers at a conserved glutamate and the other one at two conserved aspartates (DxD). This glutamate and two aspartates, cluster together to form a highly acid surface patch. The conserved glutamate may act as a general base in strand joining and as a general acid in strand cleavage by topisomerases. The DXD motif may co-ordinate Mg2+, a cofactor required for full catalytic function.	142.67	2.18E-34	66%	78%	403--521	1--120	7	BAE06274	topoisomerase II	833.56	0.00E+00	45%	60%	5--1000	35--1078
					8	cd01030	TOPRIM_TopoIIA, TOPRIM_TopoIIA: topoisomerase-primase (TOPRIM) nucleotidyl transferase/hydrolase domain of the type found in proteins of the type IIA family of DNA topoisomerases similar to Saccharomyces cerevisiae Topoisomerase II. TopoIIA enzymes cut both strands of the duplex DNA to remove (relax) both positive and negative supercoils in DNA. These enzymes covalently attach to the 5' ends of the cut DNA, separate the free ends of the cleaved strands, pass another region of the duplex through this gap, then rejoin the ends. These proteins also catenate/ decatenate duplex rings. The TOPRIM domain has two conserved motifs, one of which centers at a conserved glutamate and the other one at two conserved aspartates (DxD). The conserved glutamate may act as a general base in strand joining and as a general acid in strand cleavage by topisomerases. The DXD motif may co-ordinate Mg2+, a cofactor required for full catalytic function..	123.67	1.00E-28	53%	63%	403--521	1--115	8	BAD86854	DNA topoisomerase II	833.17	0.00E+00	43%	60%	5--1000	103--1163
					9	cd03366	TOPRIM_TopoIIA_GyrB, TOPRIM_TopoIIA_GyrB: topoisomerase-primase (TOPRIM) nucleotidyl transferase/hydrolase domain of the type found in proteins of the type IIA family of DNA topoisomerases similar to the Escherichia coli GyrB subunit. TopoIIA enzymes cut both strands of the duplex DNA to remove (relax) both positive and negative supercoils in DNA. These enzymes covalently attach to the 5' ends of the cut DNA, separate the free ends of the cleaved strands, pass another region of the duplex through this gap, then rejoin the ends. These proteins also catenate/ decatenate duplex rings. DNA gyrase is more effective at relaxing supercoils than decatentating DNA. DNA gyrase in addition inserts negative supercoils in the presence of ATP. The TOPRIM domain has two conserved motifs, one of which centers at a conserved glutamate and the other one at two conserved aspartates (DxD). The conserved glutamate may act as a general base in strand joining and as a general acid in strand cleavage by topisomerases. The DXD motif may co-ordinate Mg2+, a cofactor required for full catalytic function.	82.10	2.96E-16	44%	57%	403--510	1--105	9	AAN85208	DNA topoisomerase II	831.63	0.00E+00	45%	62%	5--1000	27--1065
					10	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..	62.57	2.23E-10	27%	44%	237--382	14--168	10	XP_751245	DNA topoisomerase II	828.17	0.00E+00	44%	61%	2--996	55--1094
N559R	219871--220755	295	33,581	10.47	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.	55.39	9.05E-09	41%	61%	181--275	1--100	1	AAA88832	unknown	194.90	2.78E-48	40%	55%	1--289	1--240
					2	smart00497	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished). .	48.98	7.50E-07	42%	60%	239--291	1--53	2	AAC96973	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	190.66	5.25E-47	39%	53%	1--290	1--224
					3	pfam07453	NUMOD1, NUMOD1 domain..	44.65	1.50E-05	48%	67%	239--272	1--34	3	AAC96655	PBCV-1 33kd peptide	173.33	8.67E-42	35%	52%	3--290	6--248
					4	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	40.06	3.56E-04	32%	54%	1--72	1--70	4	AAC96862	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	48.14	4.21E-04	42%	67%	179--234	97--151
					5	COG1693	COG1693, Uncharacterized protein conserved in archaea [Function unknown]	35.64	7.69E-03	29%	48%	108--186	13--90	5	NP_899393	SegD	85.11	3.10E-15	29%	46%	1--241	1--197
					6	pfam01995	DUF128, Domain of unknown function DUF128. This archaebacterial protein family has no known function. The domain is found duplicated in one member.	35.33	1.03E-02	27%	48%	108--186	14--91	6	YP_293795	putative endonuclease	72.02	2.72E-11	44%	58%	2--84	3--85
					7	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. .	33.98	2.43E-02	33%	56%	1--59	1--64	7	AAC49244	ORF301	71.63	3.55E-11	28%	48%	4--170	76--236
					8	COG3069	DcuC, C4-dicarboxylate transporter [Energy production and conversion].	31.79	1.05E-01	26%	43%	23--65	187--229	8	NP_074951	orf305	70.86	6.06E-11	32%	49%	4--204	74--262
					9	pfam06342	DUF1057, Protein of unknown function (DUF1057). This family consists of several Caenorhabditis elegans specific proteins of unknown function..	31.43	1.41E-01	31%	48%	118--202	200--284	9	AAC49248	ORF211	55.84	2.02E-06	28%	47%	129--290	49--208
														10	AAT53588	group I intron GIY-YIG endonuclease	49.68	1.45E-04	25%	43%	4--205	5--197
N561R	220777--221190	138	15,751	11.06	1	cd02698	Peptidase_C1A_CathepsinX, Cathepsin X; the only papain-like lysosomal cysteine peptidase exhibiting carboxymonopeptidase activity. It can also act as a carboxydipeptidase, like cathepsin B, but has been shown to preferentially cleave substrates through a monopeptidyl carboxypeptidase pathway. The propeptide region of cathepsin X, the shortest among papain-like peptidases, is covalently attached to the active site cysteine in the inactive form of the enzyme. Little is known about the biological function of cathepsin X. Some studies point to a role in early tumorigenesis. A more recent study indicates that cathepsin X expression is restricted to immune cells suggesting a role in phagocytosis and the regulation of the immune response..	26.82	9.48E-01	68%	73%	9--35	40--62	1	AAC96929	A577L	108.61	6.33E-23	55%	62%	26--130	2--104

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
N562R	221235--221453	73	8,350	10.39	1	COG2433	COG2433, Uncharacterized conserved protein [Function unknown].	26.80	5.99E-01	19%	47%	7--54	415--462	0	No Hit Found	No Hit Found						
N564R	221522--222049	176	19,167	9.91		No Hit Found								1	AAC96743	A375R	234.19	1.39E-60	64%	80%	4--171	6--172
N565L	223773--222040	578	59,995	10.57		No Hit Found								1	AAC96773	Pro-, Lys-rich, PAPK (30x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	198.75	5.00E-49	37%	56%	301--576	214--496
														2	BAE02830	surface protein	46.98	2.43E-03	30%	53%	234--304	638--707
														3	AAC96508	PBCV-1 surface protein	46.98	2.43E-03	32%	50%	234--304	829--896
														4	BAD22850	surface protein	46.60	3.18E-03	25%	49%	234--326	675--767
														5	AAC96752	Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	49.68	3.75E-04	30%	54%	233--304	201--272
														6	BAD12236	surface protein	46.98	2.43E-03	32%	50%	234--304	829--896
														7	T17636	proline-rich protein A145R - Chlorella virus PBCV-1	46.98	2.43E-03	32%	50%	234--304	38--105
N567L	224456--223893	188	21,369	3.11		No Hit Found								1	AAC96772	Glu-, Asn-rich	80.11	4.02E-14	65%	77%	1--57	2--59
N569L	224896--224627	90	10,063	10.28		No Hit Found								1	AAC96771	A403R	94.74	9.54E-19	52%	69%	2--88	6--93
N570L	225626--224934	231	25,002	6.63		No Hit Found								1	AAC96770	A402R	231.88	1.31E-59	48%	70%	1--229	1--227
N572L	226594--225719	292	32,559	5.48		No Hit Found								1	AAC96769	A401R	337.04	4.45E-91	60%	75%	4--287	2--274
														2	NP_772491	hypothetical protein bll5851	141.35	3.58E-32	35%	53%	34--284	7--245
														3	ZP_00863599	conserved hypothetical protein	130.95	4.84E-29	33%	52%	34--284	11--249
														4	AAW60029	Hypothetical protein GOX0246	119.40	1.46E-25	31%	48%	28--287	5--253
														5	ZP_01078938	hypothetical protein RS9917 04490	95.90	1.73E-18	28%	46%	34--281	8--217
														6	YP_616914	hypothetical protein Sala_1869	69.71	1.33E-10	24%	41%	23--289	2--263
N574L	226993--226637	119	13,074	10.03	1	cd02810	DHOD_DHPD_FMN, Dihydroorotate dehydrogenase (DHOD) and Dihydropyrimidine dehydrogenase (DHPD) FMN-binding domain. DHOD catalyzes the oxidation of (S)-dihydroorotate to orotate. This is the fourth step and the only redox reaction in the de novo biosynthesis of UMP, the precursor of all pyrimidine nucleotides. DHOD requires FMN as co-factor. DHOD divides into class 1 and class 2 based on their amino acid sequences and cellular location. Members of class 1 is cytosolic enzymes and multimers while class 2 enzymes are membrane associated and monomeric. The class 1 enzymes can be further divided into subtypes 1A and 1B which are homodimers and heterotetrameric proteins, respectively. DHPD catalyzes the first step in pyrimidine degradation: the NADPH-dependent reduction of uracil and thymine to the corresponding 5,6-dihydropyrimidines. DHPD contains two FAD, two FMN and eight [4Fe-4S] clusters, arranged in two electron transfer chains that pass its homodimeior interface twice. Two of the Fe-S clusters show a hitherto unobserved coordination involving a glutamine residue	27.45	4.37E-01	26%	63%	6--33	96--123	1	AAC96768	similar to PBCV-1 ORF A214, corresponds to GenBank Accession Number U42580	107.07	1.86E-22	45%	68%	5--117	4--116
N575L	227516--227019	166	18,181	9.92	1	pfam00075	RnaseH, RNase H. RNase H digests the RNA strand of an RNA/DNA hybrid. Important enzyme in retroviral replication cycle, and often found as a domain associated with reverse transcriptases. Structure is a mixed alpha+beta fold with three a/b/a layers..	73.40	1.46E-14	30%	47%	16--136	2--124	1	AAC96767	A399R	149.83	2.87E-35	51%	64%	4--158	17--180
					2	COG0328	RnhA, Ribonuclease HI [DNA replication, recombination, and repair].	57.29	1.05E-09	23%	37%	17--147	3--153	2	XP_662792	hypothetical protein AN5188.2	53.53	2.80E-06	32%	44%	18--136	228--373
														3	CAA41925	unnamed protein product	52.76	4.77E-06	32%	51%	15--140	238--369
														4	S37765	hypothetical protein - fruit fly (Drosophila miranda) transposon TRIM	52.76	4.77E-06	32%	51%	15--140	143--274
														5	1RDH_B	Chain B, Hiv-1 Reverse Transcriptase (Ribonuclease H Domain) (E.C.2.7.7.49)	51.22	1.39E-05	27%	42%	9--136	1--141
														6	ZP_01217503	hypothetical protein PCNPT3 00010	50.83	1.81E-05	32%	44%	21--136	634--755
														7	EAL40632	ENSANGP00000006819	50.45	2.37E-05	29%	48%	13--137	249--372
														8	EAL40633	ENSANGP00000020605	49.68	4.04E-05	29%	48%	13--137	274--397
														9	AAB65093	Lian-Aa1 retrotransposon protein	49.68	4.04E-05	31%	52%	15--147	910--1046
														10	XP_750465	ribonuclease H1	49.29	5.28E-05	31%	46%	18--136	239--383
N578R	227622--227981	120	12,941	10.45		No Hit Found								1	AAC96766	A398L	147.52	1.24E-34	61%	77%	1--118	1--118
N579L	228450--227986	155	17,543	11.74	1	COG4656	RnfC, Predicted NADH:ubiquinone oxidoreductase, subunit RnfC [Energy production and conversion].	28.75	3.10E-01	19%	41%	45--131	431--517	1	AAU06284	hypothetical protein	77.41	1.55E-13	33%	51%	6--152	22--150
														2	XP_650245	hypothetical protein 233.t00006	47.75	1.31E-04	31%	54%	45--152	361--461
														3	AAW42894	chromosome associated protein, putative	44.67	1.11E-03	25%	47%	8--144	314--457
														4	AAH87521	LOC496093 protein	43.90	1.90E-03	29%	54%	49--127	50--128
														5	ZP_01179929	NLP/P60	43.51	2.48E-03	25%	44%	5--149	160--291
														6	XP_654272	hypothetical protein 58.t00022	41.59	9.42E-03	29%	44%	50--152	340--438
														7	EAR81596	conserved hypothetical protein	43.13	3.24E-03	28%	46%	46--152	77--180
														8	XP_742542	hypothetical protein	42.36	5.52E-03	24%	41%	19--152	105--235
														9	CAG74282	TolA protein	41.97	7.21E-03	23%	42%	2--151	88--237
														10	XP_726544	erythrocyte binding protein	41.97	7.21E-03	28%	47%	50--152	1225--1329
N582L	228760--228479	94	10,557	10.42	1	pfam01986	DUF123, Domain of unknown function DUF123. This archaebacterial domain has no known function. It is attached to an endonuclease domain in one member. The domain contains several conserved cysteines and histidines. This suggests that the domain may be a zinc binding nucleic acid interaction domain (Bateman A unpubl.)..	26.85	6.94E-01	30%	41%	16--60	23--84	1	AAC96763	A395R	109.00	4.82E-23	64%	78%	11--92	1--82
N584R	228842--229471	210	23,746	8.20	1	COG1403	McrA, Restriction endonuclease [Defense mechanisms].	28.37	7.04E-01	20%	29%	117--193	44--120	1	AAC96845	Lys-, Arg-rich; contains eukaryotic putative RNA-binding region RNP-1 signature; similar to PBCV-1 ORF A267L, corresponds to GenBank Accession Number U42580	164.85	1.62E-39	42%	58%	19--207	106--294
					2	pfam03302	VSP, Giardia variant-specific surface protein..	28.17	7.74E-01	29%	42%	62--107	275--320	2	AAC96857	Lys-, Glu-rich	48.14	2.20E-04	27%	44%	5--97	189--288
														3	YP_142777	unknown	50.45	4.44E-05	22%	40%	2--165	365--524
														4	AAC96635	A267L	93.97	3.50E-18	32%	49%	13--207	59--247
N586L	230036--229458	193	21,469	3.72		No Hit Found								1	AAC96762	A394R	97.44	2.58E-19	47%	64%	6--110	14--116
