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 B: Gene Names M001L through M807R

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

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

Xiao Li
University of Massachusetts–Lowell

Tamara Feldblyum
The Institute for Genomic Research, Rockville, MD

James Hartigan
Agencourt Bioscience Corporation, Beverly, MA

See next page for additional authors

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

Part of the Virology Commons

Fitzgerald, Lisa A.; Graves, Michael V.; Li, Xiao; Feldblyum, Tamara; Hartigan, James; and Van Etten, James L., "Supplementary Data for "Sequence and annotation of the 314-kb MT325 and the 321-kb FR483 viruses that infect *Chlorella* Pbi": Appendix B: Gene Names M001L through M807R" (2007). *Virology Papers.* Paper 3.
http://digitalcommons.unl.edu/virologypub/3

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

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

This article is available at DigitalCommons@University of Nebraska - Lincoln: http://digitalcommons.unl.edu/virologypub/3

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

SUPPLEMENTARY DATA FOR

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

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

[a]Department of Chemistry, University of Nebraska–Lincoln, Lincoln, NE 68588-0304
[b]Department of Biological Sciences, University of Massachusetts–Lowell, Lowell, MA 01854
[c]The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850
[d]Agencourt Bioscience Corporation, 500 Cummings Center, Suite 2450, Beverly, MA 01915
[e]Deparment of Plant Pathology, University of Nebraska–Lincoln, Lincoln, NE 68583-0722
[f]Nebraska Center for Virology, University of Nebraska, Lincoln, NE 68588-0666
*Corresponding author. Email: jvanetten@unlnotes.unl.edu

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

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

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

Appendix B: Gene Names M001L through M807R

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
M001L	1052--384	223	26,253	6.70		No Hit Found								1	ZP_00575949	RepA / Rep+ protein KID	55.45	1.41E-06	25%	59%	14--113	61--160
														2	NP_705165	hypothetical malaria antigen	51.60	2.03E-05	24%	51%	1--136	1993--2129
														3	YP_161395	hypothetical protein BGP110	49.68	7.73E-05	27%	54%	27--137	91--194
M003L	1521--1147	125	13,988	9.87		No Hit Found									No Hit Found	No Hit Found						
M005L	2294--1755	180	20,424	10.16		No Hit Found								1	NP_048429	A81L	110.92	1.71E-23	40%	60%	33--158	21--163
M007L	2839--2330	170	19,447	6.12		No Hit Found								1	NP_048432	A84L	73.94	1.97E-12	33%	57%	11--135	10--149
M009L	3736--2930	269	31,504	6.02	1	pfam03016	Exostosin, Exostosin family. The EXT family is a family of tumour suppressor genes. Mutations of EXT1 on 8q24.1, EXT2 on 11p11-13, and EXT3 on 19p have been associated with the autosomal dominant disorder known as hereditary multiple exostoses (HME). This is the most common known skeletal dysplasia. The chromosomal locations of other EXT genes suggest association with other forms of neoplasia. EXT1 and EXT2 have both been shown to encode a heparan sulphate polymerase with both D-glucuronyl (GlcA) and N-acetyl-D-glucosaminoglycan (GlcNAC) transferase activities. The nature of the defect in heparan sulphate biosynthesis in HME is unclear	45.82	4.20E-06	25%	51%	159--235	221--292	1	NP_048423	A75L	233.42	5.34E-60	40%	61%	8--268	7--276
M010L	4146--3790	119	13,199	9.97	1	cd01285	nucleoside_deaminase, Nucleoside deaminases include adenosine, guanine and cytosine deaminases. These enzymes are Zn dependent and catalyze the deamination of nucleosides. The zinc ion in the active site plays a central role in the proposed catalytic mechanism, activating a water molecule to form a hydroxide ion that performs a nucleophilic attack on the substrate. The functional enzyme is a homodimer. Cytosine deaminase catalyzes the deamination of cytosine to uracil and ammonia and is a member of the pyrimidine salvage pathway. Cytosine deaminase is found in bacteria and fungi but is not present in mammals; for this reason, the enzyme is currently of interest for antimicrobial drug design and gene therapy applications against tumors. Some members of this family are tRNA-specific adenosine deaminases that generate inosine at the first position of their anticodon (position 34) of specific tRNAs; this modification is thought to enlarge the codon recognition capacity during protein synthesis. Other members of the family are guanine deaminases which deaminate guanine to xanthine as part of the utilization of guanine	57.97	9.98E-10	35%	51%	4--106	1--94	1	NP_048547	contains cytidine and deoxycytidine deaminase Zn-binding region signature	161.38	7.13E-39	62%	83%	1--118	1--118
					2	COG0590	CumB, Cytosine/adenosine deaminases [Nucleotide transport and metabolism / Translation, ribosomal structure and biogenesis]	52.65	3.33E-08	31%	47%	2--106	10--105	2	AAR26853	FirrV-1-A29	50.83	1.36E-05	27%	50%	22--111	24--105
					3	pfam00383	dCMP_cyt_deam, Cytidine and deoxycytidylate deaminase zinc-binding region.	43.43	2.04E-05	30%	47%	4--106	7--101	3	AAX51127	tRNA-specific adenosine deaminase	48.52	6.74E-05	31%	49%	4--108	10--104
					4	cd01284	Riboflavin_deaminase-reductase, Riboflavin-specific deaminase. Riboflavin biosynthesis protein RibD (Diaminohydroxyphosphoribosylaminopyrimidine deaminase) catalyzes the deamination of 2,5-diamino-6-ribosylamino-4(3H)-pyrimidinone 5'-phosphate, which is an intermediate step in the biosynthesis of riboflavin.The ribG gene of Bacillus subtilis and the ribD gene of E. coli are bifunctional and contain this deaminase domain and a reductase domain which catalyzes the subsequent reduction of the ribosyl side chain	36.76	0.002101	26%	45%	4--107	1--95	4	AAC68441	cytosine deaminase	48.14	8.80E-05	31%	49%	4--108	10--104
M011L	4524--4237	96	10,637	10.49		No Hit Found								1	NP_048546	A199R	73.17	2.59E-12	43%	68%	1--79	1--82
