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

#### **Virology Papers**

Virology, Nebraska Center for

2-20-2007

# 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, lisa.fitzgerald@nrl.navy.mil

Michael V. Graves University of Massachusetts-Lowell, Michael\_Graves@uml.edu

Xiao Li University of Massachusetts–Lowell

Tamara Feldblyum The Institute for Genomic Research, Rockville, MD

James Hartigan Agencourt Bioscience Corporation, Beverly, MA

See next page for additional authors

Follow this and additional works at: https://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.

https://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

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<sup>a</sup>, Michael V. Graves<sup>b</sup>, Xiao Li<sup>b</sup>, Tamara Feldblyum<sup>c</sup>, James Hartigan<sup>d</sup>, and James L. Van Etten<sup>e, f, \*</sup>

<sup>a</sup>Department of Chemistry, University of Nebraska–Lincoln, Lincoln, NE 68588-0304

<sup>b</sup>Department of Biological Sciences, University of Massachusetts-Lowell, Lowell, MA 01854

<sup>d</sup>Agencourt Bioscience Corporation, 500 Cummings Center, Suite 2450, Beverly, MA 01915

<sup>e</sup>Department of Plant Pathology, University of Nebraska–Lincoln, Lincoln, NE 68583-0722

<sup>f</sup>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 314,335bp 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, 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

<sup>&</sup>lt;sup>c</sup>The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850

Gene Name	Genome Position	A.A. lenath	Peptide Mw		CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identitv P	% Positive	Query from-to	Hit from- BL to N		Hit Accession	BLASTp Definition	Bit Score E-value	% dentitv F	% ositive	Query from-to	Hit from- to
N003I	2472490	661	74,180	9.20	1	COG2015	COG2015, Alkyl sulfatase and related hydrolases [Secondary	700.55	0.00E+00	45%	64%	22654	15652	1	NP 896208	alkyl sulfatase family protein	752.67 0.00E+00	58%	73%	34657	50679
			,		2		Lactamase B, Metallo-beta-lactamase superfamily	48.00	3.49E-06	22%	36%	129-232	1102	2		metallo-beta-lactamase family protein	741.50 0.00E+00	57%	74%	34657	45675
					3	COG1237	superfamily II (General function prediction only)	40.68	6.24E-04	27%	42%	138224	26107	3	CAG19676	hypothetical beta-lactamase	731.48 0.00E+00	56%	73%	34658	48678
					4	COG4783	[General function prediction only].	38.01	3.33E-03	33%	54%	464-531	378445	4	AAV96272	metallo-beta-lactamase family protein	721.47 0.00E+00	56%	73%	34657	34663
					5	COG0491	GloB, Zn-dependent hydrolases, including glyoxylases [General function prediction onlv].		1.66E-02	29%	44%	138-223	30109	5	YP_593031	beta-lactamase-like	671.77 0.00E+00	52%	70%	34658	38667
					6	COG1234	ElaC, Metal-dependent hydrolases of the beta-lactamase superfamily III IGeneral function prediction onlv1. TPR, Tetratricopeptide repeat domain; typically contains 34 amino acids	35.80	1.93E-02	24%	48%	163-215	4498	6	ZP_00753431	COG2015: Alkyl sulfatase and related hydrolases	620.93 4.82E-176	48%	67%	30658	26660
							TPR: Tetratricopeptide repeat domain: typically contains 34 amino acids (IVEF-X2)-[IVI](GAS)X2(Y)-[YVF](X4)(ASES(X3)(Y-1Y),X2)-(ASE)- X(4)-[PKE] is the consensus sequence: found in a (FV1)-X2)-(ASE)- x4(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 chapterone, cell-cycle.														
					7		transciption, and protein transport complexes; the number of TPR motifs varies among poteins (1,3-11,13 (5,16)); 5: chandem repeats generate a right-handed helical structure with an amphipathic channel that is thought accomodate an adphahetik of a larget potein; r. Itas been proposed that TPR proteins preferably interact with WD-40 repeat poteins, but in many instances several TPR-proteins enclude. Cdc16p, Cdc23a and Cdc27a, components of the acrohosmadBCC that.	33.90	7.00E-02	34%		464-508	4-48			COG2015: Alkyl sulfatase and related hydrolases	620.54 6.29E-176	48%	67%	30658	
					8	COG2333	ComEC, Predicted hydrolase (metallo-beta-lactamase superfamily) IGeneral function nerdiction onlvi RsmC, 165 RNA 61207 methylase RsmC [Translation, ribosomal structure and biogenesis].	32.23	2.16E-01	25%	43%	137-224	57138			COG2015: Alkyl sulfatase and related hydrolases	619.77 1.07E-175	48%	67%	31658	27660
					9	COG2813	structure and biogenesis). COG0595. Predicted hydrolase of the metallo-beta-lactamase	31.39	3.98E-01	23%	45%	508600	126212	9		conserved hypothetical protein	619.00 1.83E-175	48%	67%	31658	27660
					10	COG0595	COG0595, Predicted hydrolase of the metallo-beta-lactamase superfamily [General function prediction only].	30.60	5.81E-01	23%	42%	135189	2383	10	ZP_00913757	Twin-arginine translocation pathway signal	605.52 2.09E-171	49%	66%	34657	49678
N007L	31222577	182	20,422	10.22		No Hit Found								1	AAC96449	A81L	108.61 9.58E-23	41%	62%	45158	37163
N010L	36663151	172	19,292	6.65		No Hit Found								1	AAC96452	A84L	71.63 1.10E-11	32%	55%	14135	13149
							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 19 have been associated with the autosomal dominant disorder known as hereditary multiple exostoses (HME). This is the most common known kelled indexiaia. The chromeronal lorations of other														
N012L	45633751	271	31,655	6.29	1	ptam03016	common known skeletal dysplasia. The dromosomal locations of other EXT genes suggest association with other forms of neoplasia. EXT and EXT2 have both been shown to encode a heparan subhate polymerase with both D-glucorarny (IGAA) and N-acety-D-glucosaminoglyau (IGAAAC) transferase activities. The mature of the defect in heparan DSBA, DSBA+ terbinedoxin dromain. This family contains a diverse set of proteins with a thioredoxin-like structure plant0085. This family action includes 2-hydroxytchrome-2-actionaly late (DSA) family also includes 2-hydroxytchrome-2-actionaly contains a diverse actionaly includes 2-hydroxytchrome-2-actionalytic (DSA) family also includes 2-hydroxytchrome-2-actionalytic (DSA) family also includes 2-hydroxytchrome-2-actionalytic (DSA) familytic (		6.04E-06	25%	51%	159–235	221-292	1	AAC96443	A75L	234.57 2.74E-60	41%	61%	8268	7276
					2	pfam01323	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 HCAC isomerases was not previously recognised. Some members of this family may have been mis-annotated in protein senuence databases.	29.98	3.06E-01	17%	34%	43161	8138	2	NP_195517		45.82 1.80E-03	27%	48%	162-268	284-384
							nucleoside_deaminase_Nucleoside deaminases include adenosine, guaraine 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 nucleophin attack on the substrate. The functional enzyme is a homodimer. Cytosine deaminase catalyzes the deamination of cytosine to uncat and ammoni							3 4 5	ABA18110	unknown protein exostosin family protein	44.67 4.02E-03 43.51 8.95E-03	26% 26% 30%	49% 49% 44%	174–266 174–266 169–257	471-571 148-248 206-308
N014L	50254663	121	13,198	10.46	1		and is a member of the pyrimidine salvage pathway. Cytosine deaminase is found in bacteria and fungi but is not present in marmals; 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 IRNA-specific adenosine deaminases that generate inosine at the first position of their anticodom (position 34) of specific RNAs; this modification is thought to enlarge the contence of the lamily are guarine deaminases protein synthesis. Other members of the lamily are guarine deaminases Currents, Cytosine/idenosine deaminases (Pulceholde transport and metaholism). Chrossine/idenosine deaminases (Pulceholde transport and metaholism). Chrossine/idenosine deaminases	58.35	2.32E-10 6.92E-09	37%	51%	4106 2106	1-94 10105	1	AAC96568 AAR26853	contains cytldine and deoxycytidine dearninase Zn-binding region signature	n 159.84 2.40E-38 50.83 1.57E-05	61%	83%	1118 22111	1118 24105
					3	pfam00383	dCMP_cyt_deam, Cytidine and deoxycytidylate deaminase zinc-binding region.	44.20	4.27E-06	31%	47%	4106	7101	3	AAX51127	tRNA-specific adenosine deaminase	50.06 2.67E-05	32%	49%	4108	10104
					4	cd01284	Bibdiavin, deaminase-reductase, Ribdiavin-specific deaminase. Ribdiavin biosynthesis protein RibDi (Daiminohydroxyhosphothosylaminopyrimidine deaminase) catalyzes the deamination of 2.5-diamino-chitosylamino-d'All-pyrimidineo 55agos-phosphate, which is an intermediate step in the biosynthesis of ribdiavin. The hol Gener di Bacillus sublis and the hol Gener Of E.coli are bifunctional and contain this deaminase domain and a reductase domain which catalyzes the subsequent reduction of the ribosy side	37.92	3.09E-04	27%	45%	4107	1–95	4	AAC68441	cytosine deaminase	49.68 3.49E-05	32%	49%	4108	10104
					5	cd01286	decovychidylate_deaminase. Decovychidylate deaminase domain. Decovychidylate_deaminase. catalyzes the deamination of dCMP to dUMP, providing the nucleotide substate for thymidylate synthase. The erzyme bind 2:rut + 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	31.78	2.25E-02	37%	51%	50108	68117	5	ZP_00679087	Cytidine/deoxycytidylate dearninase, zinc-binding region	45.44 6.58E-04	32%	44%	4106	19111
					6	COG0117	coli with dCTP acting as an activator and dTTP as an inhibitor RibD, Pyrimidine deaminase [Coenzyme metabolism]. cytidine deaminase-like, Cytidine and deoxycytidylate deaminase zinc-	31.76	2.36E-02	24%	43%	7109	13104	6	GAA00608	unnamed protein product	44.67 1.12E-03	30%	46%	4108	36130
					7		binding region. The family contains cytidine dearninases, nucleoside dearninases, decxycitidytate dearninases and hofbatiwi dearninases. 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 tatkot on the substrate.	-	5.89E-02	20%	36%	6108	395		NP_502546		44.28 1.47E-03	32%	48%	2108	11–111
					8	pfam02041 COG2131	Auxin BP, Auxin binding protein ComEB, Deoxycytidylate deaminase [Nucleotide transport and	28.08 26.85	3.02E-01 6.80E-01	21% 45%	50% 60%	2379 88108	58126 106126	8		cytidine/deoxycytidylate deaminase family protein	44.28 1.47E-03 43.90 1.92E-03	29% 29%	46% 46%	4108 6108	10104 9101
					9	0002131	metabolism1.	20.65	0.00E-01	4076	00%	00108	100120	9 10		cytidine/deoxycytidylate deaminase family protein CMP/dCMP deaminase, zinc-binding	43.50 1.92E-03 43.51 2.50E-03	29% 31%	40%	6108	9101 9101
N015L	54075111	99	10.616	10.54		No Hit Found								1	AAC96567		70.86 1.48E-11	47%	66%	171	172
N016R	54195907	163	18,543	8.04		No Hit Found								1	AAC96564		184.88 7.72E-46	61%	79%	23156	18151
NOTON	54.8-5001	103	10,040	0.04		. to rate oullu									,		104.00 1.122-40	0170	, 0 /0	20.100	10 101

Gene Name	Genome Position	A.A.	Peptide Mw	pl	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identitv F	% Positive		Hit from- E	3LASTp Hit Number		BLASTp Definition	Bit Score E-value	% Identitv P		Query rom-to	Hit from-
namo	r comon	lonati			Humber		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				ooniro				, according to the second seco						10
N019R	59586755	266	29,728	5.00	1	cd0057	central cavity, providing a DNA-bound platform for the attachment of the polymerase. The timeric 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 (catabid) subunits in T4 and RB69). This structural correspondence further substantiates the mechanistic connection between evkaryotic and prokaryotic DNA replication that has been suggested on biochemical grounds. PCNA is a laso involved with proteins involved in cell cycle processes such as DNA repair and apoptosis. Many of these proteins contain a highly conserved molif known as the PIP-box (PCNA	<sup>:</sup> 146.96	2.13E-36	30%	55%	12261	1245	1	AAC96561	similar to human PCNA, corresponds to Swiss-Prot Accession Number P12004	368.24 1.53E-100	68%	85%	9261	7259
					2	pfam0070	interacting protein box) which contains the sequence Qxx(LM)xx(FY). 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 duelox DNA.	92.25	5.33E-20	28%	57%	9132	1125	2	XP_534355	PREDICTED: similar to proliferating cell nuclear antigen	158.69 1.84E-37	32%	55%	5264	202461
					3	pfam0274	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 experience during DNA	79.98	2.97E-16	34%	55%	137-261	2127	3	CAA55669	proliferative cell nuclear antigen	157.53 4.10E-37	34%	56%	9264	1256
					4	COG059	DnaN DNA polymerase sliding clamp subunit (PCNA bomolog) (DNA	56.08	4.36E-09	19%	37%	21261	72321	4	AAB27811		156.76 6.99E-37	32%	57%	9264	1256
					5	COG109	renization recombination and renarif COG102, Precident SAM-dependent methytransferases [General function prediction on/V.	31.44	1.30E-01	20%	31%	106–202	251–358	5 6 7 8 9 10 M	AAG10077 CAA37243 BAB28355 NP 035175	PCNA1 (PROLIFERATING CELLUAR NUCLEAR ANTIGEN); DNA bindino / DNA oolvmerase oncoessivitv factor proliferating coll nuclear antigen umamed protein product umamed protein product proliferating cell nuclear antigen hypothetical protein LOS/5499	156.76         6.99E-37           155.99         1.19E-36           155.61         1.56E-36           155.61         1.56E-36           155.61         1.56E-36           155.61         2.04E-36	32% 32% 32% 32% 32%	57% 55% 55% 55% 55%	9264 9264 9264 9264 9264 9264	1256 1256 1256 1256 1256 1256
N020L	106286753	1292	141,764	11.55	1	cd0327	ABC_SMC4_euk, Eukanyotic SMC4 proteins; SMC proteins are large (approximately 110 to 170 kDa), and each is arranged into fine between species is largely confined to the antio- and actobox-terminal globular domains. The amino-terminal domain contains a &apox/Walker Adapos, nucleotide-binding domain (CostAGKS7), in the single-term amino-add code), which by mutathonal studies has been shown to be essential in several proteins. The actooxy-terminal domain contains a and a nort with homology to the signature sequence of the ATP-binding casette (ABC)-terminal domain in relatively high within the SMC1-SMC4 group, whereas SMC5 and SMC5 shows once divergence in both of these SMC1 paired with SMC5, SMC2 with SMC4, and SMC5 with SMC6 (SMC4 group, whereas RA15).	33.04	2.36E-01	17%	51%	10731156	287–376	1	AAC96557	similar to SWI/SNF chromatin remodeling complex subunit OSA2	718.77 0.00E+00	39%	51%	2–1175	101139
					2	pfam0157	(Unitery autom as has inc). Myosin_ail, Myosin bial. 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 5 amily consists of the colled-coil myosin heavy chain tail region. The colled-coil is composed of the tail from two molecules of myosin. These can then assemble into the macromolecular thick filament. The colled-coil menion mavides the structural backhoons the thick filament.	31.82	5.49E-01	34%	55%	10731155	772846	2	T17682	hypothetical protein A192R - Chlorella virus PBCV-1	81.65 2.26E-13	43%	68% 10	0751175	103–200
					3	cd0327	ABC_SMC1_evak, Eukaryotic SMC1 proteins: SMC proteins are large (approximately 110 to 170 kbb), 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 global comains. The amino-terminal domain coming is a Ageos-Walker amino-acid code, which by matalional studies has been shown to be essential in several proteins. The activo-y-terminal domain contains a sequence (the D-Aox) that resembles a Ageosy-terminal domain contains a sequence (the D-Aox) that resembles a Ageosy-terminal domain contains a sequence (the D-Aox) that resembles a Ageosy-terminal domain contains a sequence (the D-Aox) that resembles a Ageosy-terminal domain contains a sequence (the D-Aox) that resembles a Ageosy-terminal domain contains a sequence (the D-Aox) that resembles a Ageosy-terminal domain contains a sequence (the D-Aox) that resembles a Ageosy-terminal domain the carboxy-terminal domain is relatively high whim the SMC/-SMC4 group, sequences. In evaluary-terminal domain contains as heterodimers of SMC1 paired with SMC3. SMC2 with SMC4, and SMC5 with SMC6 (formerly known as Rat15).	31.95	5.73E-01	19%	44%	10631170	245353	3	AAK77699	ORF30, putative collagen	50.83 4.27E-04	23%	28%	98482	6661074
					4	pfam0651	TolA TolA protein. This family consists of several bacterial TolA proteins as well as the outlawyoic proferes or ulmknown function. Tol proteins are involved in the transforation of group A coltans. Coltains are bacterial protein toxins, which are active against Escherichia coll and other related species (See pfam01024). TolA is anchored to the cytoplasmic membrane by a single membrane spanning segment near the N- temmuns, leaving most the protein exposed to the periplasm. ABC_SMOC_auk, Likavayotic SMOC proteins SMO proteins are large	31.64	6.12E-01	29%	45%	10701168	67171	4	NP_477523	wsv001	50.83 4.27E-04	23%	28%	98482	6661074
					5	cd0327	(approximately 110 lo 170 kDa), and each is arranged into five recognizable domina. Annino add sequence homology of SMC proteins between species is largely continued to the annino- and carboxy-terminal Adapos, runcedocidis-binding domain (CostGAKST). In the single-larger annino-aid code), which by mutational studies has been shown to be second the second second second second second second and and a motif with homology to the signature sequence of the ATI-B-inding carboxy-terminal domain a relatively high within the SMCI-SMCG ayroup, whereas SMCG and SMCG shows ome divergence in both of these sequences. In eukaryotic cells, the proteins are found as heterodimens of SMCI pared with SMCS. SMCC with SMCG with SMCS with S	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	363
							(formerly known as Rad18)							6 M 7 8 9 10	XP_783728 NP 571089 AAH90694	CG40323-PB.3 PREDICTED: similar to Protein transport protein Sec24C (SEC24-related orchein C) calymmin Crnn protein PREDICTED: similar to GA11046-PA	48.91 1.62E-03 55.07 2.26E-05 52.37 1.47E-04 52.37 1.47E-04 53.14 8.60E-05	23% 25% 23% 23% 27%	33% 32% 32%	190498 140483	3426-3778 6311 230-604 222-596 124-394
N023I	13471-10658	938	106.276	7.86	1	COG041	7 PolB, DNA polymerase elongation subunit (family B) [DNA replication,	346.65	6.28E-96	30%	48%	51879	11772	1			1352.04 0.00E+00	72%	84%	30932	1913
NUZ3L	13471-10658	878	100,276	7.86	1		recombination. and repairl. DNA_pol_B, DNA polymerase family B. This region of DNA polymerase B appears to consist of more than one structural domain, possibly		6.28E-96 4.53E-93	30%	48% 55%	51879	11772	1		PBVC-1 DNA polymerase DNA polymerase	1352.04 0.00E+00 1347.03 0.00E+00	72%	84%	30932	1-913
					3	smart0048	including elongation, DNA-binding and dNTP binding activities 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	292.12	1.89E-79	33%	51%	205667	1475	3	P30320	DNA polymerase	1341.25 0.00E+00	71%	83%	30932	1913
					4	cd0014	DNA polymerase in the con, millicitionian DNA polymerases and and virus DNA polymerases . POLBc, DNA polymerase type-B family, DNA directed DNA polymerase. 5 Posseses DNA binding, polymerase and 3&apos5', exonuclease activity.		5.27E-73	33%	49%	205704	1511	4		DNA polymerase	443.74 1.58E-122	96%	97%	489718	1230
					5	pfam0310	DNA_pol_B_exo, DNA polymerase family B, exonuclease domain. This domain has 3' to 5' exonuclease activity and adopts a	194.94	3.39E-50	26%	41%	56383	1334	5	AAK28935	DNA polymerase	431.80 6.23E-119	100%	100%	497711	1215
							ribonuclease H type fold							6 7	AAK28933	DNA polymerase DNA polymerase	429.10 4.04E-118 429.10 4.04E-118	99% 99%		497–711 498–711	1215 1214

Gene Name	Genome Position	A.A. lenath	Peptide Mw		CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identitv P	% Positive	Query from-to	Hit from- to	BLASTp Hit Number	Accession	Bit Score E-value	% Identitv F		from-to	Hit from- to
														8 9 10		427.56 1.17E-117 411.38 8.71E-113 408.30 7.37E-112	99% 32% 32%	99% 49% 52%	497711 46880 46892	1215 83929 123980
N029R	1353814146	203	23,119	6.74	1	COG5283	COG5283, Phage-related tail protein [Function unknown]. COG5325, t-SNARE complex subunit, syntaxin [Intracellular trafficking	31.97	5.05E-02	16%	36%	46198	9361079	1	AAC96971 similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	139.43 6.66E-32	37%	57%	1199	1185
					2	COG5325	and secretion].	28.06	7.54E-01	27%	49%	4186	161206	2		65.86 9.36E-10	59%	70%	147	450
														3	Accession Number 042580	62.39 1.04E-08	54% 47%	69% 59%	247 157	853 460
														* 5	Number U17055 AACOGE12 similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number	57.38 3.33E-07	47%	62%	156	460
														6	042580	56.61 5.68E-07	48%	68%	147	147
N030R	1429515038	248	27,869	8.33		No Hit Found								1	AAC96377 A9R	222.63 9.12E-57	57%	79%	73243	3173
														2		133.65 5.57E-30	32%	50%	4242	10253
														3	U42580	131.34 2.77E-29 131.34 2.77E-29	31% 33%	52% 54%	1244 1227	1248 4227
														5	AAC96643 similar to PBCV-1 ORF A79R, corresponds to GenBank Accession	127.87 3.06E-28	31%	51%	1241	4248
														6 7	AAC96447 A79R AAU06304 hypothetical protein A275R	122.48 1.29E-26 78.95 1.63E-13	33% 30%	52% 52%	1221 78241	1218 4167
														8 9	AAU06301 hypothetical protein A275R AAU06302 hypothetical protein A275R	77.80 3.63E-13 64.31 4.15E-09	30% 31%	52% 54%	78241 110242	4167 1132
														10		56.23 1.13E-06	46%	68%	147	147
N031L	1553315027	169	19,307	4.42		No Hit Found								0	No Hit Found No Hit Found					
							ANK, ankyrin repeats; ankyrin repeats mediate protein-protein													
N033I	1751215860	551	63.615	6.12	1	cd00204	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).		2.03E-09	40%	57%	395466	54126	1	XP 797753 PREDICTED: similar to ankyrin repeat domain protein 17 isoform b,	61.62 9.02E-08	20%	40%	22479	451 909
NUSSE	1/512-15660	551	03,015	0.12	'	6000204	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		2.03E-09	40%	57 %	393400	34120		xP_/9//53 partial	01.02 9.02E-08	20%	4076	22-479	431-696
							contains 4 consecutive repeats							_	XP_788092 PREDICTED: similar to Ankyrin-2 (Brain ankyrin) (Ankyrin B) (Ankyrin,				22109	
					2		Arp, FOG: Ankyrin repeat [General function prediction only]. Ank, Ankyrin repeat. There's no clear separation between noise	39.10	1.28E-03	29%	42%	405482	139216	2	xP_788092 nonervthroid). partial	47.75 1.35E-03	31%	53%	22109	11171198
					3	pfam00023	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		1.12E-02	44%	69%	412444	133	3	XP_786001 PREDICTED: similar to Ankyrin-1 (Erythrocyte ankyrin) (Ankyrin R)	71.63 8.72E-11	20%	38%	22469	11161524
					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%	80176	787888	4	XP_682036 hypothetical protein AN8767.2	71.25 1.14E-10	25%	47%	237-477	580786
					5	COG1381	RecO, Recombinational DNA repair protein (RecF pathway) [DNA replication. recombination. and repair].	30.33	6.27E-01	14%	30%	199305	62184	5	XP_787823 PREDICTED: similar to Ankyrin-1 (Erythrocyte ankyrin) (Ankyrin R)	48.52 7.90E-04	32%	52%	22110	626-708
														67	AAS13899 ankyrin repeat domain protein AAY54249 ankyrin domain protein	50.06 2.72E-04 67.78 1.26E-09 67.78 1.26E-09	24% 28%	39% 44% 44%	236466 263479	732-947
														9		45.44 6.69E-03	28% 25%	43%	263479 263476	613-827 698-901
														10	AAY61232 Guanosine polyphosphate pyrophosphohydrolases/synthetases homolog	50.45 2.08E-04	26%	40%	263481	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 homolog to thymidylate synthase.	201.73	4.40E-53	37%	55%	1210	1216	1	AAC96983 similar to Synechocystis ORF s111635, corresponds to GenBank Accession Number D90903	275.02 1.17E-72	59%	79%	1213	1215
					2	COG1351	TING BUCKLER AND THE RESIDENCE BUCKLER AND THE RESIDENCE	135.92	2.82E-33	31%	44%	16213	29237	2	ZP_00588145 Thymidylate synthase complementing protein	221.09 2.00E-56	52%	68%	1214	21236
														3 4	AAZ59097 Thymidylate synthase complementing protein ThyX	220.71 2.62E-56 217.62 2.22E-55	53% 54%	68% 70%	1214 12213	3213 6210
														5	ZP_01006033 thymidylate synthase ABB27663 Thymidylate synthase complementing protein ThyX	216.47 4.94E-55 214.54 1.88E-54	55% 51%	72% 67%	19213 1214	13210 1216
														7	YP 214756 Td ZP_00531300 Thymidylate synthase (FAD)	214.16 2.45E-54 212.62 7.13E-54	58% 52%	71% 66%	19213 1213	12210 1215
														9 10	AAM72779 thymidylate synthase, flavin-dependent ABB49306 Thymidylate synthase complementing protein ThyX	212.23 9.31E-54 212.23 9.31E-54	51% 53%	67% 74%	1214 19213	1216 13210
N035R	18207-20141	645	70,502	6.81	1	0000440	GImS, Glucosamine 6-phosphate synthetase, contains amidotransferase	622.20	5.19E-179	44%	63%	51643	1597		DAD45000	731.10 0.00E+00	61%	78%	51641	1594
NUSSR	16207-20141	045	70,502	0.01	'	0000449	and phosphosugar isomerase domains [Cell envelope biogenesis, outer membrane]. GFAT. Glutamine amidotransferases class-II (Gn-AT) GFAT-type. This		5.19E-179	444 76	03%	51045	1597	1	BAD15299 glutamine:fructose-6-phosphate amidotransferase GFAT	731.10 0.00E+00	0176	/070	51041	1394
							domain is found at the N-terminus of glucosamine 6-phosphate (GlcN-6- P) curthere (GLMS or GEAT) The glutaminase domain catalyzes amide													
					2	cd00714	nitrogen transfer from glutamine to the appropriate substrate. In this process, glutamine is hydrolyzed to glutamic acid and ammonia. GFAT	262.62	4.70E-68	47%	65%	52263	1215	2	AAC96468 PBCV-1 glucosamine synthetase	727.63 0.00E+00	60%	77%	51641	1593
							catalyzes the formation of glucosamine 6-phosphate from fructose 6- phosphate and glutamine, the initiating step in the biosynthesis of UDP-													
					3	COG2222	GICN-6-P AgaS, Predicted phosphosugar isomerases [Cell envelope biogenesis, outer membrane].	182.79	9.22E-47	29%	49%	298633	2329	3	CAE35165 glucosaminefructose-6-phosphate aminotransferase	512.69 1.79E-143	46%	63%	51643	1610
					4	pfam00310	GATase 2, Glutamine amidotransferases class-II	129.65	1.01E-30	36%	59%	52183	1136	4		509.61 1.52E-142	46%	63%	51643	1610
					5	CUG0034	PurF, Glutamine phosphoribosylpyrophosphate amidotransferase Nucleotide transport and metabolism GPATase N, Glutamine amidotransferases class-II (GN-AT) GPAT-	128.79	1.93E-30	31%	53%	48278	1232	5	YP_582342 glucosaminefructose-8-phosphate aminotransferase, isomerizing	508.45 3.38E-142	45%	64%	51643	1612
							type. This domain is found at the N-terminus of glutamine phosphoribosylpyrophosphate (Prpp) amidotransferase (GPATase). The													
					6	cd00715	glutaminase domain catalyzes amide nitrogen transfer from glutamine to the appropriate substrate. In this process, glutamine is hydrolyzed to		2.45E-29	32%	53%	52275	1224	6	ZP_00942990 Glucosaminefructose-6-phosphate aminotransferase (isomerizing)	508.06 4.41E-142	45%	64%	51643	1612
							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 homotimer													
							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 Nth hydrolase superfamily and is found at the N-terminus of enzymes such as													
							glucosamine-fructose 6-phosphate synthase (GLMS or GFAT), glutamine phosphoribosylovrophosphate (Prop) amidotransferase (GPATase)													
					7	cd00352	asparagine synthetase B (AsnB), beta lactam synthetase (beta-LS) and glutamate synthase (GltS). GLMS catalyzes the formation of glucosamine	112.11	1.89E-25	34%	48%	52246	1203	7	CAD13706 probable glucosaminefructose-6-phosphate aminotransferase	504.21 6.37E-141	45%	64%	51643	1612
							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- lactamace inhibiter disultance and GIIS extensions.													
							SIS, SIS domain. SIS (Sugar ISomerase) domains are found in many physiology and physiology and													
					8	pfam01380	domains are also found in proteins that regulate the expression of genes involved in synthesis of phosphosugars. Presumably the SIS domains	91.12	3.78E-19	37%	55%	333460	3131	8	YP_438548 glucosaminefructose-6-phosphate aminotransferase, isomerizing	499.59 1.57E-139	44%	65%	51643	1605
					9	COG0367	hind to the end-nordict of the nathway AsnB, Asparagine synthase (glutamine-hydrolyzing) [Amino acid transport and metabolism].	76.63	8.19E-15	25%	44%	51279	1195	9	YP_547531 glucosaminefructose-6-phosphate aminotransferase, isomerizing	497.66 5.96E-139	44%	63%	51643	1616
							annoon alle melebuildin.													

Gene Name	Genome Position	A.A. lenath	Peptide Mw		CDD Hit Number	COGs		Bit Score	E-value	% Identitv	% Positive	Query from-to	Hit from- I to	BLASTp Hit Number		Hit BLASTp Definition	Bit Score E-val	ue % Iden	, titv Pos	% sitive	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 coversion 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	71.40	3.39E-13	29%	48%	52246	1163	10	) ZP_00	00685730 Glucosamine-fructose-6-phosphate aminotransferase, isomerising	497.66 5.96E-	139	44%	64%	51643	1605
N040R	20182-23058	959	96,991	6.54		No Hit Found	hydrolyzes olutamine to olutamin acid and ammonia							1	BA	3AB83471 Vp260 like protein	1210.67 0.00E	+00	65%	75%	1957	1955
														2	BA	3AB83467 Vp260 like protein 3AB83468 Vp260 like protein	598.59 3.95E-	169	38%	55%	2957 27568	3954
														3	E BA	3AB83469 Vo260 like protein	148.29 1.41E 529.64 2.25E-	148	27% 37%	41% 52%	10951	346-997 1930
														5	5 BA	3AB83470 Vp260 like protein AC96490 PBCV-1 Vp260 protein	65.47 1.20E 61.23 2.27E		28% 26%	47% 42%	1179 595839	640798 15252
														7	' AA	AA86307 alvconrotein Vo260	48.91 1.176	-03	32%	48%	751-863	32153
														8		AC96397 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%	484872	14387
														9	AA (	AC96382 Asn/Thr/Ser/Val rich protein AC96386 Asn/Thr/Ser/Ile rich protein; similar to Rickettsia cell surface antigen,	110.15 4.26E		22%	38%	102-858	12779
														10	) AA	AC96386 Astrinization for protein, similar to Ricketsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	109.00 9.49E	-22	22%	37%	24831	153-950
N045R	2309327454	1454	149,210	6.85		No Hit Found								1		3AB83469 Vp260 like protein 3AB83468 Vp260 like protein	1417.52 0.00E 1399.42 0.00E		53% 52%	65% 64%	101452 101452	1
														23	BA BA	AB83470 Vp260 like protein	1367.44 0.00E	+00	56%	68%	101290	1-1283
														4		3AB83467 Vp260 like protein 3AB83471 Vp260 like protein	576.63 2.54E- 545.81 4.79E-		35% 36%	51% 52%	41131 1932	51141 1940
														6	6 AA	AC96490 PBCV-1 Vp260 protein	101.68 2.39E	-19	28%	40% 40%	27464	5461011
														8	3 AA	AA86307 glycoprotein Vp260 AC96382 Asn/Thr/Ser/Val rich protein	84.34 3.95E 81.26 3.34E		26% 20%	40% 35%	455-855 16599	8398 5731231
														9		AC96386 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%	88708	6221242
														10	) AA	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%	27282	551-812
N052B	27493-31881	1463	149.949	6.74	1	nfam02061	DUF342, Protein of unknown function (DUF342). This family of bacterial	33.39	2.37E-01	29%	46%	441-595	189318	1	DA		1630.92 0.00E	+00	58%	71%	101461	1
NUSZR	2/493-31001	1403	149,949	0.74			proteins has no known function. The proteins are in the region of 500-600 amino acid residues in length COG5520, O-Glycosyl hydrolase [Cell envelope biogenesis, outer									3AB83469 Vp260 like protein						
					2	COG5520	membranel. COG1315. Predicted polymerase, most proteins contain PALM domain.	31.16	9.33E-01	30%	44%		351412	2		3AB83468 Vp260 like protein	1603.57 0.00E		56%	70%	101461	1-1462
					3	COG1315	HD hydrolase domain and Zn-ribbon domain [DNA replication, recombination, and repair].	31.06	9.75E-01	26%	42%	23128	256347	3		3AB83470 Vp260 like protein	1542.32 0.00E		60%	72%	101329	1-1328
														4		3AB83467 Vp260 like protein 3AB83471 Vp260 like protein	602.44 4.35E- 567.77 1.19E-		35% 38%	52% 53%	41128 1940	51135 1949
														6	6 AA	AC96490 PBCV-1 Vp260 protein	203.37 5.88E	-50	28%	40%	8883	42999
														7		AAA86307 glycoprotein Vp260 AC96397 Asn/Thr/Ser/Val_rich protein; similar to Rickettsia cell surface antigen,	58.92 1.79E 60.46 6.15E		22% 23%	35% 36%	7491411 469956	43677 9565
														9		corresponds to Swiss-Prot Accession Number P15921 AC96382 Asn/Thr/Ser/Val rich protein	54.30 4.41E		21%		7461451	79765
														10	) AA	AC96386 Asn/Thr/Ser/lle rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	49.29 1.42E		21%		7451461	24768
N059R	3199636417	1474	151,244	6.52	1	pfam03961		31.85	5.63E-01	29%	46%	441-595	189318	1	ВА	3AB83468 Vp260 like protein	1989.16 0.00E	+00	67%	79%	101472	1-1462
					2	COG5520	amino acid residues in lenoth COG5520, O-Glycosyl hydrolase [Cell envelope biogenesis, outer	31.93	5.74E-01	30%	44%	518-584	351412	2	, RA	3AB83469 Vp260 like protein	1955.64 0.00E	+00	67%	77%	101472	1
					-	0000020	membranel.	01.00	0.742 01	0070	4474	010 004	001 412	3		3AB83470 Vp260 like protein	1516.52 0.00E		59%	72%	101336	1-1323
														4	BA	3AB83467 Vp260 like protein	632.10 5.16E-	179	37%	54%	4	51103
														5	5 AA	AAB83471 Vp260 like protein AC96490 PBCV-1 Vp260 protein	588.57 6.55E- 192.97 8.01E	-47	38% 29%	55% 42%	91940	1949 15839
														7	' AA	AA86307 glycoprotein Vp260 AC96397 Asn/Thr/Ser/Val rich protein; similar to Rickettsia cell surface antigen,	48.14 3.19E		21%		7491472	43664
														8		AC96397 Corresponds to Swiss-Port Accession Number P15921 AC96382 Asn/Thr/Ser/Val rich protein	65.47 1.93E 95.13 2.27E		23% 25%	39% 40%	17282 4599	472-812 6221231
														9 10	) AA	AC96386 AshThr/Ser/He rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	122.09 1.74E		23%	38%	21939	15983
N067R	36531-41027	1499	156,160	6.02	1	COG1664	determination [Cell envelope biogenesis, outer membrane].	33.75	1.66E-01	19%	41%	492-610	20120	1	BA	3AB83467 Vp260 like protein	1314.67 0.00E	+00	59%	68%	11169	1-1163
					2	COG1315	COG1315, Predicted polymerase, most proteins contain PALM domain, HD hydrolase domain and Zn-ribbon domain [DNA replication,	31.83	5.64E-01	27%	40%	498-617	185317	2	BA	3AB83469 Vp260 like protein	167.93 2.81E	-39	29%	42%	10603	327-842
							recombination and renairl							3	ва	3AB83468 Vp260 like protein	113.62 6.29E	-23	29%	42%	3336	397783
														4	BA	AB83470 Vp260 like protein AB83471 Vp260 like protein	81.26 3.46E 444.12 2.03E-		35% 33%	50% 47%	3194 3944	641-813 2944
														6	5 AA	AC96490 PBCV-1 Vp260 protein	56.61 9.13E	-06	31%	46%	10166	8561018
														7	Υ ΑΑ 	AA86307 glycoprotein Vp260 AC96397 Asn/Thr/Ser/Val rich protein; similar to Rickettsia cell surface antigen,	54.68 3.47E		30%	46%	10166 10148	670-832 670-805
														8		Corresponds to Swiss-Prot Accession Number P15921 AC96382 Asn/Thr/Ser/Val rich protein	46.60 9.44E 56.23 1.19E		31% 20%	46% 35%	10148 7561490	670-805 145-828
														10	) AA	AC35172 cell surface antigen rOmpA	53.14 1.01E		24%		388627	230-470
							Capsid_Iridovir, Iridovirus major capsid protein. This family includes the															
N074R	41091-42326	412	46.415	6.11	1	nfam04451	major capsid protein of iridoviruses, chlorella virus and Spodoptera ascovirus, which are all dsDNA viruses with no RNA stage. This is the	280.66	2.04E-76	36%	54%	5406	3443	1		AC27494 putative capsid protein	476.09 1.04E-	132	97%	99%	21250	1230
107410	41051 42020	412	40,410	0.11		planotio	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	200.00	2.042.70	0070	047	0 400	0 440				410.00 1.042	102	0170	0070	21 200	1 200
					2	COG0770	dvcoprotein MurF, UDP-N-acetylmuramyl pentapeptide synthase [Cell envelope	37.60	3.15E-03	32%	47%	1068	382441	2	, AA	AC96798 PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession	294.28 5.63E	-78	40%	55%	5410	3437
					-	0000110	biogenesis. outer membranel.	07.00	0.102.00	02.10	41 /		002 441	3	BA	BAA76601 major capsid protein MCP1	291.97 2.808		40%	55%	5410	3437
														4	BA	3AA22198 major capsid protein Vp54	286.96 8.99E 285.42 2.62E	-76	40% 40%	54% 54%	5410 5410	3437 3436
														6	5 AA	AA76600 major capsid protein AC27492 major capsid protein Vp49	281.57 3.78E	-74	41%	54%	5410	3432
														7	1	1M4X_C Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model 1M3Y_D Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid	265.00 3.66E		39%	54%	27410	1413
														8		Containing. The Virus AC96379 contains aminoacyl-tRNA synthetase class-II signature	265.00 3.66E 229.57 1.70E		39% 34%	54% 53%	27410 5410	1413 3403
														9	, да ) да	AC96379 contains aminoacyl-IRNA synthetase class-II signature AC96917 similar to Chlorella virus PBCV-1 ORF A11L, corresponds to GenBank Accession Number U42580	229.57 1.70E 229.18 2.22E		34% 33%	53% 53%	5410 5410	2400
N079R	42805-43311	169	18.821	11.55		No Hit Found										Accession Number U42580		-				
																				0.401	05.05	
N081L	43657-43367	97	10,019	11.01		No Hit Found								1		AC96569 A201L	124.41 1.14E		78%	91%	2595	2494
N082R	43922-44515		22,100	8.10		No Hit Found								1		AC96669 A301L	59.69 6.39E		30%	52%	3132	2135
N084R	4454545186	214	23,739	4.57		No Hit Found								1	AA	AC96666 A298L	150.60 3.25E	-35	40%	60%	1208	1220
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 dvnamics].	28.91	2.47E-01	23%	48%	68110	442482	1	AA	AC96664 A296R	80.88 1.44E	-14	64%	78%	62126	72135
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	1215	1	BA	3AA20342 vChta-1	429.10 1.12E-	118	66%	77%	18335	5321