N589L	230850--230080	257	28,991	9.23	1	COG0433	COG0433, Predicted ATPase [General function prediction only].	29.95	3.60E-01	25%	44%	77--146	373--446	1	AAC96760	contains ATP/GTP-binding site motif A; similar to frog virus 3 ATPase, corresponds to GenBank Accession Number M80551	377.87	1.83E-103	74%	81%	1--252	1--251
					2	COG1674	FtsK, DNA segregation ATPase FtsK/SpoIIIE and related proteins [Cell division and chromosome partitioning]	29.41	4.34E-01	22%	43%	83--149	627--696	2	BAB69884	AGB-1	146.75	6.84E-34	39%	60%	32--228	38--237
														3	AAR26836	FirrV-1-A12	145.21	1.99E-33	39%	57%	32--228	36--241
														4	NP_077511	EsV-1-26	131.72	2.28E-29	36%	57%	32--228	47--252
														5	ABF82118	hypothetical protein MIV088R	130.57	5.08E-29	34%	56%	32--223	50--244
														6	AAT09674	AAA-ATPase	118.24	2.61E-25	33%	55%	32--228	43--241
														7	AAP33264	ATPase	118.24	2.61E-25	33%	55%	32--228	43--241
														8	AAL77796	putative ATPase	118.24	2.61E-25	33%	55%	32--228	43--241
														9	AAM82544	putative ATPase	117.86	3.40E-25	33%	55%	32--228	43--241
														10	AAS18149	ATPase	116.70	7.58E-25	31%	56%	32--228	46--244
N593R	230933--232633	567	59,105	7.04	1	pfam02359	CDC48_N, Cell division protein 48 (CDC48), N-terminal domain. This domain has a double psi-beta barrel fold and includes VCP-like ATPase and N-ethylmaleimide sensitive fusion protein N-terminal domains. Both the VAT and NSF N-terminal functional domains consist of two structural domains of which this is at the N-terminus. The VAT-N domain found in AAA ATPases pfam00004 is a substrate 185-residue recognition domain	31.35	3.56E-01	36%	47%	406--452	32--85	1	AAC96752	Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	93.20	2.91E-17	31%	38%	127--337	26--181
					2	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..	30.79	4.91E-01	25%	48%	515--565	1--49	2	AAC96921	A565R	56.61	3.01E-06	27%	36%	128--333	298--469
					3	smart00494	ChtBD2, Chitin-binding domain type 2; .	30.10	6.97E-01	38%	50%	515--565	3--51	3	BAA11343	DNA binding protein	110.92	1.35E-22	34%	40%	129--374	252--458
														4	AAC96700	a332L	73.94	1.82E-11	61%	80%	523--564	2--43
														5	AAC96919	similar to Chlorella virus CVK2 DNA binding protein, corresponds to GenBank Accession Number D78305	63.16	3.22E-08	40%	51%	283--374	306--403

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	BAA11342	DNA binding protein	59.69	3.56E-07	39%	50%	283–374	304–401
														7	XP_416321	PREDICTED: similar to FELE-2	45.44	6.95E-03	22%	33%	197–374	1821–1992
														8	CAA64974	QI74 protein	52.76	4.35E-05	23%	35%	129–374	403–638
														9	CAI17471	tenascin XB	55.45	6.71E-06	26%	35%	139–367	416–673
														10	CAI18078	tenascin XB	55.45	6.71E-06	26%	35%	139–367	416–673
N598R	232979–233872	298	31,574	6.93	1	pfam03643	Vps26, Vacuolar protein sorting-associated protein 26. Vacuolar protein sorting-associated protein (Vps) 26 is one of around 50 proteins involved in protein trafficking. In particular, Vps26 assembles into a retromer complex with at least four other proteins Vps5, Vps17, Vps29 and Vps35. This family also contains Down syndrome critical region 3/A..	29.48	5.15E-01	26%	37%	34–94	9–71	1	AAC97017	a593R	47.37	7.31E-04	63%	65%	87–124	1–38
N600L	234932–233850	361	40,439	6.99	1	COG0076	GadB, Glutamate decarboxylase and related PLP-dependent proteins [Amino acid transport and metabolism].	143.59	2.97E-35	24%	44%	46–310	91–380	1	AAC96937	similar to tomato histidine decarboxylase, corresponds to Swiss-Prot Accession Number P54772	466.46	6.82E-130	62%	78%	4–353	11–362
					2	pfam00282	Pyridoxal_deC, Pyridoxal-dependent decarboxylase conserved domain..	115.40	1.01E-26	27%	44%	54–289	79–352	2	P28578	Histidine decarboxylase (HDC)	217.24	7.21E-55	35%	57%	14–347	26–369
					3	COG0520	CsdB, Selenocysteine lyase [Amino acid transport and metabolism].	47.24	3.04E-06	26%	41%	63–308	75–339	3	AAA25071	histidine decarboxylase	217.24	7.21E-55	35%	57%	14–347	26–369
					4	COG1104	NifS, Cysteine sulfinate desulfinase/cysteine desulfurase and related enzymes [Amino acid transport and metabolism].	42.08	1.14E-04	23%	44%	79–226	67–216	4	ZP_00510529	Pyridoxal-dependent decarboxylase	210.69	6.75E-53	36%	56%	14–347	47–390
					5	COG2008	GLY1, Threonine aldolase [Amino acid transport and metabolism].	36.76	4.16E-03	19%	35%	54–216	35–206	5	AAR12533	histidine decarboxylase	208.76	2.56E-52	34%	56%	14–347	40–383
					6	COG1893	ApbA, Ketopantoate reductase [Coenzyme metabolism].	32.21	1.09E-01	20%	33%	140–278	45–178	6	AAO92385	histidine decarboxylase	208.76	2.56E-52	34%	56%	14–347	26–369
					7	pfam00266	Aminotran_5, Aminotransferase class-V..	31.05	2.31E-01	22%	36%	77–226	55–203	7	ABE77150	aromatic amino acid decarboxylase 1B	208.76	2.56E-52	35%	56%	10–347	86–432
					8	pfam06101	DUF946, Plant protein of unknown function (DUF946). This family consists of several hypothetical proteins from Arabidopsis thaliana and Oryza sativa. The function of this family is unknown..	30.65	2.68E-01	27%	46%	307–355	343–391	8	ZP_00106716	COG0076: Glutamate decarboxylase and related PLP-dependent proteins	207.61	5.71E-52	35%	56%	15–347	25–363
					9	COG1003	GcvP, Glycine cleavage system protein P (pyridoxal-binding), C-terminal domain [Amino acid transport and metabolism].	30.24	4.15E-01	25%	45%	82–231	134–289	9	P05034	Histidine decarboxylase (HDC)	206.45	1.27E-51	33%	55%	14–355	26–377
														10	BAE94285	histidine decarboxylase	206.07	1.66E-51	34%	56%	14–347	26–369
N603L	235413–235090	108	12,185	5.98	1	pfam07137	VDE, Violaxanthin de-epoxidase (VDE). This family represents a conserved region approximately 350 residues long within plant violaxanthin de-epoxidase (VDE). In higher plants, violaxanthin de-epoxidase forms part of a conserved system that dissipates excess energy as heat in the light-harvesting complexes of photosystem II (PSII), thus protecting them from photo-inhibitory damage	26.54	7.02E-01	35%	53%	40–77	301–341	1	AAC96939	A602L	58.54	7.63E-08	34%	64%	3–72	46–113
N604R	235450–235770	107	11,961	6.35		No Hit Found								1	AAC96940	A603R	96.29	3.31E-19	44%	64%	1–103	1–104
N606L	236442–235942	167	18,320	10.80		No Hit Found								0	No Hit Found	No Hit Found						
N609L	237578–236511	356	40,475	9.80		No Hit Found								1	AAC96722	A354R	214.16	5.97E-54	47%	63%	110–353	1–235
														2	AAC96790	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	85.50	3.20E-15	28%	47%	30–320	30–319
														3	AAC96455	A87R	51.99	3.92E-05	28%	45%	191–350	289–447
														4	ABE04297	Ribonuclease E	49.29	2.54E-04	32%	47%	71–183	582–692
N613L	238827–238048	260	30,305	10.39		No Hit Found								1	AAC96859	A492L	108.61	2.11E-22	36%	60%	85–250	19–189
														2	AAC96860	A493L	90.12	7.78E-17	39%	56%	105–242	29–174
N614R	238583–239749	389	45,772	9.12	1	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..	29.90	6.02E-01	20%	37%	181–240	118–189	1	AAC96861	A494R	422.94	9.63E-117	58%	75%	36–387	4–360
														2	YP_142783	PBCV1-A494R-like protein	154.84	4.89E-36	34%	54%	122–387	166–431
														3	NP_077586	EsV-1-101	135.19	4.01E-30	37%	54%	195–386	120–318
														4	AAR26895	FirrV-1-B20	127.10	1.09E-27	33%	52%	178–386	86–300
														5	YP_294161	hypothetical protein EhV 403	112.08	3.64E-23	33%	51%	182–383	156–362
														6	NP_149745	282R	70.86	9.29E-11	22%	42%	126–384	101–380
														7	ABF82109	hypothetical protein MIV079L	60.85	9.61E-08	20%	37%	106–383	93–390
														8	YP_142784	unknown	58.15	6.23E-07	27%	44%	266–383	131–249
														9	XP_743482	hypothetical protein PC001078.02.0	48.52	4.94E-04	22%	43%	78–302	134–355
														10	EAR82939	hypothetical protein TTHERM_01054240	48.14	6.45E-04	21%	43%	64–326	387–627
N616L	240862–239714	383	43,529	7.20	1	smart00497	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished). .	41.66	1.49E-04	34%	62%	228–279	1–51	1	AAA88832	unknown	88.20	5.49E-16	29%	46%	56–279	14–240
					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	37.06	3.77E-03	28%	38%	44–113	4–75	2	AAC96973	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	86.66	1.60E-15	32%	47%	57–279	15–223
					3	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	35.05	1.51E-02	31%	47%	42–133	2–83	3	AAC96862	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	84.34	7.93E-15	42%	59%	53–155	13–117
														4	ZP_00738535	hypothetical protein RBTH_06729	60.85	9.39E-08	25%	44%	73–247	9–166
														5	YP_293795	putative endonuclease	55.07	5.15E-06	40%	55%	57–137	16–97
														6	CAA25939	unnamed protein product	45.05	5.33E-03	26%	41%	203–358	73–226
														7	NP_049674	MobB homing endonuclease	45.05	5.33E-03	26%	41%	203–358	110–263
														8	YP_142601	HNH endonuclease	44.67	6.96E-03	31%	57%	155–220	124–189
N618R	240974–241405	144	14,855	10.16		No Hit Found								1	AAC96864	Gly-rich, KGKDIGGG (4x)	51.99	7.08E-06	38%	49%	75–141	79–145
N619L	242426–241404	341	36,454	5.08	1	COG1058	CinA, Predicted nucleotide-utilizing enzyme related to molybdopterin-biosynthesis enzyme MoeA [General function prediction only].	30.58	2.65E-01	38%	46%	21–71	38–86	1	AAC96868	A501L	63.54	1.22E-08	48%	67%	9–78	5–66
														2	AAC96867	Pro-, Ser-rich	58.54	3.91E-07	49%	67%	270–318	4–64
N621L	242725–242495	77	8,145	4.28		No Hit Found								1	AAC96858	A491R	58.15	9.97E-08	49%	72%	25–75	26–76
N624L	243735–242770	322	34,665	5.24		No Hit Found								1	AAC96855	Pro-,Gln-rich	358.22	2.19E-97	60%	72%	1–320	1–317
														2	YP_142771	unknown	60.46	9.44E-08	25%	39%	8–256	5–245
N628L	244213–243776	146	16,890	11.00		No Hit Found								1	AAC96852	A485R	176.41	2.47E-43	57%	78%	4–144	7–147
N629L	244666–244229	146	16,765	10.06	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	135.17	2.51E-33	47%	61%	1–134	1–131	1	AAD33374	pyrimidine dimer-specific glycosylase	162.54	3.69E-39	59%	70%	1–134	1–133
					2	cd01334	Lyase_I, Lyase class I family of the Lyase_I superfamily. This family contains proteins similar to class II fumarase, aspartase, adenylosuccinate lyase (ASL), argininosuccinate lyase (ASAL), and 3-carboxy-cis,cis-muconate lactonizing (CMLE) enzyme. Proteins of this family for the most part catalyze similar beta-elimination reactions in which a C-N or C-O bond is cleaved with the release of fumarate as one of the products. These proteins are active as tetramers. The four active sites of the homotetrameric enzyme are each formed by residues from three different subunits	29.13	2.28E-01	23%	36%	12–110	56–137	2	AAC96418	PBCV-1 pyrimidine dimer-specific glycosylase	162.16	4.82E-39	59%	70%	1–134	1–133

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	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..	27.79	4.86E-01	59%	76%	72--89	105--122	3	AAD33369	pyrimidine dimer-specific glycosylase	161.77	6.29E-39	58%	70%	1--134	1--133
														4	AAD33377	pyrimidine dimer-specific glycosylase	161.00	1.07E-38	58%	70%	1--134	1--133
														5	AAD33379	pyrimidine dimer-specific glycosylase	161.00	1.07E-38	58%	70%	1--134	1--133
														6	AAD33380	pyrimidine dimer-specific glycosylase	160.61	1.40E-38	58%	72%	1--134	1--133
														7	AAD33381	pyrimidine dimer-specific glycosylase	160.61	1.40E-38	58%	72%	1--134	1--133
														8	AAD33391	pyrimidine dimer-specific glycosylase	160.61	1.40E-38	58%	72%	1--134	1--133
														9	AAD33352	pyrimidine dimer-specific glycosylase	160.23	1.83E-38	58%	71%	1--134	1--133
														10	AAD33367	pyrimidine dimer-specific glycosylase	159.84	2.39E-38	58%	72%	1--134	1--133
N631L	245296--245018	93	9,582	10.67		No Hit Found								1	AAC96847	Gly-rich	94.74	9.44E-19	60%	83%	3--68	4--69
N633L	246022--245321	234	26,212	4.58	1	COG5219	COG5219, Uncharacterized conserved protein, contains RING Zn-finger [General function prediction only].	38.58	6.29E-04	28%	41%	149--198	1471--1525	1	AAC96848	A481L	207.22	3.55E-52	44%	65%	1--228	1--224
					2	COG5540	COG5540, RING-finger-containing ubiquitin ligase [Posttranslational modification, protein turnover, chaperones].	38.15	9.29E-04	30%	51%	149--196	325--372	2	NP_908582	P0551A11.21	47.75	3.60E-04	45%	60%	146--195	143--192
					3	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).	37.76	1.08E-03	30%	52%	149--195	1--45	3	BAD45600	ring-H2 zinc finger protein-like	47.75	3.60E-04	45%	60%	146--195	293--342
														4	ABF95852	Zinc finger, C3HC4 type family protein	46.21	1.05E-03	44%	57%	143--193	38--88
														5	XP_505032	hypothetical protein	45.44	1.79E-03	32%	47%	149--203	345--399
														6	BAE65491	unnamed protein product	45.05	2.33E-03	32%	46%	139--204	181--251
														7	XP_460105	hypothetical protein DEHA0E19426g	43.90	5.20E-03	36%	46%	149--204	1614--1673
														8	NP_193525	ubiquitin-protein ligase/ zinc ion binding	43.90	5.20E-03	28%	51%	144--231	1047--1138
														9	XP_550364	putative RING finger 1	43.90	5.20E-03	44%	52%	147--195	128--176
														10	P0C041	Putative RING-H2 finger protein ATL4H	43.90	5.20E-03	28%	51%	144--231	175--266