M012R	4552--5019	156	17,894	7.66		No Hit Found								1	NP_048543	A196L	185.27	4.65E-46	56%	72%	1--151	1--151
M014R	5076--5867	264	29,728	5.00	1	pfam00705	PCNA_N, Proliferating cell nuclear antigen, N-terminal domain. N-terminal and C-terminal domains of PCNA are topologically identical. Three PCNA molecules are tightly associated to form a closed ring encircling duplex DNA..	92.25	4.08E-20	28%	57%	9--132	1--125	1	NP_048540	similar to human PCNA, corresponds to Swiss-Prot Accession Number P12004	368.24	1.34E-100	68%	85%	9--261	7--259
					2	pfam02747	PCNA_C, Proliferating cell nuclear antigen, C-terminal domain. N-terminal and C-terminal domains of PCNA are topologically identical. Three PCNA molecules are tightly associated to form a closed ring encircling duplex DNA..	79.98	2.27E-16	34%	55%	137--261	2--127	2	XP_534355	PREDICTED: similar to proliferating cell nuclear antigen	158.69	1.61E-37	32%	55%	5--264	202--461
					3	COG0592	DnaN, DNA polymerase sliding clamp subunit (PCNA homolog) [DNA replication, recombination, and repair]	56.08	3.34E-09	19%	37%	21--261	72--321	3	CAA55669	proliferative cell nuclear antigen	157.53	3.59E-37	34%	56%	9--264	1--256
														4	AAB27811	PCNA	156.76	6.13E-37	32%	57%	9--264	1--256
														5	NP_172217	PCNA1 (PROLIFERATING CELLULAR NUCLEAR ANTIGEN); DNA binding / DNA polymerase processivity factor	156.76	6.13E-37	32%	57%	9--264	1--256
														6	AAG10077	proliferating cell nuclear antigen	155.99	1.05E-36	32%	55%	9--264	1--256
														7	CAA37243	unnamed protein product	155.61	1.37E-36	32%	55%	9--264	1--256
														8	BAB28355	unnamed protein product	155.61	1.37E-36	32%	55%	9--264	1--256
														9	CAA40938	proliferating cell nuclear antigen (DNA polymerase delta auxiliary protein)	155.61	1.37E-36	32%	55%	9--264	1--256
														10	AAD10528	proliferating cell nuclear antigen	155.22	1.78E-36	34%	56%	9--261	1--253
M015L	9734--5877	1286	141,489	11.47	1	pfam05917	DUF874, Helicobacter pylori protein of unknown function (DUF874). This family consists of several hypothetical proteins specific to Helicobacter pylori. The function of this family is unknown	40.05	0.000259	18%	44%	1094--1283	123--305	1	NP_048536	similar to SWI/SNF chromatin remodeling complex subunit OSA2	743.04	0.00E+00	39%	51%	2--1171	12--1139
					2	COG5178	PRP8, U5 snRNP spliceosome subunit [RNA processing and modification]	38.14	0.000753	26%	39%	1208--1285	9--91	2	T17682	hypothetical protein A192R - Chlorella virus PBCV-1	81.65	1.96E-13	30%	47%	981--1171	2--200
					3	COG0419	SbcC, ATPase involved in DNA repair [DNA replication, recombination, and repair]	37.34	0.001476	15%	38%	1037--1278	197--462	3	NP_477523	wsv001	67.78	2.93E-09	24%	31%	94--489	666--1122
					4	COG3206	GumC, Uncharacterized protein involved in exopolysaccharide biosynthesis [Cell envelope biogenesis, outer membrane]	35.44	0.006073	12%	33%	1055--1286	144--366	4	AAK77699	ORF30, putative collagen	67.78	2.93E-09	25%	32%	106--493	444--857
					5	pfam05483	SCP-1, Synaptonemal complex protein 1 (SCP-1). Synaptonemal complex protein 1 (SCP-1) is the major component of the transverse filaments of the synaptonemal complex. Synaptonemal complexes are structures that are formed between homologous chromosomes during meiotic prophase	35.17	0.006547	24%	44%	1048--1212	600--772	5	EAA46319	CG40323-PB.3	62.77	9.42E-08	25%	34%	120--468	161--492
														6	T17681	hypothetical protein a191R - Chlorella virus PBCV-1	62.00	1.61E-07	46%	72%	918--977	2--63
														7	XP_783728	PREDICTED: similar to Protein transport protein Sec24C (SEC24-related protein C)	59.31	1.04E-06	23%	34%	207--494	6--311
														8	XP_394285	PREDICTED: similar to GA11046-PA	58.92	1.36E-06	28%	36%	209--477	124--394
														9	XP_697998	PREDICTED: similar to AT rich interactive domain 1B (SWI1-like) isoform 1	57.38	3.96E-06	25%	35%	144--460	506--844
														10	NP_571089	calymmin	57.00	5.17E-06	25%	33%	204--479	282--573
M019L	12489--9769	907	102,974	7.91	1	COG0417	PolB, DNA polymerase elongation subunit (family B) [DNA replication, recombination, and repair]	357.05	9.23E-100	30%	48%	22--850	11--772	1	NP_048532	PBVC-1 DNA polymerase	1353.96	0.00E+00	72%	84%	1--903	1--913
					2	pfam00136	DNA_pol_B, DNA polymerase family B. This region of DNA polymerase B appears to consist of more than one structural domain, possibly including elongation, DNA-binding and dNTP binding activities..	336.58	1.19E-93	38%	55%	427--850	1--439	2	BAA35142	DNA polymerase	1348.18	0.00E+00	72%	83%	1--903	1--913
					3	smart00486	POLBc, DNA polymerase type-B family; DNA polymerase alpha, delta, epsilon and zeta chain (eukaryota), DNA polymerases in archaea, DNA polymerase II in e. coli, mitochondrial DNA polymerases and and virus DNA polymerases .	306.37	1.55E-84	33%	51%	176--638	1--475	3	P30320	DNA polymerase	1342.41	0.00E+00	71%	83%	1--903	1--913
					4	cd00145	POLBc, DNA polymerase type-B family; DNA directed DNA polymerase. Posseses DNA binding, polymerase and 3'-5' exonuclease activity..	283.74	9.61E-78	33%	49%	176--675	1--511	4	AAB49748	DNA polymerase	443.74	1.33E-122	96%	97%	460--689	1--230
					5	pfam03104	DNA_pol_B_exo, DNA polymerase family B, exonuclease domain. This domain has 3' to 5' exonuclease activity and adopts a ribonuclease H type fold..	194.94	4.75E-51	26%	41%	27--354	1--334	5	AAK28935	DNA polymerase	431.80	5.23E-119	100%	100%	468--682	1--215
														6	AAK28933	DNA polymerase	429.10	3.39E-118	99%	100%	468--682	1--215
														7	AAX86472	DNA polymerase	429.10	3.39E-118	99%	100%	469--682	1--214
														8	AAK28936	DNA polymerase	427.56	9.86E-118	99%	99%	468--682	1--215
														9	XP_757605	hypothetical protein UM01458.1	412.15	4.29E-113	32%	49%	17--851	83--929
														10	CAD21389	probable DNA-directed DNA polymerase III	405.99	3.07E-111	32%	52%	17--863	123--980
M023L	12664--12446	73	8,575	8.65		No Hit Found									No Hit Found	No Hit Found						