Gene	Genome		Peptide	la	CDD Hit	COGs	COG Definition	Bit Score	F-value	%	%	Query		BLASTp Hit		BLASTp Definition	Bit Score	F-value	%	%	Query I	Hit from-
Name	Position	lenath	Mw	р.	Number		chitosanase_glyco_hydro_46, Glycosyl hydrolase family 46 chitosanase	211 00010	2 10:00	Identitv P	ositive	from-to	to	Number	Accession	DENOTP DOMINION	Dir Oboro	lo lo	lentitv Po	ositive	from-to	to
							domain. This family are composed of the chitosanase enzymes which hydrolyzes chitosan, a biopolymer of beta (1,4)-linked -D-glucosamine															
							(GIcN) residues produced by partial or full deacetylation of chitin. Chitosanases play a role in defense against pathogens such as fungi and															
					2	cd0097	are found in microorganisms, fungi, viruses, and plants. Microbial 78 chitosanases who members are the most prevalent can be divided into 3 subclosers based on the specificity of the closurage partial	203.65	2.15E-53	31%	45%	107324	1220	2	AAC96660 PBCV-1 chitosanase		426.40	.28E-118	65%	77%	18335	5321
							subclasses based on the specificity of the cleavage positions for partial acetylated chitosan. Subclass I chitosanases such as N174 can split GIcN-GIcN and GIcNAc-GIcN linkages, whereas subclass II chitosanases															
							such as Bacillus sp. no. 7-M can cleave only GicN-GicN linkages. Subclass III chitosanases such as MH-K1 chitosanase are the most															
							versatile and can split both GlcN-GlcN and GlcN-GlcNAc linkages 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															
					3	cd0098	<sup>35</sup> novel NTPase that has been shown to hydrolyze nonstandard nucleotides, such as hypoxanthine/xanthine NTP, but not standard	31.38	1.98E-01	21%	34%	56168	8120	3	AAA19865 chitosanase precursor		86.27	1.78E-15	28%	44%	93329	17266
							nucleotides PrkA, PrkA serine protein kinase. This is a family of PrkA bacterial and															
					4	pfam0679	archaeal serine kinases approximately 630 residues long. PrkA possesses the A-motif of nucleotide-binding proteins and exhibits distant	30.66	3.07E-01	25%	45%	20117	522610	4	BAA94840 chitosanase		86.27	1.78E-15	29%	47%	96304	41252
							homology to eukaryotic protein kinases. Note that many family members are hypothetical y RuvA, Holliday junction resolvasome, DNA-binding subunit [DNA							_								
					5	COG063	replication, recombination, and repairl.	30.21	3.79E-01	25%	46%	39121	102181	5	1CHK B Chain B, Streptomyces	N174 Chitosanase Ph5.5 298k		3.04E-15 8.86E-15	29% 29%	41% 46%	107328 114329	37264 12226
														7	P48846 Chitosanase precursor ZP_00908039 chitosanase		82.80 82.42	1.97E-14 2.58E-14	29% 31%	46% 48%	101329 117324	39266 38246
														9 10	ABC17783 secreted chitosanase p BAB19276 chitosanase	recursor	81.65	4.40E-14 9.79E-14	28% 27%	46% 41%	107324 107328	51274 38265
N089L	4800946927	361	39,928	8.72	2	No Hit Foun	d							1	AAC96654 A286R		362.46	1.39E-98	51%	69%	17344	22362
							Sod_Cu, Copper/zinc superoxide dismutase (SODC). superoxide															
N091R	4804848560	171	18.329	8.00	0 1	ofom0008	dismutases (SODs) catalyse the conversion of superoxide radicals to hydrogen peroxide and molecular oxygen. Three evolutionarily distinct	177.74	5.41E-46	58%	73%	16165	1152	1	contains Cu/Zn super	oxide dismutase signatures 1 and 2; similar to	261.15	0.945 60	76%	85%	10169	28187
NUSTR	4004040300	1/1	10,329	0.00	0 1	piamoood	30 families of SODs are known, of which the copper/zinc-binding family is one. Defects in the human SOD1 gene cause familial amyotrophic lateral entersity (Law Celling ensure). Churchen is on solid human family is	177.74	3.4 IE-40	36%	7376	10105	1 152		Accession Number P07	peroxide dismutase, corresponds to Swiss-Pro 7509	201.15	9.04E-09	/0%	03%	10109	20-107
							sclerosis (Lou Gehrig's disease). Structure is an eight-stranded beta sandwich. similar to the immunoalobulin fold															
							Cu-2n_Superoxide_Dismutase, Copper/zinc superoxide dismutase (SOD). superoxide dismutases catalyse the conversion of superoxide radicals to molecular oxygen. Three evolutionarily distinct families of															
					2	cd0030	SODs are known, of which the copper/zinc-binding family is one. Defects of in the human SOD1 gene causes familial amyotrophic lateral sclerosis	163.51	1.16E-41	53%	69%	18158	2144	2	Q751L8 Superoxide dismutase		177.56	1.43E-43	57%	71%	16164	1153
							(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].															
														3		mutase; some mutations are analogous to those rophic lateral sclerosis) in humans: Sod1p		2.44E-43	58%	72%	16162	1151
														4 5	XP_503850 hypothetical protein P09670 Superoxide dismutase		175.64	4.16E-43 5.44E-43	59% 58%	70% 71%	16162 18164	1151 3150
														6 7	XP_445379 unnamed protein produ	de Dismutase With The Reduced Bridge Broken Ict	174.87 174.87	9.27E-43	58% 57%	72% 70%	17162 16162	1150 1151
														8	AAS54170 AGL321Wp AAX84946 Cu/Zn superoxide dism	utase	174.10 174.10	1.58E-42	56% 57%	71% 70%	17164 18165	21172 4152
														10		ure Of Yeast H48q Cuznsod Fals Mutant Analog	173.71		57%	71%	16162	1151
N092L	4915048557	198	23,569	8.76			GpmB, Fructose-2,6-bisphosphatase [Carbohydrate transport and metabolism]. J9 AceE, Pyruvate dehydrogenase complex, dehydrogenase (E1)		6.20E-03	28%	42%	482	180	1				7.46E-49	53%	72%	8176	3171
					2	COG260	omponent lEnerov production and conversion1. GNT-I, GNT-I family. Alpha-1,3-mannosyl-glycoprotein beta-1,2-N-	31.41	7.99E-02	38%	57%	1556	1961	2	BAE64611 unnamed protein produ	ict	57.00	4.14E-07	27%	50%	8121	6119
							acetylglucosaminyltransferase (GNT-I, GLCNAC-T I) EC:2.4.1.101 transfers N-acetyl_D-glucosamine from UDP to high-mannose															
					3	pfam0307	glycoprotein N-oligosaccharide. This is an essential step in the synthesis of complex or hybrid-type N-linked oligosaccharides. The enzyme is an	28.39	6.84E-01	18%	39%	1392	60142	3	XP_750408 hypothetical protein Afr	u1g06590	52.37	1.02E-05	26%	50%	8121	26140
							integral membrane protein localised to the Golgi apparatus, and is probably distributed in all tissues. The catalytic domain is located at the C- terminus															
					4	COG206	SiXA, Phosphohistidine phosphatase SixA [Signal transduction mechanisms].	27.94	7.67E-01	21%	36%	5106	1107	4	2. 2			6.61E-05	29%	47%	8121	9123
														5	XP_748836 hypothetical protein Aft	J7g06750	48.14	1.92E-04	24%	43%	8168	22174
							PVA, Penicillin V acylase (PVA), also known as conjugated bile salt acid hydrolase (CBAH), catalyzes the hydrolysis of penicillin V to yield 6-															
N093R	4917850020	281	31,162	6.76	6 1	cd0054	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	185.80	4.42E-48	32%	47%	2275	1297	1	AAC96652 PBCV-1 amidase		314.69	2.23E-84	54%	71%	1278	1279
					2	000304	orecursor. PVA forms a homotetramer 10 COG3049, Penicillin V acylase and related amidases [Cell envelope	167.43	1.47E-42	28%	45%	1278	22323	2	ZP_00318651 COG3049: Penicillin V	and a second and a data of a second second	444 74	2.59E-32	33%	51%	39276	61304
					2	000004	biodenesis. outer membranel. CBAH, Linear amide C-N hydrolases, choloylglycine hydrolase family.	107.45	1.47 2-42	20%	4370	1-270	22525	2	21_00310031 COG3049: Peniciliin V	acylase and related annuases	141.74	2.382-32	3370	51%	35-270	01-304
					3	pfam0227	This family includes several hydrolases which cleave carbon-nitrogen 75 bonds, other than peptide bonds, in linear amides. These include choloylqlycine hydrolase (conjugated bile acid hydrolase, CBAH)	141.29	9.63E-35	31%	48%	2275	1298	3	AAP08002 Choloylglycine hydrola:	se	141.35	3.38E-32	35%	51%	49275	68298
							cnowylgtycine nydrolase (conjugated bile acid nydrolase, CBAH) EC:3.5.1.24, penicillin acylase EC:3.5.1.11 and acid ceramidase EC:3.5.1.23															
							CGH like Cholovialycine bydrolase (CGH) like This family of															
					4	cd0193	bol do la sectiona de la sectiona	107.59	1.40E-24	28%	39%	2268	1270	4	ZP_01094011 choloylglycine hydrolas	e-like protein	140.20	7.52E-32	33%	47%	2276	22327
							CGH, CGH Choloylglycine hydrolase (also known as bile salt hydrolase) is an intestinal bacterial enzyme responsible for the															
					5	cd0190	2 deconjugation and subsequent dehydroxylation of conjugated cholic acid (CA) to form deoxycholic acid (DCA). CGH has a conserved Ntn	75.04	9.35E-15	25%	44%	2261	1270	5	AAT60567 choloylglycine hydrolas	e	138.27	2.86E-31	33%	51%	49275	68298
							hydrolase fold similar to those of penicillin v acylase (PVA) and acid ceramidase (AC.)															
														6 7	ZP_01187466 Choloylglycine hydrolas ZP 00238968 choloylglycine hydrolas	e family protein	136.73	2.86E-31 8.32E-31	33% 33%	51% 51%	49275 49275	68298 68298
														8 9	NP 980098 choloylqlycine hydrolas ABB11030 Penicillin amidase		132.11	1.57E-29 2.05E-29	32% 28%	50% 45%	49275 1278	68298 1315
														10	AAL51724 CHOLOYLGLYCINE H			3.49E-29	30%	44%	2276	37337
N095L	5108850195	298	33,002	6.9	1 1	COG038	88 COG0388, Predicted amidohydrolase [General function prediction only]. CN hydrolase Cathon pitrogen hydrolase This family contains.	182.28	5.50E-47	35%	52%	1296	1272	1	AAC96446 contains ATP/GTP-bir synthetase correspond	ding site motif A; similar to rat beta-alanine Is to Swiss-Prot Accession Number Q03248	421.01	.43E-116	68%	79%	2295	4297
					2	pfam0079	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	145.10	9.22E-36	35%	53%	4177	1174	2	ZP_01112744 probable hydratase		280.80	3.93E-74	49%	64%	2295	3283
					3	00004	EC:35.112. Beta-ureidonzoitoase EC:35.16. Lnt, Apolipoprotein N-acyltransferase [Cell envelope biogenesis, outer	51.94	9.70E-08	21%	35%	35215	269451	3	AAK75046 carbon-nitrogen hydrol	ace family protein	278 87	1 49E. 72	46%	65%	5294	6289
					4	pfam0145	7 Pentidase M8 Leishmanolysin	31.87	9.54E-02	27%	47%	147-198	167212	4			276.94	5.67E-73	46%	65%	5294	6289
					5	pfam0363	39 Glyco_hydro_81, Glycosyl hydrolase family 81. Family of eukaryotic beta- 1.3-alucanases	31.17	1.79E-01	21%	35%	173-278	243353	5	BAC60037 putative carbon-nitroge ZP 01259147 putative carbon-nitroge		275.40 275.40	1.65E-72	49% 49%	65% 64%	6291 6291	8280 8280
														6 7 8	ZP 00831984 COG0388: Predicted a	midohydrolase	273.48		49% 46% 47%	64% 63% 63%	6291 2294 4295	8280 3285 9291
														8	GAD40070 Deta-alanine synthase		273.09	0.18E-72	40 / 70	03%	4295	9291

Gene Name	Genome Position	A.A. lenath	Peptide Mw	pl	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identitv	% Positive	Query from-to	Hit from-BLA toNu		Hit BLASTp Definition Accession CAG77171 putative carbon-nitrooen hydrolase ABB08352 Nitrilase/vande hydralase and acoiloporotein N-acvitransferase	Bit Score 272.71 272.32	1.07E-71	% lentitv P 47% 48%	% ositive 62%	Query from-to 2294 4295	Hit from- to 3285 7294
N099R	5114352117	325	38,567	10.78	1	COG5077	COG5077, Ubiquitin carboxyl-terminal hydrolase [Posttranslational modification. protein turnover, chaperones].	29.20	8.06E-01	21%	40%	67256	529722	1	Abbossz williadercyanide nyoratase and apolipoprotein w-acylinaristerase		2.21E-129	40% 73%	84%	4295	1302
N101L	5239452128	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%	78137	45105	1 2 3 4 5 6 7	AAU06294 hypothetical protein A253R AAU06289 hypothetical protein A253R AAU06289 hypothetical protein A253R AAU06220 hypothetical protein A253R AAU06221 hypothetical protein A253R AAU05211 hypothetical protein A253R	78.18 77.80 55.45 55.07 52.76	9.24E-14 9.24E-14 1.21E-13 6.41E-07 8.38E-07 4.16E-06 2.06E-05	35% 35% 39% 40% 38% 39%	56% 57% 56% 61% 62% 58% 60%	25147 25147 2597 2593 2597 2590	20146 20146 2091 2091 2087 2091 2084
N103R	52977-53915	313	35,392	10.08	1	cd00283	GITY-VIG_LOBERN, GITX(110-11)YIG tampy of class I homing endonucleases. Cheminus (GIV-YIG, Chem), Homing endonucleases portole the mohility of intron or inten 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 stored into forur families based on the presence of these motifs in their respective N- termin: LAGLIDBOG, Hic-Syb sox, NNH, and GIY-VIG. This CD contains several but not all members of the GIY-VIG family. The C-terminus of GIY-VIG 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-field, and a helicul-markelic. Some allo contain a zure finger (is.1-Tervi) which is not required for DNA homing domain to divert the homing site at a field distance from the intro nearbino value. The onder ducleases repeat motif; Repeat of unknown	64.64	1.55E-11	41%	61%	172–307	1–113	1	AAA88832 unknown	181.03	4.57E-44	35%	51%	1309	1242
					2		function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished)		5.75E-04	40%	63%	254309	1-53		AAC96973 similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	175.64		36%	50%	1308	1224
					3		GIYc, GIY-YIG type nucleases (URI domain); . NUMOD1, NUMOD1 domain	38.52 36.95	1.23E-03 3.61E-03	29% 48%	49% 67%	171 254288	1-70 1-34	3	AAC96655 PBCV-1 33kd peptide AAC96862 similar to PBCV-1 ORF A315L, corresponds to GenBank Accession		1.38E-24 4.77E-09	40% 37%	55% 58%	11179 170280	15192 97196
					4	pramu7453	NUMUUT, NUMUUT oomain	36.95	3.61E-03	48%	67%	254-288	1—34	5 6 7 8 9	AAU-309602 Number M/14400 AAC46248 OFE711 NP. 803933 SepD AAC46244 OFE701 AAX03955 intron encoded Bmol AAX03955 intron encoded Bmol AAX03956 intron encoded Bmol AAX03956 intron encoded Bmol AAX03926 United Bmol Code Bmol Code Bmol AAX0328 Code Bmol AAX0428 Code Bmol AAX048 Code Bmol AAX04	62.77 77.03 59.31 45.44 62.00	4.77E-09 1.81E-08 9.29E-13 2.00E-07 2.99E-03 3.09E-08 3.09E-08	37% 26% 44% 34% 30% 30%	58% 69% 63% 47% 49% 49%	120178 1300 190247 169251 139276 139276	97196 93151 1231 176240 102195 100215 75190
N106L	5459353910	228	25,018	8.41		No Hit Found								1	AAC96630 A262L AAC96631 A263L		3.20E-26 6.50E-03	47% 41%	66% 50%	111215 2591	2103 66123
N107L	5483654615	74	7,911	10.56		No Hit Found	RuvC_resolvase, Holliday junction resolvases (HJRs) are endonucleases that specifically resolve Holliday junction DNA							1	AAC96633 A265L		2.30E-17	62%	74%	271	101-170
N109L	5536654983	128	14,609	7.93	1	cd00529	intermediates during homologous recombination. HJR&aposs occur in archaea, bacteria, and in the mitchcondria of certain fungi, however this CD includes only the bacterial and mitochondrial HJR&aposs. These are referred to as the RucC family of Holiday junction resolvases, RucC being the E.coli HJR. RucC and its orthologs are homodimers and are structurely similar to RNase H and Hsn70	32.12	1.87E-02	33%	48%	346	254	1	AAC96633 A265L	68.55	7.31E-11	66%	80%	79120	142
N110R	5548957474	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	375.38	1.13E-104	36%	57%	31594	2587	1	NP_567404 HAK5; potassium ion transporter	324.32	9.33E-87	33%	53%	9592	38640
					2 3		(KUP). veast (HAK). and olant (AtKT) sequences as members Kup, K+ transporter [Inorganic ion transport and metabolism].	369.16	6.95E-103 4.67E-03	34% 17%	53% 34%	27-660 89213	14627 6133	2 4 5 6 7 8 9 10	NP 194095 TRH1 (TNY ROOT HAIR 1); potassium ion transporter GAD07337 putative potassium transporter CAD20319 putative potassium transporter CAD20319 putative potassium transporter CAD27319 putative potassium transporter CAB73319 putative potassium transporter GAB73319 putative potassium transporter BAE03236 potassium transporter BAE03348 potassium transporter	307.38 306.22 305.83 305.06 303.91	1.39E-82 9.04E-82	32% 31% 34% 31% 31% 31% 32% 32% 32% 33%	52% 48% 51% 53% 51% 51% 51% 52% 52%	28-598 26-660 28-621 21-582 28-629 28-598 27-598 21-587 26-582	11-605 73-748 21-649 85-669 74-693 26-632 20-616 107-698 14-601
N115L	5784057478	121	13,617	4.55		No Hit Found								0	No Hit Found No Hit Found						
N116R	5791658236	107	11,590	8.04	1	pfam03833	PolC_DP2, DNA polymerase II large subunit DP2 PsaN, Photosystem I reaction centre subunit N (PSAN or PSI-N). This	27.98	2.76E-01	53%	60%	7792	632647	1	YP_619251 hypothetical protein Ldb1404	42.36	5.68E-03	39%	56%	5499	58103
					2	pfam05479	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 verv likelv involved in the docking of plastocyanin	27.00	5.64E-01	38%	58%	79105	92118	2	AAV85705 At2g33790	41.97	7.42E-03	40%	54%	59102	4790
N117R	5825058678	143	16,313	11.34	1	COG1422	COG1422, Predicted membrane protein [Function unknown].	28.66	3.32E-01	25%	55%	4693	62106	1	AAC96504 a136R	57.00	2.21E-07	32%	50%	1141	1146
N119L	59913-58864	350	38,855	4.65	1	cd02180	GH16 Jaminarinase, Laminarinase, also known as glucan endo-13-beta D-glucosidase, is a glucosh yhordolase family for ember that hydrolyzes 1,3-beta-D-glucosidic inkages in 1,3-beta-D-glucans such as laminarins, curdians, paramytions, and pachymans, with very limited action on mixed- link (13-14-h-beta-D-nicrans Glyco, hydrolases, 16, The O-Glycosh hydrolases are a widespread	143.48	2.96E-35	36%	50%	88343	1236	1	AAC96462 PBCV-1 beta-1,3 glucanase	416.77 5	5.90E-115	56%	71%	9345	12361
					2	cd00413	group of enzymes that hydrolyse the glycosidic bond between two or more carbolydrates, 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 including glycoside hydrolase family 16. Family 16 includes lichenase, xyloglucan endo-beta-13-glucanase, endo-beta-13-14-glucanase, and endo-beta-al-aglucanase. (Act Th, beta-garase, kappa- carrageness, endo-beta-13-glucanase, endo-beta-13-14-glucanase, and endo-beta-glacosidase, all or which have a conserved jelly ofl 10d with a deep active site channel harboring the catalytic residues. OH16 laminamase like, A beta-13-glucanase)-like	93.52	3.36E-20	34%	47%	92–343	1216	2	P23903 Glucan endo-1.3-beta-glucosidase A1 precursor ((1-&gt.3)-beta-glucan endohydrolase) ((1-&gt.3)-beta-glucanase A1)	118.24	4.34E-25	31%	45%	64343	400-678
					3	cd02182	Grino-Janimaniase-Jinki, A Goisti, A-guidanase (utilimatinase-pixe) 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 rull fold with an active site channel. The bacterial members contain an additional C-terminal archohytrab-initing module (CRM)	76.10	6.22E-15	30%	44%	85343	1256	3	YP_435911 Beta-glucanase/Beta-glucan synthetase	115.55	2.81E-24	32%	45%	77343	322-571
					4		SKN1, Beta-glucanase/Beta-glucan synthetase [Carbohydrate transport and metabolism]		1.56E-10	25%	39%	74348	29267		AAC60453 beta-1,3-glucanase		1.40E-23	31%	44%		424681
					5		allying automatical field of the second seco	55.70	5.68E-10 7.90E-09		45% 37%	190–342 89–329	45182 2283		ZP_01113030 hypothetical protein MED297 09826 ZP_01351075 Glucan endo-1,3-beta-D-glucosidase		3.11E-23 3.11E-23	31% 29%	45% 43%	84343 68343	33297 38331

	ene ame	Genome Position	A.A. lenath	Peptide Mw	pl	CDD Hit Number	COGs	GH16 GPI glucanosyltransferase, GPI (glycosylphosphatidylinositol) -	Bit Score	E-value	% Identitv P			Hit from-BL		Hit Accession	BLASTp Definition	Bit Score	E-value Ide	% entitv Po	% sitive f	Query rom-to	Hit from- to
						7	cd0218	glucanosyltränsferase is a CPI-anchored membrane protein present in the fungal cell will that is thought to play an important role in cell wall biosynthesis. GPI-glucanosyltränsferase beforgs to a family of glycosyl 31 yktrollases that includes lichenae, xyloglucan anotaneglycosylase (XET), beta-agarase, kappa-carrageenae, vidoucan anotaneglycosylase (XET), beta-agarase, kappa-carrageenae, wido-beta-13-glucanaee, endo-beta-13-14-glucanase, and endo-beta-glacitosideae, all of which anota-beta-13-14-glucanase, and endo-beta-glacitosideae, all of which anota-beta-13-14-glucanase, and endo-beta-glacitosideae, all of which the ratiator inesitues. Uncleanse, allo known all 3-14-the-allocanaes, is child. [Lichenaes, Lichenaes, allo known all si-14-beta-glucanase), se	38.68	1.04E-03	28%	49%	190–283	58144	7	YP_56360	6 glycoside hydrolase, family 16	111.69	4.06E-23	31%	44%	76343	39325
						8	cd0217	GH16 jichenase, Lichenase, alto Knovn as 1.314,4-bet-glucanase, is a member of glycosyl hydroidas family 16, hat specifically deves 1.4-beta D-glucosidic bonds in mixed-linked beta glucans that also contain 1.3- beta-D-glucosidic linkages. Natural substrates of beta-glucanse are 5 beta-glucans from grain endosperm cell walls or linkenan from the lashadic moss. Centrain slandcar. This protein is found hot only in beta-did ut all provide the state of the state of the state of the state of the state of the state of the state of the state of the lashed moss. This protein is found hot of the beta-did ut all provides the state of the state of the formion a common ellivith lebet-state of the state of the GH16 jappa_carrageenase, Kappa-carrageenase degrades kappa- carrageenase which are the gel-forming, suffield 1.3 sylba-1.4 beta-		2.12E-03	29%	46%	126–281	30157	8	ZP_00637494	4 Giucan endo-1,3-beta-D-glucosidase	109.00	2.63E-22	31%	43%	85343	48325
						9	cd0217	galactans that make up the cell walls of marine red algae such as proficed provides. Kappa-caragenesses exist hoaderia belonging to at least three phylogenetically distant branches, including pecudoalterooms, planchorryotes, and baceroidetes. This domain active site cavity referred to ass lethern float. with a tunnel-shaped GH16_KET, Xylogucan endotransplycosylases (XETs) cleave and		1.09E-02	30%	46%	81212	3117	9	NP_76320	1 Beta-glucanase/Beta-glucan synthetase	107.84	5.86E-22	28%	43%	49343	12333
						10	cd0217	religate sylogican polymers in plant cell walls via a transplycosylation mechanism. Thus, XET is a key enzyme in all plant processes that 6 require cell wall remodeling. Even though the overall structure of XET is a curved beta-sandwich similar to other enzymes in the glycosyl hydroiase family 16, parts of its substrate binding cieft are more meminiscent of the distantiv related diversol Methodisas family 2	33.64	3.40E-02	34%	46%	190–283	50142	10	BAD6324	2 endo-beta-1,3-glucanase	107.46	7.66E-22	31%	44%	83343	31277
N124	4R	6007861757	560	63,740	8.4	9 1		5 COG1215, Glycosyltransferases, probably involved in cell wall bioenesis ICell enveloce bioenesis, outer membranel. Chitin_synth 2, Chitin synthase. Members of this family are fungal chitin synthase EC:2.4.1.16 enzymes. They catalyse chitin synthesis as	90.76	4.57E-19	25%	44%	39390	10328	1	AAD2664	1 hyaluronan synthase	832.40	0.00E+00	69%	82%	1550	8557
						2	pfam0314	12 follows: UDP-N-acetyl-D-glucosamine + {(1,4)-(N-acetyl-beta-D-glucosaminy)})(N) & <=> UDP + {(1,4)-(N-acetyl-beta-D- nkurosaminv))/N+1) Glycos transf 2, Glycosyl transferase. Diverse family, transferring sugar	54.92	2.35E-08	23%	44%	186-345	193373	2	AAC9646	6 PBCV-1 hyaluronic acid synthetase	828.55	0.00E+00	69%	82%	1550	8557
						3	pfam0053	from LIDD alvasor, LIDD N seekd, aslasteromine, CDD mersons or	47.14	5.09E-06	24%	43%	86283	2167	3		3 hyaluronan synthase 5 hyaluronan synthase 2	826.62		68% 29%	82% 48%	1550 7530	8557
															5 6 7 8 9 10	AAI0907: AAC53301 BAC3773: XP_52822: BAB6326	o Tryational synthase 2 P Hydronan synthase 2 9 Tystaronan synthase 2 3 unamed protein product 2 PREDICTED: similar to hydronan synthase 2 4 Hydroncia cid synthase 2 8 Tystaronan synthase 2	218.39 218.39 218.39 218.39 218.39 216.85	5.89E-55 5.89E-55 5.89E-55 5.89E-55	29% 29% 29% 29% 29%	48% 48% 48% 48% 48% 48%	7530 7530 7530 7530 7530 7530	12-532 12-532 12-532 24-544 12-532 12-532
N128	BR	6194462906	321	36,035	i 8.7	5 1	pfam0133	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%	50223	1192	1	AAC9647	1 PBCV-1 mRNA guanylyltransferase	326.25	9.18E-88	52%	68%	2314	12324
						2	COG522	CEG1, mRNA capping enzyme, guanylyltransferase (alpha) subunit [RNA processing and modification]		8.40E-15	21%	37%	20317	17362	2	-	B Chain B, Structure Of Guanylylated Mrna Capping Enzyme Complexed With Gtp			52%	68%	2314	12324
						3	pfam0391 COG194	9 mRNA cap C, mRNA capping enzyme, C-terminal domain 7 lspE, 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate synthase [Lioid metabolism].	50.74 29.47	2.46E-07 5.92E-01	26% 27%	44% 38%	228-315 135-183	3110 5098	3		8 putative RNA guanylytransferase 2 unnamed protein product		5.32E-11 1.55E-10	26% 24%	42% 43%	44311 9317	221-479 234-560
								Liou metabolismi.							5 6	AAT6813 AAX8237	3 mRNA capping enzyme 3 mRNA capping enzyme	68.17	3.45E-10 4.50E-10	23% 24%	44% 39%	8317 44311	234566 220478
															7	NP_974263	3 mRNA guanylyltransferase/ phosphoprotein phosphatase/ proteir tvrosine/serine/threonine ohosphatase 7 mRNA capping enzyme alpha subunit (mRNA guanylyltransferase) (GTF		5.88E-10	22%	41%	33318	337-650
															8		-RNA guanvivitransterase) (Glase) Chain B, Structure Of An Mrna Capping Enzyme Bound To The		2.92E-09 8.50E-09	25% 25%	39% 38%	47316 47316	41373 41373
															10		<ul> <li>Phosohorvlated Carboxvl-Terminal Domain Of Rna Polvmerase li</li> <li>7 RNA guanylyltransferase and 5':-phosphatase</li> </ul>		2.47E-08	23%	42%	8313	231-563
N131	1L	6328462922	121	13,791	7.4	2 1	COG517	CDC55, Serine/threonine protein phosphatase 2A, regulatory subunit ISinnal transduction mechanisms1	26.99	6.38E-01	35%	54%	5884	5985	1	AAC96969	9 A645R	80.11	2.41E-14	34%	57%	1117	1121
N132	2L	64178-63336	281	31,701	5.2	3 1	pfam0044	I3 UCH, Ubiquitin carboxyl-terminal hydrolase Peptidase_C19, Peptidase C19 contains ubiquitinyl hydrolases. They	109.71	3.76E-25	20%	40%	4275	6312	1	AAC96473	contains ubquitin carboxy-terminal hydrolase active sites; similar to 3 human ubiquitin carboxy-terminal hydrolase, coresponds to Swiss-Prot Accession Number Q09879	261.92	1.71E-68	45%	67%	1277	1280
						2	cd0225	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 addeed the ubiquition.	81.90	8.61E-17	17%	33%	4276	2320	2	XP 65473	9 ubiquitin carboxyl-terminal hydrolase	60.46	7.57E-08	23%	41%	4278	68347
								depratision, as well also recycling. Could well well and the udiquitin The ubiquiting rotation and the responsible for material problem. The the marmalian cell, and with over 50 members, family C19 is one of the larnest families of neutricases in the human nervoria Poptidase_C19R, A subfamily of peptidase C19. Peptidase C19 contains ubiquiting Tydrolases. They are intracellular peptidases that remove															
						3	cd0267	ubiquim molecules from polyubiquinated peptides by cleavage of sopeptide bonds. Thery thydrolyce bonds involving the actobay (aroug of ubiquinitation is thought to be detiling of the ubiquin conjugates, which could rescue them from degradation, as well as recycling of the tubiquin. The ubiquinitproteasme system is responsible for most protein turvorer in the mammalian cell, and with over 50 members, family C16 is one of the largest families of peptidases in the human genome.	74.61	1.15E-14	21%	35%	6276	4335	3	EAR9631	9 Ubiquitin carboxyl-terminal hydrolase family protein	55.07	3.18E-06	20%	42%	6277	425736
						4	cd0266	Peptidase_CTGE_A subfamily of Peptidase_CT19_Peptidase_CT19 contains subjuintly hydrolases. They are intracolluter peptidases that remove ubquitin molecules from polyubiquinted peptides by cleavage of isopeptide benchs. They hydrolyse books involving the acaboxy (racup of the clearing of the subjuint). The purpose of the de- ubquiutination is thought to be editing of the ubquiut on couplasets, which could rescue them from degradation, as well as recycling of the ubquitin. The ubquiutifications especial are responsible for most protein tumover in the mammalian cell, and with over 50 members, family CT9 is one of the largest families of peptidases in the human genome.	63.74	2.55E-11	19%	37%	5275	5303	4	NP_990257	7 ubiquitin specific protease 2	52.76	1.58E-05	23%	39%	6275	23348
						5	cd0266	Peptidase, C10C, A subfamily of Peptidase, C10. Peptidase C19 contains subjuintly hydrolases. They are intracellular peptidases that remove ubiquitin molecules from polyubiquinted peptidase by deavage of isopeptide bonds. They hydrolyse bonds involving the actobacy (aroup 6 13 the Carminal GV residue of ubiquitin. The purpose of the de- ubiquinting on thought to be setting of the ubiquitin consugates, which the ubiquiting constraints of the ubiquitin on consignates, which the ubiquiting constraints of the ubiquitin consignates, which the ubiquiting constraints of the ubiquitin mover of the the market and clift and the constraints of the other the ubiquiting collision coll, and with over 50 members, family C19 is one of the largest families of peptidases in the human genome.		2.15E-05	37%	63%	213–261	278332	5	EAS01964	4 Ubiquitin carboxyt-terminal hydrolase family protein	51.60	3.52E-05	21%	38%	6259	186—497

Gene	Genome	A.A.	Peptide	pl	CDD Hit	COGs	COG Definition B	lit Score	F-value	%	%	Query	lit from- BLAST		Hit	BLASTp Definition	Bit Score	F-value	%	%	Query	
Name	Position	lenath	Mw	μ.	Number		Peptidase_C19F, A subfamily of Peptidase C19. Peptidase C19 contains		L value	Identitv Po	sitive	from-to	to Num	iber /	Accession		Dir oboro	2 14.40	dentitv P	ositive	from-to	to
							ubiquiting 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-															
					6		ubiquitination is thought to be editing of the ubiquitin conjugates, which	44.00	2.20E-05	24%	43%	5106	3106	6	XP_394836	PREDICTED: similar to ENSANGP00000018711	49.29	1.75E-04	20%	38%	6275	513-835
							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															
							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															
					7	cd02660	the C-terminal Gly residue of ubiquitin. The purpose of the de- ubiquitination is thought to be editing of the ubiquitin conjugates, which	41.46	1.15E-04	26%	45%	4105	3102	7	AAC13729	ubiquitin specific protease 41	48.91	2.28E-04	21%	39%	6276	23349
							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															
							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															
					8	cd02673	isopeptide bonds. They hydrolyze bonds involving the carboxyl group of the C-terminal Gly residue of ubiquitin. The purpose of the de-	41.26	1.40E-04	24%	37%	6109	4124	8	XP 692507	PREDICTED: similar to ubiquitin specific protease 2 isoform b	47 37	6.64E-04	25%	41%	6182	244429
							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 Pentidase C10A A subfamily of Pentidase C19 Pentidase C19															
							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															
						. 100057	isopeptide bonds. They hydrolyse bonds involving the carboxyl group of the C-terminal Gly residue of ubiquitin. The purpose of the de-	00.50		000/							40.00	4 405 00	0.407	0001	0.070	50.070
					9		ubiquitination is thought to be editing of the ubiquitin conjugates, which could rescue them from degradation, as well as recycling of the ubiquitin.	39.53	4.87E-04	23%	40%	4105	2-104	9	AAC68865	ubiquitin specific protease 66	46.60	1.13E-03	21%	39%	6276	50376
							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									Ubiquitin-specific protease that deubiquitinates ubiquitin-protein moieties;						
					10	COG5560	UBP12, Ubiquitin C-terminal hydrolase [Posttranslational modification, protein turnover, chaperones].	36.53	3.84E-03	22%	40%	6170	270455	10	NP 014213	may regulate silencing by acting on Sir4p; involved in posttranscriptionally regulating Gap1p and possibly other transporters;	44 67	4.30E-03	26%	48%	5112	365-477
							protein tamorta, anaparontoj.									nrimarily located in the nucleus: Ubn10n						
N135L	6516864227	314	34,973	7.74	¥ 1	COG1405	initiation factor TEIIB [Transcription]	66.09	5.92E-12	22%	43%	44311	7282	1	AAC96475	similar to Pyrococcus woesei factor TFIIB homolog, corresponds to GenBank Accession Number X70668	207.22	5.98E-52	36%	58%	30312	1290
					2	COG5635	COG5635, Predicted NTPase (NACHT family) [Signal transduction mechanisms]	29.29	5.77E-01	22%	44%	199286	542629	2	XP_635486	transcription initiation factor IIB		7.92E-04	22%	38%	42304	26301
														3 4	AAT43923 EAM94502	transcription initiation factor IIB Transcription factor TFIIB		1.03E-03 1.77E-03	21% 20%	39% 39%	44310 44310	17306 18307
														5	XP 760941	hypothetical protein UM04794.1 putative transcription initiation factor IIB		3.01E-03 3.93E-03	20% 19%	36% 39%	42306 48305	50351 153438
														7	BAB60249	transcription initiation factor B	44.67	5.14E-03 6.71E-03	21% 22%	38% 38%	44306 44304	19304 31322
														9	NP 187644	RNA polymerase II transcription factor/ transcription factor/ zinc ion binding		8.76E-03	21%	39%	49304	15288
														10	NP_594229	hypothetical protein SPAC16E8.16	43.90	8.76E-03	22%	38%	42297	23306
							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															
N137L	6567565166	5 170	19,353	9.14	1	pfam01541	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	29.36	2.65E-01	29%	47%	3789	2-51	1	AAC96477	A109L	128.64	7.47E-29	58%	73%	66168	1103
							flanked by three helices. The most conserved and putative catalytic residues are located on a shallow, concave surface and include a metal															
							coordination site															
N140L	6757965720	620	69,038	10.90			S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosohotransferases. Serine or threonine-specific kinase subfamily	48.67	2.01E-06	33%	48%	63257	4151	1	AAC96646	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662 similar to PBCV-1 ORF A34R, corresponds to GenBank Accession		3.03E-140	45%	59%	1618	1609
					2	pfam00069	Pkinase, Protein kinase domain S_TKc, Serine/Threonine protein kinases, catalytic domain.	47.97	3.08E-06	30%	50%	63257	4151	2	AAC96650	Number U17055	486.11	1.72E-135	44%	58%	1618	15568
							Phosphotransferases of the series or threonine-specific kinase subfamily. The enzymatic activity of these protein kinases is controlled by															
					3	cd00180	phosphorylation of specific residues in the activation segment of the catalytic domain, sometimes combined with reversible conformational	45.58	2.08E-05	33%	48%	63257	5152	3	AAC96947	RPQT-like (9x)	309.69	2.20E-82	46%	63%	11344	7336
							changes in the C-terminal autoreculatory tail APH, Phosphotransferase enzyme family. This family consists of															
							bacterial antibiotic resistance proteins, which confer resistance to various aminoglycosides they include:- aminoglycoside 3'-															
					4	pfam01636	phosphotransferase or kanamycin kinase / neomycin-kanamycin phosphotransferase and streptomycin 3''-kinase or	40.90	5.30E-04	44%	67%	225251	171198	4	AAC96461	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563	56.61	3.36E-06	34%	50%	457541	30132
							streptomycin 3''-phosphotransferase. The aminoglycoside phosphotransferases inactivate aminoglycoside antibiotics via									Accession without D7 1505						
							phosphorylation. This family also includes homoserine kinase. This family is related to fructosamine kinase nfam03881															
					5	COG2334	COG234, Putative homoserine kinase type II (protein kinase fold) IGeneral function prediction only).	39.17	1.40E-03	40%	57%	225255	200230	5	AAC96984	similar to Chlorella virus PBCV-1 ORF A282L, corresponds to GenBank Accession Number U42580	58.54	8.85E-07	72%	89%	474510	541
					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%	61257	3155	6	AAC48204	Hypothetical protein ZK84.1	45.44	7.75E-03	30%	44%	387467	477-565
					7	COG0510	CotS, Predicted choline kinase involved in LPS biosynthesis [Cell envelope biogenesis.outer membrane].	36.92	7.52E-03	22%	50%	225309	156232	7	AAV63985	hydroxyproline-rich glycoprotein VSP-3	45.82	5.93E-03	35%	51%	386457	384459
							PI3Kc_related, Proteins related to phosphoinositide 3-kinase (PI3K), catalytic domain: All of the members have been found to possess lipid															
					8		kinase activity. Many show Ser/Thr protein kinase activity. Many PI3K- related proteins are involved in cell-cycle checkpoints. They share two	36.69	8.05E-03	47%	57%	223251	179209	8	AAB54132	Uncoordinated protein 89, isoform a	54.68	1.28E-05	27%	45%	381552	1387-1573
							additional domains FATC, at the very C-terminus and FAT N-terminal to the PI3K-like domain															
					9	COG0661	AarF, Predicted unusual protein kinase [General function prediction onlv].	36.46	9.58E-03	43%	57%	223259	285320	9	AAV34801	Uncoordinated protein 89, isoform g	54.68	1.28E-05	27%	45%	381552	13871573
							TEL1, Phosphatidylinositol kinase and protein kinases of the PI-3 kinase family [Signal transduction mechanisms / Cell division and chromosome	07											0.5-1			000 077
					10	COG5032	partitioning / Chromatin structure and dynamics / DNA replication, recombination, and repair / Intracellular trafficking and secretion].	35.83	1.62E-02	61%	70%	227-250	19471970	10	AAH58068	Serine/arginine repetitive matrix 1	46.21	4.54E-03	36%	58%	393470	622699
N146R	6764968707	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	14133	1	AAC96497	similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903	85.89	2.42E-15	44%	58%	9111	81188
					2	pfam04638	uncharacterized Pox_O1, Pox virus protein O1. The function of these viral proteins is not known	29.49	7.12E-01	25%	40%	289343	204257	2	AAC96920	similar to Chlorella virus PBCV-1 ORF A154L, corresponds to GenBank	70.48	1.05E-10	35%	46%	5110	112-218
							NIOTI							3	AAC96522 NP 077492	Accession Number U42580 A154L	276.17	1.26E-72 6.17E-11	40% 30%	60% 47%	9349 4151	6347 79232
														5	CAA64974	EsV-1-7 QI74 protein hypothetical protein DDB0188101	47.75	7.30E-04	25%	35%	8208	79232 182363 238386
														6	EAR82303	Ubiquitin carboxyl-terminal hydrolase family protein	50.83	2.96E-05 8.62E-05	29% 24%	37% 42%	38190 26199	661-818
														8	XP_384145	hypothetical protein FG03969.1	48.52	4.28E-04	22%	32%	5192	87268

Gene Name	Genome Position	A.A. lenath	Peptide Mw		CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identitv	% Positive	Query from-to	Hit from- Bl to	Number 9	Hit BLASTp Definition EAR99846 Neurohypophysial hormones, N-terminal Domain containing protein	Bit Score E-v 46.60 1.6	3E-03	titv Posi	itive fr 30%	om-to 40198	lit from- to 353-530
N150L	68982-68698	95	10,247	6.50		No Hit Found								10	XP 644291 hypothetical protein DDB0230011 No Hit Found No Hit Found	47.37 9.5	3E-04	26%	38%	4170	201-350
N151R	6907069546	159	17,621	5.03		No Hit Found								1	AAC96979 similar to Chlorella virus CVK2 unknown ORF, corresponds to GenBank AAC969809 similar to Chlorella virus CVK2 unknown ORF, corresponds to GenBank AAC96980 similar to Chlorella virus CVK2 unknown ORF, corresponds to GenBank	145.98 3.5 116.32 3.0			72% 67%	16153 3152	1139 22170
N152R	6957369890	106	12,326	10.98		No Hit Found								1 2 3 4 5 6 7 8 9	AAC96489 A121R EAR81541 hypothetical protein TTHERM 01719410 AAL73473 TF GFp protein EAR154267 TF GFp protein, putative EAR805551 hypothetical protein TTHERM 01666140 EAR805551 Tref5 protein, putative AAL73477 TF GFp protein EAR81408 TF GFp protein	61.23 1.1 60.08 2.5 59.69 3.3	8E-12 1E-09 4E-08 4E-08 5E-08 3E-08 3E-08 7E-06	41% 44% 42% 42% 40% 38%	77% 58% 66% 66% 66% 66% 66% 61% 68%	8104 1393 3290 3290 3290 3290 3290 3290 3290 3263	196 1390 3288 3288 3288 3288 3288 3288 3288 3288
N155R	6992274763	1614	164,975	5.26		No Hit Found								1 2 3 4	AACG8400 PECVL1 Vy208 portletin AAA8307 devoratien Vy209 portatios a leucene zigner motift similar to Bacillus subilis phage P2A AACG8401 prenek appendage protein, corresponds to Swiss-Prot Accession Number P07537 ABA83471 Vy208 like protein	318.16 1.8 83.96 5.7 176.02 1.1 66.24 1.2	6E-14 1E-41	35% 36%	49%	091090 1190 891610 14294	19739 587766 41304 552831
							KAP, Kinesin-associated protein (KAP). This family consists of several							5 6 7 8 9 10	AA833469 Vy260 like protein AA83467 Vy260 like protein AA83469 Vy260 like protein AA825470 Vy260 like protein AAN05154 RB133 ZP_00739873 Cell surface protein	47.75 4.5 70.48 6.5 48.14 3.5 49.29 1.5 108.23 2.8 78.95 1.8	9E-10 DE-03 7E-03 5E-21	21% 24% 25% 39%	36% 79 40% 40%	14185 961428 14185 14185 291570 231559	95-257 34-579 95-257 95-257 309-453 277-422
N164R	7477876070	431	48,976	10.91	1	pfam05804	Nor, Antesin-associated protein (Nor, 1 mis family contasts to servicial 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 KRPB5/95 holoenzyme associates with a KAP115 non-motor protein, forming a beterritimeric complex in vitra called the Kinessi-ill.	21.12	2.47E-01	24%	41%	265345	83169	1	AAC96984 similar to Chlorella virus PBCV-1 ORF A282L, corresponds to GenBank Accession Number U42580	209.53 1.9	3E-52	37%	53% 1	104—417	2357
														2	AAC96650 similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055 AAC96646 similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number	65.08 5.8					369-451
														3	AAC96461 similar to Chlamydia histone-like protein, corresponds to GenBank	48.14 7.4 61.23 8.5			42% 54%	36151 773	366506 49132
N165L	7661076065	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	Accession Number D71563 BAA04187 transcription elongation factor SII	281.95 6.3	2E-75	70%	82%	1180	1180
					2	smart00440	ZeE COCO COCO Zine Genere Nucleis sold hinding metif in	64.15	9.51E-12		77%		140	2	AAC96492 contains a zinc ribbon domain; similar to Chlorella virus CVU1 TFIIS-like transcription factor, corresponds to GenBank Accession Number D29631	280.80 1.4		70%	83%	1180	1180
					3	COG1594	RPB9, DNA-directed RNA polymerase, subunit M/Transcription elonoation factor TFIIS ITranscription1.	60.43	1.25E-10	54%	73%	142179	74–111	3 5 6 7 8 9 10	S47662 transcription elongation factor TFIIS homolog - Chlorella virus CV-U1 BAA0196 transcription elongation factor S18 XP 456316 involntetical accelerin DEHAOC15686q XP 445674 unamed profete product AAS54500 AGR011Wp XP 45503 unamed profete product XP 455338 transcription elongation factor TFIIS.h XP 207283 transcription elongation factor TFIIS.h	279.26 4.1 243.82 1.9 83.57 3.3 82.03 9.6 82.03 9.6 81.65 1.2 80.88 2.1 80.88 2.1	1E-63 DE-15 DE-15 DE-15 5E-14 4E-14	66% 34% 39% 37% 35% 37%	56% 52% 50% 54%	82179 66179 59179 67179	1180 1164 161288 209304 189302 174290 216331 216331
N168R	76715-77380	222	24,350	10.20		No Hit Found								1	AAC96495 A127R	291.20 1.6	BE-77	63%	76%	8217	30245
N169L	78142–77375	256	28,604	8.34		No Hit Found								1 2 3 4 5 6 7 8 9	AAC98377 AgR AAC98377 AgR Cocession Number U42580 AAC98643 Marker U42580 AAC98643 Marker U42580 AAC98643 Marker U42580 AAC98645 Marker U42580 Marker U42580 Mar		6E-18 9E-16 6E-15 3E-13 9E-12 1E-12 3E-11	28% 24% 23% 24% 25% 24%	45% 45% 46% 47% 44% 51% 51%	89254 10253 9252 9253 9232 9237 89252 89252 89253	8173 15253 11248 8246 8218 11226 4167 4167 1132
N170R	7829779217	307	33,545	5.48	1	pfam03747	ADP_ribosyLGH, ADP-ribosylg/pcohydrolase. This family includes enzymes that ADP-ribosylations, for example ADP-ribos-l-arginine. The family also includes dinitrogenase reductase activating glycohydrolase. Most surprising) the family also includes jellyfish crystallins, these proteins	110.26	4.71E-28	28%	44%	10272	2286	1	EAR97823 ADP-ribosylglycohydrolase family protein	156.38 1.1			50%		28344
							annear to have lost the nresumed active sile residues.							5 6 2 7 8 9 2	EAR82865 ADP-ribosylalycohydrolase family protein EAS94429 ADP-ribosylalycohydrolase family protein 2053136 ADP-ribosylalycohydrolase AB822956 putatve ribosylalycohydrolase EAR83155 ADP-lobosylalycohydrolase EAR83155 ADP-lobosylalycohydrolase AB82295 ADP-hobosylalycohydrolase 2P 0058429 ADP-hobosylalycohydrolase	138.27         3.3           132.11         2.3           130.18         9.0           128.26         3.4           126.72         9.9           121.71         3.2           120.94         5.4           120.55         7.1           117.09         7.9	8E-29 3E-29 3E-28 9E-28 1E-26 8E-26 6E-26	30% 31% 31% 30% 30% 30%	48% 45% 45%		38-338 41348 18-295 18-290 12283 179485 335639 12284 9257
N173R	7919979600	134	15,106	9.33		No Hit Found								1	AAC96416 A48R	95.90 4.3	1E-19	37%	62%	18132	5123
N174L N176R	79957—79595 80040—80495	121	12,049 17,368	3.55	1	No Hit Found pfam03009	GDPD. Glycerophosphoryl diester phosphodiesterase family. E. coli has too sequence related isozymes of glycerophosphoryl diester phosphodiesterase (GDPD) - periplasmic and cytosolic. This family also includes approximation esynthase, the similarity to GDPD has been noted. This family appears to have weak but not significant matches to similarity appears.	28.84	2.76E-01	37%	49%	650	1-44	1	No Hit Found No Hit Found AAZ21386 hypothetical protein SAR11_0565	94.74 9.5	DE-19	39%	62%	1117	1115
N177L	8184480465	460	51,947	5.86	1	COG0637		100.95	2.51E-22		47%		2190		ZP_01245443 Nucleotidyl transferase	211.46 5.5			67% 2		6242
					2		Cph. Predicted chosshatases (General function prediction only, Hydrolase, haload dehalognane-like hydrolase. This family are structurally different from the alpha' beta hydrolase family (pfam00561). This family includes 1-2-haloacid dehalognanese, 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 folis is composed of the core alphabeta	78.65	1.45E-15 2.34E-12		46% 39%		2191 1191		CAH09375 putative nucleotidyltransferase	206.84 1.3 200.68 9.7				220450 220450	2236 2238
					4	COG1209	domain	63.27	6.29E-11	25%	42%	223-446	4228	4 2	ZP_00371454 lipopolysaccharide biosynthesis protein, putative	196.44 1.8	5E-48	49%	63% 2	220450	1236
					5	COG1011	outer membrane). COG1011, Predicted hydrolase (HAD superfamily) [General function prediction onlv].		4.54E-10		35%		30208		AAT91796 putative nucleotidyl transferase family protein	195.67 3.1				220450	2240