N635L	247184--246084	367	41,784	10.77	1	cd00233	VIP2, VIP2; A family of actin-ADP-ribosylating toxin. A member of the Bacillus-prodiced vegetative insecticidal proteins (VIPs) possesses high specificity against the major insect pest, corn rootworms, and belongs to a classs of binary toxins and regulators of biological pathways distinct from classical A-B toxins. A novel family of insecticidal ADP-ribosyltransferses were isolated from Bacillus cereus during vegetative growth, where VIP1 likely targets insect cells and VIP2 ribosylates actin. VIP2 shares significant sequence similarity with enzymatic components of other binary toxins, Clostridium botulinum C2 toxin, C. perfringens iota toxin, C. piroforme toxin, C. piroforme toxin and C. difficile toxin..	30.29	3.84E-01	31%	56%	298--344	138--183	1	AAC96461	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563	179.49	1.71E-43	41%	61%	1--229	46--296
					2	cd03280	ABC_MutS2, MutS2 homologs in bacteria and eukaryotes. The MutS protein initiates DNA mismatch repair by recognizing mispaired and unpaired bases embedded in duplex DNA and activating endo- and exonucleases to remove the mismatch. Members of the MutS family also possess a conserved ATPase activity that belongs to the ATP binding cassette (ABC) superfamily. MutS homologs (MSH) have been identified in most prokaryotic and all eukaryotic organisms examined. Prokaryotes have two homologs (MutS1 and MutS2), whereas seven MSH proteins (MSH1 to MSH7) have been identified in eukaryotes. The homodimer MutS1 and heterodimers MSH2-MSH3 and MSH2-MSH6 are primarily involved in mitotic mismatch repair, whereas MSH4-MSH5 is involved in resolution of Holliday junctions during meiosis. All members of the MutS family contain the highly conserved Walker A/B ATPase domain, and many share a common mechanism of action. MutS1, MSH2-MSH3, MSH2-MSH6, and MSH4-MSH5 dimerize to form sliding clamps, and recognition of specific DNA structures or lesions results in ADP/ATP	29.41	7.16E-01	29%	47%	135--181	134--179	2	AAC96459	a91L	117.86	6.11E-25	49%	64%	241--364	1--126
														3	AAC96984	similar to Chlorella virus PBCV-1 ORF A282L, corresponds to GenBank Accession Number U42580	59.31	2.58E-07	56%	68%	3--61	2--60
														4	AAC96650	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055	56.23	2.18E-06	51%	68%	3--60	472--537
														5	AAC96646	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662	55.07	4.86E-06	53%	68%	6--60	516--578
N639R	247301--247579	93	10,840	4.22	1	COG3966	DltD, Protein involved in D-alanine esterification of lipoteichoic acid and wall teichoic acid (D-alanine transfer protein) [Cell envelope biogenesis, outer membrane].	26.45	7.28E-01	26%	52%	28--89	258--324	1	AAC96820	A452L	49.29	4.55E-05	38%	55%	11--91	3--79
N640R	247617--248267	217	24,309	9.28		No Hit Found								1	AAC96849	A482R	208.38	1.36E-52	46%	66%	3--214	1--213
														2	AAR26889	FirrV-1-B14	65.47	1.42E-09	25%	41%	25--190	20--216
														3	AAR26970	FirrV-1-I5	64.31	3.17E-09	30%	47%	33--130	31--146
														4	YP_294196	hypothetical protein EhV_438	63.93	4.14E-09	34%	50%	33--130	73--168
														5	NP_077581	EsV-1-96	55.45	1.47E-06	31%	46%	37--130	70--181
														6	ABF82056	hypothetical protein MIV026R	43.13	7.56E-03	28%	56%	61--138	128--204
N642L	248765--248256	170	19,947	7.55		No Hit Found								1	AAC96851	A484L	160.61	1.77E-38	51%	68%	14--167	1--154
N643R	248772--250463	564	61,365	10.37	1	cd01820	PAF_acetylesterase_like, PAF_acetylhydrolase (PAF-AH)_like subfamily of SGNH-hydrolases. Platelet-activating factor (PAF) and PAF-AH are key players in inflammation and in atherosclerosis. PAF-AH is a calcium independent phospholipase A2 which exhibits strong substrate specificity towards PAF, hydrolyzing an acetyl ester at the sn-2 position. PAF-AH also degrades a family of oxidized PAF-like phospholipids with short sn-2 residues. In addition, PAF and PAF-AH are associated with neural migration and mammalian reproduction	66.43	7.92E-12	27%	42%	254--412	35--210	1	AAC96508	PBCV-1 surface protein	58.92	6.03E-07	43%	62%	499--560	967--1028
					2	cd01828	sialate_O-acetylesterase_like2, sialate_O-acetylesterase_like subfamily of the SGNH-hydrolases, , a diverse family of lipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alpha/beta hydrolase family and unique among all known hydrolases; its active site closely resembles the Ser-His-Asp(Glu) triad found in other serine hydrolases	59.53	1.04E-09	24%	42%	254--393	2--155	2	BAD22850	surface protein	58.92	6.03E-07	43%	62%	499--560	967--1028
					3	cd01833	XynB_like, SGNH_hydrolase subfamily, similar to Ruminococcus flavefaciens XynB. Most likely a secreted hydrolase with xylanase activity. SGNH hydrolases are a diverse family of lipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alpha/beta hydrolase family and unique among all known hydrolases; its active site closely resembles the Ser-His-Asp(Glu) triad found in other serine hydrolases	47.64	3.59E-06	18%	29%	254--399	3--190	3	BAD12236	surface protein	59.69	3.54E-07	43%	62%	499--560	967--1028
					4	cd00229	SGNH_hydrolase, SGNH_hydrolase, or GDSL_hydrolase, is a diverse family of lipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alpha/beta hydrolase family and unique among all known hydrolases; its active site closely resembles the typical Ser-His-Asp(Glu) triad from other serine hydrolases, but may lack the carboxylic acid	44.06	5.21E-05	19%	31%	253--400	1--192	4	T17636	proline-rich protein A145R - Chlorella virus PBCV-1	58.92	6.03E-07	43%	62%	499--560	176--237
					5	cd01838	Isoamyl_acetate_hydrolase_like, Isoamyl-acetate hydrolyzing esterase-like proteins. SGNH_hydrolase subfamily similar to the Saccharomyces cerevisiae IAH1. IAH1 may be the major esterase that hydrolyses isoamyl acetate in sake mash. The SGNH-family of hydrolases is a diverse family of lipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alpha/beta hydrolase family and unique among all known hydrolases; its active site closely resembles the Ser-His-Asp(Glu) triad found in other serine hydrolases.	40.24	6.48E-04	20%	36%	254--408	2--198	5	BAE02830	surface protein	48.52	8.15E-04	32%	63%	503--560	1014--1071

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	cd01841	NnaC_like, NnaC (CMP-NeuNAc synthetase) _like subfamily of SGNH_hydrolases, a diverse family of lipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alpha/beta hydrolase family and unique among all known hydrolases; its active site closely resembles two of the three components of typical Ser-His-Asp(Glu) triad from other serine hydrolases. E. coli NnaC appears to be involved in polysaccharide synthesis	40.25	7.22E-04	28%	46%	304--400	47--160	6	AAH59352	MGC69176 protein	51.99	7.37E-05	27%	44%	254--414	42--227
					7	cd01835	SGNH_hydrolase_like_3, SGNH_hydrolase subfamily. SGNH hydrolases are a diverse family of lipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alpha/beta hydrolase family and unique among all known hydrolases; its active site closely resembles the Ser-His-Asp(Glu) triad found in other serine hydrolases	39.18	1.42E-03	20%	32%	254--393	2--185	7	AAH84968	LOC495446 protein	48.52	8.15E-04	28%	46%	254--400	42--205
					8	cd01822	Lysophospholipase_L1_like, Lysophospholipase L1-like subgroup of SGNH-hydrolases. The best characterized member in this family is TesA, an E. coli periplasmic protein with thioesterase, esterase, arylesterase, protease and lysophospholipase activity..	35.11	2.15E-02	21%	38%	252--408	1--176	8	BAC56453	similar to platelet-activating factor acetylhydrolase, beta subunit	47.75	1.39E-03	31%	51%	254--354	42--145
					9	cd01832	SGNH_hydrolase_like_1, Members of the SGNH-hydrolase superfamily, a diverse family of lipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alpha/beta hydrolase family and unique among all known hydrolases; its active site closely resembles the Ser-His-Asp(Glu) triad from other serine hydrolases, but may lack the carboxlic acid. Myxobacterial members of this subfamily have been reported to be involved in adventurous gliding motility..	34.86	2.57E-02	24%	35%	258--407	6--185	9	ZP_00524236	Lipolytic enzyme, G-D-S-L	45.82	5.28E-03	28%	44%	252--398	511--673
					10	cd01834	SGNH_hydrolase_like_2, SGNH_hydrolase subfamily. SGNH hydrolases are a diverse family of lipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alpha/beta hydrolase family and unique among all known hydrolases; its active site closely resembles the Ser-His-Asp(Glu) triad found in other serine hydrolases	34.92	3.06E-02	20%	35%	254--369	2--134	10	NP_071782	platelet-activating factor acetylhydrolase alpha 2 subunit	45.05	9.01E-03	30%	51%	254--351	42--142
N649R	250735--251490	252	28,277	8.34	1	cd01196	INT_VanD, VanD integrase, IntD, and related integrases, DNA breaking-rejoining enzymes, integrase/recombinases, N- and C-terminal domains. This CD includes various bacterial integrases including those similar to IntD, a putative integrase-like protein, a component of the vanD glycopeptide resistance cluster in Enterococcus faecium BM4339. Members of this CD are predominantly bacterial in origin..	28.79	6.49E-01	38%	62%	147--173	182--208	1	AAC96377	A9R	248.83	1.23E-64	67%	84%	79--249	3--173
					2	pfam01209	Ubie_methyltran, ubiE/COQ5 methyltransferase family..	28.74	7.61E-01	34%	55%	144--175	171--200	2	AAC96643	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	93.59	6.61E-18	29%	48%	3--247	5--248
														3	AAC96447	A79R	90.12	7.31E-17	28%	47%	5--226	4--217
														4	AAC96818	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	88.20	2.78E-16	28%	47%	5--248	4--246
														5	AAC96972	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	85.50	1.80E-15	29%	46%	5--248	10--253
														6	AAC96545	A177R	85.11	2.35E-15	27%	49%	3--233	5--227
														7	AAU06304	hypothetical protein A275R	63.93	5.61E-09	28%	49%	83--247	3--167
														8	AAU06301	hypothetical protein A275R	62.77	1.25E-08	28%	49%	83--247	3--167
														9	AAC96734	A366L	55.07	2.60E-06	26%	47%	3--144	10--155
														10	AAU06302	hypothetical protein A275R	51.99	2.20E-05	28%	50%	116--248	1--132
N651L	252530--251481	350	39,793	4.54	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	283.29	2.96E-77	49%	64%	43--315	4--284	1	AAC96843	contains ribonucleotide reductase (RR) signature; similar to tobacco RR small subunit, corresponds to Swiss-Prot Accession Number P49730	506.52	5.66E-142	75%	84%	27--348	3--324
					2	pfam00268	Ribonuc red sm, Ribonucleotide reductase, small chain..	274.83	1.00E-74	44%	63%	29--311	1--285	2	NP_189342	ribonucleoside-diphosphate reductase	369.01	1.41E-100	54%	72%	23--348	1--332
					3	COG0208	NrdF, Ribonucleotide reductase, beta subunit [Nucleotide transport and metabolism]	226.71	2.66E-60	32%	48%	43--348	30--348	3	AAO62422	ribonucleotide reductase small subunit	366.31	9.15E-100	53%	71%	23--348	1--333
					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 (AerCbS), the ferritin-like d	48.56	1.04E-06	19%	29%	91--237	2--140	4	CAF24073	probable ribonucleoside-diphosphate reductase small chain	363.23	7.74E-99	56%	73%	29--348	5--323
					5	COG3933	COG3933, Transcriptional antiterminator [Transcription].	29.08	7.65E-01	21%	38%	64--135	405--468	5	BAD46317	putative ribonucleotide reductase R2	355.91	1.24E-96	53%	70%	28--348	14--339
														6	XP_550581	putative ribonucleotide reductase R2	354.76	2.75E-96	53%	71%	28--348	20--345
														7	CAA63194	ribonucleotide reductase R2	354.76	2.75E-96	54%	69%	23--348	1--329
														8	CAA71741	ribonucleotide reductase (Class I)	354.76	2.75E-96	54%	70%	28--348	16--337
														9	XP_829341	ribonucleoside-diphosphate reductase small chain	353.21	8.01E-96	54%	70%	28--348	16--337
														10	O15910	Ribonucleoside-diphosphate reductase small chain (Ribonucleotide reductase small subunit) (Ribonucleotide reductase R2 subunit)	353.21	8.01E-96	54%	70%	28--348	16--337
N655L	253103--252582	174	20,085	9.95		No Hit Found								1	AAC96838	A471R	207.99	1.02E-52	52%	77%	2--172	3--173
														2	YP_142861	unknown	108.23	1.10E-22	38%	62%	16--155	43--191
														3	ZP_01257639	hypothetical protein P700755_31322	89.35	5.31E-17	34%	57%	49--172	1--135
														4	AAR26829	FirrV-1-A5	87.81	1.55E-16	31%	61%	2--140	14--153
														5	NP_077626	EsV-1-141	71.63	1.15E-11	31%	59%	3--119	20--137
N658L	253738--253139	200	22,208	4.57		No Hit Found								1	AAC96837	A470R	239.97	3.55E-62	59%	72%	1--197	1--202
														2	BAB69883	UKCH-2	90.89	2.66E-17	39%	58%	6--127	49--170
N662L	255068--253776	431	48,157	8.69	1	COG5308	NUP170, Nuclear pore complex subunit [Intracellular trafficking and secretion].	31.19	2.77E-01	23%	41%	34--117	1129--1216	1	AAC96835	A468R	429.48	1.19E-118	54%	72%	3--377	4--393
														2	AAR26870	FirrV-1-A46	47.75	9.73E-04	33%	56%	308--372	490--552
														3	NP_077530	EsV-1-45	51.99	5.16E-05	22%	40%	180--387	353--542