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
M024L	12885--12685	67	7,956	9.03		No Hit Found									No Hit Found	No Hit Found						
M025R	12808--13545	246	27,971	7.96		No Hit Found								1	NP_048357	A9R	221.09	2.33E-56	57%	79%	73--243	3--173
														2	NP_048525	A177R	137.12	4.42E-31	34%	56%	1--227	4--227
														3	NP_049005	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	135.58	1.29E-30	32%	51%	2--242	8--253
														4	NP_048807	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	133.27	6.39E-30	31%	52%	1--244	1--248
														5	NP_048629	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	131.72	1.86E-29	31%	52%	1--241	4--248
														6	NP_048427	A79R	120.94	3.28E-26	31%	52%	1--221	1--218
														7	AAU06304	hypothetical protein A275R	81.65	2.21E-14	31%	52%	78--241	4--167
														8	AAU06301	hypothetical protein A275R	80.49	4.91E-14	30%	52%	78--241	4--167
														9	AAU06302	hypothetical protein A275R	67.01	5.62E-10	32%	54%	110--242	1--132
														10	NP_049003	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	57.00	5.82E-07	46%	68%	1--47	1--47
M026L	14628--13591	346	38,210	7.11	1	COG1052	LdhA, Lactate dehydrogenase and related dehydrogenases [Energy production and conversion / Coenzyme metabolism / General function prediction only].	250.93	7.39E-68	37%	57%	16--345	3--321	1	NP_048401	contains D-isomer specific 2-hydroxyacid dehydrogenase signature; similar to E. coli D-lactate dehydrogenase, corresponds to Swiss-Prot Accession Number P52643	457.22	3.41E-127	67%	80%	9--345	25--361
					2	COG0111	SerA, Phosphoglycerate dehydrogenase and related dehydrogenases [Amino acid transport and metabolism]	189.38	2.72E-49	35%	56%	87--333	74--303	2	XP_363141	hypothetical protein MG08725.4	240.35	6.56E-62	38%	57%	16--345	3--332
					3	pfam02826	2-Hacid_dh_C, D-isomer specific 2-hydroxyacid dehydrogenase, NAD binding domain. This domain is inserted into the catalytic domain, the large dehydrogenase and D-lactate dehydrogenase families in SCOP. N-terminal portion of which is represented by family pfam00389..	168.10	7.14E-43	39%	64%	116--312	1--184	3	CAE81937	related to D-lactate dehydrogenase	228.79	1.97E-58	38%	56%	16--332	3--333
					4	pfam00389	2-Hacid_dh, D-isomer specific 2-hydroxyacid dehydrogenase, catalytic domain. This family represents the largest portion of the catalytic domain of 2-hydroxyacid dehydrogenases as the NAD binding domain is inserted within the structural domain..	55.65	4.81E-09	26%	45%	16--110	1--95	4	BAE59109	unnamed protein product	228.02	3.37E-58	40%	56%	16--332	3--317
														5	CAD01684	D-lactate dehydrogenase	226.87	7.50E-58	40%	57%	16--332	3--315
														6	AAL20565	NAD-dependent fermentative D-lactate dehydrogenase	226.48	9.80E-58	39%	57%	16--332	3--315
														7	YP_216629	D-lactate dehydrogenase	224.94	2.85E-57	39%	57%	16--332	3--315
														8	YP_408119	fermentative D-lactate dehydrogenase, NAD-dependent	223.02	1.08E-56	39%	57%	16--332	3--315
														9	CAG19621	Putative D-Lactate dehydrogenase	222.63	1.41E-56	39%	55%	16--345	13--338
														10	EAL91034	D-lactate dehydrogenase	222.25	1.85E-56	40%	55%	16--332	3--317
M030R	14798--15607	270	29,491	7.20	1	cd00333	MIP, Major intrinsic protein (MIP) superfamily. Members of the MIP superfamily function as membrane channels that selectively transport water, small neutral molecules, and ions out of and between cells. The channel proteins share a common fold: the N-terminal cytosolic portion followed by six transmembrane helices, which might have arisen through gene duplication. On the basis of sequence similarity and functional characteristics, the superfamily can be subdivided into two major groups: water-selective channels called aquaporins (AQPs) and glycerol uptake facilitators (GlpFs). AQPs are found in all three kingdoms of life, while GlpFs have been characterized only within microorganisms..	103.35	2.08E-23	26%	45%	9--262	3--228	1	ABA40763	aquaglyceroporin	463.38	3.20E-129	88%	88%	1--270	1--270
														2	AAK79288	Glycerol uptake facilitator protein, GLPF	147.52	3.88E-34	34%	51%	6--261	1--231
														3	AAU85390	glycerol uptake facilitator protein	132.88	9.90E-30	33%	48%	12--261	12--237
														4	ZP_00908482	Aquaporin	132.49	1.29E-29	32%	49%	6--256	1--227
														5	BAB80632	glycerol uptake facilitator protein	129.41	1.09E-28	34%	48%	11--259	6--230
														6	BAD65916	glycerol uptake facilitator	128.26	2.44E-28	31%	47%	11--261	6--232
														7	ZP_00540612	Major intrinsic protein	127.49	4.16E-28	34%	47%	9--261	4--231
														8	BAB82260	probable glycerol uptake facilitator protein	127.10	5.43E-28	34%	47%	9--259	5--231
														9	NP_782565	glycerol uptake facilitator protein	125.56	1.58E-27	30%	46%	6--261	1--232
														10	ABB16121	glycerol uptake facilitator protein	125.56	1.58E-27	32%	48%	9--261	4--234
M032L	16521--15835	229	26,057	10.08	1	cd00283	GIY-YIG_Cterm, GIYX(10-11)YIG family of class I homing endonucleases C-terminus (GIY-YIG_Cterm). Homing endonucleases promote the mobility of intron or intein by recognizing and cleaving a homologous allele that lacks the sequence. They catalyze a double-strand break in the DNA near the insertion site of that element to facilitate homing at that site. Class I homing endonucleases are sorted into four families based on the presence of these motifs in their respective N-termini: LAGLIDADG, His-Cys box, HNH, and GIY-YIG. This CD contains several but not all members of the GIY-YIG family. The C-terminus of GIY-YIG is a DNA-binding domain which is separated from the N-terminus by a long, flexible linker. The DNA-binding domain consists of a minor-groove binding alpha-helix, and a helix-turn-helix. Some also contain a zinc finger (i.e. I-TevI) which is not required for DNA binding or catalysis, but is a component of the linker and directs the catalytic domain to cleave the homing site at a fixed distance from the intron insertion site.	70.03	1.93E-13	42%	60%	118--225	10--113	1	NP_048671	A315L	205.30	1.15E-51	45%	61%	1--225	1--240
					2	smart00497	IENR1, Intron encoded nuclease repeat motif. Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished).	42.81	3.01E-05	40%	58%	174--227	1--53	2	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	193.74	3.47E-48	45%	62%	1--228	1--226
					3	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	39.29	0.000363	34%	55%	1--91	1--83	3	NP_048641	PBCV-1 33kd peptide	167.93	2.04E-40	41%	58%	5--225	7--247
					4	pfam07453	NUMOD1, NUMOD1 domain..	36.56	0.002303	44%	69%	174--206	1--33	4	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	131.34	2.11E-29	38%	51%	1--200	1--196
														5	AAC49244	ORF301	78.18	2.12E-13	31%	46%	9--162	80--235
														6	AAK09365	intron encoded BmoI	72.79	8.92E-12	31%	41%	4--200	5--242
														7	CAA38804	GIY COII i1 grp IB protein	68.17	2.20E-10	30%	46%	4--193	74--259
														8	AAC49248	ORF211	65.47	1.42E-09	35%	47%	55--169	50--140
														9	YP_293795	putative endonuclease	62.00	1.57E-08	40%	58%	2--81	3--81
														10	NP_899393	SegD	60.85	3.51E-08	29%	47%	1--166	1--176
M034L	17228--16587	214	24,469	8.64	1	pfam02511	Thy1, Thymidylate synthase complementing protein. Thymidylate synthase complementing protein (Thy1) complements the thymidine growth requirement of the organisms in which it is found, but shows no homology to thymidylate synthase..	204.43	7.83E-54	36%	55%	1--210	1--216	1	NP_049030	similar to Synechocystis ORF s111635, corresponds to GenBank Accession Number D90903	279.64	4.18E-74	60%	80%	1--213	1--215
					2	COG1351	THY1, Predicted alternative thymidylate synthase [Nucleotide transport and metabolism].	134.76	7.22E-33	30%	44%	16--213	29--237	2	CAF34258	thymidilate synthase	220.32	3.01E-56	53%	68%	1--214	3--213
														3	ZP_00588145	Thymidylate synthase complementing protein	219.55	5.13E-56	51%	69%	1--214	21--236
														4	ZP_01006033	thymidylate synthase	219.55	5.13E-56	54%	71%	12--213	6--210
														5	AAZ59097	Thymidylate synthase complementing protein ThyX	219.16	6.70E-56	54%	71%	12--213	6--210
														6	ZP_00531300	Thymidylate synthase (FAD)	216.47	4.34E-55	53%	68%	1--213	1--215
														7	AAP99319	Predicted alternative thymidylate synthase	214.54	1.65E-54	52%	71%	12--213	6--210
														8	AAM72779	thymidylate synthase, flavin-dependent	213.77	2.81E-54	51%	67%	1--214	1--216
														9	AAX46996	Td	213.77	2.81E-54	58%	70%	19--213	12--210
														10	ABB27663	Thymidylate synthase complementing protein ThyX	213.39	3.68E-54	50%	69%	1--214	1--216
M037R	17363--19141	593	64,754	6.02	1	COG0449	GlmS, Glucosamine 6-phosphate synthetase, contains amidotransferase and phosphosugar isomerase domains [Cell envelope biogenesis, outer membrane].	616.81	5.16E-178	44%	62%	1--593	1--597	1	BAD15299	glutamine:fructose-6-phosphate amidotransferase GFAT	729.55	0.00E+00	61%	78%	1--591	1--594
					2	cd00714	GFAT, Glutamine amidotransferases class-II (Gn-AT)_GFAT-type. This domain is found at the N-terminus of glucosamine 6-phosphate (GlcN-6-P) synthase (GLMS or GFAT).The glutaminase domain catalyzes amide nitrogen transfer from glutamine to the appropriate substrate. In this process, glutamine is hydrolyzed to glutamic acid and ammonia. GFAT catalyzes the formation of glucosamine 6-phosphate from fructose 6-phosphate and glutamine, the initiating step in the biosynthesis of UDP-GlcN-6-P..	253.13	1.55E-68	47%	66%	2--208	1--210	2	NP_048448	PBCV-1 glucosamine synthetase	728.01	0.00E+00	60%	77%	1--591	1--593
					3	COG2222	AgaS, Predicted phosphosugar isomerases [Cell envelope biogenesis, outer membrane].	179.32	2.71E-46	29%	48%	248--583	2--329	3	CAE39493	glucosamine--fructose-6-phosphate aminotransferase	511.15	4.12E-143	46%	63%	1--593	1--610
					4	pfam00310	GATase_2, Glutamine amidotransferases class-II..	128.88	3.74E-31	36%	59%	2--133	1--136	4	CAE44992	glucosamine--fructose-6-phosphate aminotransferase	508.06	3.49E-142	45%	63%	1--593	1--610
					5	COG0034	PurF, Glutamine phosphoribosylpyrophosphate amidotransferase [Nucleotide transport and metabolism].	128.02	7.24E-31	31%	54%	1--228	4--232	5	ZP_00594232	Glucosamine-fructose-6-phosphate aminotransferase, isomerising	506.52	1.02E-141	45%	64%	1--593	1--612