Gene Name	Genome Position	A.A. length	Peptide Mw	pl	CDD Hit Number	COGs	COG Definition E	Bit Score	E-value	% Identity P		Query from-to	Hit from- BLAS		Hit Accession	BLASTp Definition	Bit Score	E-value	% lentitv Po	%	Query from-to	Hit from- to
Hame	1 Galdon	lendth			6	COG1208	GCD1, Nucleoside-diphosphate-sugar pyrophosphorylase involved in lipopolysaccharide biosynthesis/translation initiation factor 2B, gamma/epsilon subunits (eIF-2Bgamma/eIF-2Bepsilon) [Cell envelope	58.43	1.77E-09	25%	42%	222-447	4228		YP_419519 dTDP-glucose pyropha	sphorylase	181.80	4.70E-44	43%		217-450	2240
					7		biogenesis, outer membrane / Translation, ribosomal structure and hiomenesis1 NTP_transferase, Nucleotidyl transferase. This family includes a wide	55.98	9.93E-09	19%	38%	223-392	3	7	AAO54248 lipopolysaccharide bios		173 71	1.28E-41	40%	57%	220450	2239
					8		rande of enzymes which transfer nucleotides onto bhosohosudars LicC, CTP:phosphocholine cytidylyltransferase involved in choline phosphorylation for cell surface LPS epitopes [Cell envelope biogenesis,	41.05	3.27E-04	45%	68%	220258	139		ZP_00207801 COG1209: dTDP-gluco		171.79	4.87E-41	42%	59%	226450	1229
							outer membranel. CTP_transf_3_Cytidytytransferase. This family consists of two main Cytidytytransferase activities: 1) 3-deoxy-manno-octulosonate cytidytytransferase, EC:2.7.7.38 catalysing the reaction- CTP + 3-deoxy-															
					9	pfam02348	D-manno-octulosonate &It=> diphosphate + CMP-3-deoxy-D-manno- octulosonate, 2) acyteuraminate cytdyltransferase CEC2.7.7.43, catalysing the reaction:- CTP + N-acytneuraminate &It=> diphosphate + CMP-N-acytneuraminate. NeuAc cytydilyttransferase of Mannheimia haemolytica has been characterised describing kinetics and regulation by	39.54	8.31E-04	19%	36%	222-392	2175	9	AAY35682 Nucleotidyl transferase		168.70	4.12E-40	38%	58%	220450	2239
					10	COG1210	substrate charge, energetic charge and amino-sugar demand GalU, UDP-glucose pyrophosphorylase [Cell envelope biogenesis, outer membrane]	37.41	3.72E-03	24%	35%	223389	8210	10	AAZ34091 lipopolysaccharide bios	synthesis protein, putative	164.85	5.95E-39	38%	55%	220450	2239
N180L	8338981878	504	57,895	5.18	3	No Hit Found								1 2 2	ZP_00520036 Methyltransferase Fkbl ZP_01305260 hypothetical protein SK	A58 11615	54.30	3.27E-09 1.29E-05	29% 27%	45% 45%	34232 37185	402595 45196
														3 4 5 2	AAV90092 hypothetical protein ZM YP 439595 methyltransferase, Fkb ZP 01090243 methyltransferase Fkb	M family domain protein	48.14	2.19E-05 9.22E-04 4.57E-03	31% 26% 27%	52% 40% 41%	282408 32237 34167	398-533 13223 38166
N183L	8399683331	222	25,840	7.70	)	No Hit Found								1	AAV90092 hypothetical protein ZM	IO1468	50.45	5.00E-05	33%	54%	69168	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	28.92	3.34E-01	27%	44%	98157	90153	1	AAC96978 A662L		210.69	1.31E-53	64%	79%	6161	16171
							to chromosome gain or loss							2 3	ABE89007 Mpv17/PMP22 NP_565983 unknown protein		71.25 64.31	1.24E-11 1.52E-09	28% 24%	51% 49%	34160 27160	92219 98232
														4 5	BAD87006 unknown protein XP_962431 hypothetical protein		50.45	3.39E-09 2.27E-05	25% 23%	50% 44%	34160 34156	112-239 129-255
														6	BAB08278 unnamed protein produ	2 kDa peroxisomal membrane protein uct	49.68	3.87E-05 3.87E-05 3.87E-05	25% 25% 25%	51% 51% 51%	34135 34135 34135	122-225 121-224 121-224
														9 10	NP_568621 unknown protein AAN46791 At5g19750/T29J13 17 AAM65990 unknown	0	49.08 47.37 47.37	1.92E-04	19% 19%	43% 43%	13161 13161	139-284 140-285
N186L	85183-84707	159	18,620	10.37	· 1	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	33.51	1.20E-02	24%	39%	894	1-83	1		e T4 intron-associated endonuclease, correspond in Number P13299		1.16E-32	52%	64%	1147	1148
					2 3	COG2989	COG2989, Uncharacterized protein conserved in bacteria [Function unknown]. LTV, Low temperature viability protein	29.19 27.39	2.78E-01 8.88E-01	28% 32%	55% 54%	12103 82149	442518 333396	2 3	AAC96973 similar to Chlorella viru Accession Number U42 YP_293795 putative endonuclease	IS PBCV-1 ORF A315L, corresponds to GenBank 2580		5.98E-05 7.81E-05	38% 34%	50% 49%	993 992	285 388
							Err, con temperature monty protein.							4	AAY63899 reticulocyte binding pro AAC96655 PBCV-1 33kd peptide	otein 1		2.51E-03	26% 30%	45% 54%	13135 1494	465583 1191
														6 7	CAA31562 unidentified protein NP_074961 orf261		41.59	9.55E-03 9.55E-03	34% 34%	51% 51%	289 289	61155 60154
N187L	8628885218	357	41,118	8.86	i 1 2		HNHc, HNH nucleases; . HNH, HNH endonuclease	30.44 29.91	3.67E-01 5.47E-01	31% 29%	46% 34%	56102 61102	2-50 2-43	1	AAC96722 A354R AAC96790 similar to Bacteriophag	e SP01 gene 31 intron, corresponds to Swiss-Prol	200.68	6.86E-50 8.76E-13	42% 29%	61% 49%	110355 54322	1235 51319
					-	plano rotti		20.01	0.472 01	2070	0470	01 102	2 40	3	Accession Number P34 AAL98037 hypothetical phage pro			2.82E-03	26%	48%	20155	11147
N189L	86927-86580	116	13,059	10.04		No Hit Found								1	AAC96507 A139L			1.39E-25	56%	78%	26108	25107
N191R	8702288587	522	58,869	8.65	5 1	COG1215	COG1215, Glycosyltransferases, probably involved in cell wall biogenesis ICell envelope biogenesis, outer membranel. Glycos_transf_2, Glycosyl transferase. Diverse family, transferring sugar	131.59	2.17E-31	20%	35%	66460	9412	1	AAC96840 similar to Acetobacter Accession Number P2	cellulose synthase, corresponds to Swiss-Pro 1877	t 800.82	0.00E+00	72%	84%	1520	1517
					2	pfam00535	from UDP-glucose. UDP-Nacetyl-galactosamine, GDP-mannose or CDP-abequese to a range of substrates including cellulose, dolichol shoenhate and technicarian shoenhate and technicarian ubranched polymers of beta 1-4 linked glucose residues, is the major component of wood and thus paper, and is synthesised by plants, most age, some bacteria and fungi, and evens some animals. The genes that	64.09	4.39E-11	18%	37%	118–280	1164	2	XP_959316 hypothetical protein		434.88	3.64E-120	44%	60%	21516	168–670
					3		synthesise cellulose in higher plants differ greatly from the well- characterized genes found in Accidotater and Agrobacterium ps. More correctly designated as &apos.cellulose synthase catalytic subunits&apos.plant cellulose synthase (CaseA) proteins are integral membrane proteins, approximately 1.000 amino acids in length. There are a number of highly conserved residues, including several motifs	43.36	6.56E-05	30%	46%	292387	517618	3	EAQ91514 conserved hypothetical	I protein	417.93	4.61E-115	42%	59%	19500	391-894
					4		COG1216, Predicted glycosyltransferases [General function prediction onlv].	42.81	1.10E-04	21%	34%	190312	82213	4	YP_482343 Cellulose synthase (UE	DP-forming)	273.09	1.83E-71	35%	54%	23453	60492
					5	pfam05601 COG0463	DUF774, Protein of unknown function (DUF774). This family consists of several uncharacterised Actinomvcete oroteins of unknown function WcaA, Glycosyltransferases involved in cell wall biogenesis [Cell	31.87	1.99E-01	32% 12%	46%	136-164	6290	5	AAO54558 cellulose synthase, cat			6.58E-37	34% 34%	51% 51%	111-367	155417 264540
					6		envelope biogenesis, outer membrane1 MdoH, Membrane glycosyltransferase [Cell envelope biogenesis, outer	30.88 29.95	4.13E-01 7.60E-01	12%	26% 34%	113395 98457	1278	6 7	NP_902348 cellulose synthase, sub AAL71842 WssB	bunit A		1.12E-36 7.28E-36	34% 34%	51%	115385 113367	264-540
						0002040	membranel.	20.00	1.002 01	2010	0470	30 407	122 000	8	AAM38361 celullose synthase NP_643825 celullose synthase		154.45 154.45	9.51E-36 9.51E-36	32% 32%	49% 49%	115405 115405	153-451 155-453
N195R	88652-91846	1065	440.000	10.84		No. 112 E								10	YP 365375 cellulose synthase cata BAE02830 surface protein	alytic subunit		1.24E-35 2.37E-21	32% 83%	49%	115405	155-453
NISSK	00032-91040	1005	112,990	10.04		No Hit Found								2	AAC96508 PBCV-1 surface protein BAD22850 surface protein	n	46.98	4.97E-03 4.97E-03	37% 37%	48% 48%	0021063 485566 485566	9491031 9491031
														4	BAD12236 surface protein	15R - Chlorella virus PBCV-1	46.60	6.49E-03 4.97E-03	37% 37%	48% 48%	485566 485566	9491031 158240
														6	AACORTED Lys-, Pro-rich, PAPK	(10x); similar to wheat Pro-, Lys-rich protein nk Accession Number X52472	51.22	2.64E-04	36%	50%	495563	195-263
														7	AAC96773 Pro-, Lys-rich, PAPK corresponds to GenBar AAC96654 A286R	(30x); similar to wheat Pro-, Lys-rich protein nk Accession Number X52472		1.31E-03 5.87E-04	32% 20%	58% 40%	638699 139416	465 68346
														9 10	YP 535207 Mucus binding protein AAT30727 conserved repeat doma	ain protein	46.98	4.97E-03 6.49E-03	28% 24%	39% 40%	734-865 173-373	723-853 741-954
N205L	92196-91882	105	11,647	6.50	) 1		COG0731, Fe-S oxidoreductases [Energy production and conversion].	28.34	1.93E-01	21%	43%	3980	2168		AAC96518 A150L			9.08E-14	41%	59%	1100	5105
N206R	92301-93668	456	50,729	8.78	1 2	COG1061	SSL2, DNA or RNA helicases of superfamily II [Transcription / DNA renlication recombination and renair] DEXDc, DEAD-like helicases superfamily; .	137.92 63.32	2.16E-33 5.82E-11	26% 23%	40% 37%	54397 94233	7379 13173	1	AAC96521 similar to phage T5 Number P11107 NP_077551 EsV-1-66	helicase, corresponds to Swiss-Prot Accession	613.99	3.69E-174 1.78E-43	66% 32%	81% 50%	4453 21426	7459 17443
							DEXHc, DEXH-box helicases. A diverse family of proteins involved in											1.78E-43				
					3	cd00269	processes. The name derives from the sequence of the Walker B motif (motif II) This domain contains the ATP-binding region DEXDc, DEAD-like helicases superfamily. A diverse family of proteins	55.86	9.68E-09	23%	34%	110-230	2143		YP_142750 VV A18 helicase				28%	48%	3430	8453
					4	cd00046	involved in ATP-dependent RNA or DNA unwinding. This domain contains the ATP-binding region	54.32	2.71E-08	25%	37%	110-230	2144	4 2	ZP_00754006 COG1061: DNA or RN	A helicases of superfamily II	146.36	2.18E-33	29%	45%	85430	419-776

Gene	Genome	A.A.	Peptide	pl	CDD Hit	COGs	COG Definition	Bit Score	E-value	%	%	Query	Hit from- BLAST	p Hit	Hit BLASTp Definition	Bit Score	E-value	%	%	Query	Hit from-
Name	Position	lenath	Мw	·	Number		DEAD, DEAD/DEAH box helicase. Members of this family include the DEAD and DEAH box helicases. Helicases are involved in unwinding			Identitv P	ositive	from-to	to Numi	ber A	Accession		lo	lentitv Po	ositive	from-to	to
					5	pfam00270	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	44.64	2.48E-05	18%	32%	102-233	19175	5	YP_437404 DNA or RNA helicase of superfamily II	140.58	1.20E-31	29%	46%	87430	421-776
					6	COG4096	subunit and related helicases [Defense mechanisms]. SNE2 N SNE2 family N-terminal domain. This domain is found in	43.81	4.13E-05	26%	40%	113-231	190321	6 Z	ZP_01111109 DNA or RNA helicase of superfamily II	139.81	2.04E-31	29%	46%	85443	419793
					7	pfam00176	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	43.73	5.08E-05	22%	38%	110-230	18156	7	NP_309332 hypothetical protein ECs1305	135.58	3.85E-30	30%	49%	81430	417-777
							chromatin unwinding (e.g., ISWI) as well as a variety of other proteins with little functional information (e on Iondear FTI 1) HELICC, Helicase superfamily c-terminal domain; associated with DEXDo, DEAD, and DEAH-box proteins; yeas 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														
					8	cd00079	wide variety of helicases and helicase related proteins; may not be an autonomously folding unit, but an integral part of the helicase. 4 helicases superformatives at present according to the organization of their segnature a distinct directional polarity; they utilize the free energy from nucleoside triphosphate hydrolysis to fuel their translocation along DNA, unwinding the direction process.	42.22	1.21E-04	23%	40%	302395	17123	8	NP_287072 putative helicase	135.58	3.85E-30	30%	49%	81430	425785
					9	COG1197		42.17	1.34E-04	26%	43%	89230	607752	9 Z	ZP_00532208 Type III restriction enzyme, res subunit:DEAD/DEAH box helicase, N-	128.26	6.14E-28	29%	46%	51430	394790
					10	COG0610	reolication. recombination. and recair / Transcription1. COG0610, Type I site-specific restriction-modification system, R (restriction) subunit and related helicases [Defense mechanisms].	38.12	2.05E-03	22%	36%	95338	258517	10 Z	Type III restriction enzyme, res subunit:DEAD/DEAH box helicase, N- terminal	127.10	1.37E-27	28%	45%	81430	481-849
N209L	94072-93716	119	12,775	3.90	1	pfam03115	Astro_capsid, Astrovirus capsid protein precursor. This product is encoded by astrovirus QRF2, one of the three astrovirus QRFs (1a, tb, 2). The 87kD precursor protein undergoes an intracellular releavage to form a 79kD protein. Subsequently, extracellular trypsin cleavage yields the three orotenis formina the infectious virion.	26.55	7.67E-01	26%	43%	71117	640686	1	AAC98525 A157L	93.59	2.12E-18	45%	56%	1117	1109
N210L	9449694128	123	14,411	4.26		No Hit Found	the time brokens forming the micelings when.							1 0	No Hit Found No Hit Found						
N211L	94977-94603	125	13,445	8.66		No Hit Found								1 0	No Hit Found No Hit Found						
N214L	9545795167	97	10,869	9.32		No Hit Found								1	AAC96533 A165L	72.79	3.93E-12	39%	53%	595	171-279
N216L	9597495516	153	17,584	10.22	1	cd01973	Nitrogenase_VFe_beta_like. Nitrogenase_VFe_beta-like: Nitrogenase VFe protein, beta subunit Nke. This group contains proteins similar to the beta subunits of the VFe protein of the variadum-dependent (V-) nitrogenase. Nitrogenase catalyzes the ATP-dependent reduction of dinitrogen NR2) to ammonia. In addition to V-nitrogenase three is a molydderum (Mo)-dependent nitrogenase and an iron only (Fe) nitrogenase. The Neuropeanse The Section Sectio	29.52	1.89E-01	41%	49%	4784	342383	1	AAC98533 A185L	85.50	5.74E-16	37%	58%	5118	20134
							(FeMoo) contained within the alpha subunit. The Fe protein which has a markedly klentinal elineture in all three subunit. The Se protein which has a markedly klentinal elineture in all three subunit. The Second alpha sub- COG5377, Phage-traited protein, predicted endonuclesse [DNA							3 Z	ZP_00133232 COG1219: ATP-dependent protease Clp, ATPase subunit ZP_00123014 COG1219: ATP-dependent protease Clp, ATPase subunit	41.97	7.28E-03 7.28E-03	29% 29%	50% 50%	31149 31149	16124 20128
N218R	96021-96887	289	33,100	5.82	1		reclication, recombination, and recair).	30.01	3.41E-01	23%	32%	30169	2148	2 3 4 5 6 7 8 9	AAC0853 PBCV-1 exonuclease PD-07546 EX-1-64 AAC38903 F12A21.19 PD-17546 EX-1-64 AAC38903 F12A21.19 CAF2748 Exonuclease CAF2748 Exonuclease CAF2748 Exonuclease CAF2748 Exonuclease XP_112778 exonuclease XP_412778 exonuclease XP_41244 OSJNBa0051L15.20	88.58 78.57 76.26 73.94 73.94 72.02 71.63 70.86	1.72E-79 2.71E-16 2.80E-13 2.80E-13 1.39E-12 6.90E-12 2.62E-12 3.42E-11 3.42E-11 5.84E-11	58% 30% 30% 29% 29% 32% 29% 29%	73% 50% 45% 45% 44% 44% 48% 45%	9253 21221 36237 36237 19223 36224 36224 36224 36183 35216 35216	1-246 1-191 95-294 116-315 3-197 3-200 3-200 3-155 134-317 141-324
N221R	97092-97394	101	10,665	4.11		No Hit Found								1	AAC96536 A168R	93.97	1.62E-18	53%	66%	995	70161
N222L	98212-97385	276	30,232	9.99	1	pfam01734	protein but it also has the enzymatic activity of lipid acyl hydrolase, catalysing the cleavage of fatty acids from membrane lipids. Members of	116.20	3.75E-27	31%	49%	16189	1179	1	AAC96541 similar to E. coli hypothetical protein, corresponds to Swiss-Prot Accession Number P39407	322.40	1.03E-86	57%	77%	5274	14288
					2	COG1752	this family have been found also in vertebrates RssA, Predicted esterase of the alpha-beta hydrolase superfamily [General function prediction only].	102.08	6.31E-23	31%	48%	16228	14229	2 Z	ZP_00240206 Patatin-like phospholipase family	106.69	8.91E-22	34%	53%	15185	7190
					3	COG4667	COG4667, Predicted esterase of the alpha-beta hydrolase superfamily [General function prediction only]	51.42	1.12E-07	24%	40%	16208	14202	3	XP_800775 PREDICTED: hypothetical protein XP_795682, partial	99.37	1.42E-19	35%	50%	14186	76265
					4	COG3621	COG3621, Patalin (General function prediction only). S_TKc, Serine/Threonine protein kinases, catalytic domain.	31.14	1.52E-01	41%	58%	566	665	5 6 7 8 9	YP_13138 hypothetical proteinin PBPRB1472 XP_789091 PREDICTED: hypothetical protein XP 783908 NP_909966 hypothetical protein PG1879 NP 149906 442 ASS4389 phospholipase, patatin family ABS37620 esterase of the alpha-beta hydrolase superfamily-like YP_142800 patatin-like phospholipase (463L)	89.74 89.35 86.66 85.89 84.73	8.63E-17 1.13E-16 1.47E-16 9.54E-16 1.63E-15 3.63E-15 1.06E-14	28% 32% 32% 29% 30% 30% 25%	48% 51% 52% 54% 50% 50% 49%	14256 14186 14185 14206 15186 15185 15238	6-261 86-275 6-194 25-215 4-194 4193 57-285
N223L	99133-98228	302	34,030	4.75	1		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 channes in the C-terminal autorenulatory trail	175.01	8.41E-45	31%	55%	36295	1256	1	AAU06282 protein kinase A248R	187.96	3.55E-46	38%	58%	21295	32305
					2	smart00220	S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosohotransferases. Serine or threonine-specific kinase subfamilv Pkinase Protein kinase domain	175.02	8.98E-45 1.48E-41	31%	52%	37295	1256		AAU06280 protein kinase A248R		6.05E-46	37% 37%	58%	21295	32305
					3		Pkinase, Protein kinase domain SPS1, Serine/threonine protein kinase [General function prediction only / Signal transduction mechanisms / Transcription / DNA replication,	164.30 117.18	1.48E-41 2.27E-27	28% 28%	52% 47%	37295 36296	1258	3	AAU06275 protein kinase A248R AAC96616 PBCV-1 protein kinase		6.05E-46 1.95E-44	37%	58% 58%	21295 24295	11284 32304
					4 5		recombination. and repairl. TyrKc, Tyrosine kinase, catalytic domain; Phosphotransferases. Tyrosine-snecific kinase subfamily.	91.39	2.27E-27 1.42E-19	28%	47% 48%	36296 39227	1279 3192	4 5	AAC96616 PBCV-1 protein kinase AAA87065 serine/threonine protein kinase		1.95E-44 7.39E-44	37%	58%	24295 24295	32304 28300
					6	cd00192	TyrKc, Tyrosine kinase, catalytic domain. Phosphotransferases; tyrosine- specific kinase subtrainiy. 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 chosphorvlation of the series of the seri	88.71	7.38E-19	27%	46%	36227	8203	6	AAU06274 protein kinase A248R	156.76	8.75E-37	38%	59%	79295	22238
					_		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 chances. COG3642, Mn2-dependent serine/threonine protein kinase [Signal														
					7	CUG3642	transduction mechanisms).	52.93	4.66E-08	28%	43%	112-211	66180	'	AAU06285 protein kinase A248R	150.60	6.27E-35	37%	58%	81295	2217

Gene	Genome	A.A.	Peptide	pl	CDD Hit	COGs	COG Definition	Bit Score	E-value	%	%	Query	Hit from-			BLASTp Definition	Bit Score	E-value	%	%	Query	
Name	Position	lenath	Mw	μ.	Number		Kdo, Lipopolysaccharide kinase (Kdo/WaaP) family. These lipopolysaccharide kinases are related to protein kinases pfam00069.	211 00010	2 10.00	Identitv	Positive	from-to	to	Number	Accession	bererp berniten	Directore	L Value lo	dentitv P	ositive	from-to	to
					8	pfam06293	This family includes waaP (rfaP) gene product is required for the addition of phosphate to 0-4 of the first heptose residue of the tipopolysaccharide (LPS) inner core region. It has previously been shown that WaaP is necessary for resistance to hydrophotic and polycationic antimicrobiats in E. coli and that it is required for virulence in invasive strains of S.	37.17	2.49E-03	21%	37%	92167	74155	8	AAC96657 similar to GenBank A	PBCV-1 serine/threonine protein kinase, corresponds to cccession Number U14660	0 129.03	1.95E-28	31%	52%	41295	22279
					9	COG4248	enterica COG4248, Uncharacterized protein with protein kinase and helix-hairpin-	33.14	4.19E-02	20%	40%	127-201	114189	9	AAW40743 serine/three	onine-protein kinase, putative	120.55	6.95E-26	31%	51%	36298	114404
					10	COG0478	helix DNA-bindina domains [General function prediction onlv]. COG0478, RIO-like serine/threonine protein kinase fused to N-terminal HTH domain [Signal transduction mechanisms]	30.59	2.57E-01	25%	47%	86174	158246	10			119.40	1.55E-25	45%	63%	146295	2152
N227L	10136299221	1 714	79,825	6.96	6 1	COG4581	COG4581, Superfamily II RNA helicase [DNA replication, recombination, and repair].	368.92	9.77E-103	40%	55%	12409	119568	1	contains A AAC96609 antiviral pro	TP-GTP binding motif; similar to Saccharomyces cervisia- otein SKI2, corresponds to Swiss-Prot Accession Numbe	e r 699.12	0.00E+00	49%	66%	4712	2725
					2	COG1204	COG1204. Superfamily II helicase [General function prediction only].	196.76	6.64E-51	31%	50%	30400	50430	2	P35207 XP_652111 DEAD/DEA	AH box helicase	306.22	2.89E-81	37%	53%	13527	96654
					3		<ul> <li>COG1202. Superfamily II helicase, archaea-specific [General function prediction onk]</li> <li>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</li> </ul>	126.23	1.03E-29 8.20E-24	33%	50%	27378	232552 3206	3				1.87E-80	38%	56%	13428	138-594 244-850
					*	pianio0270	RNA metabolism, including nuclear transcription, pre mRNA splicing, ribosome biogenesis, nucleocytoplasmic transport, translation, RNA decay and orranellar cone expression	100.00	0.2UE-24	23%	3776	5190	3200	4	EAL2//00 GA10159-P	~A	303.32	1.0/E-00	3376	50%	3342	244-030
					5		7 DEXDc, DEAD-like helicases superfamily; . COG1205 Distinct helicase family with a unique C-terminal domain.	94.52	3.52E-20	26%	42%	5175	1190		XP 970674 PREDICTE			3.20E-80	37%	53%		239-765
					6		<sup>5</sup> including a metal-binding cysteine cluster [General function prediction nnlv] Lhr, Lh-like helicases [General function prediction only].	83.14 79.14	1.06E-16 1.81E-15	24% 26%	41% 43%	3365 17374	61410 27358	6	NP_524465 twister CG1	W box helicase family protein		2.07E-79 2.71E-79	38% 37%	54% 52%	13447 3464	133-599 246-760
					8		DEXDc, DEAD-like helicases superfamily. A diverse family of proteins 5 involved in ATP-dependent RNA or DNA unwinding. This domain	78.59	2.53E-15	26%	43%	28154	1143	8				2.71E-79	37%	52%	3464	246-760
					9	COG0513	contains the ATP-binding region SrmB, Superfamily II DNA and RNA helicases [DNA replication, 3 recombination, and repair / Transcription / Translation, ribosomal	67.88	4.75E-12	21%	40%	17424	56425	9	NP_001012962 superkiller	viralicidic activity 2-like 2	298.13	7.88E-79	38%	56%	13447	127-598
					10	cd00269	structure and binonensial 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	67.41	5.63E-12	27%	45%	28154	1142	10	EAA08476 ENSANGP	00000014710	295.82	3.91E-78	39%	54%	13443	269745
N231R	101470101790	) 107	12,564	10.56	6	No Hit Found	(motif II). This domain contains the ATP-binding region							0	No Hit Found No Hit Four	nd						
N232L	103319101802	2 506	56,634	9.80	D 1	pfam06408	Homo_sperm_syn, Homospermidine synthase. This family consists of several homospermidine synthase proteins (EC:2.5.1.44). 8 Homospermidine synthase (HSS) catalyses the synthesis of the polyamine homospermidine from 2 mol putrescine in an NAD(+).	524.14	1.42E-149	38%	53%	29498	1470	1	AAC96605 PBCV-1 ho	mospermidine synthase	711.45	0.00E+00	69%	81%	23502	36516
					2	COG1748	dependent reaction 42 YS9, Saccharopine dehydrogenase and related proteins [Amino acid transport and metabolism].	33.72	5.80E-02	24%	49%	30271	2184	2	ZP_00589757 Homospern	nidine synthase	333.18	1.44E-89	38%	57%	29476	5452
					3	pfam01967	ManC. ManC family. Mambara of this family, are involved in malyhdanym	33.29	7.35E-02	21%	42%	86154	2294	3	AAM05046 homosperm	nidine synthase	300.06	1.35E-79	36%	55%	29476	13459
					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 furtion as a saccharonice reductase.	32.53	1.13E-01	22%	46%	32187	1137	4	NP_632190 homosperm	nidine synthase	295.82	2.55E-78	35%	55%	29476	6452
					5	cd00528	MoaC, MoaC family. Members of this family are involved in molybdenum cofactor (Moco) biosynthesis, an essential cofactor of a diverse group of gredox enzymes. MoaC, a small hexameric protein, converts, together with MoaA, a guanosine derivative to the precursor Z by inserting the carbon-3 of the purine between the 28aoos: and 38aoos: Thobse carbon	30.11	6.14E-01	22%	38%	86154	2294	5	AAZ71589 homosperm	nidine synthase	283.49	1.31E-74	34%	53%	29-476	6452
							atoms, which is the first of three phases of Moco biosynthesis.							7 8 9	ZP_00584810 Homospern AAU28555 homospern CAH16656 hypothetica YP_421629 Homospern	nidine synthase al protein midine synthase	241.89 240.74 234.19	1.78E-63 4.37E-62 9.74E-62 9.12E-60	33% 34% 34% 32%	51% 53% 53% 51%	32476 28476 28476 35476	14446 13446 13446 34462
N237L	103707103336	5 124	14,303	5.61	1	No Hit Found								10	AAO55386 homosperm No Hit Found No Hit Four		228.02	6.53E-58	34%	53%	79480	56448
							PH_SETbf, Set binding factor Pleckstrin Homology (PH) domain. Set															
N239R	103911104258	3 116	12,766	10.69	9 1	cd01235	binding factor is a myotubularin-related pseudo-phosphatase consisting 5 of a Denn domain, a Gram domain, an inactive phosphatase domain, a 5ID 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%	2056	1652	1	AAC96602 A234L		93.59	2.15E-18	44%	60%	1113	1107
N240L	104517104242	2 92	11,364	10.47	7	No Hit Found								1	AAC96601 A233R		85.89	4.40E-16	48%	70%	586	1191
N241L	104802104536	6 89	10,140	4.86		COG0695	protein turnover, chaperones].	43.41	7.16E-06	34%	53%	372	472	1				2.07E-05	33%	51%	483	31114
					2	ptam00462	2 Glutaredoxin, Glutaredoxin.	35.23	2.08E-03	26%	41%	370	3–71	2 3 4 5 6 7 8 9 10	AAZ82854 Hypothetics NP_077613 EV-1-128 ZP 01076784 Glutaredoxi ZP 01301447 hypothetica AAZ27138 glutaredoxi ZP_01166994 putative glu AAV29888 NT02FT11 YP_513693 Glutaredoxi VD_67056 ptteredoxi	in 1 protein Rarvi 01000777 n 1 faredoxin 1 63 in 1	48.91 47.75 47.37 46.21 46.21 44.67 43.51	4.61E-05 6.02E-05 1.34E-04 1.75E-04 3.90E-04 1.14E-03 2.53E-03 3.31E-03	33% 36% 34% 31% 32% 33% 36% 36% 33%	51% 52% 53% 61% 54% 48% 47% 50%	4-83 3-72 2-77 4-84 4-70 3-74 4-77 4-77 4-76	31114 874 478 685 669 575 586 586 581
N242L	105406104831	1 192	20,946	10.14	4	No Hit Found								10	YP_563095 Glutaredoxi AAC96598 A230R	IN, GIXA		6.01E-37	39%	65%	2190	3194
N243R	105431105670		8,656	5.75	5 1	COG1647	7 COG1647, Esterase/lipase [General function prediction only].	28.30	2.46E-01	28%	43%	2975	70116	0	No Hit Found No Hit Four	nd						
							DPS, DDS (DNA Protecting probin under Starved conditions) domain is a member of a broad superfamily of famili-kie dimor-arboxylate proteins. Some DPS proteins nonspecifically bind DNA, protecting it from deavage caused by reactive oxygen species such as the hydroxyl radicate produced during oxidation of Fe(II) by hydrogen perxide. These proteins assemble into doctagement structures, some from DPS-DNA oc- proteins assemble into doctagement structures, some from DPS-DNA oc-															
N244R	105688106083	3 132	14,773	10.00			crystalline complexes, and possess iron and H2O2 detoxification of gabilities. Expression of DFs is induced by oxidative or nutritional stress, including metal ion starvation. Members of the DPS family are homopolymers formed by 12 Gur-heits bundle subunits that assemble with 23 symmetry-into a hollow shell. The DFS ferroxidase site is unusual in that it is not located in a four-heits bundle as in formitin, but is shared by 2-fold symmetry-related subunits providing the iron ligands. Many DFS sequences (e.g. E. coll) digitary an Nerminal extension of variable landth that rontaines twin or threa noticitally change protectanty Dps. DNA-holling fertifini-like protein (oxidative damage protectanty)	32.92	1.38E-02			18128			AAC96595 A227L			1.64E-39	64%		10129	
N246L	109032106141	1 964	97,310	3.61	2	COG0783 No Hit Found	Inorganic ion transport and metabolism1.	29.46	1.26E-01	35%	52%	1867	2779	2	NP_049716 MobD.6 hy			7.70E-05 6.61E-23	27% 25%	50% 41%	29129 487830	33127 39493
1¥240L	.00032100141	. 904	<i>81,</i> 310	3.61		NO THE FOUND								2	BAB83469 Vp260 like BAB83468 Vp260 like BAB83470 Vp260 like	protein	115.93 114.01	7.81E-24 2.97E-23	28% 27%	45% 41%	602-915 590-915	11348 39348
														4 5 6	AAC96490 PBCV-1 Vp BAB83471 Vp260 like BAB83467 Vp260 like	potein	52.76 93.20	8.12E-05 5.42E-17 2.61E-03	28% 23% 22%	44% 41% 40%	665847 457955	8199 30527 7921063

Gene Name	Genome Position	A.A. lenath	Peptide Mw	pl	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identitv F	% Positive	Query from-to	Hit from- BLAS to Nun		Hit BLASTp Definition Accession AAA88307 glycoprotein Vp260	Bit Score 51.60	E-value 1.81E-04	% dentitv P	% ositive 44%	Query from-to 665-847	Hit from- to 8199
														8	AAC96397 Asn/Thr/Ser/Val rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921		1.43E-09	26%	40%	491-833	7349
														9	AAC96382 Asp/Thr/Ser/Val rich protein		4.01E-20	22%	39%		4911250
														10	AAC96386 Asn/Thr/Ser/IIe rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	81.65	1.63E-13	21%	39%	16760	5551304
N253L	110377109085	431	44,963	4.18	3	No Hit Found	I							1	BAB83467 Vp260 like protein BAB83471 Vp260 like protein		1.77E-05 1.60E-06	22% 24%	41% 37%	19284 22383	19281 22405
														3	BAB83468 Vp260 like protein	70.09	1.83E-10 5.33E-10	27%	44%	11209	571-812
														4 5	BAB83469 Vp260 like protein BAB83470 Vp260 like protein	63.16	2.24E-08	27% 26%	43% 42%	31239 11236	501-743 571-849
														6	AAA86307 glycoprotein Vp260 AAC96386 Asr/Thr/Ser/Ile rich protein; similar to Rickettsia cell surface antigen,		8.24E-03 1.50E-04	24% 21%	41% 37%	4261 5372	220-490 474-867
														8	AACOR202 Asn/Thr/Ser/Gly rich protein; similar to E. coli adhesin AIDA-I precursor,		8.80E-05	24%	42%	25233	94312
															ZP 01250137 hypothetical protein Bpse110 02005980	55.07	6.09E-06	25%	38%	17378	267-653
														10	AAC96382 Asn/Thr/Ser/Val rich protein	54.68	7.96E-06	22%	40%	31377	154548
N254L	111615110404	404	45,168	6.32	2 1	pfam0445	Capsid_Indovir, Indovirus major capsid protein. This family includes the major capsid protein of indoviruses, 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	251.38	1.27E-67	31%	50%	3398	5443	1	AAC96917 similar to Chlorella virus PBCV-1 ORF A11L, corresponds to GenBank Accession Number U42590	244.20	6.49E-63	35%	55%	3402	4400
							alvconrotein							2	AAC96379 contains aminoacyl-tRNA synthetase class-II signature		6.72E-60	35%	54%	2402	4403
														3	AAC96378 similar to PBCV-1 major capsid protein, corresponds to Swiss-Prot Accession Number P30328	220.71	7.69E-56	33%	56%	3400	6399
														4	AAC96798 PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052 BAA76601 major capsid protein MCP1		7.19E-54 1.60E-53	33% 32%	50% 50%	3402 3402	5437 5437
														6	BAA76600 major capsid protein BAA22198 major capsid protein Vp54	210.69	7.95E-53 1.50E-51	32% 32%	50% 49%	3402 3402	5436 5437
														8	AAC27492 major capsid protein Vp49 1M4X_C Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	204.53	5.70E-51 1.66E-50	31% 34%	50% 50%	3402 52402	5432 31413
														10	1M3Y_D Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid Containing. Dna Virus		1.66E-50	34%	50%	52-402	31-413
N255R	111871112290	140	15,361	9.27	,	No Hit Found								1	AAC96922 A567L	77.02	2.02E-13	37%	52%	12127	10123
1423511	1110/1-112200	140	13,301	0.21		Normano									AND SUSEZ ADDIE	11.05	2.021-15	3776	52.70	12 127	10-125
N259L	114151112286	622	67,016	7.97	7 1	cd0327;	ABC_SMC3, euk, Eukaryotic SMC3 proteins; SMC proteins are large (approximately 110 to 170 kba), and each is arranged into five recognizable domains. Amino-acid sequence homology of SMC proteins between species is largely confined to the amino- automy-terminal globular domains. The amino-terminal domain contains a ' Walker Adapos, nucleotide-binding domain (GoxGKRSF), the single-telfer amino-acid code), which by mutational studies has been shown to be exeguance (the D-Abox) that resemble a &aposyn3them Edapos and a molf with homology to the signature sequence of the ATP-inding cassette (ABC) family of ATPasses. The sequence homology within the catbox/terminal domain is relatively high within the SMC1-SMCM group, hereras SMC2 and SMC2 shows once divergence in both of these sequences. In eukaryotic cells, the proteins are found as heterodimers of SMC1 paried with SMC3. SMC2 with SMC4 and SMC5 with S	39.63	1.03E-03	20%	46%	29176	212352	1	AAC98921 A585R	45.44	7.78E-03	47%	54%	365-406	253–293
					2	000437	(formerly known as Rad18) 2 COG4372, Uncharacterized protein conserved in bacteria with the	36.64	8.29E-03	21%	44%	38182	77214	2	BAA11343 DNA binding protein	45.44	7.78E-03	47%	54%	365406	252 202
							mvosin-like domain IFunction unknownl. Zein, Zein seed storage protein, Zeins are seed storage proteins. They								AAC96752 Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein,						
					3		are unusually rich in glutamine, proline, alanine, and leucine residues and their sequences show a series of tandem repeats		6.21E-02	29%	44%	97248	75226	3	AAC96752 Lyst, Fronkit, FAR (10x), similar to write From protein, corresponds to GenBank Accession Number X52472	63.93	2.11E-08	45%	53%	251312	2999
					4	COG5064	SRP1, Karyopherin (importin) alpha [Intracellular trafficking and secretion].	33.47	8.80E-02	24%	53%	112191	37117	4	NP_620240 golgi-associated protein GCP360	57.38	1.98E-06	19%	42%	29268	218-480
					5	cd0327	ABC_SMC3, euk, Eukaryotic SMC5 proteins: SMC proteins are large (approximately 110 to 170 kba), and each is arranged into five recognizable domains. Amino-acid sequence homology of SMC proteins between species is largely continue to the amino-auto actows/terminal globular domains. The amino-terminal domain contains a '/Walker Adapos, nucleotic-beinding domain (GucCAGKST). The is angle-letter sequence (the D-Abco) that resembles a '/Walker B&apos. molecules assertial in several proteins. The carboxy-terminal domain contains a sequence (the D-Abco) that resembles a '/Walker B&apos. molecules and a molf with homology to the signature sequence do the ATP-inding cassette (ABC) tarrily of ATPasses. The sequence homology within the carboxy-terminal domain is relatively high within the SMC1-SMC4 group, whereas SMC5 and SMC5 shows some divergence. In both of these sequences. In eukaryotic cells, the proteins are found as heterofiners of SMC1 paried with SMC3. SMC2 with SMC4.		9.29E-02	14%	43%	45183	171299	5	EAS38015 hypothetical protein CIMG_01369	51.22	1.42E-04	23%	45%	32227	556-762
					6	COG049	(formerly known as Rad18) 7 RecN, ATPase involved in DNA repair [DNA replication, recombination, and explicit	33.26	9.93E-02	19%	38%	31155	279-413	6	NP_186943 unknown protein	50.45	2.42E-04	21%	50%	32-182	295440
					7		and recair). COG5280, Phage-related minor tail protein [Function unknown].	32.70	1.27E-01	19%	32%	47250	120325	7			5.39E-04	25%	46%	54185	450-592
							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														
					8	cd00584	fold correctly. The complex contains two alpha and four beta subunits, I de 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 barel assembly with six ordruting collect-olis.	32.91	1.34E-01	25%	42%	90185	12118	8	XP_730668 hypothetical protein PY02748	48.52	9.19E-04	21%	45%	27262	266-489
					9		AtpH, F0F1-type ATP synthase, delta subunit (mitochondrial oligomycin sensitivity protein) [Energy production and conversion].	32.13	2.17E-01	18%	41%	37190	8150	9	AAS10477 merozoite surface protein 3b	48.14	1.20E-03	24%	41%	30245	387-609
					10	COG1842	PspA, Phage shock protein A (IM30), suppresses sigma54-dependent transcription (Transcription / Sional transduction mechanisms).	31.79	2.42E-01	22%	43%	66189	13156	10	BAB66120 337aa long hypothetical protein	47.37	2.05E-03	26%	44%	29180	35190
															AAC96919 similar to Chlorella virus CVK2 DNA binding protein, corresponds to						
N262R	114211114780	190	20,299	4.71		No Hit Found								1	AAC96919 GenBank Accession Number D78305 BAA11342 DNA binding protein		1.38E-17 1.80E-17	43% 43%	71% 71%	65167 65167	61163 61163
							Myosin_tail_1, Myosin tail. The myosin molecule is a multi-subunit							-		01.20	1.002 17	4070	11.0	00 101	01 100
N263R	114795115994	400	42,800	10.56	8 1	pfam01576	complex made up of two heavy chains and four light chains it is a fundamental contractile profession found in all euksynetic cell types. This a maily 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 fiament. The coiled-coil region provides the structural backhone the thick fiament.	34.14	2.74E-02	29%	47%	179	477550	1	AAC96919 similar to Chlorella virus CVK2 DNA binding protein, corresponds to GenBank Accession Number D78305	53.91	1.23E-05	36%	57%	182	191–272
					2	COG281	NtpF, Archaeal/vacuolar-type H+-ATPase subunit H [Energy production and conversion].	32.63	8.67E-02	23%	46%	179	2-73	2	BAA11342 DNA binding protein	51.22	7.95E-05	35%	56%	182	191-272
							TolA, TolA protein. This family consists of several bacterial TolA proteins as well as two eukaryotic proteins of unknown function. Tol proteins are														
					3		involved in the translocation of group A colicins. Colicins are bacterial 9 protein toxins, which are active against Escherichia coli and other related species (See pfam01024). ToA 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%	582	96177	3	BAA83789 alginate lyase	97.44	9.67E-19	35%	51%	212–377	153–317
					4	COG3064	ToIA, Membrane protein involved in colicin uptake [Cell envelope biogenesis, outer membrane].	31.30	2.16E-01	27%	45%	1380	138205	4	AAC96583 PBCV-1 alginate lyase	95.90	2.81E-18	35%	50%	212-377	141-305