N665R	255164--256090	309	35,071	6.79		No Hit Found								1	AAC96834	A467L	397.51	3.07E-109	62%	80%	4--300	6--302
														2	NP_065022	Hypothetical protein	65.08	3.61E-09	29%	46%	48--221	17--218
														3	AAP99966	ABC-type antimicrobial peptide transport system permease component	43.90	8.61E-03	23%	40%	44--243	80--271
N668R	256117--256470	118	13,165	10.47	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	78.03	2.81E-16	40%	52%	18--110	2--90	1	AAC96832	PBCV-1 thiol oxidoreductase	177.95	8.60E-44	67%	80%	2--116	4--118
					2	COG5054	ERV1, Mitochondrial sulfhydryl oxidase involved in the biogenesis of cytosolic Fe/S proteins [Posttranslational modification, protein turnover, chaperones]	55.83	1.39E-09	30%	45%	1--110	70--175	2	ZP_01257237	hypothetical protein P700755_26302	77.41	1.58E-13	36%	56%	11--97	33--119
														3	YP_142722	putative thiol oxidoreductase	67.40	1.64E-10	36%	51%	10--97	3--90

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	NP_078699	Thiol oxidoreductase	62.00	6.88E-09	33%	55%	2--99	21--115
														5	XP_503294	hypothetical protein	62.00	6.88E-09	33%	50%	9--111	88--186
														6	XP_451806	unnamed protein product	60.08	2.61E-08	31%	50%	14--111	85--178
														7	CAA48192	ERV1	58.54	7.60E-08	31%	48%	14--110	19--111
														8	NP_011543	Flavin-linked sulfhydryl oxidase localized to the mitochondrial intermembrane space, has a role in the maturation of cytosolic iron-sulfur proteins; ortholog of human hepatopoietin (ALR); Erv1p	58.54	7.60E-08	31%	48%	14--110	91--183
														9	AAL98767	ORF043L	56.61	2.89E-07	34%	52%	7--99	5--96
														10	AAX82354	thiol oxidoreductase	56.23	3.77E-07	34%	52%	7--99	5--96
N670L	257268--256465	268	29,615	6.63	1	COG0571	Rnc, dsRNA-specific ribonuclease [Transcription].	172.76	3.27E-44	35%	51%	38--263	4--234	1	AAC96831	similar to Bacillus ribonuclease III, corresponds to Swiss-Prot Accession Number P51833	348.59	1.27E-94	66%	81%	16--265	25--274
					2	smart00535	RIBOc, Ribonuclease III family; .	130.75	1.46E-31	42%	58%	58--187	2--130	2	YP_445467	ribonuclease III	133.65	6.43E-30	35%	53%	33--258	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	127.32	1.54E-30	48%	62%	70--181	16--127	3	ZP_00590199	Ribonuclease III	128.26	2.70E-28	36%	54%	45--255	47--265
					4	pfam00636	Ribonuclease 3, RNase3 domain..	123.30	2.35E-29	56%	70%	76--166	1--91	4	ZP_00532592	Ribonuclease III	127.49	4.61E-28	33%	53%	16--255	21--259
					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	58.84	6.97E-10	31%	52%	192--258	1--68	5	ABB23018	Ribonuclease III	122.87	1.14E-26	35%	55%	58--255	47--252
					6	COG1939	COG1939, Uncharacterized protein conserved in bacteria [Function unknown].	34.87	9.96E-03	29%	46%	78--163	18--112	6	ZP_00591208	Ribonuclease III	122.48	1.48E-26	34%	52%	45--256	44--266
					7	pfam05948	DUF880, Protein of unknown function (DUF880). This family consists of a number of hypothetical bacterial and plant proteins. The family also contains the C terminal region of a Cysteinyl-tRNA synthetase from Staphylococcus epidermidis. The function of this family is unknown..	35.17	9.97E-03	26%	43%	78--167	11--111	7	ZP_00528534	Ribonuclease III	122.48	1.48E-26	33%	54%	41--255	58--281
					8	COG5221	DOP1, Dopey and related predicted leucine zipper transcription factors [Transcription]	34.23	1.72E-02	20%	36%	58--150	556--652	8	ZP_00511103	Ribonuclease III	121.32	3.30E-26	34%	54%	45--257	42--264
														9	ZP_00661649	Ribonuclease III	121.32	3.30E-26	32%	50%	35--257	19--254
														10	AAM73335	ribonuclease III	119.78	9.62E-26	32%	52%	30--262	15--260
N672L	257494--257279	72	7,981	9.71	1	COG2433	COG2433, Uncharacterized conserved protein [Function unknown].	28.34	2.34E-01	18%	55%	15--59	421--465	0	No Hit Found	No Hit Found						
N674R	257663--259627	655	75,058	5.23	1	COG3378	COG3378, Predicted ATPase [General function prediction only].	70.01	7.92E-13	24%	37%	281--593	146--445	1	AAC96824	contains ATP/GTP-binding site motif A	871.69	0.00E+00	65%	81%	20--648	22--650
					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..	63.73	6.04E-11	22%	39%	290--614	3--313	2	NP_077594	EsV-1-109	278.49	5.80E-73	35%	54%	190--607	141--565
														3	AAR26902	FirrV-1-B27	268.09	7.83E-70	29%	51%	35--591	6--547
														4	YP_294217	putative nucleic acid independent nucleoside triphosphatase	166.39	3.21E-39	25%	44%	187--641	168--672
														5	ZP_00123428	COG3378: Predicted ATPase	65.86	5.90E-09	28%	48%	338--523	310--495
														6	BAE05402	putative DNA primase-phage associated	61.62	1.11E-07	24%	42%	292--566	107--366
														7	AAF27348	phage phi-R73 primase-like protein	61.23	1.45E-07	24%	43%	265--523	213--458
														8	ZP_00503756	Phage/plasmid primase P4, C-terminal	60.46	2.48E-07	22%	42%	232--506	241--499
														9	YP_584804	Phage/plasmid primase P4-like protein	58.54	9.43E-07	23%	42%	321--613	307--592
														10	ZP_00111378	COG3378: Predicted ATPase	58.15	1.23E-06	30%	53%	376--517	439--581
N679R	259658--260536	293	31,095	5.47		No Hit Found								1	AAC96822	A454L	352.44	1.03E-95	64%	77%	20--290	20--288
N682L	261298--260618	227	26,441	7.60	1	pfam01108	Tissue fac, Tissue factor..	28.90	5.70E-01	25%	46%	144--223	89--176	1	AAC96923	A568L	71.25	2.88E-11	26%	53%	41--194	29--173
N684L	261677--261282	132	14,749	7.24		No Hit Found								1	AAC96924	A570L	175.25	5.46E-43	64%	74%	1--128	1--121
N685R	261742--262269	176	19,819	7.15	1	COG5098	COG5098, Chromosome condensation complex Condensin, subunit D2 [Chromatin structure and dynamics / Cell division and chromosome partitioning]	30.44	1.17E-01	21%	44%	12--135	333--444	1	AAC96926	A572R	243.82	1.75E-63	65%	85%	4--163	11--170
N687L	263042--262260	261	28,789	4.23	1	cd00577	PCNA, Proliferating Cell Nuclear Antigen (PCNA) domain found in eukaryotes and archaea. These polymerase processivity factors play a role in DNA replication and repair. PCNA encircles duplex DNA in its central cavity, providing a DNA-bound platform for the attachment of the polymerase. The trimeric PCNA ring is structurally similar to the dimeric ring formed by the DNA polymerase processivity factors in bacteria (beta subunit DNA polymerase III holoenzyme) and in bacteriophages (catalytic subunits in T4 and RB69). This structural correspondence further substantiates the mechanistic connection between eukaryotic and prokaryotic DNA replication that has been suggested on biochemical grounds. PCNA is also involved with proteins involved in cell cycle processes such as DNA repair and apoptosis. Many of these proteins contain a highly conserved motif known as the PIP-box (PCNA interacting protein box) which contains the sequence Qxx[LIM]xxF[FY]. .	142.72	3.31E-35	27%	51%	21--251	5--247	1	AAC96927	similar to Periwinkle PCNA, corresponds to GenBank Accession Number X55052	231.49	2.18E-59	44%	67%	4--259	4--264
					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..	65.73	4.87E-12	28%	49%	135--250	3--128	2	CAE67843	Hypothetical protein CBG13430	125.56	1.68E-27	29%	50%	15--257	2--260
					3	COG0592	DnaN, DNA polymerase sliding clamp subunit (PCNA homolog) [DNA replication, recombination, and repair]	59.93	2.92E-10	18%	37%	1--251	47--323	3	EAR82594	proliferating cell nuclear antigen (pcna)	117.86	3.51E-25	28%	48%	15--255	2--259
					4	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..	49.88	3.27E-07	25%	55%	25--112	12--100	4	CAA38636	proliferating cell nuclear antigen	117.09	5.99E-25	30%	49%	20--257	7--260
					5	pfam04139	Rad9, Rad9. Rad9 is required for transient cell-cycle arrests and transcriptional induction of DNA repair in response to DNA damage	34.29	1.67E-02	22%	47%	15--73	3--61	5	AAC48257	Pcna (proliferating cell nuclear antigen) homolog protein 1	117.09	5.99E-25	30%	51%	48--256	1--226
					6	COG1355	COG1355, Predicted dioxygenase [General function prediction only].	30.25	2.93E-01	17%	44%	21--153	110--227	6	XP_502661	hypothetical protein	115.16	2.28E-24	25%	52%	25--258	12--260
														7	Q9MAY3	Proliferating cell nuclear antigen (PCNA)	115.16	2.28E-24	29%	50%	25--257	12--261
														8	XP_743255	proliferating cell nuclear antigen	112.85	1.13E-23	29%	51%	21--251	7--261
														9	CAA38893	proliferating cell nuclear antigen	112.46	1.48E-23	30%	49%	25--258	12--262
														10	AAG24908	proliferating cell nuclear antigen	112.08	1.93E-23	30%	49%	25--257	12--261
N689L	263595--263089	169	19,586	10.14	1	COG1052	LdhA, Lactate dehydrogenase and related dehydrogenases [Energy production and conversion / Coenzyme metabolism / General function prediction only]	29.44	2.57E-01	22%	41%	97--148	93--147	1	AAC96928	A575L	112.46	5.42E-24	35%	66%	31--166	30--167
N690R	263679--264704	342	37,532	8.60	1	smart00494	ChtBD2, Chitin-binding domain type 2; .	38.96	8.41E-04	40%	50%	278--321	9--49	1	AAC96701	PLPRNLLL (4X), SPPPSKP (3X)	343.58	6.11E-93	70%	80%	7--220	1--213
					2	pfam03067	Chitin_bind_3, Chitin binding domain. This domain is found associated with a wide variety of cellulose binding domain. This domain however is a chitin binding domain. This domain is found in isolation in baculoviral spheroidins and spindolins, protein of unknown function..	38.43	1.20E-03	22%	29%	18--216	1--211	2	AAC96700	a332L	88.20	4.63E-16	74%	90%	280--322	1--43
					3	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	4.88E-03	41%	62%	290--321	18--47	3	EAA01148	ENSANGP00000018413	74.71	5.30E-12	27%	41%	25--217	4--202
					4	COG3397	COG3397, Uncharacterized protein conserved in bacteria [Function unknown]	35.41	1.19E-02	22%	33%	11--222	8--213	4	XP_966436	PREDICTED: similar to CG4367-PA isoform 1	72.79	2.01E-11	33%	41%	15--217	4--209

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	EAA06469	ENSANGP00000012390	69.71	1.70E-10	27%	42%	33–218	1–196
														6	XP_973033	PREDICTED: similar to CG15786-PA	68.55	3.80E-10	27%	40%	23–221	23–231
														7	EAL29006	GA18133-PA	68.55	3.80E-10	29%	44%	18–216	6–210
														8	XP_968569	PREDICTED: similar to CG4367-PA	67.01	1.10E-09	26%	38%	30–217	22–216
														9	AAM50982	RE24790p	66.63	1.44E-09	29%	42%	20–218	38–249
														10	NP_572209	CG15786-PA	66.63	1.44E-09	29%	42%	20–218	19–230
N693R	264791–265324	178	19,702	11.79	1	COG0365	Acs, Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases [Lipid metabolism].	27.95	6.76E-01	41%	52%	70–102	499–528	1	AAC96888	contains Gln-rich, neutral zinc metallopeptidase, zinc binding region signature	211.85	7.65E-54	62%	74%	1–175	212–391
N696R	265432–266775	448	50,085	5.31		No Hit Found								1	AAC96872	A505L	449.90	8.99E-125	56%	71%	48–443	73–478
N703R	266856–267593	246	27,727	5.85		No Hit Found								1	AAC96870	contains prokaryotic membrane lipoprotein lipid attachment site	184.11	3.54E-45	37%	63%	6–243	38–297
N705R	267553–267879	109	12,706	7.77	1	pfam00742	Homoserine_dh, Homoserine dehydrogenase..	27.13	4.81E-01	27%	37%	43–103	76–143	1	AAC96869	A502L	105.53	5.42E-22	55%	65%	21–107	3–95
N706R	267948–269024	359	40,925	9.95		No Hit Found								1	AAC96722	A354R	214.54	4.63E-54	45%	63%	117–356	2–235
														2	AAC96790	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	76.26	1.97E-12	29%	46%	67–324	59–321
														3	AAC96455	A87R	61.62	5.02E-08	33%	46%	197–353	289–447
N710R	269070–270233	388	44,149	6.82	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..	63.56	3.94E-11	29%	54%	28–151	6–125	1	AAC96943	contains 3 ankyrin repeat-like elements; similar to Drosophila ankyrin, corresponds to GenBank Accession Number L35601	199.90	1.32E-49	46%	64%	4–219	3–217
					2	COG0666	Arp, FOG: Ankyrin repeat [General function prediction only].	39.49	7.80E-04	17%	36%	35–218	44–234	2	AAC96944	A608R	86.66	1.63E-15	33%	52%	238–368	1–134
					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..	31.19	2.23E-01	58%	89%	131–151	2–21	3	XP_786577	PREDICTED: similar to CG6599-PA	60.08	1.63E-07	24%	44%	1–257	134–368
														4	XP_967612	PREDICTED: similar to ankyrin repeat domain 28	46.21	2.44E-03	25%	43%	67–257	224–445
														5	XP_628948	BTB/POZ domain-containing protein	55.84	3.08E-06	24%	42%	33–264	39–269
														6	XP_650635	ankyrin repeat protein	46.21	2.44E-03	24%	40%	33–257	36–288
														7	NP_039208	ORF FPV245 Ankyrin repeat gene family protein	54.68	6.86E-06	23%	42%	48–285	50–323
														8	ZP_01301013	hypothetical protein RqryI_01000336	54.68	6.86E-06	29%	46%	66–217	433–568
														9	XP_547958	PREDICTED: similar to ankyrin repeat and SOCS box-containing protein 2 isoform 1	53.91	1.17E-05	31%	46%	29–176	185–329
														10	XP_854471	PREDICTED: similar to ankyrin repeat and SOCS box-containing protein 2 (predicted) isoform 2	53.91	1.17E-05	31%	46%	29–176	233–377
N712L	271396–270224	391	43,171	5.86	1	COG1004	Ugd, Predicted UDP-glucose 6-dehydrogenase [Cell envelope biogenesis, outer membrane].	353.34	2.22E-98	36%	55%	4–375	2–392	1	AAC96945	PBCV-1 UDP-glucose dehydrogenase	463.00	8.46E-129	58%	77%	5–386	4–386