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					6	cd00715	GPATase_N, Glutamine amidotransferases class-II (GN-AT)_GPAT-type. This domain is found at the N-terminus of glutamine phosphoribosylpyrophosphate (Prpp) amidotransferase (GPATase) . The glutaminase domain catalyzes amide nitrogen transfer from glutamine to the appropriate substrate. In this process, glutamine is hydrolyzed to glutamic acid and ammonia. GPATase catalyzes the first step in purine biosynthesis, an amide transfer from glutamine to PRPP, resulting in phosphoribosylamine, pyrophosphate and glutamate. GPATase crystalizes as a homotetramer, but can also exist as a homdimer..	125.24	5.16E-30	32%	52%	2--225	1--224	6	ZP_00942990	Glucosamine--fructose-6-phosphate aminotransferase (isomerizing)	504.21	5.04E-141	45%	64%	1--593	1--612
					7	cd00352	Gn_AT_II, Glutamine amidotransferases class-II (GATase). The glutaminase domain catalyzes an amide nitrogen transfer from glutamine to the appropriate substrate. In this process, glutamine is hydrolyzed to glutamic acid and ammonia. This domain belongs to the Ntn hydrolase superfamily and is found at the N-terminus of enzymes such as glucosamine-fructose 6-phosphate synthase (GLMS or GFAT), glutamine phosphoribosylpyrophosphate (Prpp) amidotransferase (GPATase) , asparagine synthetase B (AsnB), beta lactam synthetase (beta-LS) and glutamate synthase (GltS). GLMS catalyzes the formation of glucosamine 6-phosphate from fructose 6-phosphate and glutamine in amino sugar synthesis. GPATase catalyzes the first step in purine biosynthesis, an amide transfer from glutamine to PRPP, resulting in phosphoribosylamine, pyrophosphate and glutamate. Asparagine synthetase B synthesizes asparagine from aspartate and glutamine. Beta LS catalyzes the formation of the beta-lactam ring in the beta-lactamase inhibitor clavulanic acid. GltS synthesizes L-glutamate from 2-	109.41	3.32E-25	33%	47%	2--196	1--203	7	CAD13706	PROBABLE GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE PROTEIN	500.36	7.28E-140	44%	63%	1--593	1--612
					8	pfam01380	SIS, SIS domain. SIS (Sugar ISomerase) domains are found in many phosphosugar isomerases and phosphosugar binding proteins. SIS domains are also found in proteins that regulate the expression of genes involved in synthesis of phosphosugars. Presumably the SIS domains bind to the end-product of the pathway	91.50	7.42E-20	37%	55%	283--410	3--131	8	ZP_00509192	Glucosamine-fructose-6-phosphate aminotransferase, isomerising	497.28	6.16E-139	45%	63%	1--593	1--616
					9	COG0367	AsnB, Asparagine synthase (glutamine-hydrolyzing) [Amino acid transport and metabolism].	78.94	4.88E-16	25%	44%	1--229	1--195	9	ZP_00244599	COG0449: Glucosamine 6-phosphate synthetase, contains amidotransferase and phosphosugar isomerase domains	496.89	8.05E-139	44%	62%	1--593	1--619
					10	cd00712	AsnB, Glutamine amidotransferases class-II (GATase) asparagine synthase_B type. Asparagine synthetase B catalyses the ATP-dependent conversion of aspartate to asparagine. This enzyme is a homodimer, with each monomer composed of a glutaminase domain and a synthetase domain. The N-terminal glutaminase domain hydrolyzes glutamine to glutamic acid and ammonia	72.17	4.44E-14	29%	49%	2--196	1--163	10	YP_438548	glucosamine--fructose-6-phosphate aminotransferase, isomerizing	496.12	1.37E-138	44%	64%	1--593	1--605
M042R	19188--22055	956	96,592	5.78	1	COG2911	COG2911, Uncharacterized protein conserved in bacteria [Function unknown].	35.67	0.005043	16%	30%	21--749	304--991	1	BAB83471	Vp260 like protein	1353.96	0.00E+00	71%	81%	1--956	1--955
														2	BAB83467	Vp260 like protein	694.50	0.00E+00	42%	60%	2--956	3--954
														3	BAB83468	Vp260 like protein	632.48	2.15E-179	41%	57%	10--950	1--930
														4	BAB83469	Vp260 like protein	626.32	1.54E-177	41%	57%	10--950	1--930
														5	BAB83470	Vp260 like protein	620.54	8.44E-176	40%	57%	10--950	1--931
														6	NP_048470	PBCV-1 Vp260 protein	241.89	8.16E-62	30%	42%	27--831	166--1021
														7	AAA86307	glycoprotein Vp260	215.31	8.18E-54	30%	41%	51--831	15--835
														8	NP_048362	Asn/Thr/Ser/Val rich protein	168.32	1.15E-39	25%	41%	20--850	19--961
														9	NP_048377	Asn/Thr/Ser/Val rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	166.01	5.69E-39	25%	40%	22--823	18--810
														10	AAC35172	cell surface antigen rOmpA	158.30	1.19E-36	24%	37%	28--828	255--1026
M047R	22096--26496	1467	151,716	6.61	1	COG2911	COG2911, Uncharacterized protein conserved in bacteria [Function unknown].	40.68	0.000137	20%	38%	636--897	766--1006	1	BAB83470	Vp260 like protein	1853.95	0.00E+00	68%	79%	10--1346	1--1336
														2	BAB83469	Vp260 like protein	1542.32	0.00E+00	55%	68%	10--1467	1--1464
														3	BAB83468	Vp260 like protein	1515.75	0.00E+00	53%	67%	10--1467	1--1462
														4	BAB83467	Vp260 like protein	583.56	1.83E-164	35%	52%	4--1114	5--1121
														5	BAB83471	Vp260 like protein	527.71	1.19E-147	36%	53%	1--914	1--923
														6	NP_048470	PBCV-1 Vp260 protein	190.66	3.45E-46	28%	40%	35--857	59--973
														7	AAA86307	glycoprotein Vp260	179.87	6.09E-43	30%	42%	35--691	59--748
														8	NP_048377	Asn/Thr/Ser/Val rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	120.17	5.72E-25	24%	40%	19--687	15--755
														9	NP_048362	Asn/Thr/Ser/Val rich protein	115.16	1.84E-23	21%	36%	63--1308	16--1183
														10	NP_048366	Asn/Thr/Ser/Ile rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	114.39	3.14E-23	21%	36%	65--904	19--966
M055R	26539--30921	1461	149,623	5.55	1	COG2911	COG2911, Uncharacterized protein conserved in bacteria [Function unknown].	41.45	8.60E-05	19%	34%	374--893	179--1000	1	BAB83469	Vp260 like protein	1578.53	0.00E+00	56%	69%	10--1461	1--1464
														2	BAB83470	Vp260 like protein	1574.30	0.00E+00	61%	73%	10--1335	1--1331
														3	BAB83468	Vp260 like protein	1570.44	0.00E+00	55%	69%	10--1461	1--1462
														4	BAB83467	Vp260 like protein	628.63	4.93E-178	37%	54%	4--1110	5--1118
														5	BAB83471	Vp260 like protein	570.85	1.22E-160	38%	53%	1--940	1--949
														6	NP_048470	PBCV-1 Vp260 protein	197.59	2.81E-48	28%	40%	54--957	21--922
														7	AAA86307	glycoprotein Vp260	174.10	3.33E-41	27%	40%	54--806	21--793
														8	NP_048377	Asn/Thr/Ser/Val rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	135.96	1.00E-29	25%	40%	63--815	19--805
														9	NP_048366	Asn/Thr/Ser/Ile rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	130.57	4.22E-28	23%	38%	17--854	10--951
														10	NP_048362	Asn/Thr/Ser/Val rich protein	121.32	2.56E-25	23%	37%	34--855	292--1156
M061R	31058--35458	1467	150,571	4.90	1	COG2911	COG2911, Uncharacterized protein conserved in bacteria [Function unknown].	41.06	0.0001	21%	40%	636--891	766--998	1	BAB83469	Vp260 like protein	1596.25	0.00E+00	57%	69%	10--1467	1--1464
					2	COG1315	COG1315, Predicted polymerase, most proteins contain PALM domain, HD hydrolase domain and Zn-ribbon domain [DNA replication, recombination, and repair].	37.22	0.001453	26%	38%	636--775	266--399	2	BAB83470	Vp260 like protein	1587.39	0.00E+00	62%	73%	10--1322	1--1311
					3	pfam03961	DUF342, Protein of unknown function (DUF342). This family of bacterial proteins has no known function. The proteins are in the region of 500-600 amino acid residues in length..	34.54	0.009318	30%	45%	636--784	191--313	3	BAB83468	Vp260 like protein	1582.39	0.00E+00	56%	70%	10--1467	1--1462
														4	BAB83467	Vp260 like protein	620.16	1.76E-175	36%	53%	4--1150	5--1147
														5	BAB83471	Vp260 like protein	570.08	2.09E-160	38%	53%	1--940	1--949
														6	NP_048470	PBCV-1 Vp260 protein	213.77	3.80E-53	30%	42%	54--860	21--844
														7	AAA86307	glycoprotein Vp260	194.13	3.12E-47	31%	43%	14--692	48--749
														8	NP_048377	Asn/Thr/Ser/Val rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	134.04	3.83E-29	23%	38%	63--810	19--760
														9	NP_048366	Asn/Thr/Ser/Ile rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	125.95	1.04E-26	21%	38%	65--1040	19--1089
														10	NP_048362	Asn/Thr/Ser/Val rich protein	123.25	6.76E-26	21%	36%	9--1067	166--1332
M070L	36685--36332	118	13,453	9.58		No Hit Found									No Hit Found	No Hit Found						
M071R	36740--41230	1497	156,173	5.27	1	COG1664	CcmA, Integral membrane protein CcmA involved in cell shape determination [Cell envelope biogenesis, outer membrane]	37.21	0.001544	19%	36%	714--811	34--125	1	BAB83467	Vp260 like protein	1362.05	0.00E+00	62%	72%	1--1163	1--1142
					2	COG4801	COG4801, Predicted acyltransferase [General function prediction only].	36.45	0.00259	23%	44%	514--751	40--232	2	BAB83469	Vp260 like protein	676.01	0.00E+00	33%	49%	13--1486	3--1455
					3	COG2911	COG2911, Uncharacterized protein conserved in bacteria [Function unknown].	35.67	0.004858	17%	34%	198--815	120--1006	3	BAB83468	Vp260 like protein	653.67	0.00E+00	32%	49%	13--1481	3--1448
														4	BAB83470	Vp260 like protein	626.32	2.51E-177	33%	49%	13--1345	3--1306
														5	BAB83471	Vp260 like protein	609.37	3.18E-172	39%	55%	3--950	2--954
														6	NP_048470	PBCV-1 Vp260 protein	253.83	3.39E-65	27%	41%	14--930	24--977
														7	AAA86307	glycoprotein Vp260	216.85	4.60E-54	27%	41%	14--799	24--828
														8	NP_048366	Asn/Thr/Ser/Ile rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	148.29	2.01E-33	22%	36%	18--1352	6--1328
														9	NP_048362	Asn/Thr/Ser/Val rich protein	144.82	2.22E-32	22%	36%	32--1074	92--1232
														10	NP_048377	Asn/Thr/Ser/Val rich protein; similar to Rickettsia cell surface antigen, corresponds to Swiss-Prot Accession Number P15921	131.72	1.94E-28	22%	37%	7--812	26--796