Gene	Genome	A.A.	Peptide	pl	CDD Hit	COGs	COG Definition B		-	%	%	Query I	lit from- BLAST	To Hit	Hit DI 107 D. Coltan	<b>B</b> <sup>11</sup> <b>A 1 1</b>	-	%	%	Query	Hit from-
Name	Position	lenath	Мw	рі	Number			Sit Score		Identity P	ositive	from-to	to Num	ber	Accession BLASTp Definition	Bit Score	10	lentitv P	ositive	from-to	to
					5	COG071	conversion]	29.07	9.59E-01	21%	42%	3179	58106	5	BAB19127 vAL-1		6.27E-18 6.73E-04	34% 26%	51% 44%	212-377 189-372	169-333 157-344
														7	EAL19679 hypothetical protein CNBG3070 AAW44587 hypothetical protein CNG01710	48.14	6.73E-04	26%	44%	189372	147-334
														8	AAW45360 expressed protein EAL19065 hypothetical protein CNBH1670		1.50E-03 1.50E-03	26% 26%		218388 218388	324489 324489
														10	BAE48156 hypothetical methionyl-tRNA synthetase		3.34E-03	28%		228372	82231
							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														
N266R	116058116723	222	24,410	10.29	1	pfam04443	Photobacterium phosphoreum LuxE is thought to be acylated during the transfer of the acyl group from the synthetase subunit to the reductase.	31.07	1.29E-01	25%	36%	175219	158202	1	AAC96918 A559L	80.49	4.51E-14	29%	45%	1178	1185
							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														
							THZ_kinase, 4-methyl-5-beta-hydroxyethylthiazole (Thz) kinase														
N267L	117565116714	284	32,264	8.54	1	cd0117	catalyzes the phosphorylation of the hydroxylgroup of Thz. A reaction that allows cells to recycle Thz into the thiamine biosynthesis pathway, as	29.76	3.85E-01	32%	45%	224260	4280	1	AAT60033 intein-containing protein	44.67	4.39E-03	23%	43%	94272	68244
							an alternative to its synthesis from cysteine, tyrosine and 1-deoxy-D- xylulose-5-phosphate														
							differen differen differen hydrohyner differta di MP and														
N269L	118036117614	141	14,826	4.53			dUTPase, dUTPase. dUTPase hydrolyses dUTP to dUMP and pvrophosohate	112.64	1.47E-26	48%	67%	11126	2116	1	AAW51452 deoxyuridine triphosphatase AAC96912 similar to tomato dUTP pyrophosphatase, corresponds to GenBank		4.02E-46	71%	85%	2127	3128
					2	COG071	Dcd, Deoxycytidine deaminase [Nucleotide transport and metabolism].	48.73	2.88E-07	27%	46%	32117	74158	2	AAC96912 Similar to tomato done pyrophosphatase, corresponds to GenBank Accesssion Number S40549	184.50	8.96E-46	69%	87%	2127	3128
							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														
					3	cd0145	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	27.97	5.19E-01	29%	44%	65127	2887	3	AAW51453 deoxyuridine triphosphatase	182.96	2.61E-45	69%	84%	2127	3128
							eukarvotic 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 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 renair by the														
					4	pfam0373	non-homologous end-joining pathway. This is the amino terminal alpha/beta domain. This domain only makes a small contribution to the	27.22	7.28E-01	28%	54%	73127	3585	4	NP_190278 dUTP diphosphatase/ hydrolase	148.29	7.12E-35	60%	74%	3127	29153
							dimer interface. The domain comprises a six stranded beta sheet of the														
					5	COG2433	Rossman fold COG2433, Uncharacterized conserved protein [Function unknown].	27.18	8.95E-01	23%	46%	69127	561		EAS33768 hypothetical protein CIMG 04792		7.12E-35	61%	76%	4127	27150
														6 7	XP 750039 dUTPase XP 657875 hypothetical protein AN0271.2		9.29E-35 1.21E-34	59% 59%	75% 76%	4127 4127	114-237 71194
														8 9	AAB22611 deoxyuridine triphosphatase; dUTPase; P18 XP_469212 putative deoxyuridine triphosphatase		1.21E-34 1.75E-33	60% 58%	72% 74%	3126 5126	32155 90211
														10	ABE98101 Deoxyuridine 5'-triphosphate nucleotidohydrolase, putative		1.75E-33	58%	74%	5126	36157
															expressed						
N271R	118157118981	275	30,440	8.75	1	pfam00352	2 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':		4.82E-44	41%	60%	25268	13268
														2	YP_142807 TATA-box binding protein (TBP)	49.68	1.28E-04	25%	40%	75252	138346
							Capsid_Iridovir, Iridovirus major capsid protein. This family includes the														
N274L	120289118979	437	48,753	9.11	1	pfam0445	major capsid protein of iridoviruses, chlorella virus and Spodoptera ascovirus, which are all dsDNA viruses with no RNA stage. This is the	362.71	4.67E-101	40%	57%	36431	2-443	1	AAC96917 similar to Chlorella virus PBCV-1 ORF A11L, corresponds to GenBank	509.22	1.20E-142	60%	76%	36435	1400
						p	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								Accession Number U42580						
							alvconrotein							2	AAC96379 contains aminoacyl-tRNA synthetase class-II signature	396.74	8.71E-109	49%	67%	37435	3403
														3	AAC96378 similar to PBCV-1 major capsid protein, corresponds to Swiss-Prot		6.07E-86	40%	62%	37433	4399
														4	AAC27492 major capsid protein Vp49		4.99E-72 2.10E-70	38%	53%	37435	3432
														5 6	BAA76601 major capsid protein MCP1 AAC96798 PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession	269.24	2.10E-70 2.74E-70	39% 38%	54% 54%	37435 37435	3437 3437
														7	BAA76600 major capsid protein	265.77	2.32E-69	37%	54%	37435	3436
														8 9	BAA22198 major capsid protein Vp54 1M4X C Chain C. Pbcv-1 Virus Capsid. Quasi-Atomic Model	260.00 238.81	1.27E-67 3.04E-61	37% 37%	53% 53%	37435 67435	3437 9413
														10	1M3V D Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid	238.81	3.04E-61	37%	53%	67435	9413
															<ul> <li>Containing Dia Virus</li> </ul>						
N277L	121892120225	556	62,182	9.61	1	COG0553	HepA, Superfamily II DNA/RNA helicases, SNF2 family [Transcription / DNA reolication, recombination, and repair].	168.40	1.95E-42	26%	42%	81533	314851	1	AAC96911 similar to Caenorhabditis transcription activator, corresponds to Swiss- Prot Accession Number P41877	450.28	9.12E-125	50%	71%	103552	7458
							DNA reolication, recombination, and repair! SNF2_N, SNF2 family N-terminal domain. This domain is found in proteins involved in a variety of processes including transcription														
					2	pfam00176	regulation (e.g., SNF2, STH1, brahma, MOT1), DNA repair (e.g., ERCC6, RAD16, RAD5), DNA recombination (e.g., RAD54), and	127.70	3.45E-30	26%	44%	106357	1287	2	BAC15031 helicase (swi/snf family)	167.55	1.18E-39	30%	51%	103517	6061027
							chromatin unwinding (e.g., ISWI) as well as a variety of other proteins														
					3	COG106	with little functional information (e.g. lodestar FTI 1) SSL2, DNA or RNA helicases of superfamily II [Transcription / DNA	82 84	9.09E-17	22%	37%	102-361	36286	3	CAE79539 putative helicase/SNF2 family domain protein	164 47	1 00E-38	29%	49%	98517	8561294
					4		replication. recombination. and repair!. DEXDc, DEAD-like helicases superfamily; .	79.50	1.08E-15	17%	36%	102-001	9199	4	NP_225044 SWI/SNF family helicase 2		2.23E-38	29%	47%		6981154
							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														
					5	cd00071	autonomously folding unit, but an integral part of the helicase; 4 helicase	78.04	2.56E-15	27%	44%	370497	3130	5	CAC96916 lin1685	162.16	4.96E-38	29%	46%	99548	6121064
							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 DEXHc_DEXH-box belicases. A diverse family of proteins involved in														
					6	cd0026	ATP-dependent DNA or RNA unwinding, needed in a variety of cellular processes. The name derives from the sequence of the Walker B motif	73.96	4.75E-14	25%	49%	124259	1143	6	BAA99057 SWI/SNF family helicase_2	161.00	1.11E-37	29%	46%	99532	6981154
					7	emert00.40	(motif II) This domain contains the ATP-binding region	63.71	E 64E 44	25%	46%	400 400	2 02	7	AAE72520 halfana ooro (aalfa	450.94	2 465 27	28%	47%	07 522	744 4460
					7		HELICc, helicase superfamily c-terminal domain; . Helicase_C, Helicase conserved C-terminal domain. TThis domain	63.71	5.64E-11	25%	40%	409490	382		AAF73530 helicase, Snf2 family	159.84	2.46E-37	28%	47%	97532	7111169
					8	pfam0027	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	62.54	1.44E-10	27%	47%	411-490	178	8	YP_014262 helicase, Snf2 family	159.46	3.22E-37	29%	46%	99532	6121053
							integral part of the helicase DEXDc_DEAD-like helicases superfamily. A diverse family of proteins														
					9	cd00044	b) involved in ATP-dependent RNA or DNA unwinding. This domain contains the ATP-hinding region	60.10	6.98E-10	18%	36%	125259	2144	9	ZP_00231434 helicase, Snf2 family	159.46	3.22E-37	29%	46%	99532	6121053
					10	000000	SrmB, Superfamily II DNA and RNA helicases [DNA replication,	45.92	1.14E-05	20%	40%	348-492	227364	40	AAC68303 SWF/SNF family helicase	450.00	5.49E-37	28%	47%	07 500	6931151
					10	CUG0513	recombination, and repair / Transcription / Translation, ribosomal structure and biogenesis	40.92	1.14E-05	20%	4U%	340-492	221-304	10	ANGOUGUS OVER/ONE IdENIIY REIICASE	158.69	J.49E-3/	20%	41 %o	81532	0831151
N284R	121783122544	254	27,361	4.17	· 1	COG4934	COG4934, Predicted protease [Posttranslational modification, protein	28.42	9.66E-01	21%	36%	164-217	80133	0	No Hit Found No Hit Found						
				4.0			turnover. chaperonesl.	20.42	5.00E-01	21/0	0070	104 1217	50 100								
N287L	122844122626	73	6,960	12.35		No Hit Found								0	No Hit Found No Hit Found						

Gene Name	Genome Position	A.A. lenath	Peptide Mw		CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identitv	% Positive	Query from-to	Hit from- B to	LASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value lo	% dentitv F	% Positive	Query from-to	Hit from- to
N288R	122936123199	88	9,777	10.68	1	COG2339	9 COG2339, Predicted membrane protein [Function unknown].	26.43	9.34E-01	31%	43%	2869	4082	1	AAC96886	A519L	103.99	1.58E-21	64%	75%	780	477
N289R	123228124391	388	44,114	9.57	1	COG1231	1 COG1231, Monoamine oxidase [Amino acid transport and metabolism].	52.68	7.73E-08	18%	30%	1384	1439	1	BAA83788	orf1	560.45 3	.85E-158	71%	82%	8383	19394
					2	pfam03738	GSP synth, Gkatahionykpermidine synthase. This region contains the Guitabilonykpermidine synthase serupratic activity (55.3.1.8. This the C-terminal region in bienzymes. Gkatahionykpermidine (GSP) synthasase of trypanosimotidise and Escherichia cal couple hydrolysis 8 of ATP (to ADP and P) with formation of an amide bord between spermidine and the sylcine catboxylate of gkatahione (gamma-Gluczys- Gy). In the pathegenic trypanosimotidise, this reaction is the perultimate step in the loosynthesis of the antioxidant metabolite, trypanohione (H) NA-bis-(glutathoryl)spermiding, and is a target for drug design.		3.66E-01	28%	42%	117–184	192257	2	AAC96585	similar to bovine monoamine oxidase, corresponds to Swiss-Pro Accession Number P21398	549.67 6	i.79E-155	69%	81%	8383	16391
N293R	124417125514	366	40,178	10.11		No Hit Found	1							1	AAC96583	PBCV-1 alginate lyase	416.39 8	.27E-115	64%	73%	61362	5320
														2 3 4 5 6 7 8	BAB19127 BAE48156 AAC96919 BAA11342 AAC96539 AAC96508	alqinate Ivase vAL-1 hycothetical methionyV-IRNA synthetase similar to Chorella virus CVK2 DNA binding protein, corresponds to GenBank Accession Number D78305 DNA binding protein similar to PBCV-1 ORF A41R, corresponds to GenBank Accession Number U17055 PBCV-1 surface protein surface protein	404.06 4 80.11 65.47 64.31 48.91 44.67	6.56E-112 1.25E-111 1.41E-13 3.58E-09 7.98E-09 3.47E-04 6.54E-03 6.54E-03	60% 55% 28% 30% 31% 40% 42% 42%	70% 65% 44% 50% 48% 56% 62% 62%	61362 1362 125357 230363 230363 359 352 352	5332 1348 5247 501648 500647 460 906955 906955
														10	T17636	proline-rich protein A145R - Chlorella virus PBCV-1		6.54E-03	42%	62%	352	115-164
N299R	125532126188	219	24,231	10.35		No Hit Found	1							0	No Hit Found	No Hit Found						
N303R	126293126754	154	17,595	6.55		No Hit Found	1							1 2	AAC96582 XP_767961	A214L hypothetical protein GLP 228 21235 22845		1.55E-29 7.24E-03	46% 23%	65% 49%	17149 7113	3135 145255
N304R	126830127282	151	16,662	4.56		No Hit Found	1							1	AAC96581	A213L	139.04	4.41E-32	53%	79%	1126	1125
N307L	128331127321	337	37,918	4.05		No Hit Found								0	No Hit Found	No Hit Found						
N312L	129519128398	374	41,982	5.21	1	pfam02784	Orn_Arg_deC_N, Pyridoxal-dependent decarboxylase, pyridoxal binding 4 domain. These pyridoxal-dependent decarboxylases acting on ornithine, 1 lysine, arginine and related substrates This domain has a TIM barrel fold	108 56	8.96E-52	39%	57%	25260	4246	1	AAC96575	PBCV-1 arginine decarboxylase	497.28 3	.82E-139	63%	79%	1372	1372
					2	COG0019	LysA, Diaminopimelate decarboxylase [Amino acid transport and metabolism].		7.90E-51	27%	44%	19372	29394	2	CAE02644	ornithine decarboxylase	269.63	1.29E-70	40%	59%	1364	43407
					3	pfam00278	Orn_DAP_Arg_deC, Pyridoxal-dependent decarboxylase, C-terminal 8 sheet domain. These pyridoxal-dependent decarboxylases act on	89.49	5.74E-19	30%	46%	263-365	1103	3	CAA71498	ornithine decarboxylase	266.93	8.39E-70	38%	57%	1364	48418
					4	COG1166	ornithine. Ivsine. arainine and related substrates $_{\rm 6}$ SpeA, Arginine decarboxylase (spermidine biosynthesis) [Amino acid transport and metabolism].	53.33	4.42E-08	23%	41%	45270	118374	4	XP_968571	PREDICTED: similar to Ornithine decarboxylase (ODC)	266.54	1.10E-69	39%	57%	9364	26385
					5	COG1387	7 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	1587	5	AAL87201	ornithine decarboxylase	266.16	1.43E-69	40%	59%	1364	50420
														6 7		ornithine decarboxylase putative ornithine decarboxylase 1		1.87E-69 4.16E-69	39% 40%	58% 59%	1364 1364	47417 50420
														8	NP_571876	ornithine decarboxylase 1 ornithine decarboxylase	263.85 263.46	7.10E-69	37% 38%	56% 57%	2372 1364	21409 48418
							SGNH_arylesterase_like, SGNH_hydrolase subfamily, similar to arylesterase (7-aminocephalosporanic acid-deacetylating enzyme) of A.							10		ornithine decarboxylase	263.08	1.21E-68	38%	60%	13371	54428
N313L	130282129647	212	23,020	11.49	1	cd01839	tumetaciens. SGNH hydrolases are a diverse family of lipases and 9 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(Giu) triad frund in rither serine hydrolase	28.33	7.78E-01	32%	48%	67–111	89133	1	AAC96573		176.79	4.21E-43	52%	62%	13203	9205
														2	AAC96461	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563	44.67	2.50E-03	35%	53%	126191	369
N315L	130984130319	222	23,976	4.83		No Hit Found	1							1	AAC96571	A203R	269.63	5.24E-71	63%	74%	1220	1213
N317R	131031131372	114	12,026	4.52	1		7 ZntA, Cation transport ATPase [Inorganic ion transport and metabolism]. DSPc. Dual specificity phosphatases (DSP): Ser/Thr and Tvr protein	26.73	7.52E-01	28%	52%	1338	1338	1			117.09	1.77E-25	52%	71%	3112	2112
N318L	131927131367	187	20,852	10.87	1		7 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 DSPc, Dual specificity phosphatase, catalytic domain. Ser/Thr and Tyr		4.20E-32	35%	53%	26159	4139	1	AAC96673	similar to human protein Tyr-phosphatase, corresponds to GenBank Accession Number U27193	242.66	4.62E-63	66%	85%	6168	7169
					2	pfam00782	protein phosphatases. The enzyme' tertiary fold is highly similar to that of tyrosine-specific phosphatases, except for a &auot:recognition&auot: region.	124.98	4.44E-30	35%	55%	24161	2-139	2		dual specificity phosphatase 16	87.43	2.48E-16	34%	52%	24164	159300
					3	smart00195	5 DSPc, Dual specificity phosphatase, catalytic domain; .	122.69	2.52E-29	34%	56%	24161	2139	3	XP_543810	PREDICTED: similar to Dual specificity protein phosphatase 16 (Mitoger activated protein kinase phosphatase 7) (MAP kinase phosphatase 7) (MKP-7)	84.73	1.61E-15	34%	53%	24163	159-299
					4	COG2453	mechanisms]	01.24	7.52E-08	27%	50%	58143	65147	4		DUSP16 protein		3.58E-15	34%	53%	24163	159299
					5 6	pfam00102	PTPc motif, Protein tyrosine phosphatase, catalytic domain motif; .     Y phosphatase, Protein-tyrosine phosphatase	35.01 33.42	5.22E-03 1.68E-02	22% 22%	48% 50%	99147 98143	3690 166216	5	AAI09236	Unknown (protein for IMAGE:5176724) Dual specificity phosphatase 16	83.57	3.58E-15 3.58E-15	34% 34%	53% 53%	24163 24163	110-250 159-299
					7	smart00194	PTPc, Protein tyrosine phosphatase, catalytic domain; COG2365, Protein tyrosine/serine phosphatase [Signal transduction]	31.45 31.19	7.33E-02 8.72E-02	23% 37%	47% 50%	72143 101139	168243 136174	7		Dual specificity phosphatase 16 PREDICTED: similar to KIAA1700 protein		3.58E-15 3.58E-15	34% 34%	53% 53%	24163 24163	159-299 268-408
					9		mechanismal. Glycogen_syn. Gly predistright family consists of the subarycoid of program. Syn. Gly predistright CTL (OVS and OVS3: Glycogen synthuses (GS) is the enzyme responsible for the synthesis of -14-inked glucose chanis in glycogen. It is the relate limiting enzyme in the synthesis of the polysaccharide, and tis activity is highly regulated through phosphorylation, and multiple sites and also by allocitic effectors, mainly	29.88	2.28E-01	31%	43%	78-148	174249	9	-	KIAA1700 protein		3.58E-15	34%	53%	24163	
					10	cd00047	clucose 6-bloschale (GBP). PTPC, Protein tyrosine phosphalases (PTP) catalyze the dephosphorylation of phospholyrosine peptides; they regulate phospholyrosine livels in signal ratioaction pathways. The depth of they phospholyrosine, initiad of pSer or DTr. The family has a districtive active site signature molf, HCSAGGKRG, Characterized as ether transmentorane, receptor-like or non-transmembrane (soluble) PTPs. Receptor like PTP domains tend to occur in two copies in the cytoplasmic region of the	) 29.06	3.51E-01	42%	74%	101–120	166185	10	AAH59232	Dusp16 protein	82.42	7.98E-15	35%	52%	24162	159–298
N10001	100055 10107-		40.5			No. LIN Town	transmembrane proteins, only one copy may be active								A A COORCE -	10075		2 205 22	2021	F00/	10.00	
N320L	132255131986	90	10,333	9.82		No Hit Found	NAAAR, N-acylamino acid racemase (NAAAR), an octameric enzyme							1	AAC96989	A687R	43.13	3.29E-03	36%	52%	1683	674
							that catalyzes the racemization of N-acylamino acids. NAAARs act on a broad range of N-acylamino acids rather than amino acids. Enantiopure															
N321L	132624132364	87	10,211	3.83	1	cd03317	amino acids are of industrial interest as chiral building blocks for 7 antibiotics, herbicides, and drugs, NAARA is a member of the enclase superfamily, characterized by the presence of an enclate 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 MAP> ion	27.84	2.86E-01	41%	66%	3669	4978	1	AAC96674	A306L	102.45	4.62E-21	82%	92%	3485	3586
N323L	133084132743	114	12,977	5.02		No Hit Found								1	AAC96676	A308L	46.21	3.83E-04	38%	66%	4398	2079

Gene Name	Genome Position	A.A. lenath	Peptide Mw		CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identitv	% Positive	Query from-to	Hit from- I to		Hit BLASTp Definition	Bit Score	E-value	% dentitv P	% ositive	Query from-to	Hit from- to
N325L	133669133160	170	18,092	8.10	1	cd02666	Peptidase_C19.J. A subfamily of Peptidase C19. Peptidase C19 contains ubiquitry hydrolases. They are intracelular peptidases that remove ubiquitry molecules from polyubiquitated peptides by cleavage of largeplication bands. They hydrolyze bands involving the category group of ubiquitration on though to be editing of the ubiquitry conjugates, which could rescue them from degradation, as well as recycling of the ubiquitry the ubiquitriation is trought to be editing of the ubiquitry conjugates, which could rescue them from degradation, as well as recycling of the ubiquitry in the ubiquitry of period period period by the thought of the origin the largest families of peptidases in the human genome.	30.00	1.42E-01	42%	55%	91–123	740	1	AAC90678 A310L	184.50	1.15E-45	53%	70%	1168	1170
N327L	134603133710	298	33,061	9.03	1	pfam03658	UPF0125, Uncharacterized protein family (UPF0125)	28.62	9.59E-01	25%	40%	220292	1984	1 2 3 4	AAA88833 33 kDa in vitro translation peptide NP_077561 EsV-1-76 AAR26966 FirrV-1-11 AAR26885 FirrV-1-810	83.57 58.15	5.45E-92 9.20E-15 4.14E-07 2.05E-06	71% 28% 24% 25%	82% 43% 38% 39%	56290 60291 64268 56277	1230 10259 10222 13247
N329L	134854134597	86	9,357	8.92		No Hit Found								1	AAC96681 A313L	46.98	2.31E-04	57%	69%	1143	133
N330R	134941135177	79	8,884	5.01		No Hit Found								1	AAC96682 A314R	117.86	1.05E-25	75%	80%	277	580
N331R	135231136355	375	41,182	4.87		No Hit Found								1 2 3 4 5	AM/030862 membrane protein, putative ZP_01138206 membrane protein, putative AM80642 mycohetical membrane protein AM80642 mycohetical protein AA885442 unknown	87.43 86.27 81.26	4.07E-16 9.06E-16 2.02E-15 6.50E-14 3.56E-12	26% 25% 22% 26%	46% 46% 52% 47%	130-357 130-357 130-357 126-358 119-358	18248 18248 18248 2231 14262
N333R	136377136733	119	12,895	10.56	1	COG5169	HSF1, Heat shock transcription factor [Transcription].	27.68	3.71E-01	27%	45%	51106	112167	1	AAC96688 A320R		9.56E-11	31%	47%	1115	
N335R	136751137095	115	11,866	4.87		No Hit Found								1	AAC96689 A321R		5.87E-13	33%	61%	4113	
N336L	137697137194	168	19,111	5.50		No Hit Found								1	AAC96690 A322L	121.32	1.14E-26	38%	56%	1159	1170
N339L	138996137725	424	46,766	5.00	1		DUF893, Bacterial membrane protein of unknown function (DUF893). This family consists of several putative bacterial membrane proteins of unknown function		4.37E-01	29%	46%	104-196	239323	1	AAC96692 A324L	367.08	7.07E-100	49%	60%	11422	13453
					2		SIN3, Histone deacetylase complex, SIN3 component [Chromatin structure and dynamics].		7.46E-01	26%	33%	319394	334407	2			2.26E-13	32%	55%	118254	
					3	pfam06459	. RR_TM4-8. Ryanodine Receptor TM 4-6. This region covers TM regions 4-6 of the reanodine recentor 1 family	29.22	9.89E-01	20%	32%	318-419	11114	3 4 5 7 8 9	NP_07588 EaV-1-103 YP_142803 unknown YP_291422 unknown YP_291422 unknown A199780 ORF064L AAT971869 ORF074L AAT92360 ORF57L NP 149530 067R	56.23 48.14 47.75 44.67 44.67	8.04E-03	31% 26% 26% 25% 25% 26%	54% 52% 48% 50% 50% 50% 48%	121-256 140-250 122-258 117-253 117-253 117-253 120-261	183-315 62196 62196
N342L	139681139040	214	24,766	8.76	1	COG0494	MutT, NTP pyrophosphohydrolases including oxidative damage repair enzymes [DNA replication, recombination, and repair / General function	34.96	8.28E-03	21%	34%	10180	5161	1	AAC96694 A326L	234.96	1.31E-60	61%	81%	14188	1175
					2		nerdiction only NUDIX, NUDIX domain ProRS_core_arch_euk, Proyl-IRNA synthetase (ProRS) class II core catalytic domain. ProRS is a homodimer. It is responsible for the attachment of proline to the 3&apors, OH group of ribose of the appropriate IRNA. This domain is primarily responsible for ATP-	32.13	5.60E-02	23%	41%	17155	4119	2	CAH13320 hypothetical protein	43.90	4.29E-03	29%	53%	17109	423–513
					3	cd00778	appropriate movements donated to primary responsave to Arri- dependent formation of the enzyme bound aminoacyl-adenytake. 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 eukarotes and some bacteria.		8.67E-01	23%	47%	102–144	104147	3			4.29E-03	27%	48%	15134	14141
														4			7.31E-03	27%	51%	17147	
N346L N351R	140751139690	354	40,496	9.78		No Hit Found								1	AAC96696 A328L		2.83E-64	37% 71%	58% 82%	1352	1350
NJSTR	141062141790	243	26,762	5.32		No Hit Found								1 2 3 4	AAA88833 33 kDa in vitro translation peptide NP_077561 EsV-1.76 AAR26966 FirrV-1.11 AAR26885 FirrV-1.810	84.73 58.54	3.81E-92 2.88E-15 2.21E-07 4.93E-07	28% 26% 26%	43% 40% 40%	1235 5236 9213 1222	10259 10222
Leu Arg Gly Asn Asn Tyr Intron Lrys	141848141920 141945142028 142053142125 142129142199 142223142296 142320142393 142416142501 142453142465 142504142576 142578142809	73 bs 73 bs 85 bs 13 bs 72 bs					anticodon: TAT anticodon: TAA anticodon: TCC anticodon: GTT anticodon: GTT anticodon: GTA Intino (142453-142465) anticodon: CTT anticodon: CTT anticodon: CGT														
N354L	143383142856	176	20,240	6.76		No Hit Found								1	AAC96705 A337L	66.24	5.00E-10	48%	59%	100173	578
N357R	143563143880	106	11,635	8.82		No Hit Found								0	No Hit Found No Hit Found						
N359L	145663143981	561	60,560	9.73		No Hit Found								1	AAC96710 A342L	736.49	0.00E+00	68%	78%	16550	37561
N364R	145777147096	440	48,944	4.51	1		Girco Jiydro 70n, Giycoyi hydrolase family 70, N-terminal domain, Family of endo-beth-N-glucomotalse, or hoparanase. Heparan silled proteoglycans (HSPGs) jalya key role in the self-assembly, insolubility and barrier properties of basement membranes and extracellate matrices. Hence, cleavage of heparan suffate (HS) affects the integrity and functional state of tissues and thereby fundamental normal and pathological phenomena involving cell migration and response to chanages in the extracellular micro-environment. Heparanase degrades	40.37	1.68E-04 4.12E-04	25%	41%	57-247 146-237	66256 149243	1	CAD86595 cellulase precursor		3.68E-06 2.64E-04	23%	35%	54367 57236	282-574 273-416
							HS at specific intra-chain sites. The enzyme is synthesised as a latent approximately 65 kDa protein that is processed at the Netminus into a highly active approximately 50 kDa form. Experimental evidence suggests that heparanase may facilitate both tumoru cell invasion and neovascularization, both critical steps in cancer progression. The enzyme is also involved in cell migration associated with infammation and automounting ATH. [Fehalose and maltose hydrolases (possible phosphorylases)		7.005.00		100								00%	57,000	405 400
					3	COG1809	Arin', intellable and inalide inpludease (publicate publication) areas ( Cool 1989, Uncharacterized conserved profile Function unknown). Cool 1989, Uncharacterized conserved profiles ( domain. This family of glycosy) hydrolase family 65 central catalytic domain. This family of glycosy) hydrolase actions vacuular acid intentiase and matice phosphorybase Matters phosphorase (MF) as dimeric enzyme that catalyses the conversion of matices and inorganic phosphale into beta-bglucose-1 hoopshate and glucose. The central	31.43	7.22E-02 2.03E-01	48%	76%	62162 205230	2247	4	ZP_0076677 similar to Alpha-L-arabinofuranosidase BAA10985 cellulosa-binding protein	48.14	3.45E-04 7.68E-04	24% 23%	37%	57290	105-466 677-879
					5	pfam03632	phosphate into bear-t-phicose-t-phicsphate and glocose. The certinal domain is the catalytic domain, which binds a phosphate ion that is proximal the the highly conserved GNu. The arrangement of the phosphate and the glutamate is though to cause nucleophile attack on the anometic carbon atom. The catalytic domain also forms the majority or tha dimension interfance		2.61E-01	29%	46%	62167	76159	5	AAC45377 endoglucanase F precursor	47.37	1.31E-03	23%	41%	143–290	728-879
N367L	147566147138	143	16,043	10.44		No Hit Found								1	BAA22200 URF14.2	158.30	7.04E-38	68%	81%	14122	1109

Gene Name	Genome Position	A.A. Ienath	Peptide Mw	pl	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identitv P	% Positive	Query from-to	Hit from- B to	LASTp Hit Number 3 4	Accession BLASTP Definition B BAA22202 URF14.2 BAA22201 URF14.2 BAA22199 URF14.2	157.53 156.38	9.20E-38 1.20E-37 2.68E-37	% dentity P 68% 67% 66%	0sitive 82% 81% 81%	Query I from-to 14122 14122 14122	Hit from- to 1109 1109 1109 74141
N369L	149048148017	344	38,430	8.32	: 1	cd0031	CyL C5_DNA, methylase. Cytosine-C5 specific DNA methylases: Methy transfer reactions play an important role in many aspects of biology. Cytosine-specific DNA methylases are found both in prokaryotes and eukaryotes. DNA methylasiton, or the covalent addition of a methyl group to cytosine within the context of the CpG diructedide, has profound 16 effects on the marimalian genome. These effects include transcruptional repression via mitholiton of transcruption factor binling of the recultance factors, X chromesome inactivation, imprinting and the suppression of aparastic DNA equences. DNA methylation is allo sessential for proper sential via sequences. DNA methylation is allo sessential for proper	l 162.40	6.83E-41	30%	43%	5259	4271	5	AAC96717 A349L AAC96884 M.CviAll cytosine DNA methytkransferase		7.56E-24 3.96E-124	61%	86%	1683 1340	/4141
							embryonic development and is an important player in both DNA repair and genome stability														
					2	pfam0014 COG021	45 DNA methylase, C-5 cytosine-specific DNA methylase Dcm, Site-specific DNA methylase [DNA replication, recombination, and	156.24 108.24	4.66E-39 1.21E-24	29% 28%	44% 44%	5264 1191	4267 3197	2	AAC96897 M.CviAIV cytosine DNA methyltransferase AAV84097 CviPII m5C DNA methyltransferase		2.40E-121 1.29E-82	61% 47%	76% 61%	1328 2340	2332 16357
					4	pfam0541	<sup>6</sup> recardi, pominio, Top pomining are an emerging family of problem Pomomorp the multiplean emethanes problem diable a novel (pologies) Mouse prominin and human prominin (mouse)-kika 1 (PROML1) are prodiced to contain five membranes spanning domains, with an N- 78 terminal domain exposed to the extracellular space followed by four, alternating small cytoplasmic and large extracellular, loops and optoplasmic Cherninal domain. The space function of prominin is orominin cause related interpretation. FindoN1: the gene coding for committing cause related interpretation. FindoN1: the gene coding for committing cause related interpretation. FindoN1: the gene coding for the space related interpretation. FindoN1: the space coding for the space related interpretatio	30.65 a	3.11E-01	37%	57%	205–240	200235	4	AAC64006 cytosine methyltransferase	286.19	1.17E-75	44%	58%	2342	3358
														5	AAC55063 cytosine methyltransferase AAC96987 nonfunctional M.CviAV cytosine DNA methyltransferase	261.92	1.06E-68 2.36E-68	40% 40%	56% 56%	5342 5342	6362 6362
														7	AAR23218 gp67 NP 818425 gp127	101.29 92.43	5.34E-20 2.48E-17	33% 35%	49% 51%	2162 5155	6192 6165
														9 10	NP_150145 putative DNA methylase AAC98421 methyl transferase		3.24E-17 3.24E-17	32% 33%	44% 47%	5222 5170	4226 4178
N372L	149874149227	216	23,956	3.74	1	COG196	S6 CstA, Carbon starvation protein, predicted membrane protein [Signal transduction mechanisms]	28.31	8.02E-01	28%	53%	537	2-34	1 2	AAC96720 Asp/Glu rich; DAEDDDIYxxET (2X) negative charge cluster YP_142843 unknown		3.30E-67 2.75E-05	63% 32%	71% 58%	1214 135214	1207 140221
N373L	150935149967	323	36,800	4.01		No Hit Foun	id							1	AAC96725 A357L	164.08	6.10E-39	50%	72%	140300	86249
N377L	151591150974	206	23,714	5.64		No Hit Foun								0	No Hit Found No Hit Found						
N380R	151657152352	232	26,010	9.70	1	smart0049	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown 97 function, but possibly DNA-binding via helix-turn-helix motif (Ponting,	1 , 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%	1227	1224
					2		Instabilishedi Girk, Gir-YIG Ope nucleases (URI domain), Girk-Gir-YIG Ope nucleases (URI domain), torning endonuclease endonucleases activity in the (URI domain), torning endonuclease activity of the Control of the second second second second homologous allele that lacks the sequence. They catalyze a double strand break in the DNA rear the mission side of that element to facilitae homologous allele that lacks the sequence. They catalyze a double strand break in the DNA rear the mission side of that element to facilitae homologous of the test of these motifs in their respective N-	-	1.26E-05	34%	54%	191	183	2		201.06	2.50E-50	44%	60%	1228	1242
					3	cd0028	3 termit: LAGLIDADG, His-Oys box, HH4; and GIV-YIG. This CD contains several but not all members of the GIV-YIG family. The C-terminus of GIV-YIG is a DNA-binding domain which is separated from the N- terminus by a long. flexible links: The DNA-binding domain consists of a minor growe binding alpha-belix, and a helik-turn-helix. Some also contain a zinc funger (i.e. 1-rolly which is not required for DNA binding or catalysis, but is a component of the linker and directs the catalytic diversion the the homing set at a fixed distance from the inflow reserves.	41.14	1.33E-04	46%	61%	111–226	19113	3	AAC98655 PBCV-1 33kd pepidde	172.56	9.51E-42	41%	54%	5227	7248
					4	pfam074	53 NUMOD1, NUMOD1 domain	37.33	1.52E-03	43%	63%	175-205	131	4	AAC96862 similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	141.35	2.35E-32	41%	55%	1201	1196
					5	pfam0154	GIY-YIG, GIY-YIG catalytic domain. This domain called GIY-YIG is found in the amino terminal region of excinuclease abs subunct (cwrC), bacteriophage T4 endonucleases seg4, seg8, seg0, seg0 and seg2; it is also found in putative endonucleases encoded by group introns of fungi 11 and phage. The structure of I-TevI a GIY-YIG endonuclease, reveals a novel alphagebat-Jold with a central three-strended antibaralle beta-sheet	s 36.68	2.71E-03	28%	47%	188	189	5	YP_293795 putative endonuclease	72.40	1.34E-11	39%	55%	2108	3110
							flanked by three helices. The most conserved and putative catalytic residues are located on a shallow, concave surface and include a metal concrimation site 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 protors														
					6	pfam0043	through membrane (inner membrane in mitchondria, thydakoid 30 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 synthese) which are frund in framm00006	29.44	3.90E-01	34%	53%	131	1648	6			8.41E-06	34%	53%	4104	34133
														7	CAA25939 unnamed protein product NP_049674 MobB homing endonuclease	50.06	7.12E-05 7.12E-05	32% 32%	54% 54%	155-227 155-227	75143 112180
														9 10	AC96502 similar to bacteriophage T4 intron-associated endonuclease, corresponds to Swiss-Prot Accession Number P13299 NP_899393 SeqD		1.59E-04 1.59E-04	34% 31%	52% 43%	289 1173	995 1157
N383R	152377155754	1126	123 633	10.87	1	smart/0049	90 HELICc, helicase superfamily c-terminal domain; .	37.52	1.02E-02	36%	60%	848890	3880	1	AAC96731 similar to chicken vitellogenin II, corresponds to Swiss-Prot Accession Number P02845	835.48	0.00E+00	61%	75%	3521028	1651
Noont	102017 100704	1120	120,000	10.07			Helicase_C, Helicase conserved C-terminal domain, This domain r1 family is found in a wide variety of helicases and helicase related	1													
					2		proteins, it may be that this is hot an autonomously toxing unit, but an integral part of the helicase. 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 warehe of helicases and helicase ratelated noteins; may not be an	1	1.96E-02	24%	51%	841-890	2776	2	AAC96728 A360R	240.74	2.50E-61	55%	68%	4225	14239
					3	cd0007	79 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 durinker in the organisa.	31.82	4.33E-01	26%	57%	848-890	79121	3	AAC96729 A361R		4.76E-28	70%		248331	285
														4 5 6 7	ABF82117 hypothetical protein MI/V087L YP_142731 putative NTPase I YP_142917 helicase conserved C-terminal domain protein CAJ57278 putative helicase	49.68 46.21	1.35E-06 8.18E-04 9.04E-03 9.04E-03	25% 24% 30% 26%	46% 51%	785938 833941 832912 842965	414556 608721 437519 318442

Gene	Genome	A.A.	Peptide	pl	CDD Hit	COGs	COG Definition	Bit Score	E-value	%	%	Query	Hit from- B			BLASTp Definition	Bit Score	E-value	%	%	Query	
Name	Position	lenath	Mw	•	Number		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			Identitv F	ositive	from-to	to	Number	Accession				Identitv P	ositive	from-to	to
N387R	155817156737	307	35,901	10.26	1	cd00283	homing at that site. Class I homing endowcleases are sorted into four families based on the presence of these motifs in their respective N- termin: LAGLADAG, His-Cys box, NH, and GNY-YIG. This CJ contains several but not all members of the GNY-NIG 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-felix, and a beitx-un-beic. Score also	56.55	4.35E-09	46%	66%	126–187	160	1	AAA88832	unknown	84.34	5.68E-15	34%	51%	29198	1175
							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									similar to PBCV-1 ORF A315L, corresponds to GenBank Accession						
					2	smart00465	GIVc, GIV-YIG type nucleases (URI domain); . GIV-YIG, GIV-YIG catalytic domain. This domain called GIY-YIG is found in the amino terminal region of excinuclease abc subunit c (uvrC), bacteriophage T4 endonucleases encoded by group intros of fungi also found in putative endonucleases encoded by group intros of fungi		1.69E-01	32%	57%	33117	6-83	2		Number M74440		1.40E-13	28%	45%	29276	1218
					3	pfam01541	also totato in plutative encontrol dealess enclosed by gloup introl so in tang- and phage. The structure of 1-Favi a GIV-YTG enclonuclease, reveals a novel alphabeta-foid with a central three-stranded antiparallel beta-sheet finked 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%	30112	2-85	3		similar to Chicrella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580		5.88E-12	28%	41%	29276	1220
														4 5	AAC40244	PBCV-1 33kd peptide ORF301	66.24	1.31E-11 1.60E-09	28% 30%	44% 50%	39276 38186	15244 83240
														6	AAU16837	GIY-YIG catalytic domain containing protein; possible intron encoded endonuclease		1.04E-08	29%	46%	28180	2155
														7	NP_074951 CAC51107	putative GIY-YIG endonuclease	57.00		28% 25%	52% 40%	42178 27274	85227 5225
														9 10	AAC49248 AAK09365	ORF211 intron encoded Bmol		2.16E-06 8.23E-06	44% 23%	68% 41%	104–161 40–249	86139 14242
N389L	157485156868	206	23,558	10.99		smart00465	GIYc, GIY-YIG type nucleases (URI domain);	34.28	1.14E-02		38%	76172	1-83	1	AAC96747			1.54E-39	44%	60%	12203	5203
					2	COG0322	UvrC, Nuclease subunit of the excinuclease complex [DNA replication, recombination, and repair].	28.33	7.66E-01	41%	47%	72111	1650	2		putative probable transcription repressor HOTR putative transcription repressor HOTR		6.11E-04 1.04E-03	25% 25%	40% 40%	42182 42198	29193 26206
							GIY-YIG_Clerm, GIYX(10-11)YIG family of class I homing endocuclasses C-terminus (GIY-YIG_Clerm). Homing endonucleases promete the mobility of intron or inten 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 facilate homing at that site. Class I homing endorucleases are sorted into four families based on the presence of these motifs in their respective N-							-								
N391R	157646158395	250	28,257	8.78	1	cd00283	tamiles based on the presence of these motifs in their respective N- terment based on the presence of these motifs in their respective N- terment based on therebase of the OX*10G trans). The Coronna of GIV-YIG is a DIA-binding domain which is separated from the N- terminus by along fixeble initer. The DNA-binding domain consists of a micro-groove binding alpha-telix, and a helix-turn-heik. Some also contain a zic finger (i.e. 1-TeV) which is not required for DNA binding or catalysis, but is a component of the linker and directs the catalytic domain to glaves the homing site at a feed distance from the into	66.95	2.30E-12	46%	63%	101–172	1–72	1	AAA88832	unknown	246.90	4.59E-64	51%	70%	1244	1240
					2	smart00497	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting,	42.04	6.10E-05	35%	52%	193246	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%	1247	1226
					3	pfam07453	unpublished). NUMOD1, NUMOD1 domain GIY-YIG, GIY-YIG catalytic domain. This domain called GIY-YIG is found in the amino terminal region of excinuclease abc subunit c (uvrC).	37.33	1.74E-03	44%	63%	193-225	1–33	3	AAC96655	PBCV-1 33kd peptide	198.75	1.43E-49	46%	61%	5244	7247
					4	pfam01541	backrichphage T4 endorrulcases segA, segB, segC, segD and segE; it is also found in putwise endorulcases encoded by group infrints of fungi and phage. The structure of I-Tevi a GIY-YIG endonuclease, reveals a novel alphabeta-fold with a central three-stranded antiparatelle beta-sheet finaled by three helices. The most conserved and putative catalytic residues are located on a shallow, concave surface and include a metal coordination ste.	36.29	3.93E-03	26%	46%	188	1-88	4	AAC96862	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	171.79	1.88E-41	45%	62%	1219	1196
							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 hydroxylase 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 oxygenase inabuling a ielly															
					5	pfam05118	roll strand core and residues binding iron and 2-oxoplutarate, consistent with divergent evolution within the extended family. This family represent the arginine, asparagine and proline hydroxylases. The apartyl/saparaginyl beta-hydroxylase (EC:11.41.16) 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%	2091	7-69	5	AAK09365	intron encoded Bmol	82.42	1.50E-14	34%	48%	4219	5242
					6	COG4678	COG4678, Muramidase (phage lambda lysozyme) [Carbohydrate transport and metabolism].	28.40	9.87E-01	23%	42%	61116	2484	6		SegD unnamed protein product		2.56E-14 8.22E-13	33% 31%	46% 49%	1196 14204	1205 124311
														8	AAC49244	ORF301	75.49	8.22E-13 1.83E-12 2.65E-11	31% 32% 34%	49% 52% 53%	14204 9159 4162	124-311 80240 74236
														10	NP_074951 YP_293795	putative endonuclease		4.99E-10	37%	55%	2109	3110
N395R	158437159858	474	53,895	6.11	1	pfam04451	Capsid Indovir, Iridovirus major capsid protein. This family includes the major capsid protein of indoviruses, chlorella virus and Spotopter accovirus, which are all dSDN viruses with no RNA stage. This is the most abundant structural protein and can account for up to 45% of virion protein. In Chiorella virus PRO-V1 the major capsid protein is a	182.43	8.07E-47	29%	45%	81470	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%	148405	1257
					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 refromer complex with at least four other proteins Vps5, Vps17, Vps26 and Vps29.	29.51	9.05E-01	21%	39%	86144	484546	2	BAE06835	hypothetical major capsid protein	169.09	3.29E-40	31%	49%	81446	3401
							Vps35 contains a central region of weaker sequence similarity, thought to indicate the presence of at least three domains							3	BAA76601	major capsid protein MCP1		4.31E-32	30%	45%	81438	3380
														4 5	AAC27492	major capsid protein Vp54 major capsid protein Vp49	139.81 139.43	2.14E-31 2.79E-31	29% 30%	42% 44%	81469 81438	3410 3371
														6	AAC96798	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	139.43	2.79E-31	30%	42%	81469	3410
														7	1M4X C	major capsid protein Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	115.93	3.65E-31 3.31E-24	29% 28%		81469 103469	3409 1386
														9	1M3Y_D	Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid Containing. Dna Virus	115.55	4.32E-24	28%		103469	1386
														10	YP_142795	capsid protein	67.40	1.35E-09	31%	46%	284430	310-465