					2	COG0677	WecC, UDP-N-acetyl-D-mannosaminuronate dehydrogenase [Cell envelope biogenesis, outer membrane].	125.73	8.51E-30	28%	44%	1–364	8–393	2	AAK02860	unknown	418.70	1.83E-115	53%	73%	3–389	2–389
					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.	121.49	1.47E-28	32%	48%	3–176	1–186	3	AAC67251	UDP-glucose dehydrogenase	417.54	4.08E-115	53%	73%	3–389	2–389
					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..	91.02	2.26E-19	39%	55%	193–285	2–96	4	AAK17922	UDP-glucose dehydrogenase FcbC	410.22	6.51E-113	52%	73%	3–389	2–389
					5	COG1893	ApbA, Ketopantoate reductase [Coenzyme metabolism].	46.08	7.36E-06	15%	37%	4–249	2–233	5	AAK17904	UDP-glucose dehydrogenase DcbC	405.60	1.60E-111	53%	71%	3–389	2–389
					6	COG0240	GpsA, Glycerol-3-phosphate dehydrogenase [Energy production and conversion].	44.77	2.15E-05	32%	57%	3–87	2–83	6	CAG21035	putative UDP-glucose dehydrogenase	404.06	4.67E-111	51%	73%	5–389	3–388
					7	COG2084	MmsB, 3-hydroxyisobutyrate dehydrogenase and related beta-hydroxyacid dehydrogenases [Lipid metabolism].	41.36	2.14E-04	28%	45%	3–232	1–204	7	ZP_00851874	UDP-glucose 6-dehydrogenase	400.21	6.74E-110	51%	71%	5–389	3–388
					8	COG1250	FadB, 3-hydroxyacyl-CoA dehydrogenase [Lipid metabolism].	38.70	1.20E-03	24%	46%	1–88	2–95	8	ZP_01235021	Putative nucleotide sugar dehydrogenase	399.05	1.50E-109	50%	71%	5–389	3–388
					9	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.	37.96	2.16E-03	17%	38%	302–383	1–87	9	ZP_00761397	COG1004: Predicted UDP-glucose 6-dehydrogenase	397.51	4.37E-109	50%	71%	5–389	3–388
					10	COG0654	UbiH, 2-polyprenyl-6-methoxyphenol hydroxylase and related FAD-dependent oxidoreductases [Coenzyme metabolism / Energy production and conversion].	36.98	3.89E-03	39%	68%	1–31	1–32	10	ZP_00880860	UDP-glucose 6-dehydrogenase	396.36	9.73E-109	51%	71%	5–389	3–388
N715L	273341–271449	631	72,186	6.93	1	COG1215	COG1215, Glycosyltransferases, probably involved in cell wall biogenesis [Cell envelope biogenesis, outer membrane].	110.79	4.76E-25	19%	37%	81–472	14–402	1	AAC96590	similar to cellulose synthase catalytic subunit (UDP-forming)	892.11	0.00E+00	68%	81%	23–629	68–677
					2	pfam03552	Cellulose_synt, Cellulose synthase. Cellulose, an aggregate of unbranched polymers of beta-1,4-linked glucose residues, is the major component of wood and thus paper, and is synthesised by plants, most algae, some bacteria and fungi, and even some animals. The genes that synthesise cellulose in higher plants differ greatly from the well-characterized genes found in Acetobacter and Agrobacterium sp. More correctly designated as 'cellulose synthase catalytic subunits', plant cellulose synthase (CesA) proteins are integral membrane proteins, approximately 1,000 amino acids in length. There are a number of highly conserved residues, including several motifs shown to be necessary for processive glycosyltransferase activity.	41.05	4.22E-04	24%	43%	287–492	499–698	2	AAL44127	cellulose synthase	305.06	5.55E-81	41%	57%	107–499	91–478
					3	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..	40.20	8.32E-04	15%	33%	131–298	10–168	3	YP_467623	probable cellulose synthase protein	291.58	6.35E-77	34%	49%	17–584	4–558
					4	pfam01604	7tm_5, 7TM chemoreceptor. This large family of proteins are related to pfam00001. They are 7 transmembrane receptors. This family does not include all known members, as there are problems with overlapping specificity with pfam00001. This family is greatly expanded in the nematode worm C. elegans.	30.90	4.80E-01	27%	44%	350–398	105–153	4	BAB54246	mlr7873	288.89	4.12E-76	38%	54%	118–556	112–546
														5	CAC48842	putative cellulose synthase protein	285.42	4.55E-75	38%	53%	115–551	97–528
														6	ZP_00958843	cellulose synthase-like protein	283.49	1.73E-74	36%	52%	118–582	102–561
														7	ZP_00913218	similar to cellulose synthase	279.64	2.50E-73	40%	56%	118–496	166–538
														8	AAQ87082	Cellulose synthase catalytic subunit	278.87	4.26E-73	37%	52%	118–560	102–539
														9	AAX76085	glycosyl transferase, group 2 family protein	277.33	1.24E-72	37%	53%	115–548	93–516
														10	ABA79331	cellulose synthase-like protein	276.56	2.11E-72	35%	52%	118–584	110–568
N719L	273697–273335	121	13,025	10.35	1	smart00317	SET, SET (Su(var)3-9, Enhancer-of-zeste, Trithorax) domain; Putative methyl transferase. based on outlier plant homologues .	59.64	9.59E-11	29%	38%	6–113	2–125	1	AAC96946	PBCV-1 histone H3-Lys 27 methyltransferase (vSET)	140.20	1.96E-32	57%	73%	2–114	1–113
					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..	47.37	4.54E-07	23%	38%	6–112	8–132	2	ZP_00588496	Nuclear protein SET	65.86	4.71E-10	34%	52%	7–114	39–150
					3	COG2940	COG2940, Proteins containing SET domain [General function prediction only].	41.63	2.25E-05	22%	34%	1–110	328–452	3	ZP_00511449	Nuclear protein SET	60.85	1.51E-08	35%	50%	7–114	38–149
					4	cd01187	INT_SG4, INT_SG4, DNA breaking-rejoining enzymes, integrase/recombinases subgroup 4, N- and C-terminal domains. The CD contains mainly predicted bacterial integrase/recombinases for which not much biochemical characterization is available..	27.17	6.02E-01	22%	33%	37–92	172–227	4	ZP_00661322	Nuclear protein SET	60.08	2.58E-08	35%	51%	6–117	37–152
														5	ZP_00531791	Nuclear protein SET	58.54	7.52E-08	31%	49%	7–116	39–152
														6	ABB23988	Nuclear protein SET	58.15	9.82E-08	35%	51%	6–114	38–150
														7	NP_701503	hypothetical protein PFL0690c	56.61	2.86E-07	30%	44%	3–111	29–174
														8	AAZ71510	hypothetical protein Mbar_A2600	55.84	4.87E-07	31%	51%	7–106	3–109
														9	AAM72187	conserved hypothetical protein	53.91	1.85E-06	31%	49%	7–117	39–153

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	NP_634869	hypothetical protein MM2845	53.53	2.42E-06	32%	54%	16--106	19--116
N720L	275415--273727	563	63,379	11.37	1	COG0515	SPS1, Serine/threonine protein kinase [General function prediction only / Signal transduction mechanisms / Transcription / DNA replication, recombination. and reoair].	43.61	7.23E-05	13%	28%	84--456	11--365	1	AAC96947	RPQT-like (9x)	460.69	6.86E-128	44%	62%	2--561	6--577
					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..	42.89	1.04E-04	30%	46%	53--238	6--151	2	AAC96646	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662	291.20	7.20E-77	39%	60%	1--401	9--424
					3	pfam00069	Pkinase, Protein kinase domain..	41.04	3.63E-04	28%	45%	53--236	5--148	3	AAC96650	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055	285.03	5.16E-75	36%	56%	1--453	23--465
					4	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	40.90	4.29E-04	41%	74%	207--233	171--198	4	XP_944436	PREDICTED: similar to CG10953-PA	62.39	5.44E-08	33%	55%	349--446	76--173
					5	COG2334	COG2334, Putative homoserine kinase type II (protein kinase fold) [General function prediction only]	39.94	7.47E-04	33%	52%	207--240	200--233	5	XP_932523	PREDICTED: similar to serine/arginine repetitive matrix 2	55.84	5.09E-06	35%	52%	375--456	97--178
					6	smart00220	S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosohotransferases. Serine or threonine-specific kinase subfamily. .	39.81	1.02E-03	25%	48%	53--238	5--150	6	XP_729588	dentin phosphoryn	47.75	1.39E-03	36%	51%	348--429	315--396
					7	COG0661	AarF, Predicted unusual protein kinase [General function prediction only]	38.39	2.48E-03	31%	59%	204--254	284--333	7	AAK54495	neurofilament triplet H1-like protein	57.00	2.29E-06	36%	50%	327--456	93--221
					8	smart00090	RIO, RIO-like kinase; .	38.29	2.62E-03	28%	45%	166--250	122--211	8	XP_742899	hypothetical protein PC108912.00.0	47.75	1.39E-03	28%	55%	362--431	42--111
					9	pfam01163	RIO1, RIO1 family. This family of proteins are related to eukaryotic type protein kinases..	37.50	3.98E-03	28%	50%	162--252	78--170	9	XP_427855	PREDICTED: similar to p87, partial	45.44	6.88E-03	29%	50%	375--446	86--157
					10	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	37.08	6.09E-03	38%	49%	209--252	147--192	10	AAC96402	similar to E. coli LPS core biosynthesis protein, corresponds to Swiss-Prot Accession Number P27240	50.06	2.79E-04	20%	38%	26--313	29--307
N724R	275496--276452	319	37,146	10.16		No Hit Found								1	AAC96948	similar to Variola virus orf E10L, corresponds to Swiss-Prot Accession Number P33801	284.65	3.03E-75	44%	65%	1--310	1--314
														2	YP_142754	S/T protein kinase, similar to Paramecium bursaria chlorella virus 1 A617R	58.15	4.61E-07	29%	51%	130--246	230--358
N725L	276823--276443	127	13,956	3.77		No Hit Found								1	AAC96949	A618L	75.87	4.60E-13	54%	79%	56--123	65--125
N727L	277600--276950	217	24,102	4.26	1	cd00829	SCP-x_thiolase, Thiolase domain associated with sterol carrier protein (SCP)-x isoform and related proteins; SCP-2 has multiple roles in intracellular lipid circulation and metabolism. The N-terminal presequence in the SCP-x isoform represents a peroxisomal 3-ketacyl-Coa thiolase specific for branched-chain acyl CoAs, which is proteolytically cleaved from the sterol carrier protein..	31.44	8.80E-02	31%	52%	118--172	7--61	1	AAC96950	A619L	51.60	2.13E-05	24%	38%	1--215	1--237
N731L	277945--277649	99	11,140	8.64		No Hit Found								1	AAC96951	similar to Synechocystis orf 90, corresponds to GenBank Accession Number D90902	64.31	1.39E-09	34%	50%	1--95	1--81
														2	AAC96962	A635R	50.45	2.07E-05	30%	51%	1--91	1--82
N732L	278380--278027	118	12,691	8.65		No Hit Found								1	AAC96952	A621L	130.18	2.06E-29	52%	71%	4--116	5--117
N733R	278451--281159	903	99,656	6.17	1	COG0488	Uup, ATPase components of ABC transporters with duplicated ATPase domains [General function prediction only]	280.22	6.30E-76	27%	45%	300--848	2--530	1	AAC96981	Chlorella virus CVK2 translation elongation factor-3 homolog, refer to GenBank Accession Number D16505	1106.66	0.00E+00	65%	77%	37--901	55--918
					2	cd03221	ABCF_EF-3, ABCF_EF-3 Elongation factor 3 (EF-3) is a cytosolic protein required by fungal ribosomes for in vitro protein synthesis and for in vivo growth. EF-3 stimulates the binding of the EF-1:GTP:aa-tRNA ternary complex to the ribosomal A site by facilitated release of the deacylated tRNA from the E site. The reaction requires ATP hydrolysis. EF-3 contains two ATP binding sequence (NBS) motifs. NBSI is sufficient for the intrinsic ATPase activity. NBSII is essential for the ribosome-stimulated functions	174.24	5.16E-44	39%	56%	302--492	1--191	2	A48779	translation elongation factor EF-3 homolog - Chlorella virus CVK2	1011.52	0.00E+00	61%	74%	37--885	272--1120
					3	pfam00005	ABC_tran, ABC transporter. ABC transporters for a large family of proteins responsible for translocation of a variety of compounds across biological membranes. ABC transporters are the largest family of proteins in many completely sequenced bacteria. ABC transporters are composed of two copies of this domain and two copies of a transmembrane domain pfam00664. These four domains may belong to a single polypeptide, or belong in different polypeptide chains	120.40	8.69E-28	30%	48%	327--491	1--182	3	XP_445123	unnamed protein product	757.67	0.00E+00	46%	62%	35--900	137--1043
					4	cd00267	ABC_ATPase, ABC (ATP-binding cassette) transporter nucleotide-binding domain; ABC transporters are a large family of proteins involved in the transport of a wide variety of different compounds, like sugars, ions, peptides and more complex organic molecules. The nucleotide binding domain shows the highest similarity between all members of the family. ABC transporters are a subset of nucleotide hydrolases that contain a signature motif, Q-loop, and H-loop/switch region in addition to the Walker A motif/P-loop andWalker B motif commonly found in a number of ATP- and GTP-binding and hydrolyzing proteins..	114.78	3.59E-26	29%	47%	306--491	4--207	4	BAA33959	translation elongation factor3	756.13	0.00E+00	46%	62%	35--900	137--1043
					5	cd03230	ABC_DR_subfamily_A, This family of ATP-binding proteins belongs to a multisubunit transporter involved in drug resistance (BcrA and DrrA), nodulation, lipid transport, and lantibiotic immunity. In bacteria and archaea, these transporters usually include an ATP-binding protein and one or two integral membrane proteins. Eukaryote systems of the ABCA subfamily display ABC domains that are quite similar to this family. The ATP-binding domain shows the highest similarity between all members of the ABC transporter family. ABC transporters are a subset of nucleotide hydrolases that contain a signature motif, Q-loop, and H-loop/switch region in addition to the Walker A motif/P-loop andWalker B motif commonly found in a number of ATP- and GTP-binding and hydrolyzing proteins	112.09	2.44E-25	29%	47%	311--488	10--205	5	CAA78282	translation elongation factor 3	753.82	0.00E+00	45%	62%	29--901	136--1050
					6	COG1131	CcmA, ABC-type multidrug transport system, ATPase component [Defense mechanisms].	103.13	1.29E-22	28%	48%	312--487	16--208	6	XP_711404	translation elongation factor 3	753.44	0.00E+00	45%	62%	29--901	136--1050