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
M078R	41286--42515	410	46,456	6.32	1	pfam04451	Capsid_Iridovir, Iridovirus major capsid protein. This family includes the major capsid protein of iridoviruses, chlorella virus and Spodoptera ascovirus, which are all dsDNA viruses with no RNA stage. This is the most abundant structural protein and can account for up to 45% of virion protein. In Chlorella virus MT325 the major capsid protein is a glycoprotein	286.05	2.10E-78	37%	54%	5--406	3--443	1	AAC27494	putative capsid protein	483.80	4.34E-135	99%	100%	21--250	1--230
					2	COG0770	MurF, UDP-N-acetylmuramyl pentapeptide synthase [Cell envelope biogenesis, outer membrane]	37.21	0.001475	32%	47%	10--68	382--441	2	NP_048787	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	300.83	5.21E-80	41%	55%	5--410	3--437
														3	BAA76601	major capsid protein MCP1	298.13	3.38E-79	40%	55%	5--410	3--437
														4	BAA22198	major capsid protein Vp54	293.51	8.32E-78	40%	54%	5--410	3--437
														5	BAA76600	major capsid protein	292.35	1.85E-77	40%	54%	5--410	3--436
														6	AAC27492	major capsid protein Vp49	289.27	1.57E-76	41%	54%	5--410	3--432
														7	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	271.55	3.39E-71	39%	54%	27--410	1--413
														8	1M3Y_D	Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid Containing, Dna Virus	271.55	3.39E-71	39%	54%	27--410	1--413
														9	NP_048359	contains aminoacyl-tRNA synthetase class-II signature	233.03	1.33E-59	34%	53%	5--410	3--403
														10	NP_048914	similar to Chlorella virus PBCV-1 ORF A11L, corresponds to GenBank Accession Number U42580	230.34	8.65E-59	32%	53%	5--410	2--400
M081R	42549--43169	207	23,405	11.45		No Hit Found									No Hit Found	No Hit Found						
M083R	43344--43598	85	9,762	7.49		No Hit Found									No Hit Found	No Hit Found						
M085R	43716--45284	523	57,632	9.71	1	smart00636	Glyco_18, Glyco_18 domain..	234.83	5.55E-63	38%	57%	16--372	1--335	1	NP_048613	PBCV-1 chitinase	627.48	3.37E-178	74%	82%	1--390	1--392
					2	pfam00704	Glyco_hydro_18, Glycosyl hydrolases family 18..	209.60	2.35E-55	35%	50%	16--372	1--323	2	CAB44709	chitinase	225.33	3.86E-57	34%	52%	1--372	19--384
					3	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane]	68.25	8.23E-13	45%	51%	397--484	52--136	3	AAZ32788	chitinase	224.94	5.04E-57	34%	52%	1--372	19--384
					4	pfam05887	Trypan_PARP, Procyclic acidic repetitive protein (PARP). This family consists of several Trypanosoma brucei procyclic acidic repetitive protein (PARP) like sequences. The procyclic acidic repetitive protein (parp) genes of Trypanosoma brucei encode a small family of abundant surface proteins whose expression is restricted to the procyclic form of the parasite. They are found at two unlinked loci, parpA and parpB; transcription of both loci is developmentally regulated	66.16	2.84E-12	37%	56%	394--469	47--122	4	EAA73155	hypothetical protein FG03591.1	222.63	2.50E-56	35%	52%	14--372	30--377
					5	pfam05616	Neisseria_TspB, Neisseria meningitidis TspB protein. This family consists of several Neisseria meningitidis TspB virulence factor proteins..	58.60	6.68E-10	29%	34%	405--492	323--410	5	AAP04616	chitinase	221.86	4.26E-56	35%	51%	1--372	19--384
					6	pfam06735	DUF1210, Protein of unknown function (DUF1210). This family reoresents a conserved region within plant proline-rich proteins..	50.88	1.32E-07	41%	46%	401--481	124--202	6	AAY32603	chitinase	217.62	8.04E-55	33%	51%	1--372	19--384
					7	COG3147	DedD, Uncharacterized protein conserved in bacteria [Function unknown]	46.94	1.95E-06	31%	35%	399--481	65--149	7	AAT37496	putative chitinase	214.93	5.21E-54	33%	50%	12--372	2--352
					8	pfam02318	RPH3A_effector, Rabphilin-3A effector domain. This is a family of proteins involved in protein transport in synaptic vesicles. Rabphilin-3A has been shown to contact Rab3A, a small G protein important in neurotransmitter release, in two distinct areas	46.99	2.05E-06	27%	36%	396--475	165--246	8	CAA45468	chitinase	213.39	1.52E-53	33%	49%	3--389	25--398
					9	pfam05539	Pneumo_att_G, Pneumovirinae attachment membrane glycoprotein G..	45.54	5.82E-06	20%	31%	395--485	228--323	9	AAF19616	42 kDa endochitinase	213.00	1.98E-53	34%	48%	11--389	34--399
					10	pfam02993	MCPVI, Minor capsid protein VI. This minor capsid protein may act as a link between the external capsid and the internal DNA-protein core. The C-terminal 11 residues may function as a protease cofactor leading to enzyme activation..	45.40	6.34E-06	29%	37%	396--484	110--205	10	AAF19613	42 kDa endochitinase	213.00	1.98E-53	34%	48%	11--389	34--399
M086R	45344--45931	196	22,100	8.10		No Hit Found								1	NP_048655	A301L	59.69	5.64E-08	30%	52%	3--132	2--135
M088R	45967--46605	213	23,543	4.68		No Hit Found								1	NP_048652	A298L	158.30	1.39E-37	42%	59%	1--208	1--220
M090L	47179--46685	165	18,423	12.17		No Hit Found								1	NP_048650	A296R	80.49	1.94E-14	64%	76%	80--144	72--135
M091R	47173--48201	343	39,379	9.13	1	pfam01374	Glyco_hydro_46, Glycosyl hydrolase family 46. This family are chitosanase enzymes..	237.25	1.06E-63	39%	50%	117--324	1--215	1	BAA20342	vChta-1	412.54	9.50E-114	64%	73%	18--335	5--321
					2	cd00978	chitosanase_glyco_hydro_46, Glycosyl hydrolase family 46 chitosanase domain. This family are composed of the chitosanase enzymes which hydrolyzes chitosan, a biopolymer of beta (1,4)-linked -D-glucosamine (GlcN) residues produced by partial or full deacetylation of chitin. Chitosanases play a role in defense against pathogens such as fungi and are found in microorganisms, fungi, viruses, and plants. Microbial chitosanases who members are the most prevalent can be divided into 3 subclasses based on the specificity of the cleavage positions for partial acetylated chitosan. Subclass I chitosanases such as N174 can split GlcN-GlcN and GlcNAc-GlcN linkages, whereas subclass II chitosanases such as Bacillus sp. no. 7-M can cleave only GlcN-GlcN linkages. Subclass III chitosanases such as MH-K1 chitosanase are the most versatile and can split both GlcN-GlcN and GlcN-GlcNAc linkages..	202.49	2.90E-53	30%	44%	107--324	1--220	2	NP_048646	PBCV-1 chitosanase	408.30	1.79E-112	62%	73%	18--335	5--321
														3	BAA94840	chitosanase	84.34	5.93E-15	29%	47%	96--304	41--252
														4	CAB14630	chitosanase	83.96	7.74E-15	29%	42%	107--328	37--264
														5	AAA19865	chitosanase precursor	81.65	3.84E-14	28%	44%	93--329	17--266
														6	P48846	Chitosanase precursor	81.26	5.02E-14	28%	46%	101--329	39--266
														7	ZP_00908039	chitosanase	79.72	1.46E-13	31%	48%	117--324	38--246
														8	1CHK_B	Chain B, Streptomyces N174 Chitosanase Ph5.5 298k	79.34	1.91E-13	29%	45%	117--329	15--226
														9	ABC17783	secreted chitosanase precursor	78.95	2.49E-13	27%	44%	107--324	51--274
														10	BAB19276	chitosanase	76.26	1.61E-12	28%	44%	107--328	38--265
M093L	49565--48435	377	43,656	9.08		No Hit Found								1	NP_048711	A354R	191.82	3.03E-47	42%	60%	135--376	1--235
														2	NP_048779	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	72.79	2.05E-11	28%	47%	85--345	58--319
														3	NP_048435	A87R	68.94	2.96E-10	31%	46%	194--373	270--447
M097L	50625--49549	359	39,960	8.73		No Hit Found								1	NP_048640	A286R	363.23	7.11E-99	51%	69%	17--344	22--362
M099R	50664--51170	169	18,220	8.23	1	pfam00080	Sod_Cu, Copper/zinc superoxide dismutase (SODC). superoxide dismutases (SODs) catalyse the conversion of superoxide radicals to hydrogen peroxide and molecular oxygen. Three evolutionarily distinct families of SODs are known, of which the copper/zinc-binding family is one. Defects in the human SOD1 gene cause familial amyotrophic lateral sclerosis (Lou Gehrig's disease). Structure is an eight-stranded beta sandwich. similar to the immunoglobulin fold..	179.67	2.21E-46	58%	74%	16--165	1--152	1	NP_048593	contains Cu/Zn superoxide dismutase signatures 1 and 2; similar to Neurospora Cu/Zn-superoxide dismutase, corresponds to Swiss-Prot Accession Number P07509	261.92	5.00E-69	76%	85%	10--169	28--187
														2	XP_503850	hypothetical protein	180.64	1.46E-44	60%	71%	16--162	1--151
														3	Q751L8	Superoxide dismutase	179.87	2.50E-44	57%	72%	16--164	1--153
														4	P09670	Superoxide dismutase	179.49	3.26E-44	59%	71%	18--164	3--150
														5	AAT99430	copper-zinc superoxide dismutase	179.10	4.26E-44	58%	72%	16--162	1--151
														6	AAX84946	Cu/Zn superoxide dismutase	177.95	9.49E-44	58%	70%	18--165	4--152
														7	CAG58285	unnamed protein product	177.18	1.62E-43	57%	71%	16--162	1--151
														8	O46412	Superoxide dismutase	177.18	1.62E-43	58%	70%	18--165	4--152
														9	AAB88116	superoxide dismutase	177.18	1.62E-43	58%	71%	18--165	4--152
														10	1YSO	Yeast Cu, Zn Superoxide Dismutase With The Reduced Bridge Broken	177.18	1.62E-43	58%	72%	17--162	1--150
M100L	51754--51179	192	22,991	7.15		No Hit Found								1	NP_048651	A297L	195.28	8.13E-49	53%	72%	4--172	3--171
														2	BAE64611	unnamed protein product	56.23	5.89E-07	27%	50%	4--117	6--119
														3	EAL88370	phosphoglycerate mutase family protein, putative	53.14	4.99E-06	25%	50%	4--143	26--158
														4	EAL86798	phosphoglycerate mutase family protein, putative	49.29	7.21E-05	24%	43%	4--164	22--174
														5	BAB49344	mlr2143	49.29	7.21E-05	27%	47%	4--117	9--123
M101R	51793--52629	279	31,108	6.29	1	cd00542	PVA, Penicillin V acylase (PVA), also known as conjugated bile salt acid hydrolase (CBAH), catalyzes the hydrolysis of penicillin V to yield 6-amino penicillanic acid (6-APA), an important key intermediate of semisynthetic penicillins. PVA has an N-terminal nucleophilic cysteine as do other Ntn Hydrolases which is exposed by prossessing of the PVA precursor. PVA forms a homotetramer	191.58	5.57E-50	33%	48%	2--275	1--297	1	NP_048638	PBCV-1 amidase	334.34	2.38E-90	57%	74%	1--278	1--279