Gene Name	Genome Position	A.A. length	Peptide Mw		CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identitv P	% Positive		Hit from-	BLASTp Hit Number		BLASTp Definition	Bit Score	E-value	% Identity P	% ositive	Query from-to	Hit from- to
							VIP2, VIP2; A family of actin-ADP-ribosylating toxin. A member of the Bacillus-prodiced vegetative insecticidal proteins (VIPs) possesses high	n														
N400L	161546159861	562	62,960	11.03	1	cd00233	because-produce regulative insections proteins (vr s) possesses ingli 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	) t - 22.20	8.72E-02	36%	55%	479540	128183	1	AAC96461	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563	57.77	1.34E-06	40%	54%	61150	385
							growth, where VIP1 likely targets insect cells and VIP2 incosylates adm. VIP2 shares significant sequence similarly with enzymatic components of other binary toxins, Closificium bolulium C2 toxin, C, perfingens iola toxin, C, prioferna toxin, C, prioforue toxin and C. difficie toxin. Tektin, Tektin family. Tektins are cytoskeletal proteins. They have been demonstrated in such cellular at less contriciols, basel booles, and along profoliments, organised as torgbudnal proteins thethin between with axial periodicily matching budnits. Tektin polypetides consist of the admit and the toxin. Tektin polypetides consist of the admit periodicily matching budnits. Tektin polypetides consist of	5 3 1 1 8														
					2		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	. 31.40 e f r	2.94E-01	33%	48%	218-300	27106	2	AAC96650	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055	73.94	1.80E-11	37%	46%	94238	396-556
							components							3	AAC96459	a91L similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number and a second second second second second second second		7.06E-24	48%		436560	1126
														4	AAC96646 AAC96984	similar to Chlorella virus PBCV-1 ORF A282L, corresponds to GenBank		8.11E-04 1.87E-08	31% 61%	43% 71%	115-238 58114	465597 255
														0		Accession Number U42580						
N403R	161664162374	237	27,164	4.53	1	COG1623	COG1623, Predicted nucleic-acid-binding protein (contains the HHH domain) IGeneral function prediction only Fungal_ODC_AZ, Fungal ormithine decarboxylase antizyme. This family consists of several fungal ornithine decarboxylase antizyme proteins. The	y e	4.27E-01	22%	48%	248	301347	1	YP_214596	hypothetical protein PSSM4_035	56.23	1.04E-06	27%	44%	6205	306511
					2	pfam06329	polyamine biosynthetic enzyme omithine 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 (421). This family is specific to funcal snecies but is related to the dam02100 family	29.24	4.63E-01	37%	50%	195–227	110140	2	CAJ74836	Conserved Hypothetical Protein	45.05	2.40E-03	27%	41%	33146	48187
N404R	162413163957	515	59,192	8.27	1	pfam01171	ATP_bind_3, PP-loop family. This family of proteins belongs to the PP- loop superfamily PP-ATPase, N-terminal domain of predicted ATPase of the PP-loop family implicated in cell cycle control [Cell division and chromosome		1.62E-29	28%	48%	210413	1204	1	AAC96914	similar to MesJ cell cycle protein	523.86	5.86E-147	52%	67%	15508	1497
					2		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	y 123.41	6.24E-29	29%	48%	210395	1185	2	T18059	hypothetical protein A557L - Chlorella virus PBCV-1	102.45	4.21E-20	52%	66%	40132	14108
					3	COG0037	apnaibeeraapna tool which binds to Adenosine group. I nis domain nas a stronolv conserved midf SGGXD at the Nerminus MesJ, Predicted ATPase of the PP-loop superfamily implicated in cell cvcle control (Call division and chromosome nathlinnino) Alpha_NH_like_II, This is a subfamily of Adenine nucleotide alpha	3	2.32E-23	24%	44%	195403	6218	3	YP_039960	hypothetical protein SAR0510	87.04	1.83E-15	23%	46%	211-509	14320
					4	cd01993	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	t 65.65	1.28E-11	23%	41%	210381	1180	4	AAV42170	) putative cell cycle	86.66	2.39E-15	24%	45%	210-472	23293
					5	pfam06041	stronolv conserved motif SGGKD at the N terminus DUF924, Bacterial protein of unknown function (DUF924). This family consists of several hypothetical bacterial proteins of unknown function	y 46.83	6.42E-06	24%	39%	16119	4117	5	YP_415957	hypothetical protein SAB0458	86.27	3.12E-15	23%	46%	211-509	14320
					6	cd01712	Thil, Thil is required for thiazole synthesis in the thiamine biosynthesis pathway. It belongs to the Adenosine Nucleotide Hydrolysis superfamily	/ 38.65	1.66E-03	23%	41%	215370	6153	6	BAB56671	conserved hypothetical protein	85.89	4.08E-15	23%	46%	211509	14320
					7	cd01713	and predicted to bind to Adenosine nucleoide. Phases reductions in the total of the phases produce the phases phases and the phases in the total of the adenite nucleoide subplotnarisense. PAPS reductase is part of the adenite nucleoide adpha hydrolases. A highly modified version of the P hoses and ATP subplotnylases. A highly modified version of the P hoses and ATP subplotnylases. A highly modified version of the P hoses and ATP subplotnylases. A highly modified version of the P hoses and ATP subplotnylases. A highly modified version of the P hose of the advase of the protein, which appears to be a positively charged cleft PAPS inductase than to ATTses activity, it shows a thirking similarity to the structure of the ATP prophosphatase (ATP PPase) domain of GMP synthesis, nickating that both enzyme families have evolved from a common ancestral nucleoide-binding total. The enzyme uses thioradoxin as an allection domor for the reduction of PAPS to phospho- P from Ritzobium mellioti which has ATP subplurylase activity (subplate sidentified the structure of the reduction of the reductin of the reduct	S 9 9 1 37.34 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9	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	14320
					8	COG0175	CysH, 3'-phosphoadenosine 5'-phosphosulfate sulfotransferase (PAPS reductase)/FAD synthetase and related enzymes	e s 35.41	1.85E-02	19%	35%	192378	23199	8	EAN10266	PP-loop	84.34	1.19E-14	24%	47%	206433	28260
					9	COG0301	metabolismi	04.07	2.22E-02	24%	45%	210370	177329	9	CAG42241	conserved hypothetical protein	83.96	1.55E-14	23%	46%	211-509	14320
					10	cd01990	Apha_ANH_like.], This is a subfamily of Adenine nucleotide alpha hydrolases subgrafmily. Adenine nucleotide alpha hydrolases subgrafmily includes N type ATP PPases and ATP subphurylases. It forms a aphabetalapha 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 multi SGGKD at the Netminus.	ly a 34.82 f	2.69E-02	27%	38%	211–384	1154	10	YP_428986	<sup>2</sup> tRNA(lie)-lysidine synthetase-like	81.26	1.00E-13	27%	51%	210408	20217
N410L	164225163956	90	9,651	10.15		No Hit Found								0	No Hit Found	No Hit Found						
N412R	164354164968	205	23,554	7.25	1	COG4942	COG4942, Membrane-bound metallopeptidase [Cell division and chromosome partitioninn]	d 29.16	4.07E-01	18%	42%	65115	4696	1		similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number 1142580 Hypothetical protein		1.25E-25 8.71E-03	34% 36%	53% 54%	8201 32110	7185 114201
N413R	165118165882	255	28,764	8.33		No Hit Found								1	AAC96377	7 A9R 3 similar to PBCV-1 ORF A79R, corresponds to GenBank Accession		7.66E-54	57%	78%	88253	8173
														2		similar to PBCV-1 ORF A/9R, corresponds to GenBank Accession Number U17055 similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank		7.73E-22	28%	46%	3251	5248
														3		Accession Number U42580 3 similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number		2.49E-20 9.45E-20	28% 27%	46% 46%	3252 3252	8253
														4	AAC96545	042580 5 A177R	99.75	9.45E-20	28%	48%	3240	5230
														6	AAC96447 AAU06304	hypothetical protein A275R	81.26	1.28E-16 3.48E-14 7.74E-14	28% 31%	46% 50%	3231 88251	2218 4167
														8 9 10	AAU06302	hypothetical protein A275R 2 hypothetical protein A275R 9 hypothetical protein A275R	68.17	7.74E-14 3.05E-10 6.15E-03	30% 32% 41%	50% 52% 52%	88251 120252 189251	4167 1132 159
							mRNA_triPase, mRNA capping enzyme, beta chain. The beta chain of	F						10	,000299	- mponoval protein Azron	43.80	5.102-00		5270	100-201	
N416L	166454165900	185	20,724	7.72	1	pfam02940	mRNA capping enzyme has triphosphatase activity. The function of the capping enzyme also depends on the guanylyltransferase activity	42.21	4.00E-05	34%	55%	94156	191256	1	AAC96817	7 PBCV-1 RNA triphosphatase	185.65	6.48E-46	55%	69%	5183	11193
							conferred by the alpha chain (see pfam01331).							2	XP_636333	hypothetical protein DDB0188303	42.74	6.79E-03	39%	56%	94156	155-219
N418R	166478167233	252	28,265	7.95	1	smart00702	P4Hc, Prohl 4-hydroxylase alpha subunit homologues. Mammalian enzymes catalyse hydroxylation of collagen, for example. Prokaryotic enzymes might catalyse hydroxylation of antibiotic peptides. These are 2- oxoglutratel dependent dioxygenases, requiring 2-oxoglutarate and rinorunen as cossibilitates and fermus iron as a cnfactor.	c - 113.63	2.26E-26	29%	44%	60248	1178	1	AAC96453	8 PBCV-1 prolyl 4-hydroxylase	224.94	1.90E-57	48%	63%	19250	3241

Gene Name	Genome Position	A.A. Ienath	Peptide Mw	pl	CDD Hit Number	COGs	COG Definition 20G-Fell Oxy. 20G-Fell) oxygenase superfamily. This family contains	Bit Score	E-value	% Identitv	% Positive	Query from-to	Hit from- I to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% dentitv P		Query from-to	Hit from- to
					2	pfam03171	members of the 2-acaptitate (2OG) and Fell)-dependent organisas superfamity. This family includes the C-terminal of group 4-bydrochysics alpha subunit. The holenczyme has the admity EC1.14.112 catalysing the reaction: Procellagen L-proline + 2-acaptitatent + O2 & H-zdq procolagen trans- 4+ydroxy-L-proline + succinate + O2 athready compare consists of a alpha2 bea2 complex with the alpha subunit contributing most of the parts of the active site. The family also includes how thortfanses: incomercillin servitances and ABR	41.64	1.00E-04	27%	39%	145–249	2-96	2		<sup>1</sup> Procollagen-proline,2-oxoglutarate-4-dioxygenase		6.61E-18	33%	50%	60248	90274
															EAP73718 YP_585239 YP_550510 ZP_00984285 ZP_01271629 YP_142947	putative prohf 4-hydroxylase alpha subunit homologue oxidoreductase ordein 9 Prohf 4-hydroxylase alpha subunit 2 OG-FeIIII 0 oxygenase 9 Procollagen-proline 2-oxoglutarite 4-dioxygenase 1 Procollagen-proline Bol/A 01003928 2 OG-FeIII 0 oxygenase 2 OG-FeIII 0 oxygenase 3 Unknown protein	90.12 87.81 85.89 84.73 84.34 81.65	4.28E-17 7.31E-17 3.63E-16 1.38E-15 3.07E-15 4.01E-15 2.60E-14 2.20E-13	32% 32% 31% 30% 31% 28% 32%	47% 46% 46% 46% 52% 44% 43%	60-248 60-248 58-248 60-248 60-248 60-248 56-248 60-247	97281 97281 99285 90272 73257 79263 56237 45223
N420R	167270167593	108	11,872	10.26	6 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 secarated from the noise.	61.78	1.78E-11	32%	49%	29105	33108	1	AAC96816	contains cytochrome C family heme-binding site signature; similar to <sup>3</sup> maize protein disulphide isomerase, correspond to Swiss-Prot Accession Number P52588	127.87	1.02E-28	55%	73%	4103	6105
					2	COG3118	3 COG3118, Thioredoxin domain-containing protein [Posttranslational modification. protein turnover. chaperones]. TrxA, Thiol-disulfide isomerase and thioredoxins [Posttranslational	00.00	6.93E-05	20%	39%	13103	39129	2	AAQ23042	2 transglutaminase	67.01	2.14E-10	35%	58%	10104	6105
					3	COG0526	<ul> <li>modification, protein turnover, chaperones / Energy production and conversionl.</li> </ul>	33.34	6.49E-03	14%	33%	27106	42126	3		2 hypothetical protein		2.14E-10	46%	62%	29102	60134
					4	pfam01216	Calsequestrin, Calsequestrin	28.43	2.02E-01	30%	59%	5499	95139	4	XP_419952	2 PREDICTED: similar to Protein disulfide isomerase A6 precursor (Protein disulfide isomerase P5) (Thioredoxin domain containing protein 7)	45.44	6.68E-04	32%	48%	29100	196-271
														5 6 7 8 9 10	CAG03659 ABC75564 AAH67155 AAC24752 AAH44507	7 protein disulfide isomerase ER-60 9 unnamed protein product 1 protein disulfide isomerase-related protein P5 precursor 1 Hypothecial protein zgc.77086 2 uransplutaminae precursor 7 Protein disulfide isomerase-related protein (provisional)	50.45 48.52 63.93 63.54	3.66E-10 2.08E-05 7.89E-05 1.82E-09 2.37E-09 2.71E-05	39% 35% 31% 36% 35% 35%	59% 49% 48% 56% 56% 47%	23101 2397 23100 23104 10104 23100	43119 185261 167246 45124 35134 186265
N421R	167673169067	465	52,190	5.95	5 1	COG0661	AarF, Predicted unusual protein kinase [General function prediction onlv] ABC1, ABC1 family. This family includes ABC1 from yeast and AarF from E. coli. These oroteins have a nuclear or mitochondrial subcellular	176.29	5.77E-45	27%	42%	24385	37423	1	AAC96813	3 similar to Clostridium pasteurianum ORF, corresponds to GenBank Accession Number 728353	615.92	9.91E-175	64%	84%	13463	11-462
					2	pfam03109	location in eukaryotes. The exact molecular functions of these proteins is not clear, however yeast ABC suppresses a cytochrome b mRNA translation defect and is essential for the electron transfer in the bc1 complex and E. coli Aart is required for ubiquinone production. It has been suggested that members of the ABC transporter proteins. These proteins are unrelated to the ABC transporter proteins.	113.77	4.19E-26	41%	57%	96209	7–119	2	YP_477188	3 ABC1 domain protein	139.81	2.08E-31	26%	48%	49422	68475
					3	COG4670	COG4670, Acyl CoA:acetate/3-ketoacid CoA transferase [Lipid metabolism].	30.19	5.76E-01	31%	47%	250-366	368476	3 4 5 6 7 8 9 10	YP_399885 BAB66733 ZP 00673531	7 universally conserved protein 3 ABC-1 4 alr4515	134.04 132.88 132.11 131.34 130.57	6.69E-30 1.14E-29 2.54E-29 4.34E-29 7.40E-29 1.26E-28 2.15E-28 2.81E-28	26% 24% 26% 25% 25% 25%	48% 47% 46% 43% 46% 46% 46%	49392 49392 12457 49384 12460 37392 37392 54394	79453 83452 9459 70431 9467 71452 82463 102470
N426R	169131169475	5 115	13,217	5.50	0 1	pfam03179	V-ATBase_G, Vacubar (H+)-ATBase G suburit. This family propendity the eukarydic vacubar (H+)-ATBase G suburit. V-ATBase) generate an actic environment in several intracellular compartments. Correspondingly, they are found as membrane-attached proteins is several organelles. They are also found in the plasma membranes of some specialized cells. V-ATBase consist of periodenal (VI) and some specialized cells. V-ATBase consist of periodenal (VI) and part of the VI suburit, but is also thought to be strongly attached to the V0 complex. It may be involved in the coupling of ATB egradation to H+	28.73	1.72E-01	21%	49%	2798	2697	1	AAC96812	2 A444L	71.25	1.11E-11	38%	55%	4112	299
					2	COG1322	translocation COG1322, Uncharacterized protein conserved in bacteria [Function unknown]	28.42	1.89E-01	14%	46%	22100	107185	2	AAQ96243	3 LRRGT00030	43.90	1.89E-03	30%	53%	13107	144-234
					3	COG0484	DnaJ, DnaJ-class molecular chaperone with C-terminal Zn finger domain IPosttranslational modification. protein turnover. chaperones1.	28.27	2.43E-01	38%	65%	2963	1751	3	XP_215260	PREDICTED: similar to LRRGT00030	43.90	1.89E-03	30%	53%	13107	149-239
N428R	169736171154	473	51,078	10.86	6	No Hit Found								0	No Hit Found	d No Hit Found						
N430L	171587171159	143	15,847	4.4	1	No Hit Found								1	AAC96809	9 A441L	140.97	1.16E-32	51%	67%	6141	2137
N433R	171678172577	300	33,203	5.57	7	No Hit Found								1	AAC96811	1 A443R	182.57	1.47E-44	34%	55%	6297	8306
N434L	172956172621	112	12,602	6.50	0	No Hit Found								1	AAC96807			7.23E-27	58%	81%	19103	21105
N436R	172977173240	88	10,083	11.02	2 1		<ul> <li>GrxC, Glutaredoxin and related proteins [Posttranslational modification, protein turnover. chaperones].</li> <li>Glutaredoxin, Glutaredoxin</li> </ul>	52.66 39.85	1.13E-08 7.56E-05	34% 25%	55% 49%	372 172	572 273	1		3 similar to E. coli glutaredoxin, corresponds to Swiss-Prot Accession Number P37687 Okumber P37687		7.35E-19 1.44E-06	58% 31%	84% 57%	372 170	574 978
					3	COG0278		27.10	4.98E-01	25%	49%	1074	2=73	2		Glutaredoxin GrxC     probable peroxiredoxin/glutaredoxin family protein	50.83	1.44E-06	38%	56%	174	978 173-241
														9	AAN30771 YP_465226 YP_154197 ZP_00374087 AAZ41226	5 GLUTAREDOXIN 1 glutaredoxin 3 Glutaredoxin 3 Glutaredoxin GAC 1 glutaredoxin family protein 3 glutaredoxin family protein 3 glutaredoxin 5 COC06955: Glutaredoxin and related proteins	50.83 49.29 45.44 45.44 44.67	1.59E-05 1.59E-05 4.63E-05 6.68E-04 6.68E-04 1.14E-03 1.14E-03	33% 33% 33% 33% 32% 28%	55% 59% 51% 50% 55% 54%	1-78 1-78 3-71 1-74 1-78 3-70 1-70	878 474 872 473 480 673 469
N438R	173248173814	189	21,593	5.53	3 1	cd01673	dNK, Deoxyribonucleoside kinase (dNK) catalyzes the phosphorylation of deoxyribonucleosides to yield corresponding monophosphate (dNMPs). This family consists of various deoxynucleoside kinases including deoxyribo-cyldine (EC 27.174), guanosime (EC 27.113), adenosine (EC 27.176), and trymdine (EC 27.112) kinases. They are key enzymes in the salvage of deoxyribonucleosides originating from	102.24	3.21E-23	31%	52%	3165	1178	1	AAC96784	contains ATP/GTP-binding site motif A: similar to Bacilius subitus 24.1 kDa protein, corresponds to Swiss-Prot Accession Number P37530	191.05	1.64E-47	50%	68%	3181	2187
					2	COG1428	extra-or intracellular breaktnwn of DNA COG1428, Deoxynucleoside kinases [Nucleotide transport and metabolism]. dNK, Deoxynucleoside kinase. This family consists of various		6.18E-20	34%	54%	1160	4178	2	ABF82059	9 hypothetical protein MIV029R	91.28	1.77E-17	31%	56%	3183	4189
					3	pfam01712	deoxynucleoside kinases cytidine EC:2.7.1.74, guanosine EC:2.7.1.113, 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	80.31	1.44E-16	31%	49%	53183	1141	3	NP_149606	<sup>3</sup> 143R	88.20	1.50E-16	31%	52%	1152	1157
					4	COG0125	from a deoxynucleoside. Using ATP and yielding ADP in the process 5 Tmk, Thymidylate kinase [Nucleotide transport and metabolism].	53.35	1.90E-08	23%	45%	4184	6204	4	AAU10875	5 deoxynucleoside kinases	76.64	4.52E-13	31%	55%	3152	4161
					5	cd01672	TMPK, Thymidine monophosphate kinase (TMPK), also known as thymidytale kinase, catalyzes the phosphorylation of thymidine monophosphate (TMP) to thymidine diphosphate (TDP) utilizing ATP as its preferred photphoryl donor. TMPK represents the rate-limiting step in either de novo or salvage biosynthesis of thymidine triphosphate (TTP).		6.22E-08	19%	42%	3181	2–198	5	CAC84481	1 thymidine kinase	75.49	1.01E-12	26%	46%	4182	13208

Gene	Genome	A.A.	Peptide	pl	CDD Hit	COGs	COG Definition	Bit Score	Evoluo	%	%	Query	Hit from- BLA	STp Hit	t Hit	BLASTp Definition	Bit Score	Evolue	%	%	Query	Hit from-
Name	Position	lenath	Мw	рі	Number	0005	NDUO42, NADH:Ubiquinone oxioreductase, 42 kDa (NDUO42) is a family of proteins that are highly similar to deoxyribonucleoside kinases	Bit Score	E-value	Identitv P	ositive	from-to	to Nu	Imber	Accession	BLASTP Definition	Bit Score	E-value	dentitv P	ositive	from-to	to
					6	cd02030	(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	43.31	2.00E-05	27%	45%	3153	1184	6	YP_040010 putative deoxyadenosi	ine kinase protein	72.79	6.53E-12	31%	47%	3181	11-211
					7	cd02019	matrix to the inter membrane scace. NRC, Nucleosidehuciedolde kinase (MX) is a protein superfamily and are functionally related to the catalysis of the reversitie 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 methodism, the biosynthesis	42.11	3.79E-05	21%	38%	3159	1165	7	YP_416005 deoxypurine kinase su	bunit	72.79	6.53E-12	31%	48%	3181	11211
					8	pfam02223	of coenzymes and aromatic compounds, as well as the metabolism of sunar and sulfate Thymidylate kin, Thymidylate kinase	40.65	1.12E-04	19%	37%	6177	1186	8	YP 499111 hypothetical protein SA	AOUHSC 00539		1.90E-11	34%	51%	3153	11172
					9 10		Udk, Uridine kinase [Nucleotide transport and metabolism]. 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 steo in CoA biosynthesis	35.22 35.14	5.22E-03 6.14E-03	22% 25%	44% 40%	3159 3134	10179 1142	9 10	CAC84464 thymidine kinase AAW53624 deoxynucleoside kinas	se family protein		3.24E-11 3.24E-11	26% 30%	51% 48%	4174 3181	9196 11211
N441L	174871173786	362	40,197	8.53	1 2		COG1910, Periplasmic molybdate-binding protein/domain [Inorganic ion transport and metabolism].	35.64 34.61	9.41E-03 1.78E-02	42% 33%	58% 53%	5083 2577	97130 1572	1				3.97E-53 7.60E-04	36% 26%	56% 42%	1348 115346	1372 210468
					2		Rad17, Rad17 cell cycle checkpoint protein SCF, Stem cell factor. Stem cell factor (SCF) is a homodimer involved in	34.01	1.70E-02	3376	53%	2311	15-72	2	YP_142765 unknown		47.75	7.00E-04	20%	42.70	115340	210-408
N445L	175247174900	) 116	13,136	10.69	) 1		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%	568	66129	1	AAC96788 A420L		69.71	3.32E-11	48%	65%	45114	170
N446R	175027175434	136	15,397	10.41	1	COG1177	PotC, ABC-type spermidine/putrescine transport system, permease component II fAmino acid transport and metabolisml.	27.44	5.85E-01	31%	53%	368	195265	1	AAC96789 A421R		96.67	2.51E-19	52%	71%	47128	891
N447L	175712175458	85	9,585	7.92	2	No Hit Found								0	No Hit Found No Hit Found							
N449R	175739176215		17,816	4.79			AcnA, Aconitase A [Energy production and conversion].	28.74	3.61E-01	33%	54%	99147	519567	1	AAC96791 A423R		76.64	2.68E-13	34%	62%	39154	36155
N451R	176270176593		12,209	11.33		No Hit Found								0	No Hit Found No Hit Found							
N452R	176630176983	118	13,306	4.82	2	No Hit Found	Thioredoxin, Thioredoxin. Thioredoxins are small enzymes that							1	AAC96794 A426R		97.06	1.93E-19	43%	66%	6116	4114
N453L	177331176978	118	13,316	5.26	5 1	pfam00085	participate in redox reactions, via the reversible oxidation of an active centre disulfide bond. Some members with only the active site are not seoarated from the noise	43.67	5.19E-06	24%	48%	25105	25107	1		ctive site-like sequence; similar to Synechocystis n, corresponds to Swiss-Prot Accession Number		2.06E-21	43%	66%	2115	7119
					2	COG3118	COG3118, Thioredoxin domain-containing protein [Posttranslational modification protein turnover chanerones] TrxA, Thiol-disulfide isomerase and thioredoxins [Posttranslational	32.19	1.52E-02	26%	46%	33101	55129	2	NP_072786 thioredoxin (trxA)		56.23	3.77E-07	30%	53%	4103	2102
					3		modification, protein turnover, chaperones / Energy production and conversion).		2.28E-02	26%	49%	25105	36124		NP_857580 THIOREDOXIN TRXC			2.45E-06	31%	54%	17105	21112
					4	COG0259	PdxH, Pyridoxamine-phosphate oxidase [Coerzyme metabolism]. Peptidase_C19E, A subfamily of Peptidase C19. Peptidase C19 contains ubiquitryh 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	28.30	2.43E-01	56%	75%	8096	5672	4	ZP 01043911 Thioredoxin domain-cc	ntaining protein	51.99	7.12E-06	31%	50%	1101	6107
					5	cd02661	the C-terminal Gy residue of ubiquitin. The purpose of the de- ubiquinitation is thought to be defining of the ubiquitin conjugates, which could rescue them from degradation, as well as recycling of the ubiquitin. The ubiquitiproteasme 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%	2973	134177	5	YP_521223 hypothetical protein DS YP_521223 hypothetical protein DS	SY4990	51.60	9.29E-06	28%	52%	8107	9108
														7	ZP_01192583 Thioredoxin AAT27628 thioredoxin ZP_01204863 Thioredoxin AAG21048 thioredoxin; TrxA1 2 BAE56042 unnamed protein prod	uct	51.22	1.21E-05 1.59E-05	32% 23% 31% 30% 29%	53% 56% 54% 51% 51%	17105 25104 17105 11101 290	18109 21100 19110 23114 192
N457L	177735177394	114	12,710	5.76	i 1	COG3118	modification, protein turnover, chaperones].	40.66	4.01E-05	38%	56%	5589	81115	1	contains thioredoxin a AAC96795 thioredoxin-like protein P52232	ctive site-like sequence; similar to Synechocystis n, corresponds to Swiss-Prot Accession Number	52.76	4.09E-06	28%	52%	1105	7110
					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 secarated from the noise.	39.05	1.33E-04	31%	53%	40100	44103	2	ZP_00673542 Thioredoxin		42.74	4.24E-03	26%	61%	40106	41107
					3	COG3640	CooC, CO dehydrogenase maturation factor [Cell division and chromosome partitioning]	28.64	1.56E-01	23%	47%	3076	1356		ABE04083 Thioredoxin ZP 01228494 thioredoxin			9.43E-03 9.43E-03	42% 34%	64% 61%	4081 49100	5596 50101
															ZP_00857352 Thioredoxin			9.43E-03	39%	60%	4797	87137
N458L	178083177766	5 106	11,770	7.34	i -	No Hit Found								1	contains thioredoxin a AAC96795 thioredoxin-like protein P52232	ctive site-like sequence; similar to Synechocystis n, corresponds to Swiss-Prot Accession Number	53.14	3.12E-06	32%	53%	295	9106
N460L	179433178102	2 444	51,180	5.46	5 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 antiparalle helices and a beta-haripin,	43.53	5.90E-05	27%	55%	26121	395	1	AAC96797 A429L		210.69	9.06E-53	33%	52%	7437	34466
							repeats are stacked in a superhelical arrangement; this alignment contains 4 consecutive repeats							2				8.31E-06	25%	43%	51237	61250
														3	ABF95306 ankyrin repeat family p XP 780211 PREDICTED: similar f	orotein, putative, expressed to Ankyrin-2 (Brain ankyrin) (Ankyrin B) (Ankyrin	51.22 51.22	9.19E-05 9.19E-05	23% 23%	43% 41%	49237 11271	205-396 337-595
														5	BAB03143 ankyrin-like protein		50.83	1.20E-04 2.67E-04	22% 22%	43% 43%	49253 49237	599-805 114-305
														7	NP_187842 protein binding ABE93792 Ankyrin AAM62711 ankyrin-like protein			4.56E-04	22% 24% 21%	43% 43% 43%	49237 30237 49237	27242 58249
														9		hate pyrophosphohydrolases/synthetases homolog		8.60E-03	21%	43%		730-914
														10	NP_192256 protein binding		46.98	1.73E-03	22%	41%	27281	198-454
N462R	179472179855		14,974	9.10		No Hit Found								0	No Hit Found No Hit Found							
N464R	179846180130	) 95	10,649	10.71	l	No Hit Found								1 2	AAC96938 A601R AAC96975 A656L			1.60E-18 2.24E-04	53% 40%	64% 64%	1191 4589	12101 51100
N465R	180130181275	382	42,891	10.09	)	No Hit Found								1	AAC96599 contains ATP/GTP-bin similar to Chlorella viru Accession Number U4	us PBCV-1 ORF A231L, corresponds to GenBank	489.57 305.45	8.15E-137 2.18E-81	61% 44%	76% 62%	6380 14340	3383 104428
N469R	181355181765	i 137	15,763	6.80	)	No Hit Found								0	No Hit Found No Hit Found							
N470L	183182181881	434	47,713	6.94	i 1	pfam04451	Capsid_indovir, Indovirus major capsid protein. This family includes the major capsid protein of indoviruses, 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 divconnotain.	428.58	5.73E-121	53%	64%	1428	1443	1	AAC27492 major capsid protein V	<sup>1</sup> P49	738.03	0.00E+00	86%	87%	1432	1432

Gene Name	Genome Position	A.A. length	Peptide Mw	pl	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity P	%	Query from-to	Hit from-	BLASTp Hit Number		Score E-value	% Identity	% Positive	Query from-to	Hit from- to
Name	FUSILION	ienuur	WW		Number						OSILIVE	110111-10	10	2	DDC)/ 1 major consid protois \/sE4_corresponds to ConBonk Association	618.23 1.82E-175	72%	78%	1432	1437
														3	BAA76601 major capsid protein MCP1	613.22 5.86E-174	72%	78%	1432	1437
														4	BAA22198 major capsid protein Vp54 BAA76600 major capsid protein	611.30 2.23E-173 606.29 7.17E-172	72% 71%	77% 77%	1432 1432	1437 1436
														6	<ul> <li>1M4X_C Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model</li> <li>1M3Y_D Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid</li> </ul>	571.24 2.56E-161 571.24 2.56E-161	71% 71%	77% 77%	25432 25432	1413 1413
														7	Containing. Dha virus	571.24 2.56E-161 305.83 2.00E-81	71% 40%	77% 53%	25432 1432	1413 1440
														9	AAC06017 similar to Chlorella virus PBCV-1 ORF A11L, corresponds to GenBank	236.11 1.95E-60	35%	50%	3432	2400
														10		233.80 9.67E-60	36%	50%	1432	1403
							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													
N471L	184168183284	295	35.080	8.32	2 1	nfam0154	also found in putative endonucleases encoded by group I introns of fungi 1 and phage. The structure of I-TevI a GIY-YIG endonuclease, reveals a		1.03E-02	20%	36%	28112	288	1	AAC96862 similar to PBCV-1 ORF A315L, corresponds to GenBank Accession	72.02 2.72E-11	30%	49%	30185	4166
	104100 100204	200	00,000	0.01	- '	planoro4	novel alpha/beta-fold with a central three-stranded antiparallel beta-sheet	00.14	1.002.02	20,0	0070	20 112	2 00		Number M74440	12.02 2.722 11	0070	4070	00 100	4 100
							flanked by three helices. The most conserved and putative catalytic residues are located on a shallow, concave surface and include a metal													
					2	smart0046	coordination site 5 GIYc, GIY-YIG type nucleases (URI domain); .	33.89	2.28E-02	34%	54%	30116	483	2	AAA88832 unknown	64.31 5.67E-09	29%	45%	30212	4180
							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													
					3	cd0327	<ul> <li>essential in several proteins. The carboxy-terminal domain contains a sequence (the DA-box) that resembles a 'Walker B' motif,</li> </ul>	32.27	8.15E-02	18%	38%	123293	8851060	3	AAC96973 similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	57.38 6.93E-07	39%	51%	30132	4105
							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) eIF3_gamma, Eukaryotic initiation factor 3, gamma subunit. eIF-3 is a													
					4	pfam0418	o multi-subunit complex that stimulates translation initiation in vitro at	21.52	1.16E-01	23%	47%	150242	75165	4	AAC96655 PBCV-1 33kd peptide	56.61 1.18E-06	28%	45%	40199	19192
							elF3													
							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													
					5	cd0327	7 esemble in several proteins. The carboxy ethnicial 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	31.17	1.70E-01	22%	48%	130240	194308	5	YP_293795 putative endonuclease	56.61 1.18E-06	39%	55%	38118	1495
							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)							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 CAG08671 unnamed protein product	44.28 6.07E-03 44.67 4.65E-03	25% 33%	45% 52%	123292 123238	9721136 338441
														9	AAC26930 Hypothetical protein F36H12.3	51.22 4.97E-05	29%	51%	123248	95217
														10	XP 654034 hypothetical protein 64.t00001	50.06 1.11E-04	25%	54%	123233	151-261
N472R	184277186826	850	95,405	6.12	2 1	pfam0053	Glycos_transf_2, Glycosyl transferase. Diverse family, transferring sugar 5 from UDP-glucose, UDP-N-acetyl- galactosamine, GDP-mannose or	49.83	1.32E-06	30%	43%	256383	2124	1	AAC96482 A114R	639.80 0.00E+00	64%	77%	373847	5477
							CDP-abequose, to a range of substrates including cellulose, dolichol phosphate and teichoic acids													
					2	COG046	WcaA, Glycosyltransferases involved in cell wall biogenesis [Cell envelope biogenesis. outer membrane].		7.61E-06	24%	45%	251-380	2121	2		489.96 1.72E-136	58%	75%	1372	1379
					3	COG121	Diodenesis ICell envelope biodenesis. Outer memoranel.	44.10	7.72E-05	24%	38%	251-367	53167	3	NP_872956 possible glycosyltransferase	90.51 3.05E-16	32%	46%	254426	6189
					4	COG121		38.19	4.16E-03	25%	46%	252-369	3116	4	CAG34747 hypothetical protein	87.81 1.97E-15	29%	45%	1225	1225
							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													
					5	pfam0301	EXT genes suggest association with other forms of neoplasia. EXT1 and	35.03	3.45E-02	23%	43%	657-818	177325	5	5 ZP_00202013 COG0463: Glycosyltransferases involved in cell wall biogenesis	86.27 5.74E-15	32%	49%	254426	48231
							EXT2 have both been shown to encode a heparan sulphate polymerase with both D-glucuronyl (GlcA) and N-acetyl-D-glucosaminoglycan													
							(GIcNAC) transferase activities. The nature of the defect in heparan subhate biosynthesis in HMF is unclear													
														6	AU37201 unknown ZP_00154867 COG1216: Predicted glycosyltransferases	81.26 1.85E-13 78.95 9.17E-13	26% 29%	40% 44%	254-495 254-427	6259 29215
														8	AAU03775 conserved hypothetical protein ZP_00340129 hypothetical protein RakaH01000503	59.31 7.52E-07 58.54 1.28E-06	25% 28%	50% 51%	58206 58206	103-250 107-254
														10		57.77 2.19E-06	23%	43%	3201	2204
N477R	186874187335	154	16,607	8.18	В	No Hit Found	i							1	AAC96800 Lys-rich	114.01 1.50E-24	46%	67%	1109	1109
N480R	187361187714	118	12,745	11.93	3	No Hit Found	1							1	AAC96925 A571R	156.76 2.05E-37	79%	89%	1101	1101
														2	AAC96803 Arg-rich AAC96573 A205R	56.23 3.77E-07 45.05 8.70E-04	61% 57%	70% 75%	144 739	447 174206
															Lys-, Arg-rich; contains eukaryotic putative RNA-binding region RNP-1					
N482R	187960188919	320	36,440	8.67	7 1	smart0050	7 HNHc, HNH nucleases; .	29.67	5.70E-01	36%	45%	127168	1052	1	AAC96845 signature; similar to PBCV-1 ORF A267L, corresponds to GenBank Accession Number U42580	188.73 2.27E-46	35%	47%	9312	3305
					2	cd0008	HNHc, HNH nucleases; HNH endonuclease signature which is found in 5 viral, prokaryotic, and eukaryotic proteins. The alignment includes	28.89	8.81E-01	30%	43%	127-172	1157	2	AAC96857 Lys-, Glu-rich	177.56 5.24E-43	34%	47%	9312	3305
					2	00000	members of the large group of homing endonucleases, yeast intron 1 protein. MutS. as well as bacterial colicins. pvocins. and anaredoxins	20.05	0.012-01	30%	4370	127-172	11-57	2	Acoust Eys., Glu-Inti	111.30 3.246-43	3470	47.75	8-512	3303
					3	COG111	0 COG1110, Reverse gyrase [DNA replication, recombination, and repair].	28.74	9.54E-01	30%	52%	427	692715		YP_142777 unknown	73.17 1.39E-11	38%	56%		251-349
														4	AAC96635 A267L	85.11 3.54E-15	34%	47%	114293	73238
N484L	189482189174	103	10,874	10.74	4 1	pfam0585	MC1, Non-histone chromosomal protein MC1. This family consists of 4 archaeal chromosomal protein MC1 sequences which protect DNA	42.35	1.23E-05	37%	53%	568	2-70	1	AAC96805 similar to Methanothrix chromosomal protein MC1A, corresponds to	134.81 8.23E-31	69%	81%	599	6100
							against thermal denaturation						-	2	Swiss-Prot Accession Number P15251 P15249 Chromosomal protein MC1a	54.30 1.41E-06	42%	50%	790	484
														3	P15250 Chromosomal protein MC1b	50.83 1.56E-05	45%	60%	766	462
														4	ZP 01153404 Non-histone chromosomal MC1 5 P15251 Chromosomal protein MC1c	49.68 3.48E-05 48.14 1.01E-04	42% 45%	57% 58%	466 667	667 462
														6 7	<ul> <li>P12770 Chromosomal protein MC1</li> <li>A25343 nonhistone chromosomal protein HMb - Methanosarcina barkeri</li> </ul>	45.82 5.03E-04 43.51 2.50E-03	35% 40%	50% 54%	493 467	190 169
														8	AAZ70150 chromosomal protein	43.13 3.26E-03	40%	54%	467	270
N485R	189557189775	73	7,645	10.67	7	No Hit Found	1							0	No Hit Found No Hit Found					

Gene Name	Genome Position	A.A. lenath	Peptide Mw		CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identitv I	% Positive	Query from-to	Hit from- BLASTp to Numbe		Hit BLASTp Definition	Bit Score	E-value	% dentitv P	% ositive	Query I from-to	Hit from- to
N487L	190133189810	108	12,183	4.88	1	COG3164	COG3164, Predicted membrane protein [Function unknown].	27.59	3.50E-01	33%	54%	327	11351159	1 2 3	AAC96778 A410L AAC96776 A408L NP_077527 EsV-1-42	78.18 50.83	1.59E-05	48% 33% 30%	74% 61% 50%	4105 3102 3101	8110 40139 6110
N488R	190231190728	166	18.734	5.30		No Hit Found								4	AAR26867 Firv-1-A43 AAC96779 A411R		3.54E-05 4.76E-14	33% 37%	51% 56%	3103 1131	5110 1133
N491R	190757191299	181	20.356	6.54	1	COG5221	DOP1, Dopey and related predicted leucine zipper transcription factors	28.45	5.72E-01	22%	39%	65174	575682	1	AAC96780 A412R		2.08E-46	51%	67%	1178	1179
N493L							l I ranscriptioni.								AAC96781 Gin-rich, QQQQM(4x); similar to human transcription factor TFIID,			37%	50%	2245	
N493L	192037191294	248	26,968	8.99		No Hit Found								1	AAC96781 Corresponds to Swiss-Prot Accession Number P20226	146.75	6.36E-34	37%	50%	2245	3244
N494R	192130192828	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 tocosyl residue. This tucosyltransferase adds this residue	28.81	6.83E-01	21%	44%	12144	250393	1	AAC96776 A408L	209.53	7.09E-53	51%	75%	1187	37233
														2 3 4	NP_077527 EsV-1-42 AAC96778 A410L AAR26867 FirrV-1-A43	62.39 58.54 56.23	2.02E-07	26% 29% 24%	49% 53% 51%	5127 497 4130	7142 7108 5143
N496L	193660192842	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	72157	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 dentatorukra-prolifolovjasian atrophy (IDRPLA) gene. DRPLA CMIM: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 interactions with other proteins. The expansion of a CAG repeat is also the undenlying detect in six other neurodegenerative disorders, including to consider in six other neurodegenerative disorders, including the undenlying detect in six other neurodegenerative disorders.		3.22E-01	33%	42%	172-239	503570	2 2	2P_01147863 Initiation factor 2:Small GTP-binding protein domain	80.88	5.13E-14	37%	44%	123252	198–336
							HuntingtonSapots disease. One interaction of expanded polyuluarine repeats that is thought to be pathogenic is that with the soluti guitamine repeat in the transcriptional coactivator CREB binding protein. CBP. This interaction draws CBP away from its usual nuclear location to the expanded polyularinine repeat protein aggregates that are characteristic of the polyularine neurodegenerative disorders. This interferes with CBP-mediated transcription and causes cylotoxidy:														
														3 4 5 6 7 8 9 10	AB445731 collagen type III alpha 1 AA59633 proceediagen D PV4258 Collagen alpha-1(III) chain XP_98840 PREDICTED: similar to Collagen alpha 1(III) chain precursor isoform 1 XP_983522 PREDICTED: similar to Collagen alpha 1(III) chain precursor isoform 7 XP 88438 PREDICTED: similar to Collagen alpha 1(III) chain precursor isoform 7 XP 88438 PREDICTED: similar to Collagen alpha 1(III) chain precursor isoform 7 XP 88438 PREDICTED: similar to Collagen alpha 1(III) chain precursor isoform 7	55.45 58.15 48.91 56.23 48.91 47.37	1.35E-06	38% 36% 32% 31% 32% 34% 32%	39% 42% 38% 42% 36% 42% 38% 42%	123-185 129-258 153-257 124-213 131-257 124-213 153-257 124-213	94231 11331218 1111
N500R	193784194935	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%	116253	51167	0	No Hit Found No Hit Found						
N504R	194966195598	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%	1085	299377		AAC96775 A407L YP 142822 unknown		5.37E-83 3.80E-04	70% 23%	82% 42%	1209 12208	1209 41225
N506L	196264195644	207	23,029	7.81	1	smart00731	SprT, SprT homologues. Predicted to have roles in transcription elongation. Contains a conserved HExxH motif, indicating a	35.66	4.74E-03	28%	40%	122-201	56146	1	AAC96888 contains GIn-rich, neutral zinc metallopeptidase, zinc binding region signature	216.85	3.46E-55	62%	79%	1167	1165
					2	COG0501	metallionratease function HtpX, Zn-dependent protease with chaperone function [Posttranslational modification. protein turnover. chaperones].	33.36	2.26E-02	38%	53%	100-138	130170	2	YP_142679 metal-dependent hydrolase	56.23	7.81E-07	27%	46%	4181	3191
					3	COG0503	Apt, Adenine/guanine phosphoribosyltransferases and related PRPP- binding proteins INucleotide transport and metabolismi. 9 Peptidase M48, Peptidase family M48	31.42 30.83	8.08E-02 1.39E-01	28% 40%	43% 53%	52108 122166	54108 5191	3 4	NP_149599 136R ABF82072 hypothetical protein MIV042R	48.52 45.44		34% 33%	59% 60%	100162 100163	98163 81145
N507R	196327196899	191	20,846	9.33		No Hit Found	replace with, replace any with	00.00	1.002 01	40,0	0070	122 100	01 01	1	AAC96890 A523R		2.47E-59	70%	80%	30186	7168
N512R	196920197366	149	16,359	8.39		No Hit Found								1	AAC96893 A526R	129.03	4.60E-29	50%	62%	18147	1129
N513R	197365197649	95	10,969	10.78		No Hit Found								1	AAC96894 A527R	76.64	2.64E-13	60%	74%	669	569
N514L	197904197689	72	8,035	10.35		No Hit Found								1	AAC96898 A531L	48.91	5.92E-05	54%	80%	2262	2465
N515L	198165197926	80	8,518	9.37		No Hit Found								1	AAC96899 A532L	96.67	2.50E-19	60%	77%	172	174
N516R	198415200001	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%	94124	5686	1	AAC96900 A533R		1.28E-120	58%	73%	1362	1365
														2	AAA66400 unknown protein AAC96901 a534R	154.45	6.33E-120 9.68E-36 5.18E-05	58% 77%	73% 88%	1362 438526	1365 16104
														4 5	NP_077576 EsV-1-91 AAX13509 EtpA	47.37	1.67E-03	27% 24%	45% 40%		8158 11751447
N518L	200210199992	73	8,156	4.08		No Hit Found								6	XP 654515 hypothetical protein 51.t00023 AAC96902 A535L	46.60	2.84E-03 3.23E-11	25% 50%	35% 68%	112-299 1-71	39227 171
N510L	200210199992 200500200258	81	9,159	4.08		No Hit Found								1	AAC96903 A536L	60.85		49%	66%	1779	1773
N52UL	200500200258	01	9,159	10.91		NO HIL FOUND	Bact_CD, Bacterial cytosine deaminase and related metal-dependent								WC20202 4230F	60.65	1.32E-06	4976	00%	1779	17-73
N521L	201271200522	250	28,688	5.31	1	cd01293	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 reliated to the adenosine deaminases (ADA). The bacterial enzymes are inrin dinendent and hexamence:	29.43	4.18E-01	30%	43%	73127	92148	1	AAC98904 A537L	103.22	8.20E-21	30%	46%	4248	10263
N524R	201389201898	170	18,987	10.49	1	smart00465	5 GIYc, GIY-YIG type nucleases (URI domain); .	32.35	3.52E-02	38%	54%	2968	544	1 2 3	AAC96906 A539R AAC96502 similar to bacteriophage T4 intron-associated endonuclease, corresponds to Swiss-Prot Accession Number P13299 AAC98682 similar to PBCV-1 ORF A315L, corresponds to GenBank Accession	50.45	2.82E-44 2.58E-05 5.39E-03	51% 53% 29%	70% 68% 44%	1168 2874 37145	1173 1155 15123
N525L	205835201867	1323	133,329	4.77		No Hit Found								1 2	AAC96062 Number M74440 AAC96907 A540L ABE69167 A540L	227.64	2.62E-57 3.82E-16	29% 26% 24%	44% 40% 36%	28897 28452	138123 1381055 172543
														3 4 2	YP_589431 hypothetical protein Acid345 0352 ZP 00950302 outer membrane protein	50.45 48.14	5.72E-04 2.84E-03	26% 22%	41% 35%	28194 8371205	701-864 58-446
														5 6	AAA66404 unknown protein YP 214529 possible T4-like proximal tail fiber	150.21 94.74	5.31E-34 2.65E-17	37%	53%	10851312 5691223	9273 29598
														8 2	ZP_00533161 Hep_Hag ZP 01257209 haemaoluttinin family protein	109.00 50.06	1.36E-21 7.47E-04	26% 26%	42% 39%	582-959 32210	71447 156334
														9 2	ZP_01059081 hypothetical protein MED217 15260 ZP_00532602 Hep Hag	49.29	1.27E-03 3.71E-03	24% 24%	38% 37%	30208 9210	221-394 655-868