					7	cd03225	ABC_cobalt_transport_domain1, Domain I of the ATPase component of a cobalt transport family found in both bacteria and archaea. This ABC transporter subfamily is involved cobalt transport as part of the cobalamin biosynthetic pathway. Cobalamin is derived from uroporphyrinogen III, a precursor of heme, siroheme and chlorophylls, and a cobalt ion is chelated in the center of the molecules. The genes necessary for cobalamin production are organized into a single operon in S. typhimurium. In addition to genes known to encode enzymes catalyzing steps of the cobalamine biosynthetic pathway, the products of cbiQ, cbiO and cbiN, were proposed to constitute a cobalt uptake system since CbiN and CbiQ are integral membrane proteins and CbiO is an ABC ATPase. However, direct evidence supporting this idea is lacking..	102.50	1.87E-22	28%	46%	306--491	4--209	7	CAA22654	SPCC417.08	751.90	0.00E+00	45%	63%	35--898	142--1046
					8	COG1121	ZnuC, ABC-type Mn/Zn transport systems, ATPase component [Inorganic ion transport and metabolism].	101.06	5.09E-22	29%	50%	306--493	9--216	8	XP_711356	translation elongation factor 3	751.13	0.00E+00	45%	62%	29--901	136--1050

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	cd03214	ABC_Iron-Siderophores_B12_Hemin, ABC transporters involved in the uptake of siderophores, heme, and vitamin B12 are widely conserved in bacteria and archea. Only very few species lack representatives of the siderophore family transporters. The E. coli BtuCD protein is an ABC transporter mediating vitamin B12 uptake. The two ATP-binding cassettes (BtuD) are in close contact with each other, as are the two membrane-spanning subunits (BtuC); this arrangement is distinct from that observed for the E. coli lipid flippase MsbA. The BtuC subunits provide 20 transmembrane helices grouped around a translocation pathway that is closed to the cytoplasm by a gate region whereas the dimer arrangement of the BtuD subunits resembles the ATP-bound form of the Rad50 DNA repair enzyme. A prominent cytoplasmic loop of BtuC forms the contact region with the ATP-binding cassette and represent a conserved motif among the ABC transporters. .	100.96	6.39E-22	28%	48%	306--487	4--206	9	CAA77567	elongation factor 3	748.04	0.00E+00	45%	62%	29--901	136--1049
					10	cd03235	ABC_Metallic_Cations, ABC-type Mn/Zn transport systems, ATPase component; This family is comprised of systems involved in the uptake of various metallic cations such as iron, manganese and zinc. The ATPases of these systems are stronly related to those of iron-siderophore uptake systems suggesting that they rose from a common ancestor..	99.82	1.31E-21	28%	48%	306--488	4--204	10	XP_455632	unnamed protein product	747.66	0.00E+00	46%	62%	35--901	137--1044
N737L	282723--281155	523	57,890	5.46	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 PBCV-1 the major capsid protein is a glycoprotein	369.64	4.21E-103	42%	55%	173--517	78--443	1	AAC27493	putative capsid protein	987.25	0.00E+00	92%	93%	1--521	1--521
					2	COG0173	AspS, Aspartyl-tRNA synthetase [Translation, ribosomal structure and biogenesis]	30.52	4.64E-01	36%	59%	490--512	85--107	2	AAC96953	similar to Simulium iridescent virus capsid protein, corresponds to Swiss-Prot Accession Number P22166	745.73	0.00E+00	69%	79%	1--521	1--520
														3	BAE06835	hypothetical major capsid protein	116.70	2.20E-24	67%	84%	1--79	1--79
														4	BAA76600	major capsid protein	102.83	3.29E-20	65%	78%	1--76	1--76
														5	AAC27492	major capsid protein Vp49	103.61	1.93E-20	65%	78%	1--76	1--76
														6	BAA76601	major capsid protein MCP1	103.61	1.93E-20	65%	78%	1--76	1--76
														7	AAC96798	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	103.61	1.93E-20	65%	78%	1--76	1--76
														8	1M3Y_D	Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid Containing, Dna Virus	62.77	3.77E-08	57%	73%	25--76	1--52
														9	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	62.77	3.77E-08	57%	73%	25--76	1--52
														10	BAA22198	major capsid protein Vp54	103.61	1.93E-20	65%	78%	1--76	1--76
N741R	282995--283372	126	14,030	7.52	1	COG4852	COG4852, Predicted membrane protein [Function unknown].	38.35	2.87E-04	25%	39%	8--121	10--128	1	AAC96955	A624R	129.80	2.69E-29	50%	72%	5--117	8--120
														2	YP_485867	hypothetical protein RPB_2251	53.53	2.45E-06	31%	47%	6--108	9--115
														3	ZP_00234461	conserved hypothetical protein	51.60	9.32E-06	25%	44%	6--121	9--129
														4	YP_579998	hypothetical protein Pcryo_0731	50.45	2.08E-05	31%	47%	6--108	8--117
														5	CAC98668	lmo0589	50.06	2.71E-05	25%	43%	6--121	9--129
														6	AAZ18596	conserved hypothetical protein	49.68	3.54E-05	27%	45%	6--108	8--117
														7	CAC95830	lin0598	48.91	6.04E-05	24%	43%	8--121	11--129
														8	YP_013223	hypothetical protein LMOf2365_0618	48.52	7.89E-05	24%	43%	6--121	9--129
														9	AAU26313	membrane protein	46.98	2.30E-04	28%	47%	1--105	1--118
														10	CAH11414	hypothetical protein	46.98	2.30E-04	28%	47%	1--105	1--118
N742L	284255--283347	303	34,057	10.78		No Hit Found								1	AAC96459	a91L	63.93	7.77E-09	33%	53%	177--291	2--112
														2	AAC96461	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563	46.60	1.28E-03	26%	47%	8--118	142--259
N745R	284297--284851	185	20,162	3.49		No Hit Found								1	AAC96977	A659L	65.08	1.28E-09	39%	47%	1--107	1--92
N746R	284991--285428	146	16,290	6.80	1	pfam06738	DUF1212, Protein of unknown function (DUF1212). This family represents a conserved region within a number of hypothetical proteins of unknown function found in eukaryotes, bacteria and archaea. Some family members are membrane proteins..	33.27	1.39E-02	28%	49%	19--89	73--141	1	AAC96975	A656L	65.08	8.01E-10	37%	56%	9--92	8--100
N747R	285467--286063	199	21,772	5.85	1	pfam00583	Acetyltransf_1, Acetyltransferase (GNAT) family. This family contains proteins with N-acetyltransferase functions	45.24	5.71E-06	31%	52%	113--163	30--82	1	AAC96974	A654L	204.91	1.25E-51	51%	68%	1--194	1--194
					2	COG1247	COG1247, Sortase and related acyltransferases [Cell envelope biogenesis, outer membrane].	34.86	7.18E-03	35%	49%	115--162	89--138	2	EAO24365	GCN5-related N-acetyltransferase	63.93	3.44E-09	28%	46%	2--191	11--217
					3	COG3153	COG3153, Predicted acetyltransferase [General function prediction only].	34.47	9.00E-03	42%	62%	113--162	81--126	3	ZP_00675288	hypothetical protein TeryDRAFT_0585	62.39	1.00E-08	25%	45%	10--184	15--209
					4	COG0456	RimI, Acetyltransferases [General function prediction only].	34.60	9.23E-03	28%	46%	113--173	97--158	4	XP_392876	PREDICTED: similar to arylalkylamine N-acetyltransferase	51.99	1.35E-05	26%	43%	11--173	58--237
					5	COG1670	RimL, Acetyltransferases, including N-acetylases of ribosomal proteins [Translation, ribosomal structure and biogenesis].	34.17	1.25E-02	20%	41%	97--184	85--173	5	BAC87874	arylalkylamine N-acetyltransferase	50.45	3.93E-05	21%	40%	10--173	44--225
					6	COG2388	COG2388, Predicted acetyltransferase [General function prediction only].	32.97	2.62E-02	31%	66%	115--144	47--76	6	XP_394768	PREDICTED: similar to ENSANGP00000021991	49.29	8.76E-05	33%	54%	95--184	116--205
					7	COG1246	ArgA, N-acetylglutamate synthase and related acetyltransferases [Amino acid transport and metabolism].	30.61	1.35E-01	26%	47%	114--163	72--119	7	EAA05105	ENSANGP00000021991	47.37	3.33E-04	32%	59%	112--173	132--193
					8	COG3053	CitC, Citrate lyase synthetase [Energy production and conversion].	30.28	1.57E-01	33%	52%	114--169	63--117	8	ZP_01182660	hypothetical protein BcerKBAB4DRAFT_5424	46.98	4.35E-04	25%	41%	12--197	11--204
					9	COG4552	Eis, Predicted acetyltransferase involved in intracellular survival and related acetyltransferases [General function prediction only].	29.93	2.12E-01	20%	44%	114--183	77--143	9	CAE65892	Hypothetical protein CBG11057	45.82	9.68E-04	23%	42%	4--191	6--201
					10	COG1444	COG1444, Predicted P-loop ATPase fused to an acetyltransferase [General function prediction only].	29.92	2.41E-01	43%	67%	110--140	534--564	10	ZP_00740288	hypothetical protein RBTH_05448	45.82	9.68E-04	26%	42%	12--197	11--204
N748L	287110--286058	351	39,512	7.86	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	147.76	1.87E-36	35%	52%	1--187	1--183	1	AAV84097	CviPII m5C DNA methyltransferase	478.79	1.27E-133	65%	77%	1--349	16--363
					2	pfam00145	DNA_methylase, C-5 cytosine-specific DNA methylase..	142.76	5.07E-35	31%	46%	1--195	1--196	2	AAC96884	M.CviAII cytosine DNA methyltransferase	312.38	1.57E-83	45%	65%	1--343	2--342
					3	COG0270	Dcm, Site-specific DNA methylase [DNA replication, recombination, and repair].	95.53	8.19E-21	25%	43%	1--175	4--183	3	AAC64006	cytosine methyltransferase	298.13	3.07E-79	44%	60%	1--349	3--362
					4	pfam00398	RrnaAD, Ribosomal RNA adenine dimethylase..	30.21	3.95E-01	25%	37%	1--52	32--89	4	AAC55063	cytosine methyltransferase	279.64	1.13E-73	42%	57%	4--349	6--366
					5	cd00653	RNA_pol_B_RPB2, RNA polymerase beta subunit. RNA polymerases catalyse the DNA dependent polymerization of RNA. Prokaryotes contain a single RNA polymerase compared to three in eukaryotes (not including mitochondrial. and chloroplast polymerases). Each RNA polymerase complex contains two related members of this family, in each case they are the two largest subunits.The clamp is a mobile structure that grips DNA during elongation	29.54	6.87E-01	32%	38%	4--52	905--952	5	AAC96987	nonfunctional M.CviAV cytosine DNA methyltransferase	276.56	9.57E-73	41%	57%	4--349	6--366
														6	AAC96897	M.CviAIV cytosine DNA methyltransferase	266.93	7.59E-70	44%	58%	1--331	3--332
														7	NP_813725	gp9.1	96.29	1.77E-18	36%	51%	1--157	1--158
														8	AAR23218	gp67	88.20	4.83E-16	31%	48%	1--161	6--192
														9	BAB77127	site-specific DNA-methyltransferase	83.96	9.11E-15	34%	46%	4--161	5--165
														10	AAV83360	DNA cytosine methylase	82.42	2.65E-14	33%	50%	4--158	102--259
N751L	288505--287126	460	51,385	10.96		No Hit Found								1	AAV84098	CviPII top-strand DNA nicking endonuclease	217.62	7.73E-55	44%	59%	179--455	9--302
														2	AAC55064	restriction endonuclease	126.72	1.80E-27	35%	50%	173--419	25--256
N754R	288709--289167	153	18,318	4.37		No Hit Found								0	No Hit Found	No Hit Found						
N755L	289726--289202	175	19,326	9.68		No Hit Found								1	AAC96968	A644R	117.09	2.42E-25	72%	79%	89--172	2--84
														2	YP_142763	unknown	47.37	2.36E-04	26%	47%	45--165	53--176
N756L	291196--289751	482	54,577	11.33	1	COG4487	COG4487, Uncharacterized protein conserved in bacteria [Function unknown].	30.00	6.02E-01	20%	45%	230--315	90--177	1	AAC96967	Gln-rich; KQQ (6X)	201.45	6.12E-50	79%	93%	196--317	1--122
														2	AAC96966	A642R	58.92	4.91E-07	65%	80%	19--53	21--55

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	AAA36349	nonmuscle myosin heavy chain (NMHC)	44.67	9.58E-03	21%	44%	86–314	325–537
N759R	291268–291675	136	15,341	9.58		No Hit Found								1	AAC96961	A634L	168.32	6.78E-41	58%	78%	2–129	4–131
N762L	292048–291662	129	8	14674.85		No Hit Found								0	No Hit Found	No Hit Found						
N763R	292132–292551	140	15,655	4.20	1	COG5407	SEC63, Preprotein translocase subunit Sec63 [Intracellular trafficking and secretion]	27.35	7.54E-01	20%	49%	24–74	551–602	0	No Hit Found	No Hit Found						
N764L	292919–292560	120	13,206	9.29	1	pfam07340	Herpes_IE1, Cytomegalovirus IE1 protein. Expression from a human cytomegalovirus early promoter (E1.7) has been shown to be activated in trans by the IE2 gene product. Although the IE1 gene product alone had no effect on this early viral promoter, maximal early promoter activity was detected when both IE1 and IE2 gene products were present. The IE1 protein from cytomegalovirus is also known as UL123..	30.82	4.85E-02	31%	52%	17–76	272–336	1	AAC96960	A633R	157.53	1.19E-37	61%	80%	1–114	2–115
N766L	295247–292947	767	85,429	7.75	1	cd01679	RNR_1, RNR, class I. Ribonucleotide reductase (RNR) catalyzes the reductive synthesis of deoxyribonucleotides from their corresponding ribonucleotides. It provides the precursors necessary for DNA synthesis. RNRs are separated into three classes based on 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, as	736.63	0.00E+00	51%	68%	176–746	1–578	1	AAC96959	similar to Schizosaccharomyces ribonucleotide reductase M1 chain, corresponds to Swiss-Prot Accession Number P36602	1217.99	0.00E+00	78%	88%	9–765	14–771
					2	pfam02867	Ribonuc_red_lgC, Ribonucleotide reductase, barrel domain..	637.33	0.00E+00	48%	66%	224–748	1–532	2	NP_001026008	ribonucleoside-diphosphate reductase M1 chain	764.61	0.00E+00	50%	69%	9–759	1–754
					3	COG0209	NrdA, Ribonucleotide reductase, alpha subunit [Nucleotide transport and metabolism].	516.10	5.70E-147	37%	53%	59–764	1–696	3	AAH46846	RRM1 protein	757.67	0.00E+00	50%	69%	9–759	1–754
					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..	304.43	2.73E-83	34%	51%	196–743	2–521	4	AAH74185	RRM1 protein	756.13	0.00E+00	50%	69%	9–759	1–754