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					2	COG3049	COG3049, Penicillin V acylase and related amidases [Cell envelope bioogenesis. outer membrane].	172.05	3.72E-44	29%	46%	1--278	22--323	2	ZP_00318651	COG3049: Penicillin V acylase and related amidases	145.98	1.20E-33	31%	48%	2--276	4--304
					3	pfam02275	CBAH, Linear amide C-N hydrolases, choloylglycine hydrolase family. This family includes several hydrolases which cleave carbon-nitrogen bonds, other than peptide bonds, in linear amides. These include choloylglycine hydrolase (conjugated bile acid hydrolase, CBAH) EC:3.5.1.24, penicillin acylase EC:3.5.1.11 and acid ceramidase EC:3.5.1.23.	147.84	7.49E-37	32%	50%	2--275	1--298	3	AAP08002	Choloylglycine hydrolase	143.67	5.96E-33	31%	48%	1--275	1--298
					4	cd01935	CGH_like, Choloylglycine hydrolase (CGH)_like. This family of choloylglycine hydrolases-like proteins includes conjugated bile acid hydrolase (CBAH), penicillin acylase and acid ceramidase which cleave carbon-nitrogen bonds, other than peptide bonds, in linear amides..	116.83	1.54E-27	29%	42%	2--268	1--270	4	ABB11030	Penicillin amidase	141.35	2.96E-32	29%	46%	1--278	1--315
					5	cd01902	CGH, CGH Choloylglycine hydrolase (also known as bile salt hydrolase) is an intestinal bacterial enzyme responsible for the deconjugation and subsequent dehydroxylation of conjugated cholic acid (CA) to form deoxycholic acid (DCA). CGH has a conserved Ntn hydrolase fold similar to those of penicillin v acylase (PVA) and acid ceramidase (AC)..	80.82	1.27E-16	27%	46%	2--261	1--270	5	AAT60567	choloylglycine hydrolase	140.58	5.04E-32	30%	47%	1--275	1--298
														6	ZP_00238968	choloylglycine hydrolase family protein	138.27	2.50E-31	29%	47%	1--275	1--298
														7	AAL51724	CHOLOYLGLYCINE HYDROLASE	133.65	6.17E-30	29%	45%	2--276	37--337
														8	AAN30379	choloylglycine hydrolase family protein	133.65	6.17E-30	29%	45%	2--276	3--303
														9	YP_414856	Choloylglycine hydrolase	133.65	6.17E-30	29%	45%	2--276	3--303
														10	AAT33011	choloylglycine hydrolase family protein	132.88	1.05E-29	28%	47%	1--275	1--298
M103L	53704--52817	296	32,914	7.01	1	COG0388	COG0388, Predicted amidohydrolase [General function prediction only].	181.89	4.09E-47	34%	52%	1--296	1--272	1	NP_048426	contains ATP/GTP-binding site motif A; similar to rat beta-alanine synthetase. corresponds to Swiss-Prot Accession Number Q03248	420.24	3.63E-116	68%	79%	2--295	4--297
					2	pfam00795	CN_hydrolase, Carbon-nitrogen hydrolase. This family contains hydrolases that break carbon-nitrogen bonds. The family includes: Nitrilase EC:3.5.5.1, Aliphatic amidase EC:3.5.1.4, Biotidinase EC:3.5.1.12. Beta-ureidodoropionase EC:3.5.1.6..	143.56	1.70E-35	34%	53%	4--177	1--174	2	ZP_00403767	COG0388: Predicted amidohydrolase	278.87	1.31E-73	46%	65%	5--294	6--289
					3	COG0815	Lnt, Apolipoprotein N-acyltransferase [Cell envelope biogenesis, outer membrane].	51.17	1.10E-07	23%	36%	35--215	269--451	3	BAC60037	putative carbon-nitrogen hydrolase	278.10	2.23E-73	49%	65%	6--291	8--280
														4	AAK99627	Beta-alanine synthase or beta-ureidopropionase	276.94	4.96E-73	46%	65%	5--294	6--289
														5	CAB45873	beta-alanine synthase	276.17	8.46E-73	47%	63%	4--295	9--291
														6	ABB08352	Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase	273.86	4.20E-72	47%	62%	4--295	7--294
														7	ZP_00831984	COG0388: Predicted amidohydrolase	273.86	4.20E-72	46%	63%	2--294	3--285
														8	ABA40443	beta-alanine synthase-like protein	273.86	4.20E-72	46%	63%	4--295	9--291
														9	CAG77171	putative carbon-nitrogen hydrolase	271.94	1.60E-71	46%	62%	2--294	3--285
														10	AAS63658	putative carbon-nitrogen hydrolase	271.94	1.60E-71	46%	63%	2--294	3--285
M107R	53804--54709	302	36,038	10.66		No Hit Found								1	NP_048591	A243R	483.03	4.72E-135	76%	87%	1--302	1--302
M109R	55071--55517	149	16,949	8.37		No Hit Found								1	NP_048604	A253R	72.02	5.73E-12	34%	53%	25--147	20--146
														2	AAU06294	hypothetical protein A253R	72.02	5.73E-12	35%	53%	25--147	20--146
														3	AAU06289	hypothetical protein A253R	70.86	1.28E-11	33%	53%	25--149	20--148
														4	AAU06293	hypothetical protein A253R	51.22	1.05E-05	38%	57%	25--97	20--91
														5	AAU06292	hypothetical protein A253R	50.83	1.37E-05	39%	57%	25--93	20--87
														6	AAU06291	hypothetical protein A253R	50.06	2.33E-05	38%	56%	25--97	20--91