Gene Name	Genome Position	A.A. lenath	Peptide Mw	pl CDD Hit Number	COGs COG Definition	Bit Score	E-value	% Identity Po		Query H from-to	Hit from- E	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% lentitv Po		Query from-to	Hit from- to
N531L	207897205900	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	50.06	3.42E-04	28%	45%	174291	30134
												2		P40602 DNA binding protein	51.22	1.54E-04	26%	36%	133289	95261
N533L	208755207985	257	29,166	8.44	No Hit Found							1	AAC9637	7 A9R	222.25	1.28E-56	63%	79%	85250	8173
												2	AAC96972	Asimal to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580		1.48E-28	31%	51%	1251	4255
												3	AAC9681	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number	118.24	2.61E-25	29%	50%	5249	2246
												4		similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055		8.39E-24	28%	49%	5248	5248
												5	AAC9654 AAC9644	A79R	91.66	2.61E-17 3.41E-17	25% 27%	44% 45%	5252 5228	5245 2218
												7	AAC9697	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number LI42580		1.43E-15 1.88E-07	64% 27%	90% 46%	555	252
												8 9	AAU0630	hypothetical protein A275R hypothetical protein A275R	57.77	4.18E-07	26%	46%	85248 85248	4167 4167
												10		2 hypothetical protein A275R		7.13E-07	28%	47%	117251	1134
N535L	213044209160	1295	130,191	6.38 1	COG0196 RibF, FAD synthase [Coenzyme metabolism].	35.26	5.65E-02	25%	39%	10261143	164274	1	AAC9690 ABE6916	7 A540L	50.45	2.76E-19 5.59E-04	24% 26%	36% 41%	9170	5231052 389544
												3 4	AAA66404	hypothetical protein Acid345_0352 unknown protein	137.50	1.47E-04 3.48E-30	23% 37%		28198 10801292	438-599 50280
												5	ZP_01257209 ZP_00533161	haemagluttinin family protein Hep Hag	60.46	5.04E-13 5.40E-07	23% 22%	35% 45%	30465 12209	24412 54249
												7	ZP 00945173	possible T4-like proximal tail fiber     Collagen alpha 2(I) chain precursor	72.40	2.05E-14 1.37E-10	21% 25%	35% 33%		362-960 7361083
												9 10	ZP 00532602 ZP_00950302	2 Hep Hag 2 outer membrane protein	49.68 51.22	9.53E-04 3.28E-04	24% 25%	38% 45%	40209 28178	684-867 640-784
N540L	214063213113	317	36.864	10.55 1	COG0436 COG0436, Aspartate/tyrosine/aromatic aminotransferase [Amino acid	29.12	7.54E-01	28%	44%	3387	124174			No Hit Found						
IN340L	214003213113	317	30,004	10.55 1	transport and metabolismI.	29.12	7.34E-01	20%	444 70	3367	124174									
N542L	215171214032	380	41,731	10.39 1	unknown].	29.53	7.19E-01	37%	47%	282348	74133	1	AAC9698	5 STKPP (11x); similar to Gossypium Pro-rich wall protein, corresponds to GenBank Accession Number U04267	181.80	3.60E-44	46%	60%	159378	163384
					deoxycytidylate_deaminase, Deoxycytidylate deaminase domain. Deoxycytidylate deaminase catalyzes the deamination of dCMP to															
N546L	215578215141	146	16,059	7.00 1	dUMP, providing the nucleotide substrate for thymidylate synthase. The cd01286 enzyme binds Zn++, which is required for catalytic activity. The activity of the activity of the activity of the second seco	128.08	3.69E-31	39%	55%	9135	4131	1	AAC9693	similar to Vibrio fischeri dCMP deaminase, corresponds to Swiss-Prot Accession Number P33968	201.45	7.16E-51	65%	81%	4144	2142
					the enzyme is allosterically regulated by the ratio of dCTP to dTTP not only in eukaryotic cells but also in Teven phage-infected Escherichia coli with dCTP action as an activator and dTTP as an inhibitor															
				2	COG2131 ComEB, Deoxycytidylate deaminase [Nucleotide transport and	104.27	5.66E-24	36%	54%	9141	12148	2	YP_22395	deoxycytidylate deaminase	124.02	1.45E-27	52%	65%	9139	110-237
				3	pfam00383 dCMP_cyt_deam, Cytidine and deoxycytidylate deaminase zinc-binding region.	86.57	9.86E-19	41%	60%	8113	5100	3	YP_42206	Deoxycytidylate deaminase	109.38	3.70E-23	44%	62%	21141	20139
					cytidine_deaminase-like, Cytidine and deoxycytidylate deaminase zinc- binding region. The family contains cytidine deaminases, nucleoside															
				4	deaminases, deoxycytidylate deaminases and riboflavin deaminases, cd00786 Also included are the apoBec family of mRNA editing enzymes. All	58.87	2.71E-10	29%	50%	14113	6-92	4	70 0006296	COG2131: Deoxycytidylate deaminase	100.28	3.70E-23	44%	62%	21141	20139
				•	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	56.67	2.712-10	2070	50%	14-115	0-32	-	21 _0003200.	COG2131: Deoxycylidylate deaniinase	109.00	3.70E-23	4476	02.70	21-141	20-138
					form a hydroxide ion that performs a nucleophilic attack on the substrate. nucleoside_deaminase, Nucleoside deaminases include adenosine,															
					guarine and cytosine dearninase. These enzymes are Zn dependent and catalyze the dearnination 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															
				5	cd01285 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	56.43	1.36E-09	32%	55%	10120	1100	5	NP_975562	dCMP deaminase	99.37	3.83E-20	40%	56%	9142	13147
					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 quanine to vanthine as part of the utilization of quanine : Riboflavin_deaminase-reductase, Riboflavin-specific deaminase.															
					Riboflavin biosynthesis protein RibD (Diaminohydroxyphosphoribosylaminopyrimidine deaminase) catalyzes															
				6	the deamination of 2,5-diamino-6-ribosylamino-4(3H)-pyrimidinone cd01284 5'-phosphate, which is an intermediate step in the biosynthesis of	55.25	2.72E-09	38%	56%	28113	2193	6	YP_42437	probable deoxycytidylate deaminase	99.37	3.83E-20	40%	56%	9142	6140
					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 COG0590 CumB, Cytosine/adenosine deaminases [Nucleotide transport and															
				7	COG0590 Collimit of Statistical Collimit of Collimit o	54.57 54.10	5.04E-09 7.39E-09	34% 28%	50% 51%	1113 1113	3104 1100			Cytidine/deoxycytidylate deaminase, zinc-binding region     hypothetical protein LOC549019		5.01E-20 6.54E-20	45% 37%	68% 56%	21113 8140	22115 28162
												9 10		MGC81193 protein putative deoxycytidylate dearninase		8.54E-20 1.46E-19	38% 50%	56% 67%	8140 21113	28162 35134
					S_TKc, Serine/Threonine protein kinases, catalytic domain.														-	
N548R	215695216513	273	30.481	8.08 1	Phosphotransferases of the serine or threonine-specific kinase subfamily. cd00180 The enzymatic activity of these protein kinases is controlled by	150.74	1.47E-37	29%	50%	15265	1256	1	A A CORRE	similar to PBCV-1 serine/threonine protein kinase, corresponds to	100 70	7 88E-31	33%	49%	4265	6279
N048K	215695216513	2/3	30,481	8.08 1	phosphorylation of specific residues in the activation segment of the catalytic domain, sometimes combined with reversible conformational	150.74	1.4/E-3/	29%	50%	15265	1256	1	AAC9665	GenBank Accession Number U14660	136.73	7.88E-31	33%	49%	4205	6279
				2	changes in the C-terminal autoregulatory tail	144.97	8.97E-36	28%	48%	16265	1256	2	XP 38729	hypothetical protein FG07121.1	98.21	3.10E-19	27%	46%	14269	285556
				3	Phosphotransferases. Serine or threonine-specific kinase subfamily pfam00069 Pkinase, Protein kinase domain	143.89	1.70E-35	30%	49%	16265	1258	3		7 mitogen-activated protein kinase kinase 2		9.03E-19	29%	47%	19268	11265
				4	SPS1, Serine/threonine protein kinase [General function prediction only / COG0515 Signal transduction mechanisms / Transcription / DNA replication,	106.78	2.54E-24	26%	44%	16269	2282	4	CAC0796	putative mitogen-activated protein kinase kinase 2	95.52	2.01E-18	29%	47%	19268	11265
				5	recombination and renair! smart00219 Tyrkc, Tyrosine kinase, catalytic domain; Phosphotransferases.	74.44	1.48E-14	24%	44%	17258	2250	5	AAU0628	protein kinase A248R	93.59	7.65E-18	25%	46%	14268	46308
					Tyrosine-specific kinase subfamily. 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 cd00192 common to both serine/threonine and tyrosine protein kinases. Enzymatic															
				6	cd00192 activity of tyrosine protein kinases is controlled by phosphorylation of specific tyrosine residues in the activation segment of the catalytic	68.68	7.32E-13	24%	43%	13261	6264	6	AAU0627	5 protein kinase A248R	93.59	7.65E-18	25%	46%	14268	25287
					domain or a C-terminal tyrosine (tail) residue with reversible															
				7	COG3894 Uncharacterized metal-binding protein [General function prediction onlv].		7.13E-03	21%	35%	130247	105227	7	XP_57864	PREDICTED: similar to hypothetical protein 4930509O22	92.43	1.70E-17	30%	46%	22265	734-976
					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		1 105 0	2007	E001	110 157	450 400		ND 10055	5 CIPK11; ATP binding / kinase/ protein kinase/ protein serine/threonine	04.07	2.91E-17	28%	47%	00.004	07 070
				8	pfam01636 phosphotransferase and streptomycin 3' & apos; kinase or streptomycin 3' & apos; -phosphotransferase. The aminoglycoside		1.10E-01	36%	52%	116157	156198	8	NP_180595	kinase/ protein-tyrosine kinase	91.66	2.91E-17	28%	47%	22264	21-276
					phosphotransferases inactivate aminoglycoside antibiotics via phosphorylation. This family also includes homoserine kinase. This family is related to fourteensing kinase a form2030.															
				9	is related to functinsamine kinase nfam03881 COG2334, Putative homoserine kinase type II (protein kinase fold) IGeneral function prediction onlvi	30.70	1.92E-01	27%	43%	114158	183227	9	AAL16166	At2g30360/T9D9.17	91.66	2.91E-17	28%	47%	22264	27276

Gen	e Genome	A.A.	Peptide	la	CDD Hit	COGs	COG Definition	Bit Score	E-value	%	%	Query	Hit from- BL/	ASTp Hit	Hit BLASTp Definition	Bit Score	Evalue	%	%	Query	Hit from-
Nam	e Position	lenath	Мw	р	Number 10	COG3173	COG3173, Predicted aminoglycoside phosphotransferase [General	29.61	3.86E-01	Identitv F 42%	ositive 52%	from-to 125-157	to N 195226	umber 10	ACCession AAU06282 protein kinase A248R		3.79E-17	dentitv P 25%	ositive 45%	from-to 14268	to 46308
					10		function prediction only1	20.01	3.002-01	42.76	5270	123-137	155-220	10		51.20	3.782-17	2370	4576	14200	40500
N549R	216631219810	1060	117,600	9.53	1		TOP2c, TopoisomeraseII; Eukaryotic DNA topoisomerase II, GyrB, ParE .	453.09	7.69E-128	32%	49%	50610	1589	1	AAU95770 topoisomerase II	1889.39	0.00E+00	90%	92%	11058	1-1058
					2	smart00434	TOP4c, DNA Topoisomerase IV; Bacterial DNA topoisomerase IV, GyrA, ParC	392.63	1.11E-109	33%	50%	6261017	1423	2	AAC96932 PBCV-1 DNA topoisomerase II	1338.55	0.00E+00	63%	75%	31056	2-1061
							TOP4c, DNA Topoisomerase, subtype IIA; domain A'; bacterial DNA topoisomerase IV (C subunit, ParC), bacterial DNA gyrases (A														
					3	cd00187	subunit, GyrA),mammalian DNA toposiomerases II. DNA topoisomerases are essential enzymes that regulate the conformational changes in DNA	362.96	8.47E-101	33%	51%	6441017	1401	3	NP_584718 DNA TOPOISOMERASE II	877.09	0.00E+00	47%	64%	5-1017	81030
							topology by catalysing the concerted breakage and rejoining of DNA strands during normal cellular growth.														
					4		GyrB, Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), B subunit IDNA reolication. recombination. and reoair1. DNA topoisolV, DNA gyrase/topoisomerase IV, subunit A.	319.39 271.27	1.17E-87 3.49E-73	27% 30%	43% 45%	2626 6461017	8635 1397	4	NP_189031 TOPII (TOPOISOMERASE II); ATP binding / DNA binding / D topoisomerase (ATP-hvdrolvzina) AAW40881 DNA topoisomerase II, putative		0.00E+00 0.00E+00	46% 45%	63% 61%	51000 51000	351072 1121153
					6	COG0188	GyrA, Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV),	251.33	3.96E-67	23%		6171015	2428	6			0.00E+00	45%	60%	5-1000	321097
							A subunit IDNA replication, recombination, and repairl, TOPRIM_TopolIA_TopolIA, TOPRIM_TopolIA: 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'														
					7	cd02265	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.	142.67	2.18E-34	66%	78%	403-521	1120	7	BAE06274 topoisomerase II	922.56	0.00E+00	45%	60%	51000	25 1079
						000000	Ihese proteins also catenate/ decatenate duplex rings. The TOPRIM domain has two conserved motifs, one of which centers at a conserved	142.07	2.102.04	00%	1070	400 021	1 120		bi 2002) - Tupulaumerase m	000.00	0.002.00	4070	00%	0 1000	00 1010
							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 catalutic function														
							TOPRIM_TopolIA, TOPRIM_TopolIA: 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'														
					8	cd01030	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.	123.67	1.00E-28	53%	63%	403-521	1115	8	BAD86854 DNA topoisomerase II	833.17	0.00E+00	43%	60%	51000	1031163
							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.														
							TOPRIM_TopolIA_GyrB, TOPRIM_TopolIA_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. TopolIA 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														
					9	cd03366	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	82.10	2.96E-16	44%	57%	403-510	1105	9	AAN85208 DNA topoisomerase II	831.63	0.00E+00	45%	62%	5-1000	271065
							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 and the strand placement by the strangeneous the DVD mattice and the strangeneous base of the strangeneous the strangeneous base of														
							acid in strand cleavage by topisomerases. The DXD motif may co- ordinate Mc2+ a cofactor required for full ratatytic function DNA gyraseB, DNA gyrase B. This family represents the second														
					10	pfam00204	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	14168	10	XP_751245 DNA topoisomerase II	828.17	0.00E+00	44%	61%	2996	551094
							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-														
N559R	219871220755	295	33,581	10.47	1	cd00283	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	55.39	9.05E-09	41%	61%	181-275	1100	1	AAA88832 unknown	194.90	2.78E-48	40%	55%	1289	1240
							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.														
					2	smart00497	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting,	48.98	7.50E-07	42%	60%	239-291	1-53	2	AAC96973 similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBa Accession Number U42580	nk 190.66	5.25E-47	39%	53%	1290	1224
					3		unpublished) NUMOD1, NUMOD1 domain	44.65	1.50E-05	48%	67%	239-272	1-34	3	AACOPPEE DDDU/ 4 001-1 11-1	172.22	8.67E-42	35%	52%	3290	6248
					4	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	40.06	3.56E-04	32%	54%	172	170	4	AAC96865 PBUV-1 33k0 peptide AAC96862 similar to PBCV-1 ORF A315L, corresponds to GenBank Access Number M74440	on 48.14	4.21E-04	42%	67%	179234	97151
					5	COG1693	COG1693, Uncharacterized protein conserved in archaea [Function unknown]	35.64	7.69E-03	29%	48%	108186	1390	5			3.10E-15	29%	46%	1241	1197
					6	pfam01995	DUF128, Domain of unknown function DUF128. This archaebacterial protein family has no known function. The domain is found duplicated in	35.33	1.03E-02	27%	48%	108186	1491	6	YP_293795 putative endonuclease	72.02	2.72E-11	44%	58%	284	385
							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														
					7	pfam01541	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	33.98	2.43E-02	33%	56%	159	1-64	7	AAC49244 ORF301	71.63	3.55E-11	28%	48%	4170	76236
							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														
					8	COG3069	DcuC, C4-dicarboxylate transporter [Energy production and conversion].	31.79	1.05E-01	26%	43%	2365	187229	8	NP_074951 orf305	70.86	6.06E-11	32%	49%	4204	74262
					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%	118202	200284	9	AAC49248 ORF211	55.84	2.02E-06	28%	47%	129290	49208
														10	AAT53588 group I intron GIY-YIG endonuclease	49.68	1.45E-04	25%	43%	4205	5197
							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														
N561R	220777221190	138	15.751	11.06	4	c403604	shown to preferentially cleave substrates through a monopeptidyl carboxypeptidase pathway. The propeptide region of cathepsin X, the	26.82	9.48E-01	68%	73%	935	4062		AAC96929 A577L	100 04	6.33E-23	55%	63%	26130	2104
NJOIR	220111-221190	130	13,751	11.00		0002098	shortest among papain-like peptidases, is covalently attached to the active site cysteine in the inactive form of the enzyme. Little is known bend the indexine fractions of enterprise of the source and attaches the source of the source o	20.02	0.40E-U1	00%	/ 376	830	-1002	1	AND ADD ADD ADD ADD ADD ADD ADD ADD ADD	106.61	0.336-23	3370	0276	20-130	2104
							about the biological function of cathepsin X. Some studies point to a role in early tumorigenesis. A more recent study indicates that cathepsin X supractices is restricted to immune cells curgenting a role in phase order in														
							expression is restricted to immune cells suggesting a role in phagocytosis and the regulation of the immune response														

Gene Name N562R	Genome Position 221235221453	A.A. lenath 73	Peptide Mw 8,350	pl CDD Hit Number 10.39 1	COGs COG Definition COG2433 COG2433, Uncharacterized conserved protein (Function unknown		E-value 5.99E-01	% Identitv P 19%	% ositive 47%	Query from-to 754	Hit from- E to 415-462	BLASTp Hit Number		BLASTp Definition	Bit Score	E-value Id	% lentitv Po	% ositive	Query from-to	Hit from- to
N564R	221522222049	176	19,167	9.91	No Hit Found							1	AAC96743 A375R		234.19	1.39E-60	64%	80%	4171	6172
N565L	223773222040	578	59,995	10.57	No Hit Found							1 2 3 4	2 BAE02830 surface protein 3 AAC96508 PBCV-1 surface pro 4 BAD22850 surface protein		46.98 46.98	5.00E-49 2.43E-03 2.43E-03 3.18E-03	37% 30% 32% 25%	56% 53% 50% 49%	301-576 234-304 234-304 234-326	214496 638707 829896 675767
												5 6 7	5 AAC96752 Lys-, Pro-rich, PA corresponds to Ger 5 BAD12236 surface protein 7 T17636 proline-rich protein	PK (10x); similar to wheat Pro-, Lys-rich protein nBank Accession Number X52472 A145R - Chlorella virus PBCV-1	46.98 46.98	3.75E-04 2.43E-03 2.43E-03	30% 32% 32%	54% 50% 50%	233-304 234-304 234-304	201272 829896 38105
N567L	224456223893	188	21,369	3.11	No Hit Found							1	AAC96772 Glu-, Asn-rich			4.02E-14	65%	77%	157	259
N569L	224896224627	90	10,063	10.28	No Hit Found							1	AAC96771 A403R			9.54E-19 1.31E-59	52%	69%	288	693
N570L	225626224934 226594225719	231 292	25,002 32,559	6.63 5.48	No Hit Found							1	AAC96770 A402R			1.31E-59	48% 60%	70% 75%	1229	1227
NJZL	220394-223713	2.92	32,338	0.40	DHOD_DHPD_FMN, Dillydroordate dehydrogenase (DHOD Dillydropyrmidine dehydrogenase (DHPD) FMN-binding domain.	DHOD						2 3 4 5 6	2 NP_772491 hypothetical protein	tical protein n GOX0246 n RS9917 04490	141.35 130.95 119.40	3.58E-32 4.84E-29 1.46E-25 1.73E-18	35% 33% 31% 28% 24%	53% 52% 48% 46% 41%	34-284 34-284 28-287 34-281 23-289	7245 11249 5253 8217 2263
N574L	226993226637	119	13,074	10.03 1	catalyzes the oxidation of (5)-dihydroordate to orotate. This is the step and the only redox reaction in the donovo biosynthesis of UM precursor of all pryimidine nucleotides. DHCD requires FNN ac- sequences and outliair location. Members of class 1 is or obsequences and outliair location. Members of class 1 is or obsequences and outlimers while class 2 enzymes are men into subtypes 1A and 1B which are homodimers and hetrofielt proteins, respectively. DHCD catalyzes the first step in prin the corresponding 5,6-dihydropyrmidines. DHFD contains two FA FNN and eight (F4-5) clusters, armograd in two electron transfer	P, the factor. o acid tosolic hbrane 27.45 americ midine hine to D, two	4.37E-01	26%	63%	633	96123	1	I AAC96768 similar to PBCV-1 Number U42580	ORF A214, corresponds to GenBank Accession	107.07	1.86E-22	45%	68%	5117	4116
N575L	227516227019	166	18,181	9.92 1	that pass its homodimeicr interface twice. Two of the Fe-S clusters a hitherto uncheaned coordination invulving a duitamine residue RnaseH, RNase H. RNase H digests the RNA strand of an RNA nfam00075 hybrid. Important enzyme in retroviral replication cycle, and often	v/DNA	1.46E-14	30%	47%	16136	2124	1	AAC96767 A399R		149.83	2.87E-35	51%	64%	4158	17180
				2	<ul> <li>as a domain associated with reverse transcriptases. Structure is a aloha+beta fold with three a/b/a lavers</li> <li>COG0328 RnhA, Ribonuclease HI [DNA replication, recombination, and repa</li> </ul>		1.05E-09	23%	37%	17147	3153	2	2 XP_662792 hypothetical protein	AN5188.2	53.53	2.80E-06	32%	44%	18136	228-373
												3 4	GAA41925 unnamed protein pr	roduct	52.76	4.77E-06 4.77E-06	32% 32%	51% 51%	15140 15140	238369
												5		Reverse Transcriptase (Ribonuclease H Domain		1.39E-05 1.81E-05	27% 32%	42% 44%	9136 21136	1141 634755
												7	3         ZP_01217503         hypothetical protein           7         EAL40632         ENSANGP0000000           3         EAL40633         ENSANGP00000000	D6819	50.45	2.37E-05 4.04E-05	29% 29%	44 % 48% 48%	13-137	249-372 274-397
												9 10	AAB65093 Lian-Aa1 retrotrans	sposon protein		4.04E-05	31% 31%	52% 46%	15147 18136	9101046 239383
N578R	227622227981	120	12,941	10.45	No Hit Found							1	I AAC96766 A398L		147.52	1.24E-34	61%	77%	1118	1118
N579L	228450227986	155	17,543	11.74 1	COG4656 RnfC, Predicted NADH:ubiquinone oxidoreductase, subunit IEnerav production and conversion1.	RnfC 28.75	3.10E-01	19%	41%	45131	431517	1				1.55E-13 1.31E-04	33% 31%	51% 54%	6152 45152	22150 361461
												3 4	AAW42894 chromosome assoc AAH87521 LOC496093 protein	siated protein, putative	44.67 43.90	1.90E-03	25% 29%	47% 54%	8144 49127	314457 50128
												5	5 ZP_01179929 NLP/P60 5 XP_654272 hypothetical protein	58.t00022	41.59	2.48E-03 9.42E-03 3.24E-03	25% 29% 28%	44% 44% 46%	5149 50152 46152	
												8	<ul> <li>EAR81596 conserved hypothel</li> <li>XP 742542 hypothetical protein</li> <li>CAG74282 ToIA protein</li> </ul>	ucal protein 1	42.36	5.52E-03 7.21E-03	28% 24% 23%	40% 41% 42%	40152 19152 2151	105-235
												10		protein	41.97	7.21E-03	28%	47%	50152	1225-1329
N582L	228760228479	94	10,557	10.42 1	DUF123, Domain of unknown function DUF123. This archaebest domain has no known function. It is attached to an endounclease de pfam01986 in one member. The domain contains several conserved cysteine histidines. This suggests that the domain may be a zinc binding n acid interaction domain (Bateman A unoubl.).	omain es and 26.85	6.94E-01	30%	41%	1660	2384	1	I AAC96763 A395R		109.00	4.82E-23	64%	78%	1192	182
N584R	228842229471	210	23,746	8.20 1	COG1403 McrA, Restriction endonuclease [Defense mechanisms].	28.37	7.04E-01	20%	29%	117-193	44120	1	I AAC96845 signature; similar 1	tains eukaryotic putative RNA-binding region RNP-1 to PBCV-1 ORF A267L, corresponds to GenBank	164.85	1.62E-39	42%	58%	19207	106-294
				2	pfam03302 VSP, Giardia variant-specific surface protein	28.17	7.74E-01	29%	42%	62107	275320	2	Accession Number AAC96857 Lys-, Glu-rich YP 142777 unknown	U42580		2.20E-04 4.44E-05	27% 22%	44% 40%	597 2165	189-288 365-524
												4	4 AAC96635 A267L			3.50E-18	32%	49%	13207	59247
N586L	230036229458	193	21,469	3.72	No Hit Found							1				2.58E-19	47%	64%	6110	14116
N589L	230850230080	257	28,991	9.23 1	COG0433 COG0433, Predicted ATPase [General function prediction only]. FtsK, DNA segregation ATPase FtsK/SpolIIE and related protein	29.95 Is [Cell 29.41	3.60E-01		44% 43%	77146	373446 627696	1	corresponds to Ger	-binding site motif A; similar to frog virus 3 ATPase, nBank Accession Number M80551		1.83E-103	74% 39%	81% 60%	1252	1251
				2	division and chromosome nartitioning	29.41	4.34E-01	22%	43%	83149	627696	2	3 AAR26836 FirrV-1-A12		145.21	6.84E-34 1.99E-33	39%	57%	32228 32228	36241
												4 5	ABF82118 hypothetical protein AAT09674 AAA-ATPase	MIV088R	130.57	2.28E-29 5.08E-29 2.61E-25	36% 34% 33%	57% 56% 55%	32228 32223 32228	47252 50244 43241
												7	AAP33264 ATPase AAP37796 putative ATPase		118.24	2.61E-25 2.61E-25 2.61E-25	33% 33%	55% 55%	32228 32228 32228	43241 43241 43241
												9 10	AAM82544 putative ATPase		117.86	3.40E-25 7.58E-25	33% 31%	55% 56%	32228 32228	43241
N593R	230933232633	567	59,105	7.04 1	domains of which this is at the N-terminus. The VAT-N domain for AAA ATPases pfam00004 is a substrate 185-residue reco	TPase s. Both uctural 31.35 und in gnition	3.56E-01	36%	47%	406452	3285	1	AACONTED LYS-, Pro-rich, PA	PK (10x); similar to wheat Pro-, Lys-rich protein Bank Accession Number X52472		2.91E-17	31%			26181
				2	nonain CBM_14. Chilto binding Pertrophin-A domain. This domain is call Pertrophin-A domain and is found in chilto binding proteins partie pertrophic matrix proteins of insects and animal chiltomase. Copies pfam01607 domain are also found in some baculoviruses. Relevant reference becribe proteins with this domain include. It is an extracellular dr domain are also found in some baculoviruses.	of the es that 30.79 omain	4.91E-01	25%	48%	515565	1-49	2	2 AAC96921 A565R		56.61	3.01E-06	27%	36%	128–333	298-469
					that contains six conserved cysteines that probably form three disu bridges. Chitin binding has been demonstrated for a protein cont only two of these domains	taining														
				3	smart00494 ChtBD2, Chitin-binding domain type 2; .	30.10	6.97E-01	38%	50%	515565	351	3 4	BAA11343 DNA binding protein AAC96700 a332L		73 04	1.35E-22 1.82E-11	34% 61%	40% 80%	129374 523564	252458 243
												5	5 AAC96919 similar to Chlorella GenBank Accession	a virus CVK2 DNA binding protein, corresponds to n Number D78305	63.16	3.22E-08	40%		283374	306-403

Gene Name	Genome Position	A.A. lenath	Peptide Mw		CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identitv F		Query from-to	Hit from-BL to N		Hit BLASTp Definition BAA1132 DNA binding protein XP 416321 PREDICTED: similar to FELE-2 CAA46476 / Q1/4 protein CA117471 tenascin XB CA11975 tenascin XB	45.44 52.76	3.56E-07 6.95E-03 4.35E-05 6.71E-06	% dentitv P 39% 22% 23% 26% 26%	% 50% 33% 35% 35% 35%	from-to 283-374	Hit from- to 304-401 1821-1992 403-638 416-673 416-673
N598R	232979233872	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 trafficing. In particular, Vps26 assortibles into a retromer complex with at least four other proteins Vps5, Vps17, Vps29 and Vps35. This family also contains Down syndrome critical eigon 3/A.	29.48	5.15E-01	26%	37%	3494	971	1	AAC97017 a593R	47.37	7.31E-04	63%	65%	87124	138
N600L	234932233850	361	40,439	6.99	1 2 3 4 5 6 7 8	pfam00283 COG0520 COG1104 COG2000 COG1893 pfam00260 pfam00610	6 GadB, Glutamate decarboxylase and related PLP-dependent proteins (Amino acid transocit and metabolism). 9. Prytöxol.46c-pr.9. prytöxol.46c-pendent decarboxylase conserved domain. 0. GsdB. Selencoxyteine lyase (Amino acid transport and metabolism). NRS, Oyteine sulfnaet desulthose/vyteine desultrase and related enzymes (Amino acid transport and metabolism). 3. ApbA. Ketopantoate reductase (Coercyme metabolism). 3. ApbA. Ketopantoate reductase (Coercyme metabolism). 1. Aminotran 5. Aminotranifisma et al. Set J. (CPM). This family 1. consists of serveral hypothetical proteins for Arabidopsis thaliana and Ovza sativa: The function of the acid minis unknown.	36.76 32.21 31.05 30.65	2.97E-35 1.01E-26 3.04E-06 1.14E-04 4.16E-03 1.09E-01 2.31E-01 2.68E-01	27% 26% 23%	44% 44% 44% 35% 33% 36%	46-310 54-289 63-308 79-226 54-216 140-278 77-226 307-355	91380 79352 75339 67216 35206 45178 55-203 343391	5 6 7	AAC96037 similar to tomato hisidine decarboxylese, corresponds to Swiss-Prot Accession Number P54772 P28578 Hisidine decarboxylese (HDC) AAA20507 hisidine decarboxylese ZP_00510529 Pyridoxal-dependent decarboxylase AAR12533 hisidine decarboxylase AA02258 hisidine decarboxylase AA02258 hisidine decarboxylase A20218 Abiddine decarboxylase A20218 Abiddine decarboxylase A20218 COG0076: Glutamate decarboxylase and related PLP-dependent proteins	217.24 217.24 210.69 208.76 208.76 208.76	6.82E-130 7.21E-55 7.21E-55 6.75E-53 2.56E-52 2.56E-52 2.56E-52 5.71E-52	62% 35% 36% 34% 34% 35%	78% 57% 56% 56% 56% 56%	4353 14347 14347 14347 14347 14347 10347	11362 26369 26369 47390 40383 26369 86432 25363
N603L	235413235090	108	12,185	5.98	9		Gor-P. Clycine cleavage system protein P (cyridoar-binding), C-terminal domain (Amino acid transcort and metabolismi). VDE, Violaxanthin de-epoxidase (VDE). This family represents a conserved region approximately 350 residues long within plant epoxidase forms part of a conserved system that dissipate encress energy as heat in the light-harvesting complexes of holosystem (IFR).	26.54	4.15E-01 7.02E-01	25%	45% 53%	82231 4077	134289 301341	9 10 1	P05034 Histidine decarboxylase (HDC) BAE94285 histidine decarboxylase AAC96639 A662L	206.07	1.27E-51 1.66E-51 7.63E-08	33% 34% 34%	55% 56% 64%	14355 14347 372	26377 26369 46113
N604R	235450235770	107	11,961	6.35		No Hit Found	thus protecting them from photo-inhibitory damage							1	AAC96940 A603R	96.29	3.31E-19	44%	64%	1103	1104
N606L	236442235942	167	18,320	10.80		No Hit Found	1							0	No Hit Found No Hit Found						
N609L N613L	237578236511	356	40,475	9.80		No Hit Found								1 2 3 4	AAC96722 A354R AAC96790 similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot AAC96455 A87R AAC96455 A87R AAE904297 Rhomuclease E AAC98659 A492L	85.50 51.99 49.29	5.97E-54 3.20E-15 3.92E-05 2.54E-04 2.11E-22	47% 28% 28% 32%	63% 47% 45% 47% 60%	110-353 30320 191350 71183 85250	289447 582692
N614R	238583239749		45,772	9.12	1	cd02661	Peptidase_C19D, A subfamily of Peptidase C19. Peptidase C19 contains ubiquiting hydrolases. They are intracellular peptidases that remove ubiquitin molecules from polyubiquintated peptidase by davage of isopeptide bonds. They hydrolyse bonds involving the activoxy group of the C-terminal Gy residue of ubiquitin. The purpose of the de- uod feasour them from degradation, as well as recycling of the hubinum. The ubiquininproteasome system is responsible for most protein turnover in the mannalism Cell, and with over 50 members, family C191s one of	29.90	6.02E-01	20%	37%	181–240	118189	2	AAC96880 A493L		7.78E-17 9.63E-117	39%	56%	105–242 36–387	29174 4360
							the largest families of poptidases in the human genome							2 3 4 5 6 7 8 9 10	YP         142783         PBCV1-A494R-like protein           NP_07786         ExV-1-101           ARA2886         Friv-1-820           YP_284161         hsponhetical protein ENV 403           NP_672472         Starshidd protein MIV079L           YP_274342         Mishonhetical protein PC01078 02.0           EAR82893         hypothetical protein TTHERM 01054240	135.19 127.10 112.08 70.86 60.85 58.15	4.89E-36 4.01E-30 1.09E-27 3.64E-23 9.29E-11 9.61E-08 6.23E-07 4.94E-04 6.45E-04	34% 37% 33% 22% 20% 27% 22% 22% 21%	54% 52% 51% 42% 37% 44% 43%	122-387 195-386 178-386 182-383 126-384 106-383 266-383 78-302 64-326	86300 156362 101380 93390 131249 134355
N616L	240862239714	383	43,529	7.20	1	smart00497	IENR1, Intron encoded nuclease repeat motif. Repeat of unknown f function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unoublished). GIY-YIG, GIY-YIG catalytic domain. This domain called GIY-YIG is found in the amino terminal region of excinuclease abc subunit (uvrC), bacteriophage 14 endonucleases seqA. segB. segC. segD and segE; it it		1.49E-04	34%	62%	228–279	151	1	AAA88832 unknown	88.20	5.49E-16	29%	46%	56-279	14240
					2	pfam0154	also found in putative endonucleases encoded by group i Introns of fungi and phage. The structure of I-revia GIV-YIG endonuclease, reveals a novel alphabeta-toid with a central three stranded antiparalie beta-sheet lanked by three helices. The most conserved and putative catalytic residues are located on a shallow, concave surface and include a metal covortination site	37.06	3.77E-03	28%	38%	44113	475	2	Accession Number U42580		1.60E-15	32%	47%	57-279	15223
					3	smart0046	5 GIYc, GIY-YIG type nucleases (URI domain): .	35.05	1.51E-02	31%	47%	42-133	283	3 4 5 6 7 8	AAC99692 similar to PECV-1 ORF A315L, corresponds to GenBank Accession ZP 0073833 hypothetical protein R8TH 06729 YP_233756 bytailve endonuclease CAA29393 unnamed protein product NP_049674 ModB homing endonuclease YP_142601 HNH endonuclease	60.85 55.07 45.05 45.05	7.93E-15 9.39E-08 5.15E-06 5.33E-03 5.33E-03 6.96E-03	42% 25% 40% 26% 26% 31%	59% 44% 55% 41% 41% 57%	53155 73247 57137 203358 203358 155220	13117 9166 1697 73226 110263 124189
N618R	240974241405	144	14,855	10.16		No Hit Found								1	AAC96864 Gly-rich, KGKDIGGG (4x)	51.99	7.08E-06	38%	49%	75141	79145
N619L	242426241404		36,454	5.08	1		8 CinA, Predicted nucleotide-utilizing enzyme related to molybdopterin- biosvnthesis enzyme MoeA (General function prediction onlv).	30.58	2.65E-01	38%	46%	2171	3886	1 2		58.54	1.22E-08 3.91E-07	48% 49%	67% 67%	978 270318	566 464
N621L	242725242495	77	8,145	4.28		No Hit Found								1			9.97E-08	49%	72%	2575	
N624L	243735242770	322	34,665	5.24		No Hit Found	1							1 2	AAC96855 Pro-,Gin-rich YP_142771 unknown		2.19E-97 9.44E-08	60% 25%	72% 39%	1320 8256	
N628L	244213243776	146	16,890	11.00		No Hit Found	I Pyr_excise, Pyrimidine dimer DNA glycosylase. Pyrimidine dimer DNA							1	AAC96852 A485R	176.41	2.47E-43	57%	78%	4144	7147
N629L	244666244229	146	16,765	10.06	1	pfam0301	glycosylases excise pyrimidine dimers by hydrolysis of the glycosylic 3 bond of the Skapos: pyrimidine, followed by the intra-pyrimidine phosphodiester bond. Pyrimidine dimers are the major UV-lesions of NA Lyase_L Lyase class I family of the Lyase_J superfamily. This family contains proteins similar to class II fumarse, espartage, aden/josuccinate lyase (ASL), argininosuccinate lyase (ASL), and 3- carboxy-cis.cis.microante lactorizing (OMLE) enzyme. Proteins of this		2.51E-33	47%	61%	1134	1–131	1	AAD33374 pyrimidine dimer-specific glycosylase	162.54	3.69E-39	59%	70%	1134	1133
					2	cd01334	4 family for the most part catalyze similar beta-elimination reactions which a C. Nor 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%	12110	56137	2	AAC86418 PBCV-1 pyrimidine dimer-specific glycosylase	162.16	4.82E-39	59%	70%	1134	1133

Gene Name	Genome Position	A.A. lenath	Peptide Mw	pl	CDD Hit Number	COGs	COG Definition GFAT, Glutamine amidotransferases class-II (Gn-AT)_GFAT-type. This domain is found at the N-terminus of glucosamine 6-phosphate (GicN-6-	Bit Score	E-value	% Identitv	% Positive		Hit from- to	BLASTp Hi Number		on BLASTp Definition	Bit Score	E-value	% dentitv P		Query from-to	Hit from- to
					3	cd00714	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-	27.79	4.86E-01	59%	76%	7289	10512	2 3	3 AAD3336	169 pyrimidine dimer-specific glycosylase	161.77	6.29E-39	58%	70%	1134	1133
							GicN-6-P.							4 5 6 7 8 9 10	5 AAD3333 5 AAD3338 7 AAD3338 3 AAD3338 9 AAD3338	77 pyrimidine dimer-specific dycosylase 79 pyrimidine dimer-specific dycosylase 80 pyrimidine dimer-specific dycosylase 81 pyrimidine dimer-specific dycosylase 91 pyrimidine dimer-specific dycosylase 92 pyrimidine dimer-specific dycosylase 97 pyrimidine dimer-specific dycosylase	161.00 160.61 160.61 160.61 160.23	1.40E-38 1.40E-38	58% 58% 58% 58% 58% 58%	70% 70% 72% 72% 72% 71% 72%	1134 1134 1134 1134 1134 1134 1134	1133 1133 1133 1133 1133 1133 1133
N631L	245296245018	93	9,582	10.67		No Hit Found								1	AAC9684	47 Gly-rich	94.74	9.44E-19	60%	83%	368	469
N633L	246022245321	234	26,212	4.58	1	COG5219	[General function prediction only].	38.58	6.29E-04	28%	41%	149198	1471152	5 1	AAC9684	48 A481L	207.22	3.55E-52	44%	65%	1228	1224
					2	COG5540 cd00162	modification. protein turrover. chaperoneal. RING, RING-GrineGrine (Realt) Interesting New Gene) domain, a specialized type of Zn-finger of 40 to 60 residues that binds two atoms of zinc; defined by the daposcross-brace' molf C-XC=XC;43;90;C-X(1-3); H-X(2-3);NIC(H):X2=C:X(4-49)C:X2=C; probably involved in mediating protein-protein interactions; identified in a proteins with a wide range of	38.15 37.76	9.29E-04 1.08E-03	30%	51%		32537			82 P0551A11.21 00 ring-H2 zinc finger protein-like		3.60E-04 3.60E-04	45%	60% 60%	146-195	143192 293342
							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).								405050		10.01	1.05E-03		57%		
														4	5 XP 50503	<ul> <li>Inc finger, C3HC4 type family protein</li> <li>hypothetical protein</li> <li>unnamed protein product</li> </ul>	45.44	1.05E-03 1.79E-03 2.33E-03	44% 32% 32%	57% 47% 46%	143-193 149-203 139-204	3888 345399 181251
														7	7 XP 46010	105 hypothetical protein DEHA0E19426g 125 ubiquitin-protein ligase/ zinc ion binding	43.90	5.20E-03 5.20E-03	36%	46%	149204	1614-1673 1047-1138
														9	XP_55036	64 putative RING finger 1 141 Putative RING-H2 finger protein ATL4H	43.90	5.20E-03 5.20E-03	44% 28%	52% 51%	147-195 144-231	128-176 175-266
							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															
N635L	247184246084	367	41,784	10.77	1	cd00233	a classs of binary toxins and regulators of biological pathways distinct from classical A-B toxins. A novel family of insecticidal ADP-	30.29	3.84E-01	31%	56%	298-344	13818	3 1	I AAC9646	61 similar to Chlamydia histone-like protein, corresponds to GenBank	179.49	1.71E-43	41%	61%	1229	46296
							mbog/transferses were isolated from Bacillus cereus during vegetative growth, where VIPI lakely targets transfer due doel and VIP2 attransfer domponents VIP2 attrans significant segurance similarity with enzymatic components toxin. C priodrem toxin. C proformer boxin and C difficiel toxin. ABC MutS2 MutS2 homotogs in bacteria and eukaryotes. The MutS protein initiates DNA mismatch repair by recognizing mispared and unpaired bases embedded in duplex DNA and activating endo- and exonucleases to DNA mismatch repair by recognizing mispared and possess a conserved ATPase activity that belongs to the ATP binding in mest prokervice and all eukaryotic constinues examined. Prokarvotes									<ul> <li>Accession Number D71563</li> </ul>						
					2	cd03280	In most protection and an extending of tighter that the extending of the transmission	29.41	7.16E-01	29%	47%	135–181	13417		2 AAC964	59 a91L similar to Chlorella virus PBCV-1 ORF A282L, corresponds to GenBank		6.11E-25	49%		241–364	1126
														3		Accession Number U42580 sto similar to PBCV-1 ORF A24R, corresponds to GenBank Accession		2.58E-07	56% 51%	68% 68%	361 360	260 472537
														4		Number U17055 smilar to bovine cylicin I, corresponds to Swiss-Prot Accession Number		2.18E-06 4.86E-06	53%	68%	660	472-537
														-		P35662						
N639R	247301247579		10,840 24,309	4.22	1	COG3966 No Hit Found	DID, 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%	2889	25832	۱ ۱ ۱	I AAC968	120 A452L 149 A482R		4.55E-05	38%	55% 66%	1191 3214	379
			,											2	2 AAR268	89 FirtV-1-B14 70 FirtV-1-I5	65.47	1.42E-09 3.17E-09	25% 30%	41% 47%	25190 33130	20216 31146
														4	YP_29419 NP 07758	96 hypothetical protein EhV_438 i81 EsV-1-96	63.93 55.45	4.14E-09 1.47E-06	34% 31%	50% 46%	33130 37130	73168 70181
														6	6 ABF820	156 hypothetical protein MIV026R		7.56E-03	28%	56%	61138	128204
N642L	248765248256	170	19,947	7.55		No Hit Found	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							1	I AAC968	51 A484L	160.61	1.77E-38	51%	68%	14167	1154
N643R	248772250463	564	61,365	10.37	1	cd01820	independent phospholipase A2 which exhibits strong substrate specificity towards PAF, hydrolyzing an acoevel teet at the ser 2 position. PAF-A4 also degrades a family of oxidized PAF-like phospholipids with shot sn-2 residues. In a didlion, PAF and PAF-A4 are associated with neural minration and mammalian remodurition salateacoevelsetares_like_ sublamily	66.43	7.92E-12	27%	42%	254-412	3521	) 1	I AAC9650	08 PBCV-1 surface protein	58.92	6.03E-07	43%	62%	499–560	9671028
					2	cd01828	of the SGNH-hydrolases, a diverse family of lipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alphabeta hydrolase family and unique among all known hydrolases (fa active site closely resembles the Ser-His-Asp(Glu) triad found in other snina hydrolase. SGNH-hydrolase subfamily, similar to Ruminococcus	59.53	1.04E-09	24%	42%	254-393	215	5 2	2 BAD228	150 surface protein	58.92	6.03E-07	43%	62%	499560	9671028
					3	cd01833	Bavefaciens XynB. Most likely a secreted hydrolase with xylanase activity. SONH hydrolases are a diverse family of lipases and esterases. The tertiary lot of the enzyme is substantially different from that of the alphabeta hydrolase family and unique among all known hydrolases; its active site closed reservities are Fil-Fils-RpG(II) und found in other servine hydrolase, SGNH_hydrolase, or GDSL_hydrolase, is a diverse	47.64	3.59E-06	18%	29%	254-399	319	) 3	3 BAD122	136 surface protein	59.69	3.54E-07	43%	62%	499–560	9671028
					4	cd00229	family of lipases and esterases. The tertiary told of the enzyme is substantially different from that of the alphabeta hydrolase family and unique among all known hydrolases; its active site closely resembles the typical Ser-His-Asp(Slu) triad from other serine hydrolases, but may lack the carthoxilic acid	44.06	5.21E-05	19%	31%	253-400	1—19	2 4	4 T176	136 proline-rich protein A145R - Chlorella virus PBCV-1	58.92	6.03E-07	43%	62%	499–560	176–237
					5	cd01838	IsoamyLacetate_hydrolase_like, IsoamyLacetate hydrolyzing esterate- like proteins. SSOH hydrolase sublemily similar to the Saccharomyce- cerevises IAHT. IAHT may be the major esterate that hydrolyses devise family of piases and esterates. The tertary fork of the enzyme is substantially different from that of the alphabeta hydrolase family and unique among all known hydrolases; it sactive site loosely resembles the Ser-His-Asy(Glu) triad found in other serine hydrolases.	40.24	6.48E-04	20%	36%	254-408	219	3 5	5 BAE028	30 surface protein	48.52	8.15E-04	32%	63%	503–560	1014–1071