					5	pfam00317	Ribonuc_red_lgN, Ribonucleotide reductase, all-alpha domain..	68.35	3.45E-12	40%	57%	159–222	10–78	5	XP_968671	PREDICTED: similar to ribonucleotide reductase M1	748.81	0.00E+00	49%	67%	4–759	3–766
					6	cd00576	RNR_PFL, RNR_PFL. Ribonucleotide reductase (RNR) and pyruvate formate lyase (PFL) have a structurally similar ten-stranded alpha-beta barrel active site domain and are believed to have diverged from a common ancestor. RNRs are found in all organisms and provide the only mechanism by which nucleotides are converted to deoxynucleotides, while PFL, an essential enzyme in anaerobic bacteria, catalyzes the conversion of pyruvate and CoA to acteylCoA and formate. Both RNR and PFL are glycyl radical enzymes..	56.65	9.52E-09	21%	35%	249–678	70–461	6	AAD49743	ribonucleotide reductase large subunit	748.43	0.00E+00	49%	67%	9–759	1–755
					7	pfam03477	ATP-cone, ATP cone domain..	50.37	8.39E-07	25%	40%	9–110	1–89	7	XP_322797	ribonucleoside-diphosphate reductase large chain (un-24gene)	748.43	0.00E+00	49%	67%	9–759	1–755
														8	AAH85906	Ribonucleotide reductase M1	746.50	0.00E+00	49%	68%	9–759	1–754
														9	ABF71875	ribonucleotide reductase large subunit	746.12	0.00E+00	49%	66%	9–759	1–755
														10	AAH16450	Ribonucleotide reductase M1	745.35	0.00E+00	49%	68%	9–759	1–754
N773L	296759–295458	434	48,499	10.93	1	cd00016	alkPPc, Alkaline phosphatase homologues; alkaline phosphatases are non-specific phosphomonoesterases that catalyze the hydrolysis reaction via a phosphoseryl intermediate to produce inorganic phosphate and the corresponding alcohol, optimally at high pH. Alkaline phosphatase exists as a dimer, each monomer binding 2 zinc atoms and one magnesium atom, which are essential for enzymatic activity..	30.27	4.92E-01	28%	36%	116–174	67–125	1	AAC96957	similar to Chlorella virus PBCV-1 ORF A231L, corresponds to GenBank Accession Number U42580	465.31	1.97E-129	53%	72%	25–432	31–436
														2	AAC96599	contains ATP/GTP-binding motif A	296.98	9.30E-79	44%	62%	93–422	4–337
N777L	297109–296774	112	12,158	9.86	1	COG1293	COG1293, Predicted RNA-binding protein homologous to eukaryotic snRNP [Transcription].	57.30	4.62E-10	33%	53%	4–97	445–541	1	AAC96548	similar to Streptococcus pyogenes fibronectin protein, corresponds to GenBank Accession Number L28919	114.78	8.84E-25	52%	70%	1–106	1–105
					2	pfam05670	DUF814, Domain of unknown function (DUF814). This domain occurs in proteins that have been annotated as Fibronectin/fibrinogen binding protein by similarity. This annotation comes from a sequence, where the N-terminal region is involved in this activity. Hence the activity of this C-terminal domain is unknown. This domain contains a conserved motif D/E X-W/Y-X-H that may be functionally important	53.29	7.46E-09	39%	60%	4–87	3–90	2	ZP_00504390	Protein of unknown function DUF814:Fibronectin-binding A, N-terminal	59.69	3.37E-08	34%	50%	5–109	469–576
														3	AAD35450	fibronectin-binding protein, putative	56.23	3.73E-07	32%	53%	6–106	440–542
														4	CAA08863	putative fibronectin-binding protein	55.45	6.36E-07	32%	50%	5–106	439–542
														5	YP_004218	fibronectin/fibrinogen-binding protein	54.68	1.09E-06	31%	53%	8–97	405–496
														6	BAD70431	probable RNA-binding protein	54.68	1.09E-06	31%	53%	8–97	405–496
														7	CAB13438	yloA	54.30	1.42E-06	31%	49%	5–109	457–562
														8	ZP_01361871	conserved hypothetical protein	54.30	1.42E-06	29%	48%	1–109	470–581
														9	ZP_00886600	RNA-binding protein homologous to eukaryotic snRNP	53.53	2.42E-06	33%	49%	5–106	468–572
														10	AAK80069	Fibronectin-binding protein	53.14	3.16E-06	32%	51%	5–97	460–554
N779R	297218–299638	807	87,716	6.99	1	smart00637	CBD_II, CBD_II domain..	64.96	3.83E-11	28%	43%	12–93	4–85	1	BAA78554	vChti-1	1110.52	0.00E+00	66%	74%	1–802	1–833
					2	pfam00704	Glyco_hydro_18, Glycosyl hydrolases family 18..	57.07	8.98E-09	22%	35%	538–714	7–195	2	AAC96549	PBCV-1 chitinase	1098.19	0.00E+00	66%	74%	1–802	1–827
					3	COG3469	COG3469, Chitinase [Carbohydrate transport and metabolism].	34.63	5.39E-02	33%	46%	179–259	90–169	3	ZP_00570566	Cellulose-binding, bacterial type	141.35	1.41E-31	32%	51%	531–802	238–536
					4	cd00116	LRR_RI, Leucine-rich repeats (LRRs), ribonuclease inhibitor (RI)-like subfamily. LRRs are 20-29 residue sequence motifs present in many proteins that participate in protein-protein interactions and have different functions and cellular locations. LRRs correspond to structural units consisting of a beta strand (LxxLxLxxN/CxL conserved pattern) and an alpha helix. This alignment contains 12 strands corresponding to 11 full repeats, consistent with the extent observed in the subfamily acting as Ran GTPase Activating Proteins (RanGAP1)	32.18	2.56E-01	29%	47%	653–701	155–200	4	NP_627029	secreted sugar hydrolase	146.75	3.36E-33	35%	52%	531–783	197–456
														5	NP_826429	sugar hydrolase	147.52	1.97E-33	33%	54%	531–783	197–456
														6	ZP_01208817	Chitinase., Cellulase	122.48	6.79E-26	31%	47%	530–805	948–1240
														7	ZP_01191795	Glycoside hydrolase, family 18:Glycoside hydrolase, family 5:Cellulose-binding, family II, bacterial type:Na-Ca exchanger/integrin-beta4	109.77	4.56E-22	29%	44%	530–783	949–1228
														8	ZP_00687414	Chitinase	133.27	3.85E-29	30%	46%	531–801	32–334
														9	CAH35762	putative exported chitinase	127.49	2.11E-27	31%	45%	531–801	36–339
														10	YP_621070	Chitinase	132.11	8.57E-29	31%	46%	531–801	32–334
N784R	299716–300531	272	30,760	5.95	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..	191.19	1.06E-49	30%	53%	10–268	1–256	1	AAU06280	protein kinase A248R	251.14	2.85E-65	48%	69%	4–267	41–304
					2	smart00220	S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-soecific kinase subfamily. .	187.34	1.44E-48	31%	49%	11–268	1–256	2	AAU06275	protein kinase A248R	251.14	2.85E-65	48%	69%	4–267	20–283
					3	pfam00069	Pkinase, Protein kinase domain..	183.18	2.61E-47	32%	53%	11–268	1–258	3	AAU06282	protein kinase A248R	244.97	2.04E-63	47%	68%	4–267	41–304
					4	COG0515	SPS1, Serine/threonine protein kinase [General function prediction only / Signal transduction mechanisms / Transcription / DNA replication, recombination, and repair].	115.64	6.00E-27	28%	47%	10–268	1–278	4	AAC96616	PBCV-1 protein kinase	236.88	5.55E-61	44%	67%	4–267	38–303

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	smart00219	TyrKc, Tyrosine kinase, catalytic domain; Phosphotransferases. Tyrosine-specific kinase subfamily	114.50	1.07E-26	26%	44%	12--264	2--253	5	AAA87065	serine/threonine protein kinase	235.34	1.62E-60	44%	67%	4--267	34--299
					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	107.58	1.42E-24	25%	43%	9--263	7--263	6	AAU06274	protein kinase A248R	210.31	5.57E-53	50%	70%	52--267	22--237
					7	COG3642	COG3642, Mn2+-dependent serine/threonine protein kinase [Signal transduction mechanisms]	35.59	6.48E-03	26%	44%	86--152	67--139	7	AAU06285	protein kinase A248R	201.45	2.59E-50	47%	68%	54--267	2--216
					8	smart00750	KIND, kinase non-catalytic C-lobe domain; It is an interaction domain identified as being similar to the C-terminal protein kinase catalytic fold (C lobe). Its presence at the N terminus of signalling proteins and the absence of the active-site residues in the catalytic and activation loops suggest that it folds independently and is likely to be non-catalytic. The occurrence of KIND only in metazoa implies that it has evolved from the catalytic protein kinase domain into an interaction domain possibly by keeping the substrate-binding features	34.58	1.21E-02	29%	51%	88--204	7--101	8	AAU06286	protein kinase A248R	150.98	4.01E-35	47%	70%	119--267	1--151
					9	COG0478	COG0478, RIO-like serine/threonine protein kinase fused to N-terminal HTH domain [Signal transduction mechanisms]	31.75	9.29E-02	24%	50%	102--148	200--246	9	AAU06270	protein kinase A248R	137.89	3.51E-31	45%	67%	124--267	1--146
					10	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..	30.50	2.15E-01	35%	56%	110--151	155--198	10	AAC96657	similar to PBCV-1 serine/threonine protein kinase, corresponds to GenBank Accession Number U14660	136.73	7.82E-31	32%	55%	9--267	16--278
N787R	300578--301717	380	39,916	8.96		No Hit Found								1	AAC96746	A378L	133.65	1.13E-29	55%	69%	178--281	69--173
														2	CAA64974	QI74 protein	48.14	6.22E-04	20%	33%	96--377	200--461
														3	AAC96921	A565R	48.14	6.22E-04	25%	33%	84--281	265--466
														4	BAA11343	DNA binding protein	46.60	1.81E-03	26%	33%	84--281	265--458
N791R	301947--302597	217	24,656	7.65	1	cd00890	Prefoldin, Prefoldin is a hexameric molecular chaperone complex, found in both eukaryotes and archaea, that binds and stabilizes newly synthesized polypeptides allowing them to fold correctly. The complex contains two alpha and four beta subunits, the two subunits being evolutionarily related. In archaea, there is usually only one gene for each subunit while in eukaryotes there two or more paralogous genes encoding each subunit adding heterogeneity to the structure of the hexamer. The structure of the complex consists of a double beta barrel assembly with six protruding coiled-coils.	28.74	5.89E-01	28%	48%	38--105	63--124	1	AAC96971	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	120.55	3.73E-26	34%	50%	11--213	8--185
					2	COG1508	RpoN, DNA-directed RNA polymerase specialized sigma subunit, sigma54 homolog [Transcription]	28.35	6.82E-01	22%	52%	57--111	258--312	2	AAC96972	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	60.85	3.51E-08	62%	77%	11--50	14--53
														3	AAC96545	A177R	59.31	1.02E-07	51%	70%	7--53	7--53
														4	AAC96447	A79R	58.92	1.33E-07	59%	69%	9--50	6--47
														5	AAC96818	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	52.76	9.55E-06	51%	76%	12--50	9--47
														6	AAC96643	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	48.52	1.80E-04	47%	64%	9--50	9--50
N792R	302755--303510	252	28,357	7.98		No Hit Found								1	AAC96377	A9R	249.98	5.52E-65	64%	83%	76--247	3--174
														2	AAC96818	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	136.35	8.89E-31	33%	55%	1--245	1--246
														3	AAC96972	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	132.11	1.68E-29	34%	54%	4--245	10--253
														4	AAC96643	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	121.71	2.27E-26	30%	54%	1--244	4--248
														5	AAC96545	A177R	120.55	5.05E-26	31%	50%	1--230	4--227
														6	AAC96447	A79R	114.78	2.77E-24	32%	53%	1--223	1--217
														7	AAU06304	hypothetical protein A275R	73.94	5.42E-12	27%	51%	78--244	1--167
														8	AAU06301	hypothetical protein A275R	72.79	1.21E-11	26%	51%	78--244	1--167
														9	AAC96971	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	67.01	6.62E-10	59%	74%	1--47	1--47
														10	AAU06302	hypothetical protein A275R	61.23	3.63E-08	28%	52%	113--245	1--132
N794R	303528--304043	172	20,438	4.52		No Hit Found								1	AAC96458	contains phenyl group binding site (CAAX box)	58.92	7.38E-08	32%	53%	6--118	1--105
N795L	305535--304069	489	51,812	10.74	1	COG3509	LpqC, Poly(3-hydroxybutyrate) depolymerase [Secondary metabolites biosynthesis, transport, and catabolism]	37.27	5.09E-03	26%	37%	201--331	60--189	1	AAC96409	contains Pro-rich Px motif, PAPK (8X); similar to Thermoproteus virus protein TPX. corresponds to Swiss-Prot Accession Number P19275	429.10	1.84E-118	65%	80%	173--486	99--410
					2	COG0596	MhpC, Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily) [General function prediction only]	32.75	1.06E-01	20%	38%	185--306	4--108	2	AAC96539	similar to PBCV-1 ORF A41R, corresponds to GenBank Accession Number U17055	412.54	1.78E-113	63%	80%	176--486	77--385
					3	cd01915	CODH, Carbon monoxide dehydrogenase (CODH) is found in acetogenic and methanogenic organisms and is responsible for the synthesis and breakdown of acetyl-CoA, respectively. CODH has two types of metal clusters, a cubane [Fe4-S4] center (B-cluster) similar to that of hybrid cluster protein (HCP) and a Ni-Fe-S center (C-cluster) where carbon monoxide oxidation occurs. Bifunctional CODH forms a heterotetramer with acetyl-CoA synthase (ACS) consisting of two CODH and two ACS subunits while monofunctional CODH forms a homodimer. Bifunctional CODH reduces carbon dioxide to carbon monoxide and ACS then synthesizes acetyl-CoA from carbon monoxide, CoA, and a methyl group donated by another protein (CoFeSP), while monofunctional CODH oxidizes carbon monoxide to carbon dioxide. CODH and ACS each have a metal cluster referred to as the C- and A-clusters, respectively	31.34	2.98E-01	35%	63%	69--109	229--269	3	BAE02830	surface protein	56.61	2.49E-06	45%	58%	41--102	638--699
														4	NP_916095	P0481E12.18	65.08	6.99E-09	23%	41%	181--463	72--352
														5	BAD86968	hypothetical protein	65.08	6.99E-09	23%	41%	181--463	95--375