M111R	55571--56287	239	27,298	9.91	1	cd00283	GIY-YIG_Cterm, GIYX(10-11)YIG family of class I homing endonucleases C-terminus (GIY-YIG_Cterm). Homing endonucleases promote the mobility of intron or intein by recognizing and cleaving a homologous allele that lacks the sequence. They catalyze a double-strand break in the DNA near the insertion site of that element to facilitate homing at that site. Class I homing endonucleases are sorted into four families based on the presence of these motifs in their respective N-termini: LAGLIDADG, His-Cys box, HNH, and GIY-YIG. This CD contains several but not all members of the GIY-YIG family. The C-terminus of GIY-YIG is a DNA-binding domain which is separated from the N-terminus by a long, flexible linker. The DNA-binding domain consists of a minor-groove binding alpha-helix, and a helix-turn-helix. Some also contain a zinc finger (i.e. I-TevI) which is not required for DNA binding or catalysis, but is a component of the linker and directs the catalytic domain to cleave the homing site at a fixed distance from the intron insertion site.	61.17	9.29E-11	38%	59%	110--235	3--113	1	NP_048671	A315L	200.29	4.06E-50	43%	61%	1--237	1--242
					2	smart00497	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished). .	39.35	0.000336	40%	63%	182--237	1--53	2	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	192.59	8.47E-48	44%	59%	1--236	1--224
					3	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	37.75	0.001287	33%	49%	1--89	1--83	3	NP_048641	PBCV-1 33kd peptide	132.88	7.96E-30	36%	53%	11--236	15--248
					4	pfam07453	NUMOD1, NUMOD1 domain..	36.95	0.001811	50%	68%	182--210	1--29	4	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	128.64	1.50E-28	37%	53%	1--208	1--196
														5	NP_899393	SegD	66.63	7.01E-10	28%	45%	1--228	1--231
														6	AAC49244	ORF301	65.08	2.04E-09	28%	43%	11--171	84--236
														7	YP_293795	putative endonuclease	65.08	2.04E-09	41%	58%	2--79	3--81
														8	CAA38813	GIY ND1 i4 grp IB protein b	63.54	5.93E-09	29%	41%	11--198	77--247
														9	NP_074961	orf261	63.54	5.93E-09	29%	41%	11--198	76--246
														10	AAC49248	ORF211	60.08	6.56E-08	33%	46%	53--177	50--140
M114L	56971--56294	226	24,793	8.60		No Hit Found								1	NP_048616	A262L	122.48	9.53E-27	47%	66%	111--215	2--103
M116L	57762--56998	255	29,094	9.41		No Hit Found								1	NP_048619	A265L	198.36	1.74E-49	56%	68%	85--254	1--170
M120L	58145--57789	119	13,637	4.55		No Hit Found									No Hit Found	No Hit Found						
M121R	58225--58503	93	10,134	9.41		No Hit Found									No Hit Found	No Hit Found						
M122R	58523--58945	141	16,343	11.34		No Hit Found								1	NP_048484	a136R	58.92	5.04E-08	31%	48%	1--141	1--146
M124L	60186--59143	348	38,758	4.79	1	cd02180	GH16_laminarinase, Laminarinase, also known as glucan endo-1,3-beta-D-glucosidase, is a glycosyl hydrolase family 16 member that hydrolyzes 1,3-beta-D-glucosidic linkages in 1,3-beta-D-glucans such as laminarins, curdlans, paramylons, and pachymans, with very limited action on mixed-link (1,3-1,4-)-beta-D-glucans.	142.71	2.61E-35	36%	51%	88--343	1--236	1	NP_048442	PBCV-1 beta-1,3 glucanase	410.22	4.83E-113	54%	70%	9--345	12--361
					2	cd00413	Glyco_hydrolase_16, The O-Glycosyl hydrolases are a widespread group of enzymes that hydrolyse the glycosidic bond between two or more carbohydrates, or between a carbohydrate and a non-carbohydrate moiety. A glycosyl hydrolase classification system based on sequence similarity has led to the definition of more than 95 different families inlcuding glycoside hydrolase family 16. Family 16 includes lichenase, xyloglucan endotransglycosylase (XET), beta-agarase, kappa-carrageenase, endo-beta-1,3-glucanase, endo-beta-1,3-1,4-glucanase, and endo-beta-galactosidase, all of which have a conserved jelly roll fold with a deep active site channel harboring the catalytic residues..	94.29	1.14E-20	34%	48%	92--343	1--216	2	P23903	Glucan endo-1,3-beta-glucosidase A1 precursor ((1->3)-beta-glucan endohydrolase) ((1->3)-beta-glucanase A1)	117.86	4.95E-25	31%	46%	74--343	410--678
					3	cd02182	GH16_laminarinase_like, A beta-1,3-glucanase (laminarinase)-like protein exists in the bacterial genus Streptomyces as well as the fungal class Sordariomycetes. The laminarinases belong to glycosyl hydrolase family 16 all of which have a conserved jelly roll fold with an active site channel. The bacterial members contain an additional C-terminal carbohydrate-binding module (CBM)	79.95	2.18E-16	29%	43%	85--343	1--256	3	YP_435911	Beta-glucanase/Beta-glucan synthetase	117.09	8.45E-25	32%	46%	79--343	322--571
					4	COG2273	SKN1, Beta-glucanase/Beta-glucan synthetase [Carbohydrate transport and metabolism].	64.69	8.99E-12	26%	39%	74--348	29--267	4	AAC60453	beta-1,3-glucanase	115.16	3.21E-24	31%	44%	84--343	424--681
					5	pfam00722	Glyco_hydro_16, Glycosyl hydrolases family 16..	59.47	3.46E-10	28%	45%	190--342	45--182	5	EAN71367	Glycoside hydrolase, family 16	113.24	1.22E-23	32%	45%	85--343	48--325
					6	cd02179	GH16_beta_GRP, Beta-GRP (beta-1,3-glucan recognition protein) is one of several pattern recognition receptors (PRRs), also referred to as biosensor proteins, that complexes with pathogen-associated beta-1,3-glucans and then transduces signals necessary for activation of an appropriate immune response. Their structures adopt a jelly roll fold with a deep active site channel harboring the catalytic residues, like those of other glycosyl hydrolase family 16 members.	58.78	4.92E-10	24%	37%	89--329	2--283	6	ZP_00637494	Glucan endo-1,3-beta-D-glucosidase	111.31	4.64E-23	31%	43%	85--343	48--325