Gene	Genome Position	A.A. lenath	Peptide Mw	pl	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity F	% Positive	Query I from-to	lit from-BLA		Hit BLASTp Defin	nition	Bit Score E-v		% ntitv Po	%	Query I	Hit from- to
Name	FUSICION	lendun	WW		Number		NnaC_like, NnaC (CMP-NeuNAc synthetase) _like subfamily of SGNH_hydrolases, a diverse family of lipases and esterases. The tertiary				ositive	110111-10	10 N	umber	ACCESSION			lue	nutv PO	SILIVE	1011-10	10
					6	cd01841	fold of the enzyme is substantially different from that of the alphabeta 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 invalued in entheraptic authorize	40.25	7.22E-04	28%	46%	304-400	47160	6	AAH59352 MGC69176 protein		51.99 7.3	7E-05	27%	44%	254414	42227
					7	cd01835	involved in notesancharide swithesis SGNH_hydrolase like 3, SGNH_hydrolase subfamily. SGNH hydrolases are a diverse family of ipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alphabeta hydrolase family and unique among all known hydrolases, its active site	39.18	1.42E-03	20%	32%	254-393	2185	7	AAH84968 LOC495446 protein		48.52 8.1	5E-04	28%	46%	254400	42205
					8	cd01822	closely resembles the Ser-His-Asp(Glu) triad found in other serine hydrolaxes Lysophospholipase_L1_like, Lysophospholipase L1-like subgroup of SGNH-hydrolaxes. The best characterized member in this family is TesA, an E. coli periplasmic protein with thiosetrase, esterase, arytestrase,	35.11	2.15E-02	21%	38%	252-408	1176	8	BAC56453 similar to platelet-activating factor acetylhy	ydrolase, beta subunit	47.75 1.3	9E-03	31%	51%	254354	42145
							protease and Ivsophospholipase activity 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															
					9	cd01832	unique among all known hydrolases; its active site closely resembles the Ser-Hie-AsyClub irtiad from other serine hydrolases, but may lack the carboxic acid. Myxobacterial members of this subfamily have been reported to be involved in adventurous gliding motility SGNH_hydrolase_linke_2. SGNH_hydrolase subfamily. SGNH	34.86	2.57E-02	24%	35%	258-407	6185	9	ZP_00524236 Lipolytic enzyme, G-D-S-L		45.82 5.2	8E-03	28%	44%	252398	511-673
					10	cd01834	hydrolases are a diverse family of ipases and estimases. The tertiary (bid of the enzyme is substantially different from that of the alphabeta hydrolases family and unique among all known hydrolases; its active site loosely resembles the Ser-His-Asp(Glu) triad found in other serine hydrolases.	34.92	3.06E-02	20%	35%	254-369	2134	10	NP_071782 platelet-activating factor acetylhydrolase a	lpha 2 subunit	45.05 9.0	1E-03	30%	51%	254351	42142
N649R	250735251490	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 olvcoenclide resistance cluster in Enterococcus faectum BM4339.	28.79	6.49E-01	38%	62%	147–173	182208	1	AAC96377 A9R		248.83 1.2	3E-64	67%	84%	79249	3173
					2	ofam01200	Members of this CD are predominantly bacterial in origin Ubie_methyltran, ubiE/COQ5 methyltransferase family	28.74	7.61E-01	34%	55%	144-175	171200	2	AAC96643 similar to PBCV-1 ORF A79R, corresp Number U17055	onds to GenBank Accession	93.59 6.6	1E 19	29%	48%	3247	5248
					2	planto1205		20.74	7.012-01	3476	3378	144-175	171-200	3	AAC96447 A79R		90 12 7 3	1E-17	28%	47%	5226	4217
														4	AAC96818 similar to PBCV-1 ORF A275R, encoded b U42580 AAC96972 similar to Chlorella virus PBCV-1 ORF A4				28%	47%	5248	4246
														5	ACCession Number 042580 AAC96545 A177R	our , concepting to concentrative	85.11 2.3	5E-15	29% 27%	46% 49%	5248 3233	10253 5227
														7	AAU06304 hypothetical protein A275R AAU06301 hypothetical protein A275R		63.93 5.6 62.77 1.2	5E-08	28% 28%	49% 49%	83247 83247	3167 3167
														9 10	AAC96734 A366L AAU06302 hypothetical protein A275R		55.07 2.6 51.99 2.2		26% 28%	47% 50%	3144 116248	10155 1132
							RNRR2, Ribonucleotide Reductase, R2beta subunit (RNRR2) is a member of a broad superfamily of ferritin-like dirion-carboxylate proteins. The RNR protein catalyzes the conversion of ribonucleotides to deoxyrbonucleotides and is found in all eukaryotes, many prokaryotes, several viruses, and few archaea. The catalytically acties from of RNR is a proposed alpha2-beta2 tetramer. The homodimeric alpha subunit (R1) contains the active site and redox, active crysteines as well as the															
N651L	252530251481	350	39,793	4.54	1	cd01049	allosteric binding sites. The beta subunit (R2) contains a dirion duster that, in its reduced state, reacts with discogen to form a stable tyrosyi radical and a diron(III) cluster. This essential tyrosyi radical is proposed to generate a thij radical, located on a cysteine resolution in the R1 active set 0.14 helices, the discogen blucker. The beta subusit is composed on the stable of the stable of the stable of the stable of the some have 2 addition beta strandow. Yeast is unique in that it assembles	283.29	2.96E-77	49%	64%	43315	4284	1	AAC96643 contains ribonucleotide reductase (RR) si small subunit, corresponds to Swiss-Prot /	gnature; similar to tobacco RR Accession Number P49730	506.52 5.66	E-142	75%	84%	27348	3324
					2		both homodimers and haterodimers of RNRP2. The usast heterodimer Ribonuc red sm, Ribonucleotide reductase, small chain. NrdF, Ribonucleotide reductase, beta subunit [Nucleotide transport and	274.83	1.00E-74	44%	63%	29311	1285		NP_189342 ribonucleoside-diphosphate reductase		369.01 1.41		54%	72%	23348	1332
					3	COG0208	metaholism! Ferrifin like, Ferrifin-like, diron-carboxylate proteins participate in a range of functions including iron regulation, mono-oxygenation, and reactive raicida production. These proteins are characterized by the fact that they catalyze dioxygen-dependent oxidation-hydroxylation reactions within diron centers; one exception is manganese catalase, which catalyzes peroxide-dependent oxidation-reduction within a dimanganese center. Diron-oxaboxylate proteins are further characterized by the	226.71	2.66E-60	32%	48%	43348	30348	3	AAO62422 ribonucleotide reductase small subunit		366.31 9.15	E-100	53%	71%	23348	1333
					4	cd00657	presence of duplicate metal ligands, glutamates and históries (CxH) and two additional glutamates within 6 tor-heits bundle. Cubise of these conserved residues there is tiltle obvious homology. Members include backnofertini, fuertyrhin, ramati can ad kare monocoygenese hydroxylases (AAH), flocmuleotide reductase R2 (RNR2), scyt-ACP- desaturase (AA), ACP-Desat), manganese (Min), cataliases, derpli oxyubiational hydroxyla, (AAC), INAA, postening at the Fe- cretarion at the advector (AA), (AAC). INAA postening the Fe- cretarion at them of the aarobic originas secteral (ACP). In the formitian's	48.56	1.04E-06	19%	29%	91237	2140	4	CAF24073 probable ribonucleoside-diphosphate redu	uctase small chain	363.23 7.7	4E-99	56%	73%	29–348	5323
					5	COG3933	COG3833, Transcriptional antiferminator (Transcription).	29.08	7.65E-01	21%	38%	64135	405468	5 6 7 8 9	BAD46317 putative ribonucleotide reductase R2 XP_550581 putative ribonucleotide reductase R2 CAA63194 ribonucleotide reductase (Class I) XP 829341 ribonucleotide reductase (Class I) XP 829341 ribonucleoside-diphosphate reductase O15910	small chain (Ribonucleotide	355.91 1.2 354.76 2.7 354.76 2.7 354.76 2.7 353.21 8.0 353.21 8.0	5E-96 5E-96 5E-96 1E-96	53% 53% 54% 54% 54%	70% 71% 69% 70% 70%	28348 28348 23348 28348 28348 28348	14339 20345 1329 16337 16337 16337
N655L	253103252582	174	20,085	9.95	i	No Hit Found								1	ACO6838 A471R YP 142861 unknown ZP 01257639 hypothetical protein P700755 31322	reductase R2 subunit)	207.99 1.0 108.23 1.1 89.35 5.3	2E-52 0E-22	52% 38% 34%	77% 62% 57%	2172 16155 49172	3173 43191 1135
N658L	253738253139	200	22,208	4.57		No Hit Found								4 5 1	AAR26829 FirrV-1-A5 NP_077626 EsV-1-141 AAC96837 A470R		87.81 1.5 71.63 1.1 239.97 3.5	5E-11 5E-62	31% 31% 59%	61% 59% 72%	2140 3119 1197	14153 20137 1202
							NUD170 Nuclear new complex a burn fragment							2	BAB69883 UKCH-2		90.89 2.6		39%	58%	6127	49170
N662L	255068253776	431	48,157	8.69	1	COG5308	NUP170, Nuclear pore complex subunit [Intracellular trafficking and secretion].	31.19	2.77E-01	23%	41%	34117	1291216	1 2 3	AAC96835 A468R AAR26870 FirrV-1-A46 NP_077530 EsV-1-45		429.48 1.19 47.75 9.7 51.99 5.1	3E-04	54% 33% 22%	72% 56% 40%	3377 308372 180387	4393 490552 353542
N665R	255164256090	309	35,071	6.79		No Hit Found	Furt Air Frut / Air family. Biogenetic of FAIR obvious involves a							1 2 3	AAC96834 A467L NP_065022 Hypothetical protein AAP99966 ABC-type antimicrobial peptide transport s	system permease component	397.51 3.07 65.08 3.6 43.90 8.6	1E-09	62% 29% 23%	80% 46% 40%	4300 48221 44243	6302 17218 80271
N668R	256117256470	118	13,165	10.47	1	pfam04777	Evr1_A/r. Evr1 / A/r family. Biogenesis of Fe/S clusters involves a number of essential mitochondrial proteins. Evr1 pol Saccharomyces cerevisiae mitochondria is required for the maturation of Fe/S proteins in the cytosol. The ALR (aggement of liver regeneration) represents a are located in the mitochondrial intermembrane and at thought to operate drexentsman of the mitochondrial ARC transnoter.	78.03	2.81E-16	40%	52%	18110	2-90	1	AAC96832 PBCV-1 thiol oxidoreductase		177.95 8.6	0E-44	67%	80%	2116	4118
					2	COG5054	Devisition of the minor control ARL transporter ERV1, Mitochondrial sulfhydryl oxidase involved in the biogenesis of cytosolic Fe/S proteins [Posttranslational modification, protein turnover,	55.83	1.39E-09	30%	45%	1110	70175	2	ZP_01257237 hypothetical protein P700755_26302		77.41 1.5	8E-13	36%	56%	1197	33119
							chaoerones].								YP_142722 putative thiol oxidoreductase		67.40 1.6	4E-10	36%	51%	1097	390

Gene Name	Genome Position	A.A. lenath	Peptide Mw		CDD Hit Number	COGs COG Definition	Bit Score	E-value	% Identitv F	% Positive		Hit from- BL to P		Accession DECGT Definition NP_078699 Thiol oxidoreductase XP_503294 hypothetical protein XP_451806 unnamed protein product	62.00 60.08	6.88E-09 6.88E-09 2.61E-08	% lentitv P 33% 33% 31%	55% 50% 50%	from-to 299 9111 14111	Hit from- to 21115 88186 85178
													8	CAA48192 ERV1 Flavin-linked sulfhydryl oxidase localized to the mitochondrial NP_011543 intermembrane space, has a role in the maturation of cytosolic iron-sulfur oroteins: ortholog of human hepatopoietin (ALR): Erv1o AAL98767 ORF043.	58.54	7.60E-08	31% 31% 34%	48% 48% 52%	14110 14110 799	19111 91183
													10	AAX82354 thiol oxidoreductase	56.23	2.89E-07 3.77E-07	34%	52%	799	596 596
N670L	257268256465	268	29,615	6.63	1 2	COG0571 Rnc, dsRNA-specific ribonuclease [Transcription]. smart0055 RiBCo, Ribonuclease III family: Ribonuclease III Cerminal domain. This group consists of eukayotic, bacterial and archeal ribonuclease III (RNAse III) proteins RNAse III is a double stranded RNA-specific endonuclease Protayoti RNAse III is in portant in post-transcriptional control of mRNA stability and translational efficiency. It is imovide in the processing of ribosoma	s. c y	3.27E-44 1.46E-31	35% 42%	51% 58%	38263 58187	4234 2130	1	AAC96831 similar to Bacillus ribonuclease III, corresponds to Swiss-Prof Accession NMMbr P51833 YP_445467 ribonuclease III		1.27E-94 6.43E-30	66% 35%	81% 53%	16265 33258	25–274 9–241
					3	and transactional endertry. It is involved in the processing of nocostine cotops3 of RNA precursors. Protocy for RNAe III also plays a role in the maturation transcripts. Eukaryotic RNAe III also plays a role in the maturation transcripts. Eukaryotic RNAes III also plays a role in through direc deavage) in rRNA processing, in processing of small nucleater RNA (snoRNAs) and snRNA&poss components of the splicesome). In eukaryotes RNAse III or RNAesell like enzymes such as Dioer are involved in RNAi (RNA inteference) and miRNA (micro-RNA) gen	n 127.32 ct s n e	1.54E-30	48%	62%	70181	16127	3	ZP_00590199 Ribonuclease III	128.26	2.70E-28	36%	54%	45255	47265
					4	eilanvinn pfam00636 flibonuclease 3, RNase3 domain DSRM, Double-stranded RNA binding motif. Binding is not sequence		2.35E-29	56%	70%	76166	191	4	ZP_00532592 Ribonuclease III	127.49	4.61E-28	33%	53%	16255	21259
					5	specific but is highly specific for double stranded RNA. Found in a variety cd00048 of proteins including dsRNA dependent protein kinase PKR, RNA helicases, Drosophila staufen protein, E. coli RNase III, RNases H1, and	ς 58.84	6.97E-10	31%	52%	192-258	1-68	5	ABB23018 Ribonuclease III	122.87	1.14E-26	35%	55%	58255	47252
					6	dsRNA denendent adenosine deaminases COG1939 COG1939, Uncharacterized protein conserved in bacteria [Function unknown].	n 34.87	9.96E-03	29%	46%	78163	18112	6	ZP_00591208 Ribonuclease III	122.48	1.48E-26	34%	52%	45256	44266
					7	DUF880, Protein of unknown function (DUF880). This family consists pfam05948 a number of hypothetical bacterial and plant proteins. The family also contains the C terminal region of a Cysteinyl-RNA synthetase from	0 35.17	9.97E-03	26%	43%	78167	11111	7	ZP_00528534 Ribonuclease III	122.48	1.48E-26	33%	54%	41255	58281
					8	Stabhylococcus eoidermidis. The function of this family is unknown DOP1, Dopey and related predicted leucine zipper transcription factors Transcription	s 34.23	1.72E-02	20%	36%	58150	556652		ZP_00511103 Ribonuclease III ZP_00661649 Ribonuclease III		3.30E-26 3.30E-26	34%	54% 50%	45257	42264 19254
													10	AAM73335 ribonuclease III		3.30E-26 9.62E-26	32% 32%	50% 52%	30262	15260
N672L	257494257279 257663259627	72 655	7,981 75,058	9.71 5.23	1	COG2433 COG2433, Uncharacterized conserved protein [Function unknown]. COG3378 COG3378, Predicted ATPase [General function prediction only].	28.34 70.01	2.34E-01 7.92E-13	18% 24%	55% 37%	1559 281593	421465 146445	0	No Hit Found No Hit Found AAC96824 contains ATP/GTP-binding site motif A	971.00	0.00E+00	65%	81%	20648	22650
N074R	25/06325962/	655	75,056	5.23	2	Pox_D5, Poxvirus D5 protein-like. This family includes D5 from pfam03288 Poxviruses which is necessary for viral DNA replication, and is a nucleic	n ° 63.73	6.04E-11	24%	39%	201-595	3313		NP_077594 EsV-1-109		5.80E-73	35%	54%	190-607	141-565
					L	pleino200 acid independent nucleoside triphosphatase. Members of this family are also found outside of ocxviruses.	e 03.73	0.042-11	22 /0	3878	250-014	5-515	3 4 5 6 7 8 9	AA26002 FirV-1-827 YP 294217 Datke nucleic acid independent nucleoside triphosphatase 27_012428 COG3378: Fredicted ATPase BAE50402 Joulave INA primase-hage associated AA272348 phage pla-173 primase-like protein 27_0603376 Phage/plasmid primase PL - Cleminal YP 594040 Phage/plasmid primase PL - Clemin	268.09 166.39 65.86 61.62 61.23 60.46 58.54	7.83E-70 3.21E-39 5.90E-09 1.11E-07	29% 25% 28% 24% 22% 23% 30%	51% 44% 48% 42% 42% 42% 53%	35-591 187-641 338-523 292-566 265-523 232-506 321-613 376-517	6-547 168-672 310-495 107-366 213-458 241-499 307-592 439-581
N679R	259658260536	293	31,095	5.47		No Hit Found							1	AAC96822 A454L		1.03E-95	64%	77%	20290	20288
N682L	261298260618	227	26,441	7.60	1	pfam01108 Tissue fac, Tissue factor	28.90	5.70E-01	25%	46%	144-223	89176	1			2.88E-11	26%	53%	41194	29173
N684L N685R	261677261282 261742262269	132	14,749 19,819	7.24	1	No Hit Found COG5098, Chromosome condensation complex Condensin, subunit D: COG5098 [Chromatin structure and dynamics / Cell division and chromosom	2 e 30.44	1.17E-01	21%	44%	12135	333444	1	AAC96924 A570L AAC96926 A572R		5.46E-43 1.75E-63	64% 65%	74%	1128 4163	1121
N687L	263042262260	261	28,789	4.23	1	nartitioninni PCNA, Proliferating Cell Nuclear Antigen (PCNA) domain found ir eukaryotes and archaea. These polymerase processivity factors play, role in DNA replication and repar. PCNA encircles duplex DNA in its central cavity, providing 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 pocessivity factors to tractine (bet cotof)? suburit DNA polymerase pocessivity factors to tacking (bet cotof)? suburit DNA polymerase and the cotomer polymerase and the cotomer pokaryotic DNA replication that has been suggested on biochemic opticas. PCNA is also involved with proteins involved in cell cycle processes such as DNA repair and apoptosis. Many of these protein contrain a highly conserved motif known as the PTP-box (PCNA interacting protein box) which contains the sequence Qxc[LMI]xvF[FV]. PCNA_C, Proteining cell nuclear antigen, Clemming domain.	n a s c c a d d a a a a a a s s A A	3.31E-35	27%	51%	21–251	5247	1	AAC96627 similar to Periwinkle PCNA, corresponds to GenBank Accession Number X55052	231.49	2.18E-59	44%	67%	4259	4264
					2	pfam02747 terminal and C-terminal domains of PCNA are topologically identical Three PCNA molecules are tightly associated to form a closed ring		4.87E-12	28%	49%	135-250	3128	2	CAE67843 Hypothetical protein CBG13430	125.56	1.68E-27	29%	50%	15257	2260
					3	encirclina duolex DNA COG0592 DnaN, DNA polymerase sliding clamp subunit (PCNA homolog) [DNA reolication. recombination. and recairl. PCNA_N, Proliferating cell nuclear antigen, N-terminal domain. N		2.92E-10	18%	37%	1251	47323	3	EAR82594 proliferating cell nuclear antigen (pcna)	117.86	3.51E-25	28%	48%	15255	2259
					4	pfam00705 terminal and C-terminal domains of PCNA are topologically identical pfam00705 terminal and C-terminal domains of PCNA retopologically identical Three PCNA molecules are tightly associated to form a closed ring encircling duplex DNA.	l. 49.88	3.27E-07	25%	55%	25112	12100	4	CAA38636 proliferating cell nuclear antigen	117.09	5.99E-25	30%	49%	20257	7260
					5	pfam04139 Rad9, Rad9. Rad9 is required for transient cell-cycle arrests and transcriptional induction of DNA repair in response to DNA damage		1.67E-02	22% 17%	47%	1573	361	5			5.99E-25 2 28E-24	30%	51%	48256 25258	1226
					6	COG1355 COG1355, Predicted dioxygenase [General function prediction only].	30.25	2.93E-01	17%	44%	21153	110227	6 7 8 9 10	XP_502661 hypothetical protein ORMX3P yrolfferating cell nuclear antigen (PCNA) XP_743255 proliferating cell nuclear antigen CAA38895 proliferating cell nuclear antigen AAG24008 proliferating cell nuclear antigen	115.16 112.85 112.46	2.28E-24	25% 29% 29% 30% 30%	52% 50% 51% 49% 49%	25258 25257 21251 25258 25257	12260 12261 7261 12262 12261
N689L	263595263089	169	19,586	10.14	1	LdhA, Lactate dehydrogenase and related dehydrogenases [Energy COG1052 production and conversion / Coenzyme metabolism / General function prediction onlv1.	y n 29.44	2.57E-01	22%	41%	97148	93147	1	AAC96928 A575L	112.46	5.42E-24	35%	66%	31166	30167
N690R	263679264704	342	37,532	8.60	1	smart00494 ChtBD2, Chitin-binding domain type 2; . Chitin_bind_3, Chitin binding domain. This domain is found associated	38.96	8.41E-04	40%	50%	278321	949	1	AAC96701 PLPRNLLL (4X), SPPPSKP (3X)	343.58	6.11E-93	70%	80%	7220	1213
					2	pfam03087 with a wide variety of cellulose binding domain. This domain is to domain to baculovir spheroidins and spindolins, protein of unknown function. CBM 14, Chithi Inding Petritophin-A domain. This domain is called th Petritophin-A domain and is found in chitin binding proteins particularly petritophin-A domain and is found in chitin binding proteins particularly petritophin-A domain and is found in chitin binding proteins particularly petritophin-A domain and is found in chitin binding proteins particularly petritophin-A domain and is found in chitin binding proteins particularly petritophin-A domain and is found in chitin binding proteins particularly petritophin and proteins of the set and a manual chitasses. Coxies of the	a 38.43 e y	1.20E-03		29%	18216	1211		AAC96700 a332L		4.63E-16	74%	90%	280322	143
					3	pfam01607 domain are also found in some baculoviruses. Relevant references that describe proteins with this domain include. It is an extracellular domain that contains is conserved cystelines that probably form three disulphid bridges. Chitin binding has been demonstrated for a protein containing only two nf threes dromains	n e g	4.88E-03	41%	62%	290321	1847	3	EAA01148 ENSANGP00000018413	74.71	5.30E-12	27%	41%	25217	4202
					4	COG3397, Uncharacterized protein conserved in bacteria [Function unknown].	n 35.41	1.19E-02	22%	33%	11222	8213	4	XP_966436 PREDICTED: similar to CG4367-PA isoform 1	72.79	2.01E-11	33%	41%	15217	4209

Gene Name	Genome Position	A.A. Ienath	Peptide Mw	pl	CDD Hit Number	COGs COG Definition	Bit Score	E-value	% Identitv P	% Positive		Hit from-BL to N		Hit         BLASTp Definition           EAA09469         ENSANGP00000012300           XP 973033         PREDICTEC to Jamilar to CG15786-PA           EAL20006 GN1833-PA         EAL30006 GN1833-PA	68.55	E-value 1.70E-10 3.80E-10 3.80E-10	% dentitv P 27% 27% 29%	% ositive 42% 40% 44%	Query from-to 33218 23221 18216	Hit from- to 1196 23231 6210
													8 9 10	XP_968569 PREDICTED: similar to CG4367-PA AAM50982 RE24790p NP_572209 CG15786-PA	66.63	1.10E-09 1.44E-09 1.44E-09	26% 29% 29%	38% 42% 42%	30217 20218 20218	22216 38249 19230
N693R	264791265324	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%	70102	499528	1	AAC96888 contains Gln-rich, neutral zinc metallopeptidase, zinc binding region signature	211.85	7.65E-54	62%	74%	1175	212-391
N696R	265432266775			5.31		No Hit Found							1	AAC96872 A505L		8.99E-125	56%	71%	48443	73478
N703R	266856267593			5.85		No Hit Found							1	AAC96870 contains prokaryotic membrane lipoprotein lipid attachment site		3.54E-45	37%	63%	6243	38297
N705R	267553267879	109	12,706	7.77	1	pfam00742 Homoserine_dh, Homoserine dehydrogenase	27.13	4.81E-01	27%	37%	43103	76143	1	AAC96869 A502L AAC96722 A354R		5.42E-22	55%	65%	21107	395
N706R	267948269024	359	40,925	9.95		No Hit Found							1	AAC96722 A354R AAC96790 similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081		4.63E-54 1.97E-12	45% 29%	63% 46%	117356 67324	2235 59321
						ANK, ankyrin repeats; ankyrin repeats mediate protein-protein interactions in very diverse families of proteins. The number of ANK							3	AAC96455 A87R	61.62	5.02E-08	33%	46%	197–353	289-447
N710R	269070270233	388	44,149	6.82		repeats in a protein can range from 2 to over 20 (ankyrins, for example). cd00204 ANK repeats may occur in combinations with other types of domains. The subcutural repeat unit contains two antiparallel helices and a beta-haipin, repeats are stacked in a superhelical arrangement; this alignment contains 4 consecutive receats.	63.56	3.94E-11	29%	54%	28151	6125	1	AAC96943 contains 3 ankyrin repeat-like elements; similar to Drosophila ankyrin, corresponds to GenBank Accession Number L35601		1.32E-49	46%	64%	4219	3217
					2	COG0666 Arp, FOG: Ankyrin repeat [General function prediction only]. Ank, Ankyrin repeat. There's no clear separation between noise	39.49	7.80E-04	17%	36%	35218	44234	2	AAC96944 A608R	86.66	1.63E-15	33%	52%	238368	1134
					3	pfam00023 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	221	3	XP_786577 PREDICTED: similar to CG6599-PA	60.08	1.63E-07	24%	44%	1257	134368
						associate to form a money order suddure							4 5	XP_967612 PREDICTED: similar to ankyrin repeat domain 28 XP_628948 BTB/POZ domain-containing protein	55.84	2.44E-03 3.08E-06	25% 24%	43% 42%	67257 33264	224445 39269
													6 7	XP_650635 ankyrin repeat protein NP 039208 ORF FPV245 Ankyrin repeat gene family protein	54.68	2.44E-03 6.86E-06	24% 23%	40% 42%	33257 48285	36288 50323
													8.9	ZP 01301013 hypothetical protein RqrvI 01000336 XP_547958 PREDICTED: similar to ankyrin repeat and SOCS box-containing protein	54.68 53.91	6.86E-06 1.17E-05	29% 31%	46% 46%	66217 29176	433-568 185-329
													10	Z isoform 1     ZP_854471     PREDICTED: similar to ankyrin repeat and SOCS box-containing protein     2 (oredicted) isoform 2		1.17E-05	31%	46%	29176	233-377
N712L	271396270224	391	43,171	5.86	1	COG1004 Ugd, Predicted UDP-glucose 6-dehydrogenase [Cell envelope	353.34	2.22E-98	36%	55%	4375	2392			462.00	8.46E-129	58%	77%	5386	4386
N/ IZL	2/13902/0224	391	43,171	5.60	2	biogenesis. outer membrane1. COG0677 WecC, UDP-N-acetyl-D-mannosaminuronate dehydrogenase [Cell	125.73	2.22E-98 8.51E-30	28%	44%	4375	8392	2	AAC96945 PBCV-1 UDP-glucose dehydrogenase AAK02860 unknown		1.83E-115	53%	73%	3389	4389
					3	Person and Construction of the provided and the provid	121.49	1.47E-28	32%	48%	3176	1186	3	AAC67251 UDP-glucose dehydrogenase		4.08E-115	53%	73%	3389	2389
					4	acid without the release of an aldehvude intermediate UDPG_MGDP_dh, UDP-gluccese/GDP-mannose dehydrogenase family, central domain. The UDP-gluccese/GDP-mannose dehydrogenaseses are pfam00984 a small group of enzymes which posseses the ability to catalyse the	91.02	2.26E-19	39%	55%	193–285	2-96	4	AAK17922 UDP-glucose dehydrogenase FcbC	410.22 e	6.51E-113	52%	73%	3389	2389
					5	NAD-dependent 2-fold oxidation of an alcohol to an acid without the release of an aldehvde intermediate COG1893 ApbA, Ketopantoate reductase [Coenzyme metabolism].	46.08	7.36E-06	15%	37%	4249	2233	5	AAK17904 UDP-glucose dehydrogenase DcbC	405.60 1	1.60E-111	53%	71%	3389	2389
					6	COG0240 GpsA, Glycerol-3-phosphate dehydrogenase [Energy production and conversion].	44.77	2.15E-05	32%	57%	387	2-83	6	CAG21035 putative UDP-glucose dehydrogenase	404.06 4	4.67E-111	51%	73%	5389	3388
					7 8	COG2084 MmsB, 3-hydroxyisobutyrate dehydrogenase and related beta- hydroxyacid dehydrocenases ILioid metabolisml. COG1250 FadB, 3-hydroxyacyl-CoA dehydrogenase [Lipid metabolism]. UDPG_MGDP_dh_C, UDP-glucose/CDP-mannose dehydrogenase	41.36 38.70	2.14E-04 1.20E-03	28% 24%	45% 46%	3232 188	1204 295		ZP_00851874 UDP-glucose 6-dehydrogenase ZP_01235021 Putative nucleotide sugar dehydrogenase		6.74E-110 1.50E-109	51% 50%	71% 71%	5389 5389	3388 3388
					9	family, UDP binding domain. The UDP-glucose/GDP-mannose pfam03720 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 aldehydra intermediate	37.96	2.16E-03	17%	38%	302383	187	9	ZP_00761397 COG1004: Predicted UDP-glucose 6-dehydrogenase	397.51 4	4.37E-109	50%	71%	5389	3388
					10	UbiH, 2-polyprenyl-6-methoxyphenol hydroxylase and related FAD- COG0654 dependent oxidoreductases [Coenzyme metabolism / Energy production and conversion].	36.98	3.89E-03	39%	68%	131	1-32	10	ZP_00880860 UDP-glucose 6-dehydrogenase	396.36 9	9.73E-109	51%	71%	5389	3388
N715L	273341271449	631	72,186	6.93	1	COG1215 COG1215. Glycosyltransferases, probably involved in cell wall biocensis (Cell enveloce biocenesis) cuter methyane. Cellulose, synt, Cellulose synthase. Cellulose, an aggregate of unbranched polymers of beta 1-4-finked 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	110.79	4.76E-25	19%	37%	81472	14402	1	AAC96590 similar to cellulose synthase catalytic subunit (UDP-forming)	892.11	0.00E+00	68%	81%	23629	68677
					2	synthesise cellulose in higher plants differ greatly from the well- plam0355C characterized genes found in Acobioater and Agrobactirum sp. More correctly designated as &apos.cellulose synthese catabytic subunitsAgors; plant cellulose synthase (cesA) profiles are integral membrane proteins, approximately 1,000 amino acids in length. There are a number of highly conserved residues; including several motifs	41.05	4.22E-04	24%	43%	287-492	499698	2	AAL44127 cellulose synthase	305.06	5.55E-81	41%	57%	107499	91478
					_	shrwm to be necessary for noncessive diverset/fransferase activity Glycos transf 2, Glycosy transferase. Diverse family, transferring sugar pfam00535 from UDP-glucose, UDP-N-acetyl- galactosamine, GDP-mannose or														
					3	CDP-abequose, to a range of substrates including cellulose, dolichol phoschate and teichoic acids. 7tm 5, 7TM chemoreceptor. This large family of proteins are related to pfam00001. They are 7 transmembrane receptors. This family does not	40.20	8.32E-04	15%	33%	131–298	10168	3	YP_467623 probable cellulose synthase protein	291.58	6.35E-77	34%	49%	17584	4558
					4	pfam01604 include all known members, as there are problems with overlapping specificity with pfam00001. This family is greatly expanded in the nematorie worm C elegans	30.90	4.80E-01	27%	44%	350398	105153	4	BAB54246 mlr7873 CAC48842 putative cellulose synthase protein		4.12E-76	38%	54%	118-556	112-546 97-528
													5 6 7 8 9 10	CAC48942 putative celluiose synthase protein 2P_0058843 celluiose synthase kike protein AAQ87082 Celluiose synthase celluiose synthase catalvic subunit AAX76085 givcosi transferase, group 2 family protein ABA7931 celluiose synthase-kike protein	283.49 279.64 278.87 277.33	4.55E-75 1.73E-74 2.50E-73 4.26E-73 1.24E-72 2.11E-72	38% 36% 40% 37% 37% 35%	53% 52% 56% 52% 53% 52%	115-551 118-582 118-496 118-560 115-548 118-584	97-528 102-561 166-538 102-539 93-516 110-568
N719L	273697273335	121	13,025	10.35	1	smart00317 SET, SET (Su(var)3-9, Enhancer-of-zeste, Trithorax) domain; Putative methvl transferase. based on outlier olant homolooues . SET, SET domain. SET domains are protein lysine methyltransferase	59.64	9.59E-11	29%	38%	6113	2125	1	AAC96946 PBCV-1 histone H3-Lys 27 methyltransferase (vSET)	140.20	1.96E-32	57%	73%	2114	1113
					2	enzymes. SET domains appear to be protein-protein interaction domains. It has been demonstrated that SET domains mediate interactions with a ptam00856 family of proteins that display similarity with dua-specificity phosphatases (down and the second second second second second second second second These domains are divergent in sequence from other SET domains, but also aopear to mediate orden-orden interaction.	47.37	4.54E-07	23%	38%	6112	8132	2	ZP_00588496 Nuclear protein SET	65.86	4.71E-10	34%	52%	7114	39150
					3	COG2940, Proteins containing SET domain [General function prediction	41.63	2.25E-05	22%	34%	1110	328452	3	ZP_00511449 Nuclear protein SET	60.85	1.51E-08	35%	50%	7114	38149
					4	INT_SG4, INT_SG4, DNA breaking-rejoining enzymes, cd01187 integrase/recombinases subgroup 4, N- and C-terminal domains. The CD	27.17	6.02E-01	22%	33%	3792	172227	4	ZP_00661322 Nuclear protein SET	60.08	2.58E-08	35%	51%	6117	37152
						contains mainly predicted bacterial integrase/recombinases for which not much biochemical characterization is available		== 51						ZP_00531791 Nuclear protein SET		7.52E-08	31%	49%	7116	39152
													6 7 8 9	ABS2383 AUXdear protein SET ABS2383 AUXdear protein SET NP_701503 hypothetical protein PFL0690c AAZ7150 hypothetical protein Maar A2600 AAM72187 conserved hypothetical protein	58.15 56.61 55.84	9.82E-08 9.82E-08 2.86E-07 4.87E-07 1.85E-06	31% 35% 30% 31% 31%	49% 51% 44% 51% 49%	6116 6114 3111 7106 7117	38150 29174 3109 39153

Gene Name	Genome Position	A.A. Ienath	Peptide Mw		CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identitv Po	% ositive		Hit from- B to		Accession		Bit Score E-v 53.53 2.43	alue Ide	% entitv Po	% sitive	Query from-to 16106	Hit from- to 19-116
N720L	275415273727	563	63,379	11.37	1	COG0515	SPS1, Serine/threonine protein kinase [General function prediction only / Signal transduction mechanisms / Transcription / DNA replication, recombination, and repairl. S TKc. Serine/Threonine protein kinases, catalytic domain	43.61	7.23E-05	13%	28%	84456	11365	1	AAC96947 RPQT-like (9x)		460.69 6.86	E-128	44%	62%	2561	6577
					2	cd00180	S_INC, Sennel Interonine 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	42.89	1.04E-04	30%	46%	53238	6151	2	AAC98646 similar to bovine cylicin I, corresponds to Swiss-Prot Acces P35662	sion Number	291.20 7.20	0E-77	39%	60%	1401	9424
					3	pfam00069	chances in the C-terminal autoreculatory tail Pkinase, Protein kinase domain	41.04	3.63E-04	28%	45%	53236	5148	3	AAC96650 similar to PBCV-1 ORF A34R, corresponds to GenBar Number U17055	nk Accession	285.03 5.1	6E-75	36%	56%	1453	23465
					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 bhosphotransferase and strontomycin 3' Banos-kinase or		4.29E-04	41%	74%	207-233	171198	4	XP_944436 PREDICTED: similar to CG10953-PA		62.39 5.4	4E-08	33%	55%	349446	76173
							phosphortansferase and streptomycin 3'&apos-kraise or streptomycin 3'&apos.posphortansferase. The aminopycoside phosphortansferases inactivate aminopycoside antibiotics via phosphortansferases inactivate aminopycoside antibiotics is related to fructivasmine kinase framt3881 COC2333, Ptatket homoserine kinase type II (protein kinase fold)															
					5	COG2334			7.47E-04	33%	52%	207-240	200233	5	XP_932523 PREDICTED: similar to serine/arginine repetitive matrix 2		55.84 5.0	9E-06	35%	52%	375456	97178
					6	smart00220	Phosphotransferases. Serine or threonine-specific kinase subfamily.	39.81	1.02E-03	25%	48%	53238	5150	6	XP_729588 dentin phosphoryn		47.75 1.3	9E-03	36%	51%	348429	315396
					7	COG0661	AarF, Predicted unusual protein kinase [General function prediction onlv].	38.39	2.48E-03	31%	59%	204-254	284333	7	AAK54495 neurofilament triplet H1-like protein		57.00 2.2	9E-06	36%	50%	327456	93221
					8	smart00090	RIO RIO-like kinase:	38.29	2.62E-03	28%	45%	166-250	122211	8			47.75 1.3		28%	55%	362-431	42111
					9	pfam01163	RIO1, RIO1 family. This family of proteins are related to eukaryotic type protein kinases. PI3Kc family. Phosphoinositide 3-kinase, catalytic domain; Phosphoinositide 3-kinase isoforms participate in a variety of processes.	37.50	3.98E-03	28%	50%	162-252	78170	9	XP_427855 PREDICTED: similar to p87, partial		45.44 6.8	3E-03	29%	50%	375446	86157
					10	cd00142	including cell molity, the Ras pathway, vesicle trafficking and secretion, and apoptosis. These homologues may be either tipid kinases and/or protein kinases: the former phosphorylate the 3-position in the inostoli ing of inositol phospholipids. The ataxia telanguetesia-mutated gene product, the targets of rapamycin (TOR) and the DNA-dependent kinase have not been found to possess pild kinase activity. Some of this family	37.08	6.09E-03	38%	49%	209-252	147192	10	AAC96402 similar to E. coli LPS core biosynthesis protein, correspon Prot Accession Number P27240	nds to Swiss-	50.06 2.7	9E-04	20%	38%	26313	29-307
							nnssess PI-4 kinase activities															
N724R	275496276452	319	37,146	10.16		No Hit Found								1	Number P33801	ot Accession	284.65 3.0		44%	65%	1310	1314
														2	YP_142754 S/T protein kinase, similar to Paramecium bursaria chlo A617R	Jiella Vilus I	58.15 4.6	1E-07	29%	51%	130246	230-358
N725L	276823276443	127	13,956	3.77		No Hit Found	SCP-x_thiolase, Thiolase domain associated with sterol carrier protein							1	AAC96949 A618L		75.87 4.6	DE-13	54%	79%	56123	65125
N727L	277600276950	217	24,102	4.26	1	cd00829	(SCP)-x isoform and related proteins; SCP-2 has multiple roles in intracellular light 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 ordein.	31.44	8.80E-02	31%	52%	118–172	7–61	1	AAC96950 A619L		51.60 2.1	3E-05	24%	38%	1215	1237
N731L	277945277649	99	11,140	8.64		No Hit Found								1	Number D90902	nk Accession	64.31 1.3 50.45 2.0		34% 30%	50% 51%	195 191	181 182
N732L	278380278027	118	12,691	8.65		No Hit Found								1	AAC96952 A621L		130.18 2.0	6E-29	52%	71%	4116	5117
							Uup, ATPase components of ABC transporters with duplicated ATPase								AAC96981 Chlorella virus CVK2 translation elongation factor-3 hom	olog refer to						
N733R	278451281159	903	99,656	6.17	1		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	280.22	6.30E-76	27%	45%	300-848	2530	1	AAC96981 Children and Concentration and a concentration action of the GenBank Accession Number D16505	olog, reler to	1106.66 0.00	E+00	65%	77%	37901	55918
					2	cd03221	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. NBSI is essential for the ribosome-	174.24	5.16E-44	39%	56%	302-492	1191	2	A48779 translation elongation factor EF-3 homolog - Chlorella virus	CVK2	1011.52 0.00	E+00	61%	74%	37885	2721120
					3	pfam00005	elimitated functions 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		8.69E-28	30%	48%	327491	1182	3	XP_445123 unnamed protein product		757.67 0.00	)E+00	46%	62%	35900	1371043
							pfam00664. These four domains may belong to a single polypeptide, or belong in different nolvonentide chains ABC_ATPase, ABC (ATP-binding cassette) transporter nucleotide- binding domain: ABC transporters are a large family of proteins involved															
					4	cd00267	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, 4-000, and H-loog/switch region in addition to	114.78	3.59E-26	29%	47%	306491	4207	4	BAA33959 translation elongation factor3		756.13 0.00	)E+00	46%	62%	35900	1371043
							the Walker A motifP-loop andWalker B motif commonly found in a number of ATP- and GTP-binding and hydrolyzing proteins. ABC_DR_subfamily_A. This family of ATP-binding proteins belongs to a multisubunit transporter involved in drug resistance (BcA and DrA), notulation, tipid transport, and latticidic minumult, in bacteria and archaea, these transporters usually inclusions. Eukaryote systems of the ABCA one or two integral membrane proteins. Eukaryote systems of the ABCA															
					5	cd03230	one or two integral membrane proteins. Eukaryote systems of the ABLA sublamity display ABC domains that are quite aimlar to this family. The ATP-Rocing domain shows the highest similarity between all members of hydrolases that contain a signature motif. (-loop, and H-loopfwirkh region in addition to the Walker A motifP-loop and/Walker B motif commonly found in a number of ATP- and GTP-binding and hydrolays	112.09	2.44E-25	29%	47%	311488	10205	5	CAA78282 translation elongation factor 3		753.82 0.00	IE+00	45%	62%	29901	1361050
					6	COG1131	northeine ComA, ABC-type multidrug transport system, ATPase component IDefense mechanisms1. ABC_cobalt_transport_domain1, Domain I of the ATPase component of a cobalt transport_domain both bacteria and archaea. This ABC		1.29E-22	28%	48%	312487	16208	6	XP_711404 translation elongation factor 3		753.44 0.00	/E+00	45%	62%	29901	1361050
					7	C403335	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 chlorophills, and a cobalt ion is chelated in the center of the molecules. The genes	400.50	1.87E-22	28%	46%	306-491	4209	7	CAA22654 SPCC417.08		751.90 0.00	DE+00	45%	63%	35808	1421046
					1	6003223	S. typhimurium. In addition to genes known to encode enzymes catalyzing steps of the cobalamine biosynthetic pathway, the products of cbiQ, cbiQ and cbiN, were proposed to constitute a cobalt uptake system since CbiN and CbiQ are integral membrane proteins and CbiO is an		1.072-22	20 /8	4070	300-481	4209	,			751.80 0.00	2.00	40 /0	0378	33686	142-1040
					8	COG1121	ABC ATPase. However, direct evidence supporting this idea is lacking ZnuC, ABC-type Mn/Zn transport systems, ATPase component Ilnoroanic ion transport and metabolism).	101.06	5.09E-22	29%	50%	306493	9216	8	XP_711356 translation elongation factor 3		751.13 0.00	JE+00	45%	62%	29901	1361050