														6	AAM63817	unknown	59.31	3.84E-07	21%	40%	182--464	63--341
														7	NP_191439	unknown protein	59.31	3.84E-07	21%	40%	182--464	63--341
														8	XP_464783	unknown protein	49.68	3.04E-04	22%	41%	201--464	78--346
														9	AAC96403	contains Pro-rich Px motif, PAPK (19X); similar to Arabidopsis anter-specific Pro-rich protein, corresponds to Swiss-Prot Accession Number P40602	48.14	8.84E-04	39%	57%	42--107	382--449
														10	AAC96752	Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	45.05	7.48E-03	31%	53%	40--107	201--269
N799R	305621--306091	157	17,756	4.43	1	COG5201	SKP1, SCF ubiquitin ligase, SKP1 component [Posttranslational modification, protein turnover, chaperones]	96.28	1.51E-21	34%	56%	3--143	4--155	1	NP_567959	ASK11 (ARABIDOPSIS SKP1-LIKE 11); ubiquitin-protein ligase	113.24	2.60E-24	39%	61%	3--144	6--151
					2	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. .	82.61	2.25E-17	38%	59%	1--95	2--107	2	NP_567967	ASK12; protein binding / ubiquitin-protein ligase	112.46	4.44E-24	39%	63%	3--144	6--151
					3	pfam01466	Skp1, Skp1 family, dimerisation domain..	66.79	1.31E-12	39%	64%	69--144	2--77	3	AAX83944	Skp1	112.08	5.80E-24	38%	58%	3--144	7--155
					4	pfam03931	Skp1 POZ, Skp1 family, tetramerisation domain..	61.07	6.07E-11	42%	66%	1--61	1--66	4	AAO85510	SKP1	110.92	1.29E-23	38%	61%	3--144	4--152
					5	COG1096	COG1096, Predicted RNA-binding protein (consists of S1 domain and a Zn-ribbon domain) [Translation, ribosomal structure and biogenesis]	31.80	4.52E-02	19%	34%	57--126	19--89	5	AAC63273	SKP1-like protein	110.54	1.69E-23	38%	60%	3--144	4--152
														6	AAC96407	contains ATP/GTP-binding motif A; similar to Dictyostelium FP21 glycoprotein. corresponds to Swiss-Prot Accession Number P52285	109.38	3.76E-23	37%	59%	1--142	1--142
														7	AAT99735	SKP1	109.00	4.91E-23	39%	59%	3--144	7--154
														8	AAD34458	Skp1	109.00	4.91E-23	38%	60%	3--144	5--152
														9	ABE93442	SKP1 component	109.00	4.91E-23	38%	59%	3--144	7--154
														10	1P22_B	Chain B, Structure Of A Beta-Trcp1-Skp1-Beta-Catenin Complex: Destruction Motif Binding And Lysine Specificity On The Scfbeta-Trcp1 Ubiquitin Ligase	108.23	8.38E-23	38%	62%	2--142	3--141
N802R	306133--306387	85	9,463	10.33		No Hit Found								1	AAC96445	contains type 1 hydrophobic transmembrane region	41.97	7.19E-03	35%	51%	3--81	5--86
N803L	307143--306382	254	29,858	5.27		No Hit Found								0	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
N805R	307438--308127	230	25,806	4.92		No Hit Found								1	YP_477577	glycosyl transferase, group 1 family protein	49.68	9.12E-05	26%	41%	10--192	196--384
N806L	308866--308288	193	21,810	10.62		No Hit Found								0	No Hit Found	No Hit Found						
N809R	308948--309418	157	18,734	6.54	1	COG4425	COG4425, Predicted membrane protein [Function unknown].	29.56	1.74E-01	33%	47%	102--153	444--493	0	No Hit Found	No Hit Found						
N811R	309467--309859	131	14,507	4.33		No Hit Found								0	No Hit Found	No Hit Found						
N814R	309873--310316	148	16,736	10.33	1	COG4536	CorB, Putative Mg2+ and Co2+ transporter CorB [Inorganic ion transport and metabolism].	27.86	5.78E-01	23%	49%	40--106	286--351	0	No Hit Found	No Hit Found						
N815L	311403--310333	357	41,147	10.09	1	smart00507	HNHc, HNH nucleases; .	28.90	9.97E-01	31%	44%	60--106	1--49	1	AAC96722	A354R	206.07	1.63E-51	43%	61%	116--354	2--235
														2	AAC96790	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	81.26	6.07E-14	27%	47%	59--326	51--323
														3	AAC96455	A87R	62.77	2.23E-08	25%	40%	29--351	105--447
														4	AAL98037	hypothetical phage protein	45.05	4.82E-03	27%	50%	23--152	9--137
N819R	311726--312187	154	18,361	5.77	1	COG5280	COG5280, Phage-related minor tail protein [Function unknown].	27.69	6.93E-01	17%	31%	24--59	16--51	1	CAD50809	hypothetical protein	44.67	1.12E-03	24%	43%	1--149	429--577
N820R	312298--314097	600	65,919	10.86	1	cd01828	sialate_O-acetylesterase_like2, sialate_O-acetylesterase_like subfamily of the SGNH-hydrolases, , a diverse family of lipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alpha/beta hydrolase family and unique among all known hydrolases; its active site closely resembles the Ser-His-Asp(Glu) triad found in other serine hydrolases	78.02	3.44E-15	25%	38%	119--274	1--167	1	AAC96686	A318R	238.04	7.81E-61	54%	69%	402--598	6--211
					2	cd01841	NnaC_like, NnaC (CMP-NeuNAc synthetase) _like subfamily of SGNH_hydrolases, a diverse family of lipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alpha/beta hydrolase family and unique among all known hydrolases; its active site closely resembles two of the three components of typical Ser-His-Asp(Glu) triad from other serine hydrolases. E. coli NnaC appears to be involved in polysaccharide synthesis..	49.88	9.78E-07	25%	43%	124--276	6--167	2	AAC96684	PAPK (17X); similar to PBCV-1 ORF A41R, corresponds to Genbank Accession Number U17055	45.82	5.69E-03	64%	74%	1--31	1--31
					3	cd00229	SGNH_hydrolase, SGNH_hydrolase, or GDSL_hydrolase, is a diverse family of lipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alpha/beta hydrolase family and unique among all known hydrolases; its active site closely resembles the typical Ser-His-Asp(Glu) triad from other serine hydrolases, but may lack the catalytic acid	48.30	2.83E-06	17%	33%	119--273	1--196	3	ZP_00244751	COG2755: Lysophospholipase L1 and related esterases	51.22	1.35E-04	28%	44%	118--274	74--244
					4	cd01820	PAF_acetylesterase_like, PAF_acetylhydrolase (PAF-AH)_like subfamily of SGNH-hydrolases. Platelet-activating factor (PAF) and PAF-AH are key players in inflammation and in atherosclerosis. PAF-AH is a calcium independent phospholipase A2 which exhibits strong substrate specificity towards PAF, hydrolyzing an acetyl ester at the sn-2 position. PAF-AH also degrades a family of oxidized PAF-like phospholipids with short sn-2 residues. In addition, PAF and PAF-AH are associated with neural migration and mammalian reproduction	45.63	1.53E-05	26%	40%	120--269	35--198	4	BAD50731	conserved hypothetical protein	46.21	4.36E-03	28%	40%	120--274	54--227
					5	cd01833	XynB_like, SGNH_hydrolase subfamily, similar to Ruminococcus flavefaciens XynB. Most likely a secreted hydrolase with xylanase activity. SGNH hydrolases are a diverse family of lipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alpha/beta hydrolase family and unique among all known hydrolases; its active site closely resembles the Ser-His-Asp(Glu) triad found in other serine hydrolases	44.55	4.03E-05	20%	39%	145--260	53--182	5	CAH09444	putative acylhydrolase	45.44	7.43E-03	28%	40%	120--274	54--227
N827R	314231--314899	223	25,953	6.19		No Hit Found								1	AAC96456	A88R	61.23	2.86E-08	32%	48%	1--147	1--143
N830L	315907--315026	294	32,506	6.73	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	99.00	5.91E-22	43%	68%	29--150	2--125	1	XP_681288	hypothetical protein AN8019.2	78.57	2.89E-13	42%	56%	136--233	799--902
					2	COG0666	Arp, FOG: Ankyrin repeat [General function prediction only].	66.06	5.77E-12	30%	41%	5--148	77--224	2	CAE64680	Hypothetical protein CBG09456	65.08	3.30E-09	33%	51%	108--230	39--169
					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..	45.05	1.21E-05	59%	78%	193--225	1--33	3	EAS30828	hypothetical protein CIMG_06307	48.14	4.18E-04	28%	50%	99--230	315--458
					4	smart00248	ANK, ankyrin repeats; Ankyrin repeats are about 33 amino acids long and occur in at least four consecutive copies. They are involved in protein protein interactions. The core of the repeat seems to be an helix-loop-helix structure. .	38.02	1.38E-03	59%	78%	193--220	1--28	4	XP_749852	hypothetical protein Afu1g01020	72.02	2.70E-11	34%	49%	136--285	834--983
					5	COG0823	TolB, Periplasmic component of the Tol biopolymer transport system [Intracellular trafficking and secretion].	33.49	3.07E-02	28%	37%	193--236	292--335	5	XP_392578	PREDICTED: similar to CG7462-PB, isoform B	87.43	6.22E-16	28%	45%	38--230	83--308
														6	AAF61702	ankyrin 1	107.84	4.45E-22	34%	50%	17--230	419--637
														7	XP_581734	PREDICTED: similar to Ankyrin-1 (Erythrocyte ankyrin), partial	83.96	6.87E-15	27%	46%	5--230	223--480
														8	YP_476042	ankyrin repeat protein	134.04	5.79E-30	33%	52%	5--249	5--247
														9	AAM11327	GH01626p	129.80	1.09E-28	35%	49%	5--244	26--264
														10	NP_648148	Ank2 CG7462-PC, isoform C	64.31	5.64E-09	29%	52%	89--230	1--146
N836L	316749--315937	271	28,506	6.45	1	cd00204	ANK, ankyrin repeats; ankyrin repeats mediate protein-protein interactions in very diverse families of proteins. The number of ANK repeats in a protein can range from 2 to over 20 (ankyrins, for example). ANK repeats may occur in combinations with other types of domains. The structural repeat unit contains two antiparallel helices and a beta-hairpin, repeats are stacked in a superhelical arrangement; this alignment contains 4 consecutive repeats	123.27	2.55E-29	46%	59%	39--157	8--126	1	AAC96373	contains 4 ankyrin repeats; similar to reticulocyte ankyrin, corresponds to Swiss-Prot Accession Number P16157	179.87	8.00E-44	40%	54%	3--252	2--250
					2	COG0666	Arp, FOG: Ankyrin repeat [General function prediction only].	84.17	1.49E-17	38%	53%	22--160	57--203	2	AAC96986	contains 10 ankyrin-like repeats; similar to human ankyrin, corresponds to Swiss-Prot Accession Number P16157	129.41	1.24E-28	36%	52%	4--209	3--208
					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	41.97	7.43E-05	53%	63%	105--135	3--33	3	XP_796302	PREDICTED: similar to Ankyrin repeat domain protein 28, partial	45.05	3.08E-03	31%	42%	104--229	12--140
					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	1.47E-03	45%	62%	70--99	1--30	4	XP_794653	PREDICTED: similar to Ankyrin-1 (Erythrocyte ankyrin)	113.62	7.04E-24	39%	49%	12--189	72--249
					5	COG2378	COG2378, Predicted transcriptional regulator [Transcription].	34.21	1.88E-02	28%	47%	192--250	69--127	5	XP_746992	hypothetical protein Afu8g02140	68.17	3.39E-10	34%	47%	3--127	688--812
					6	pfam04300	FBA, F-box associated region. Members of this family are associated with F-box domains, hence the name FBA. This domain is probably involved in binding other proteins that will be targeted for ubiquitination. One member has been shown to be involved in binding to N-glycosylated proteins	29.98	3.59E-01	24%	33%	44--84	96--138	6	XP_784202	PREDICTED: similar to CG10011-PA	57.00	7.82E-07	41%	51%	9--98	220--307
					7	COG0283	Cmk, Cytidylate kinase [Nucleotide transport and metabolism].	29.02	6.95E-01	27%	47%	197--252	151--206	7	AAQ09555	inv-like protein	58.15	3.51E-07	29%	44%	75--202	68--196
														8	XP_794449	PREDICTED: similar to inversin isoform b, partial	56.61	1.02E-06	27%	42%	75--202	69--208
														9	XP_785043	PREDICTED: similar to Ankyrin-1 (Erythrocyte ankyrin) (Ankyrin R), partial	57.00	7.82E-07	27%	48%	9--156	51--199
														10	XP_782809	PREDICTED: similar to ankyrin repeat domain 28	63.54	8.36E-09	38%	53%	9--96	783--870
N840R	316897--317373	159	18,480	8.36		No Hit Found								0	No Hit Found	No Hit Found						
N845R	317707--318180	158	17,734	5.51		No Hit Found								1	CAI06222	conserved hypothetical protein	44.28	1.48E-03	50%	67%	20--59	460--499
														2	AAC96796	Pro- and Glu-rich, PENPEV (10x); similar to Streptococcus B antigen, corresponds to Swiss-Prot Accession Number P27951	43.51	2.52E-03	55%	58%	20--55	53--88
														3	XP_726599	hypothetical protein PY00072	41.59	9.59E-03	60%	60%	20--59	1683--1722
														4	ABA45585	collagen-like surface protein, putative	45.05	8.67E-04	67%	70%	21--54	1685--1718
														5	EAS36650	hypothetical protein CIMG_02004	41.97	7.34E-03	48%	65%	25--59	23--57
														6	XP_504118	hypothetical protein	45.05	8.67E-04	42%	51%	16--60	181--225
														7	ZP_01250057	hypothetical protein Bpse110_02005899	43.13	3.30E-03	52%	65%	20--59	156--195
														8	ZP_00686206	hypothetical protein BambDRAFT_5422	41.97	7.34E-03	50%	65%	20--59	570--609
														9	ZP_00462072	hypothetical protein Bcen2424DRAFT_3653	42.74	4.30E-03	55%	62%	20--59	210--249
														10	ZP_00978857	hypothetical protein BcenP_01004081	42.74	4.30E-03	55%	62%	20--59	105--144

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
N847R	319064--319798	245	27,064	8.51		No Hit Found								1	AAC96734	A366L	155.61	1.35E-36	38%	59%	6--237	14--251
														2	AAC96972	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	61.62	2.66E-08	32%	46%	2--151	8--144
														3	AAC96818	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	56.61	8.55E-07	37%	51%	1--72	1--81
														4	AAC96643	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	55.45	1.91E-06	25%	40%	1--237	4--199
														5	AAC96447	A79R	55.07	2.49E-06	27%	40%	1--237	1--189
														6	AAC96971	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	54.30	4.25E-06	48%	59%	1--47	1--47
														7	AAC96545	A177R	51.22	3.59E-05	42%	63%	1--47	4--50