Gene	Genome	A.A.	Peptide	pl	CDD Hit	COGs	COG Definition	Bit Score	E-value	%	%		Hit from-	BLASTp Hit	Hit	BLASTp Definition	Bit Score	E-value	%	%	Query	Hit from-
Name	Position	lenath	Mw	pi	Number	0003	ABC_Iron-Siderophores_B12_Hemin, ABC transporters involved in the		L-value	Identitv	Positive	from-to	to	Number	Accession	DEAD IP Deminion	Dit Ocore	L-Value	Identitv P	ositive	from-to	to
							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															
					9	cd03214	membrane-spanning subunits (BtuC); this arrangement is distinct from that observed for the E. coli lipid flippase MsbA. The BtuC subunits	100.96	6.39E-22	28%	48%	306-487	4206	9	CAA77567 elongation factor 3		748.04	0.00E+00	45%	62%	29901	1361049
							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															
							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															
							ABC_Metallic_Cations, ABC-type Mn/Zn transport systems, ATPase component; This family is comprised of systems involved in the uptake of															
					10	cd03235	various metallic cations such as iron, manganese and zinc. The ATPases of these systems are stronly related to those of iron-siderophore uptake	99.82	1.31E-21	28%	48%	306488	4204	10	XP_455632 unnamed protein prode	uct	747.66	0.00E+00	46%	62%	35901	1371044
							systems suggesting that they rose from a common ancestor															
							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															
N737L	282723281155	523	57,890	5.4	6 1	pfam04451	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	369.64	4.21E-103	42%	55%	173–517	78443	1	AAC27493 putative capsid protein	1	987.25	0.00E+00	92%	93%	1521	1521
					2	COG0173	nivconrotein 3 AspS, Aspartyl-tRNA synthetase [Translation, ribosomal structure and bionenesis]	30.52	4.64E-01	36%	59%	490512	85107	2	AAC96953 similar to Simulium irid Prot Accession Numbe	lescent virus capsid protein, corresponds to Swiss	745.73	0.00E+00	69%	79%	1521	1520
							THE R DELIDENSE							3 4	BAE06835 hypothetical major cap BAA76600 major capsid protein	osid protein		2.20E-24 3.29E-20	67% 65%	84% 78%	179 176	179 176
														5	AAC27492 major capsid protein V BAA76601 major capsid protein N	ICP1	103.61 103.61	1.93E-20 1.93E-20	65% 65%	78% 78%	176 176	
														7		protein Vp54, corresponds to GenBank Accession		1.93E-20	65%	78%	176	176
														8	1M3Y_D Chain D, The Structu Containing. Dna Virus 1M4X_C Chain C, Pbcv-1 Virus	ure Of Major Capsid Protein Of A Large, Lipid		3.77E-08 3.77E-08	57% 57%	73% 73%	2576 2576	
														10	BAA22198 major capsid protein V	/p54	103.61	1.93E-20	65%	78%	176	176
N741R	282995283372	126	14,030	7.5	2 1	COG4852	2 COG4852, Predicted membrane protein [Function unknown].	38.35	2.87E-04	25%	39%	8121	10128	1	AAC96955 A624R YP_485867 hypothetical protein RI	PB_2251	53.53	2.69E-29 2.45E-06	50% 31%	72% 47%	5117 6108	8120 9115
														3 4	ZP_00234461 conserved hypothetica YP_579998 hypothetical protein Po	al protein cryo 0731	50.45	9.32E-06 2.08E-05	25% 31%	44% 47%	6121 6108	9129 8117
														5	CAC98668 Imo0589 AAZ18596 conserved hypothetica	al protein	49.68	2.71E-05 3.54E-05	25% 27%	43% 45%	6121 6108	9129 8117
														8	CAC95830 lin0598 YP_013223 hypothetical protein LM	MOf2365_0618	48.52	6.04E-05 7.89E-05 2.30E-04	24% 24% 28%	43% 43% 47%	8121 6121 1105	11129 9129 1118
														10	AAU26313 membrane protein CAH11414 hypothetical protein			2.30E-04 2.30E-04	28%	47%	1105	
N742L	284255283347	303	34,057	10.7	8	No Hit Found								1	AAC96459 a91L	histone-like protein, corresponds to GenBank	63.93	7.77E-09 1.28E-03	33% 26%	53% 47%	177–291 8118	2112 142259
N745R	004007 004054	185	00.400	3.4											Accession Number D7	1563			39%	47%	1107	
N/45R	284297284851	105	20,162	3.4	9	No Hit Found	DUF1212. Protein of unknown function (DUF1212). This family							1	AAC96977 A659L		05.08	1.28E-09	39%	4770	1107	192
N746R	284991285428	146	16,290	6.8	0 1	pfam06738	3 represents a conserved region within a number of hypothetical proteins of unknown function found in eukaryotes, bacteria and archaea. Some	1 33.27	1.39E-02	28%	49%	1989	73141	1	AAC96975 A656L		65.08	8.01E-10	37%	56%	992	8100
							familv members are membrane proteins Acetyltransf_1, Acetyltransferase (GNAT) family. This family contains															
N747R	285467286063	199	21,772	5.8	5 1 2	pfam00583 COG1247	conteins with N-acetvItransferase functions COG1247, Sortase and related acyltransferases [Cell envelope	40.24	5.71E-06 7.18E-03	31% 35%	52% 49%	113-163 115-162	3082 89138	1	AAC96974 A654L EAO24365 GCN5-related N-acety	litransferase	204.91 63.93	1.25E-51 3.44E-09	51% 28%	68% 46%	1194 2191	1194 11217
					3	COG3153	biogenesis. outer membranel.		9.00E-03	42%	62%	113-162	81126		ZP_00675288 hypothetical protein Te		62.39		25%	45%	10184	
					4	COG0456 COG1670	Riml, Acetyltransferases [General function prediction only].	34.60 34.17	9.23E-03	28%	46% 41%	113-173 97184	97158 85173		XP 392876 PREDICTED: similar to		51.99		26%	43%	11173	
					5	COG1670	[Translation. ribosomal structure and biogenesis].		1.25E-02 2.62E-02	20% 31%	41% 66%	97184	4776	5	BAC87874 arylalkylamine N-acety XP_394768 PREDICTED: similar to			3.93E-05 8.76E-05	21% 33%	40% 54%	10173 95184	
					7	COG1246	onIvI. ArgA, N-acetylglutamate synthase and related acetyltransferases [Amino acid transport and metabolism]		1.35E-01	26%	47%	114163	72119	7				3.33E-04	32%	59%	112-173	
					8 9	COG3053 COG4552	acid transport and metabolism1.     CitC, Citrate lyase synthetase [Energy production and conversion].     Eis. Predicted acetvitransferase involved in intracellular survival and	30.28 29.93	1.57E-01 2.12E-01	33% 20%	52%	114-169	63117 77143		ZP_01182660 hypothetical protein Bo			4.35E-04	25% 23%	41% 42%	12197	11204 6201
					9	COG4552	related acetvltransferases (General function prediction onlv). COG1444, Predicted P-loop ATPase fused to an acetyltransferase	29.93	2.12E-01 2.41E-01	43%	44% 67%	114183 110140	534564	9	CAE65892 Hypothetical protein C ZP_00740288 hypothetical protein RB			9.68E-04 9.68E-04	23%	42%	4191 12197	11204
							IGeneral function prediction onlv1. 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 affecte as the meaning one approximate a finde include transmissional															
N748L	287110286058	351	39,512	7.8	6 1	cd00315	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	147.76	1.87E-36	35%	52%	1187	1183	1	AAV84097 CviPII m5C DNA meth	nyltransferase	478.79	1.27E-133	65%	77%	1349	16363
							factors, X chromosome inactivation, imprinting and the suppression of parasitic DNA sequences. DNA methylation is also essential for proper															
					2	-600145	embryonic development and is an important player in both DNA repair and nanome stability DNA methylase, C-5 cytosine-specific DNA methylase.	142.76	5.07E-35	31%	46%	1195	1196	2	AAC96884 M.CviAll cytosine DNA	a maile disconference	212.20	1.57E-83	45%	65%	1343	2342
					2	COG0270		95.53	5.07E-35 8.19E-21	25%	46%	1195	4183	2	AAC96884 M.CVIAII Cytosine DNA AAC64006 cytosine methyltransfe			1.57E-83 3.07E-79	45% 44%	60%	1343	
					4	pfam00398	RrnaAD, Ribosomal RNA adenine dimethylase RNA pol B RPB2, RNA polymerase beta subunit. RNA polymerases	30.21	3.95E-01	25%	37%	152	3289	4	AAC55063 cytosine methyltransfe	erase	279.64	1.13E-73	42%	57%	4349	6366
					5	cd00651	catalyse the DNA dependent polymerization of RNA. Prokaryotes contain a single RNA polymerase compared to three in eukaryotes (not including mitighered) and eukaryotes and provide the second	29.54	6.87E-01	32%	38%	4 52	905952	6	AAC06097 approximational M Crist	( edapine DNA melloulisensferens	276 56	9.57E-73	41%	57%	4349	6366
					5	0000653	<sup>3</sup> 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	29.54	0.6/E-U1	32%	38%	452	900952	5	AAC96987 nonfunctional M.CviAV	Cytosine DNA metnyitransferase	2/6.56	9.37E-73	+ 170	3/%	4349	0300
							DNA during elongation							6	AAC96897 M.CviAIV cytosine DN	A methyltransferase	266.93	7.59E-70	44%	58%	1331	3332
														7	NP_813725 gp9.1 AAR23218 gp67 BAB77127 site-specific DNA-meth		88.20	1.77E-18 4.83E-16 9.11E-15	36% 31% 34%	51% 48% 46%	1157 1161 4161	
														9 10	BAB77127 site-specific DNA-meth AAV83360 DNA cytosine methyla:			9.11E-15 2.65E-14	34% 33%	46% 50%	4161 4158	
N751L	288505287126	i 460	51,385	10.9	6	No Hit Found								1 2	AAV84098 CviPII top-strand DNA AAC55064 restriction endonuclea:	nicking endonuclease se		7.73E-55 1.80E-27	44% 35%	59% 50%	179455 173419	
N754R	288709289167	153	18,318	4.3	7	No Hit Found								0	No Hit Found No Hit Found							
N755L	289726289202	175	19,326	9.6	8	No Hit Found								1	AAC96968 A644R YP_142763 unknown			2.42E-25 2.36E-04	72% 26%	79% 47%	89172 45165	
N756L	291196289751	482	54,577	11.3	3 1	COG4/97	, COG4487, Uncharacterized protein conserved in bacteria [Function	30.00	6.02E-01	20%	45%	230-315	90177	1	AAC96967 GIn-rich; KQQ (6X)			6.12E-50	79%	93%	196317	
	200701		24,077		. 1	20040	unknown].	00.00		2013	-0,0	0.0		2				4.91E-07	65%	80%	1953	

Gene Name	Genome Position	A.A. lenath	Peptide Mw		CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identitv P	% ositive	Query from-to	lit from- BLASI to Num		Hit BLASTp Definition AAA38349 nonmuscle myosin heavy chain (NMHC)	Bit Score 44.67	E-value 9.58E-03	% dentitv P 21%	% Positive 44%	Query from-to 86314	Hit from- to 325-537
N759R	291268291675	136	15,341	9.58		No Hit Found								1	AAC96961 A634L	168.32	6.78E-41	58%	78%	2129	4131
N762L	292048291662	129	8	14674.85		No Hit Found								0	No Hit Found No Hit Found						
N763R	292132292551	140	15,655	4.20	1	COG5407	SEC63, Preprotein translocase subunit Sec63 [Intracellular trafficking and secretion]		7.54E-01	20%	49%	2474	551602	0	No Hit Found No Hit Found						
N764L	292919292560	120	13,206	9.29	1	pfam07340	Herpes [E1, 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.	20.92	4.85E-02	31%	52%	1776	272336	1	AAC96060 A633R	157.53	1.19E-37	61%	80%	1114	2115
N766L	295247292947	767	85,429	7.75	1	cd01675	RNR_1, RNR, class I. Ribbroucleotide reductase (RNR) catalyzes the reductive synthesis of deoxynhoulceotides from their corresponding niborulceotides. It provides the precursors necessary for DNA synthesis. RNRs are separated into three classes based on their metalloc/dated usage. Class I RNRs, found in eukaryotes, bacteria, and many vruses, use a dison-typery Indical, Class II RNRs, found in bacteria, and bacteriophages, use consymme B12 (adenosylcobalarmi, AdoCDI). Class II RNRs, nour an acrobic bacteria, bacteria presentation and bacteria and bacteriophages, use consymme B12 (adenosylcobalarmi, AdoCDI). Class II RNRs, nour and scrobic bacteria, bacteria presentation and the set of the Many organisms have more than one class of RNR present in their genomes. All three RNRs have a ten-strandid alpha-ten-bab barriel domain that is structurally similar to the domain of FPL (pyruvate formate lyase). Class I RNR is oxygen-degendert and can be subdivided into classes a (eukaryotes, prokaryotes, vruses and phages) and Ib (which is found in prokaryotes only). It is a tetrameter enzyme of two alpha and two belar.	736.63	0.00E+00	51%	68%	176–746	1578	1	AAC96659 similar to Schizosaccharomyces ribonucleotide reductase M1 chain, corresponds to Swiss-Prot Accession Number P36602	1217.99	0.00E+00	78%	88%	9765	14-771
					2	pfam02867	eubunite: this model covers the major part of the alpha or large subunit of Ribonuc red InC Ribonucleotide reductase barrel domain	637.33	0.00E+00	48%	66%	224748	1532	2 N	NP_001026008 ribonucleoside-diphosphate reductase M1 chain	764.61	0.00E+00	50%	69%	9759	1754
					3	COG0209	NrdA, Ribonucleotide reductase, alpha subunit [Nucleotide transport and metabolism].	516.10	5.70E-147	37%	53%	59764	1696		AAH46846 RRM1 protein	757.67	0.00E+00	50%	69%	9759	1754
					4	cd02888	RNR_1 Like, RNR, class Like family, Ribonudeoido reductase (RNR) catalyzes the reductive synthesis of deoxynohucedoidos from their corresponding ribonucleoidos. Il provides the precursors necessary for DNA synthesis. RNRs are separated into three classes based on their metalicochactor usage. Class I RNRs, found in eukaryotes, bacteria, and bacteria, and bacteriophage, usao concryme B12 (denosylcobalami, AdoCb). Class II RNRs, found in anserbic bacteria, bacteriophages, and archaea, usa n R-56 cluster and 5-adenosynthenionine to generate the synthesis and results.	304.43	2.73E-83	34%	51%	196743	2521	4	AAH74185 RRM1 protein	756.13	0.00E+00	50%	69%	9759	1754
					_		a glycyl radical. Many organisms have more than one class of RNR present in their genomes. All three RNRs have a ten-stranded alphabeta 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														
					5		Ribonuc red IqN, Ribonucleotide reductase, all-alpha domain 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		3.45E-12	40%	57%	159222	1078	5	XP_968671 PREDICTED: similar to ribonucleotide reductase M1		0.00E+00	49%	67%	4759	3766
					6	cd00576	common ancestor. RNRs are found in all organisms and provide the only mechanism by which nucleotides are converted to deoxynucleotides, while PFL and essential enzyme in anaeotic bacteria, catalyzes the conversion of pyruvate and CoA to actey/CoA and formate. Both RNR and PFL are dov/r datcial enzymes		9.52E-09	21%	35%	249678	70461	6	AAD49743 ribonucleotide reductase large subunit	748.43	0.00E+00	49%	67%	9759	1755
					7	pfam03477	altu Pr-L ale UNVI raucal etizimies ATP-cone, ATP cone domain alkPPc, Alkaline phosphatase homologues; alkaline phosphatases are	50.37	8.39E-07	25%	40%	9110	1—89	7 8 9 10	YP_322797 röborus-deoide-diobosphate reductase large chain (un-24gene) AAH95000 Fiborus-leoide reductase M1 ABF71875 riborus-leoide reductase large subunit AAH16450 Riborus-leoide reductase M1	746.50 746.12	0.00E+00 0.00E+00 0.00E+00 0.00E+00	49% 49% 49%	67% 68% 66% 68%	9759 9759 9759 9759	1755 1754 1755 1754
N773L	296759295458	434	48,499	10.93	1	cd00016	and r c., Alease physicalase innovages, analysis physicalases are non-specific physichromonesterases that calayses the hydrolysis reaction via a phosphoseryl intermediate to produce inorganic phosphate and the corresponding actionic, optimally at high pH. Alkaline phosphate actists 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%	116174	67125	1	AAC96957 similar to Chlorella virus PBCV-1 ORF A231L, corresponds to GenBank Accession Number U42590		1.97E-129	53%	72%	25432	31436
														2	AAC96599 contains ATP/GTP-binding motif A	296.98	9.30E-79	44%	62%	93422	4337
N777L	297109296774	112	12,158	9.86	1		COG1293, Predicted RNA-binding protein homologous to eukaryotic snRNP Transcription. DUF814, Domain of unknown function (DUF814). This domain occurs in proteins that have been annotated as Fibronectin/fibrinogen binding		4.62E-10	33%	53%	497	445541	1	AAC96548 similar to Streptococcus pyogenes fibronectin protein, corresponds to GenBank Accession Number L28919		8.84E-25	52%	70%	1106	1105
					2	pfam05670	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		7.46E-09	39%	60%	487	3—90	2 :	ZP_00504390 Protein of unknown function DUF814:Fibronectin-binding A, N-terminal AAD35450 fibronectin-binding protein, putative		3.37E-08 3.73E-07	34%	50%	5109	469-576
														9	CAA08883 putative fibromecin-binding protein VP_004218 fibromecin/binding protein BAD70431 probable RNA-binding protein CAB1348 yiAA ZP 01361871 conserved hypothetical protein ZP_0058600 RNA-binding protein bindingous te ukaryotic snRNP	55.45 54.68 54.68 54.30 54.30 53.53	6.36E-07 1.09E-06 1.09E-06 1.42E-06 1.42E-06 2.42E-06	32% 31% 31% 31% 29% 33%	50% 53% 53% 49% 48% 49%	5106 897 897 5109 1109 5106	439-542 405-496 405-496 457-562 470-581 468-572
N779R	297218299638	807	87,716	6.99	1	cmat/00637	CBD II, CBD II domain	64.96	3.83E-11	28%	43%	1293	485	10	AAK80069 Fibronectin-binding protein BAA78554 vChti-1		3.16E-06 0.00E+00	32% 66%	51% 74%	597 1802	460554
WINK	237210-233030	007	07,710	0.55	2	pfam00704	Glyco hydro 18, Glycosyl hydrolases family 18 COG3469, Chitinase (Carbohydrate transport and metabolism). LRR_RI, Leucine-rich repeats (LRRs), ribonuclease inhibitor (RI)-like subfamily. LRRs are 20-29 residue sequence molifs present in many	57.07 34.63	8.98E-09 5.39E-02	22% 33%	45% 35% 46%	538-714 179-259	7195 90169	2	ZP 00570566 Cellulose-binding, bacterial type	1098.19	0.00E+00 1.41E-31	66% 32%	74% 51%	1802 531802	1827 238536
					4	cd00116	proteins that participate in protein-protein interactions and have different functions and cellular locations. LRRs correspond to structural units consisting of a beta strand (LxxLxLxMVICxL conserved pattern) and an alpha heix. This alignment contains 12 strands corresponding to 11 full repeats, consistent with the extent observed in the subfamily acting as Ran GTPace Activatine Proteins (RanGAPI)	32.18	2.56E-01	29%	47%	653–701	155200	4	NP_627029 secreted sugar hydrolase	146.75	3.36E-33	35%	52%	531–783	197456
														6 7 8 9	NP 226429 sugar hydrolase ZP 0120817 Chilinase, Cellulase ZP_01191795 bindina, family 18.Glycoside hydrolase, family 5.Cellulose- ZP_008744 (Chilinase CAH35762 pulative exported chilinase YP 621070 Chilinase	122.48 109.77 133.27 127.49	1.97E-33 6.79E-26 4.56E-22 3.85E-29 2.11E-27 8.57E-29	33% 31% 29% 30% 31% 31%	47% 44%	531-783 530-805 530-783 531-801 531-801 531-801	
N784R	299716300531	272	30,760	5.95	1	cd00180	phosphorylation of specific residues in the activation segment of the catalytic domain, sometimes combined with reversible conformational	191.19	1.06E-49	30%	53%	10268	1256	1			2.85E-65	48%	69%	4267	41304
					2	smart00220	chances in the C-terminal autoreculatory tail S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosohotransferases. Serine or threonine-specific kinase subfamily Pkinase Protein kinase domain	187.34	1.44E-48	31%	49%	11268	1256	2			2.85E-65	48%	69%	4267	20283
					3		SPS1. Serine/threonine protein kinase [General function prediction only /	183.18	2.61E-47	32%	53%	11268	1258	3	AAU06282 protein kinase A248R		2.04E-63	47%	68%	4267	41304
					4	COG0515	Signal transduction mechanisms / Transcription / DNA replication, recombination. and reoairl.	115.64	6.00E-27	28%	47%	10268	1278	4	AAC96616 PBCV-1 protein kinase	236.88	5.55E-61	44%	67%	4267	38303

Gana	Ganama		Pontido							0/	e/	0	Lit from Pl		Hit				0/	0/	Quary	Lit from
Name	Genome Position	A.A. lenath	Peptide Mw	pl	CDD Hit Number	COGs	Total Total David State Andre Divide Inc.					from-to		Number A	Accession BLASTP Defin	nition	Bit Score			ositive	from-to	Hit from- to
					5	smart00219	Tyrrkc, Tyrosine kinase, catalytic domain; Phosphotransferases. Tyrrsine-snecific kinase suhfamilv Tyrkc, Tyrosine kinase, catalytic domain. Phosphotransferases; tyrosine-	114.50	1.07E-26	26%	44%	12264	2253	5	AAA87065 serine/threonine protein kinase		235.34	1.62E-60	44%	67%	4267	34299
							specific kinase subfamily. Enzymes with TyrKc domains belong to an extensive family of proteins which share a conserved catalytic core															
					6	cd00192	common to both serine/threonine and tyrosine protein kinases. Enzymatic activity of tyrosine protein kinases is controlled by phosphorylation of	107.58	1.42E-24	25%	43%	9263	7263	6	AAU06274 protein kinase A248R		210.31	5.57E-53	50%	70%	52267	22237
							specific tyrosine residues in the activation segment of the catalytic domain or a C-terminal tyrosine (tail) residue with reversible															
					_		conformational channes COG3642, Mn2+-dependent serine/threonine protein kinase [Signal							_								
					7	CUG3642	KIND, kinase non-catalytic C-lobe domain; It is an interaction domain	35.59	6.48E-03	26%	44%	86152	67139	7	AAU06285 protein kinase A248R		201.45	2.59E-50	47%	68%	54267	2216
							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															
					8	smart00750	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	34.58	1.21E-02	29%	51%	88204	7101	8	AAU06286 protein kinase A248R		150.98	4.01E-35	47%	70%	119267	1151
							occurrence of KIND only in metazoa implies that it has evolved from the catalytic protein kinase domain into an interaction domain possibly by															
					9	COG0478	keeping the substrate-binding features COG0478, RIO-like serine/threonine protein kinase fused to N-terminal	31.75	9.29E-02	24%	50%	102148	200246	9	AAU06270 protein kinase A248R		137 89	3.51E-31	45%	67%	124267	1146
							HTH domain ISinnal transduction mechanisms1 APH, Phosphotransferase enzyme family. This family consists of bacterial antibiotic resistance proteins, which confer resistance to various															
							aminoglycosides they include:- aminoglycoside 3'-															
					10	pfam01636	phosphotransferase or kanamycin kinase / neomycin-kanamycin phosphotransferase and streptomycin 3''-kinase or streptomycin 3''-phosphotransferase. The aminoglycoside	30.50	2.15E-01	35%	56%	110151	155198	10	AAC96657 similar to PBCV-1 serine/threonine pro GenBank Accession Number U14660	utein kinase, corresponds to	<sup>0</sup> 136.73	7.82E-31	32%	55%	9267	16278
							phosphotralsferases inactivate aminoglycoside antibiotics via obosphotralion. This family also includes homoserine kinase. This family															
							is related to fructosamine kinase ofam03881															
N787R	300578301717	380	39,916	8.96	3	No Hit Found								1	AAC96746 A378L CAA64974 QI74 protein		133.65 48.14	1.13E-29 6.22E-04	55% 20%	69% 33%	178-281 96-377	69173 200461
														3	AAC96921 A565R BAA11343 DNA binding protein			6.22E-04 1.81E-03	25% 26%	33% 33%	84281 84281	265-466 265-458
							Prefoldin, Prefoldin is a hexameric molecular chaperone complex, found								Distribute Distributing protein		40.00	1.012 00	2070	0070	04 201	200 400
							in both eukaryotes and archaea, that binds and stabilizes newly synthesized polypeptides allowing them to fold correctly. The complex															
N791R	301947302597	217	24,656	7.65	5 1	cd00890	contains two alpha and four beta subunits, the two subunits being evolutionarily related. In archaea, there is usually only one gene for each	28.74	5.89E-01	28%	48%	38105	63124	1	AAC96971 similar to Chlorella virus PBCV-1 ORF A4	50R, corresponds to GenBani	k 120.55	3.73E-26	34%	50%	11213	8185
							subunit while in eukaryotes there two or more paralogous genes encoding each subunit adding heterogeneity to the structure of the								Accession Number U42580							
							hexamer. The structure of the complex consists of a double beta barrel assembly with six protruding coiled-coils.															
					2	COG1508	RpoN, DNA-directed RNA polymerase specialized sigma subunit, sigma54 homolog (Transcription).	28.35	6.82E-01	22%	52%	57111	258312	2	AAC96972 similar to Chlorella virus PBCV-1 ORF A4 Accession Number U42580	50R, corresponds to GenBank	k 60.85	3.51E-08	62%	77%	1150	1453
														3 4	AAC96545 A177R AAC96447 A79R		59.31 58.92	1.02E-07 1.33E-07	51% 59%	70% 69%	753 950	753 647
														5	AAC96818 similar to PBCV-1 ORF A275R, encoded b U42580		r 52.76	9.55E-06	51%	76%	1250	947
														6	AAC96643 similar to PBCV-1 ORF A79R, corresp Number U17055	onds to GenBank Accession	n 48.52	1.80E-04	47%	64%	950	950
N792R	302755303510	252	28,357	7.98	3	No Hit Found								1	AAC96377 A9R		249.98	5.52E-65	64%	83%	76247	3174
														2	AAC96377 A9R AAC96818 similar to PBCV-1 ORF A275R, encoded b U42580	y GenBank Accession Number	r 136.35	8.89E-31	33%	55%	1245	1246
														3	AAC96972 similar to Chlorella virus PBCV-1 ORF A4 Accession Number U42580	50R, corresponds to GenBank	k 132.11	1.68E-29	34%	54%	4245	10253
														4	AAC96643 similar to PBCV-1 ORF A79R, corresp Number U17055 AAC96545 A177R	onds to GenBank Accession	121.71	2.27E-26 5.05E-26	30% 31%	54% 50%	1244 1230	4248 4227
														6	AAC96447 A79R		114.78	2.77E-24	32%	53%	1223	1217
														8	AAU06304 hypothetical protein A275R AAU06301 hypothetical protein A275R		72 70	5.42E-12 1.21E-11	27% 26%	51% 51%	78244 78244	1167 1167
														9	AAC96971 Accession Number U42580	50R, corresponds to GenBank		6.62E-10	59%	74%	147	147
17040						No. 112 E								10	AAU06302 hypothetical protein A275R			3.63E-08	28%	52%	113-245	1132
N794R	303528304043	172	20,438	4.52	2	No Hit Found								1	AAC96458 contains phenyl group binding site (CAAX			7.38E-08	32%	53%	6118	1105
N795L	305535304069	489	51,812	10.74	¥ 1		LpqC, Poly(3-hydroxybutyrate) depolymerase [Secondary metabolites biosvnthesis. transport. and catabolism].	37.27	5.09E-03	26%	37%	201-331	60189	1	AAC96409 contains Pro-rich Px motif, PAPK (8X); s protein TPX, corresponds to Swiss-Prot A	ccession Number P19275		I.84E-118	65%	80%	173-486	99410
					2	COG0596	MhpC, Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamilv) IGeneral function orediction onlv1. CODH, Carbon monoxide dehydrogenase (CODH) is found in	32.75	1.06E-01	20%	38%	185306	4108	2	AAC96539 similar to PBCV-1 ORF A41R, corresp Number U17055	onds to GenBank Accession	n 412.54 1	1.78E-113	63%	80%	176-486	77385
							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		0.005.04	0.50/	0001				B4500000		50.04	0.405.00		5001		
					3	cd01915	and two ACS subunits while monofunctional CODH forms a homodimer. Bifunctional CODH reduces carbon dioxide to carbon monoxide and ACS	31.34	2.98E-01	35%	63%	69109	229269	3	BAE02830 surface protein		56.61	2.49E-06	45%	58%	41102	638699
							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,															
							recortively							4	NP_916095 P0481E12.18			6.99E-09	23%	41%	181-463	72352
														5	BAD86968 hypothetical protein AAM63817 unknown			6.99E-09 3.84E-07	23% 21%	41% 40%	181-463 182-464	95375 63341
														7	NP_191439 unknown protein XP 464783 unknown protein			3.84E-07 3.04E-04	21% 22%	40% 41%	182464 201464	63341 78346
														9	contains Pro-rich Px motif, PAPK (19X); AAC96403 specific Pro-rich protein, corresponds to	similar to Arabidopsis anter- Swiss-Prot Accession Numbe	r 48.14	8.84E-04	39%	57%	42107	382449
														10	P40602 AACORTE2 Lys-, Pro-rich, PAPK (10x); similar to	wheat Pro-, Lys-rich protein,		7.48E-03	31%	53%	40107	201-269
														10	corresponds to GenBank Accession Numb	Jer X52472	43.03	7.402-03	5176	5570	40107	201-205
N799R	305621306091	157	17,756	4.43	3 1	COG5201	modification, protein turnover, chaperonesi.	96.28	1.51E-21	34%	56%	3143	4155	1	NP_567959 ASK11 (ARABIDOPSIS SKP1-LIKE 11); u	biquitin-protein ligase	113.24	2.60E-24	39%	61%	3144	6151
					2	smart00512	Skp1, Found in Skp1 protein family; Family of Skp1 (kinetochore protein required for cell cycle progression) and elongin C (subunit of RNA	82.61	2.25E-17	38%	59%	195	2-107	2	NP_567967 ASK12; protein binding / ubiquitin-protein	ligase	112.46	4.44E-24	39%	63%	3144	6151
					3	pfam01466	polymerase II transcription factor SIII) homologues Skp1, Skp1 family, dimerisation domain	66.79	1.31E-12	39%	64%	69144	2-77	3	AAX83944 Skp1			5.80E-24	38%	58%	3144	7155
					4	ptam03931 COG1096	Skp1 POZ, Skp1 family, tetramerisation domain COG1096, Predicted RNA-binding protein (consists of S1 domain and a	61.07 31.80	6.07E-11 4.52E-02	42% 19%	66% 34%	161 57126	166 1989	4	AAO85510 SKP1 AAC63273 SKP1-like protein		110.92 110.54	1.29E-23 1.69E-23	38% 38%	61% 60%	3144 3144	4152 4152
					-		Zn-ribbon domain) [Translation. ribosomal structure and biogenesis].							6	AAC96407 contains ATP/GTP-binding motif A; si	milar to Dictyostelium FP2'			37%	59%	1142	1142
														7	alvcoprotein. corresponds to Swiss-Prot A AAT99735 SKP1 AAD34458 Skp1	ccession Number P52285	109.00	4.91E-23 4.91E-23	39% 38%	59% 60%	3144 3144	7154 5152
														9	ABE93442 SKP1 component	Sheet Bala Catania Cat	109.00	4.91E-23 4.91E-23	38%	59%	3144 3144	5152 7154
														10	Chain B, Structure Of A Beta-Trcp1- 1P22_B Destruction Motif Binding And Lysine Spi Ubiquitin Lingson	ecificity On The Scfbeta-Trcp	1 108.23	8.38E-23	38%	62%	2142	3141
N802R	306133306387	85	9.463	10.33		No Hit Found								1	Ubicuitin Licase AAC96445 contains type 1 hydrophobic transmembra		44.07	7.19E-03	35%	51%	381	586
N802R N803L	306133306387		9,463 29.858	10.33		No Hit Found									AAC96445 contains type 1 hydrophobic transmembra No Hit Found No Hit Found	ie regiun	41.9/	r.19⊑-03	33%	51%	381	386
HOUSE	307 143-300382	204	23,000	5.27		wo nit Found								U	NOTING NO FIL FOUND							

Gene Name N805R	Genome Position 307438308127	A.A. lenath 230	Peptide Mw 25,806		CDD Hit Number	COGs COG Definition		Bit Score	E-value	% Identitv P	% ositive	Query from-to	Hit from-BL to		Hit BLASTp Definition YP_477577 glycosyl transferase, group 1 family protein	Bit Score 49.68	E-value 10 9.12E-05	% dentitv P	% ositive 41%	Query from-to 10192	Hit from- to 196–384
N806L	308866308288	193	21,810	10.62		No Hit Found								0	No Hit Found No Hit Found						
N809R	308948309418	157	18,734	6.54	1	COG4425 COG4425, Predicted membrane protein [Func	ion unknown].	29.56	1.74E-01	33%	47%	102153	444493	0	No Hit Found No Hit Found						
N811R	309467309859	131	14,507	4.33		No Hit Found								0	No Hit Found No Hit Found						
N814R	309873310316	148	16,736	10.33	1	COG4536 CorB, Putative Mg2+ and Co2+ transporter Co and metabolism1.	B [Inorganic ion transport	27.86	5.78E-01	23%	49%	40106	286351	0	No Hit Found No Hit Found						
N815L	311403310333	357	41,147	10.09	1	smart00507 HNHc, HNH nucleases; .		28.90	9.97E-01	31%	44%	60106	1-49	1 2 3 4	AAC96722 A354R AAC96729 similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot AAC96455 Ascisian Nummer P340R1 AAC98465 Ascisian Nummer P340R1 AAL98037 hypothetical phage protein	81.26 62.77	1.63E-51 6.07E-14 2.23E-08 4.82E-03	43% 27% 25% 27%	61% 47% 40% 50%	116354 59326 29351 23152	2235 51323 105447 9137
N819R	311726312187	154	18,361	5.77	1	COG5280 COG5280, Phage-related minor tail protein [Fu		27.69	6.93E-01	17%	31%	2459	1651	1	CAD50809 hypothetical protein	44.67	1.12E-03	24%	43%	1149	429-577
N820R	312298314097	600	65,919	10.86	1	sialate O-acetylesterase like2, sialate O-ace of the SGNH-hydrolases, a diverse family of li cd01828 tertiary fold of the enzyme is substantially of alpha/beta hydrolase family and unique among active site closely resembles the Ser-His-Asp serine hydrolases NnaC like, NnaC (CMP-NeuNAc synthet	ases and esterases. The fferent from that of the all known hydrolases; its Glu) triad found in other		3.44E-15	25%	38%	119–274	1167	1	AAC96688 A318R	238.04	7.81E-61	54%	69%	402–598	6211
					2	SGNH hydrolases, a diverse family of lipases a fold of the enzyme is substantially different fr cd01841 hydrolase family and unique among all known closely resembles two of the three compor Asc/Glu/triad from other serine hydrolases. E.	nd esterases. The tertiary m that of the alpha/beta hydrolases; its active site ents of typical Ser-His- coli NnaC appears to be	49.88	9.78E-07	25%	43%	124–276	6167	2	AAC96684 PAPK (17X); similar to PBCV-1 ORF A41R, corresponds to Genbank Accession Number U17055	45.82	5.69E-03	64%	74%	131	131
					3	involved in colvsaccharide svnthesis. SOMH hydrolase, SOMH hydrolase, or GDS family of lipases and esterases. The tertiary colo229 substantialy different from that of the alphat unique among all known hydrolases; its active prioral Ser-His-Asp(Glu) triad from other serine the rarhonic axid PAF_acetylesterase_like, PAF_acetylhydrolas	fold of the enzyme is eta hydrolase family and ite closely resembles the hydrolases, but may lack		2.83E-06	17%	33%	119–273	1196	3	ZP_00244751 COG2755: Lysophospholipase L1 and related esterases	51.22	1.35E-04	28%	44%	118–274	74244
					4	of SGNH-hydrolases. Platelet-activating factor key players in inflammation and in atherosclerc cd01820 independent phospholipase A2 which exhibits a towards PAF, hydrolyzing an acetyl ester at the also degrades a family of oxidized PAF-like pho- residues. In addition, PAF and PAF-AH an	(PAF) and PAF-AH are sis. PAF-AH is a calcium rong substrate specificity a sn-2 position. PAF-AH spholipids with short sn-2 associated with neural	45.63	1.53E-05	26%	40%	120269	35198	4	BAD50731 conserved hypothetical protein	46.21	4.36E-03	28%	40%	120274	54227
					5	micration and mammalian reproduction XymB like, SXMI-hydrolase subdhaniy, s flavefaciens XymB. Most likely a secreted activity. SXMI-hydrolases are a diverse family cd01833 The tertiary fold of the enzyme is substantially hanholten hydrolase family and unique among active site closely resembles the Ser-His-Asp serin hydrolase.	hydrolase with xylanase of lipases and esterases. different from that of the all known hydrolases; its	44.55	4.03E-05	20%	39%	145-260	53182	5	CAH09444 putative acylhydrolase	45.44	7.43E-03	28%	40%	120274	54-227
N827R	314231314899	223	25,953	6.19		No Hit Found								1	AAC96456 A88R	61.23	2.86E-08	32%	48%	1147	1143
N830L	315907315026	294	32,506	6.73	1	ANK, ankyrin repeats; ankyrin repeats interactions in very diverse families of protei repeats in a protein can range from 2 to over 2 cd00204 ANK repeats may occur in combinations with o structural repeat unit contains two antiparallel repeats are stacked in a superhelical are	<li>Is. The number of ANK (ankyrins, for example). her types of domains. The elices and a beta-hairpin,</li>	99.00	5.91E-22	43%	68%	29150	2125	1	XP_681288 hypothetical protein AN8019.2	78.57	2.89E-13	42%	56%	136233	799–902
					2	contains 4 consecutive reneats COG0666 Arp, FOG: Ankyrin repeat [General function pr		66.06	5.77E-12	30%	41%	5148	77224	2	CAE64680 Hypothetical protein CBG09456	65.08	3.30E-09	33%	51%	108230	39169
					3	Ank, Ankyrin repeat. There's no clear s pfam00023 and signal on the HMM search Ankyrin repeated, alpha, alpha, beta order of secondary associate to form a higher order structure	eparation between noise ts generally consist of a structures. The repeats	45.05	1.21E-05	59%	78%	193–225	1–33	3	EAS30828 hypothetical protein CIMG_06307	48.14	4.18E-04	28%	50%	99230	315-458
					4	ANK, ankyrin repeats; Ankyrin repeats are al smart00248 and occur in at least four consecutive copies. T protein interactions. The core of the repeat se	ney are involved in protein	38.02	1.38E-03	59%	78%	193-220	1-28	4	XP_749852 hypothetical protein Afu1g01020	72.02	2.70E-11	34%	49%	136285	834983
					5	helix structure COG0823 ToIB, Periplasmic component of the ToI biog Intracellular trafficking and secretion.	olymer transport system	33.49	3.07E-02	28%	37%	193-236	292335	5	XP_392578 PREDICTED: similar to CG7462-PB, isoform B	87.43	6.22E-16	28%	45%	38230	83308
														6 7 8 9 10	AAF81702 ankyrin 1 SP 581734 PREICITED: similar to Ankyrin-1 (Erythrocyte ankyrin), partial VP 476042 ankyrin seaset portein AAM11327 (GH1625p NP_648148 Ank2 CG7462-PC, isoform C	83.96 134.04 129.80	4.45E-22 6.87E-15 5.79E-30 1.09E-28 5.64E-09	34% 27% 33% 35% 29%	50% 46% 52% 49% 52%	17230 5230 5249 5244 89230	419637 223480 5247 26264 1146
N836L	316749315937	271	28,506	6.45	1	ANK, ankyrin repeats; ankyrin repeats interactions in very diverse families of protei interactions in a protein can range from 2 to over 2 cd00204 RNK repeats may occur in combinations with o structural repeat unit contains two antiparallel 1	<li>Is. The number of ANK (ankyrins, for example). her types of domains. The elices and a beta-hairpin,</li>	123.27	2.55E-29	46%	59%	39157	8126	1	AAC98373 contains 4 ankyrin repeats: similar to reticulocyte ankyrin, corresponds to Swiss-Prot Accession Number P16157	179.87	8.00E-44	40%	54%	3252	2250
						repeats are stacked in a superhelical arra contains 4 consecutive repeats				38%	53%		57,000		AAC96986 contains 10 ankyrin-like repeats; similar to human ankyrin, corresponds	100.11	1 24E-28	0.007	5004		0.000
					2	COG0666 Arp, FOG: Ankyrin repeat [General function pro Ank, Ankyrin repeat. There's no clear	enaration between noise	84.17	1.49E-17	38%	53%	22160	57203	2	to Swiss-Prot Accession Number P16157			36%	52%	4209	3208
					3	pfam00023 and signal on the HMM search Ankyrin reper beta, alpha, alpha, beta order of secondary associate to form a hindher order structure ANK, ankyrin repeats; Ankyrin repeats are al	out 33 amino acids long	41.97	7.43E-05	53%	63%	105–135	3-33		XP_796302 PREDICTED: similar to Ankyrin repeat domain protein 28, partial		3.08E-03	31%	42%	104-229	12140
					4	smart00248 and occur in at least four consecutive copies. T protein interactions. The core of the repeat se helix structure.	ims to be an helix-loop-	37.63	1.47E-03	45%	62%	7099	130	4	XP_794653 PREDICTED: similar to Ankyrin-1 (Erythrocyte ankyrin)	113.62	7.04E-24	39%	49%	12189	72249
					5	COG2378 COG2378, Predicted transcriptional regulator   FBA, F-box associated region. Members of t with F-box domains, hence the name FBA.	is family are associated	34.21	1.88E-02	28%	47%	192250	69127	5	XP 746992 hypothetical protein Afu8g02140		3.39E-10	34%	47%	3127	688-812
					6	pfam04300 involved in binding other proteins that will be t One member has been shown to be involved in proteins	rgeted for ubiquitination. binding to N-glycosylated	29.98	3.59E-01	24%	33%	4484	96138	6	-		7.82E-07	41%	51%	998	220307
					7	COG0283 Cmk, Cytidylate kinase [Nucleotide transport a	nd metabolism].	29.02	6.95E-01	27%	47%	197–252	151206	7 8 9 10	XP_785043 PREDICTED: similar to Ankyrin-1 (Erythrocyte ankyrin) (Ankyrin R), partial	56.61 57.00	3.51E-07 1.02E-06 7.82E-07 8.36E-09	29% 27% 27% 38%	44% 42% 48% 53%	75202 75202 9156 996	68196 69208 51199 783870
N840R	316897317373	159	18,480	8.36		No Hit Found								0	No Hit Found No Hit Found						
N845R	317707318180	158	17,734	5.51		No Hit Found								8	corresponds to Swiss-Prot Accession Number P27951	43.51 41.59 45.05 41.97 45.05 43.13 41.97 42.74	1.48E-03 2.52E-03 9.59E-03 8.67E-04 7.34E-03 8.67E-04 3.30E-03 4.30E-03 4.30E-03	50% 55% 60% 67% 48% 52% 52% 55% 55%	67% 58% 60% 65% 65% 65% 65% 62% 62%		5388 16831722 16851718 2357 181225 156195 570609 210249

Gene Name	Genome Position	A.A. lenath	Peptide Mw	pl	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	% Identitv F		Query from-to	Hit from- to
N847R	319064319798	245	27,064	8.51		No Hit Found								1	AAC96734 A366L			1.35E-36	38%	59%	6237	14251
														2	Accession N	lorella virus PBCV-1 ORF A450R, corresponds to GenBank umber U42580	01.02	2.66E-08	32%	46%	2151	8144
																CV-1 ORF A275R, encoded by GenBank Accession Number		8.55E-07	37%	51%	172	181
														4		BCV-1 ORF A79R, corresponds to GenBank Accession 055	55.45	1.91E-06	25%	40%	1237	4199
														5	AAC96447 A79R			2.49E-06	27%	40%	1237	1189
														6	AAC96971 similar to Ch Accession N	lorella virus PBCV-1 ORF A450R, corresponds to GenBank umber U42580	54.30	4.25E-06	48%	59%	147	147
														7	AAC96545 A177R		51.22	3.59E-05	42%	63%	147	450