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					7	cd02183	GH16_GPI_glucanosyltransferase, GPI (glycosylphosphatidylinositol) -glucanosyltransferase is a GPI-anchored membrane protein present in the fungal cell wall that is thought to play an important role in cell wall biosynthesis. GPI-glucanosyltransferase belongs to a family of glycosyl hydrolases that includes lichenase, xyloglucan endotransglycosylase (XET), beta-agarase, kappa-carrageenase, endo-beta-1,3-glucanase, endo-beta-1,3-1,4-glucanase, and endo-beta-galactosidase, all of which have a conserved jelly roll fold with a deep active site channel harboring the catalytic residues.	38.68	0.000569	28%	49%	190--283	58--144	7	BAD63242	endo-beta-1,3-glucanase	109.00	2.30E-22	31%	45%	83--343	31--277
					8	cd02175	GH16_lichenase, Lichenase, also known as 1,3-1,4-beta-glucanase, is a member of glycosyl hydrolase family 16, that specifically cleaves 1,4-beta D-glucosidic bonds in mixed-linked beta glucans that also contain 1,3-beta-D-glucosidic linkages. Natural substrates of beta-glucanase are beta glucans from grain endosperm cell walls or lichenan from the Islandic moss, Cetraria islandica. This protein is found not only in bacteria but also in anaerobic fungi. This domain includes two seven-stranded antiparallel beta-sheets that are adjacent to one another forming a compact, jelly-roll beta-sandwich structure.	38.74	0.000637	29%	46%	126--281	30--157	8	BAC06195	1,3-(1,3;1,4)-beta-D-glucan 3(4)-glucanohydrolase	107.84	5.12E-22	32%	44%	89--347	34--263
					9	cd02177	GH16_kappa_carrageenase, Kappa-carrageenase degrades kappa-carrageenans which are the gel-forming, sulfated 1,3-alpha-1,4-beta-galactans that make up the cell walls of marine red algae such as Rhodophyceaea. Kappa-carrageenases exist in bacteria belonging to at least three phylogenetically distant branches, including pseudoalteromonas, planctomycetes, and baceroidetes. This domain adopts a curved beta-sandwich conformation, with a tunnel-shaped active site cavity, referred to as a jellyroll fold.	35.47	0.005539	30%	47%	81--212	3--117	9	AAO08191	Beta-glucanase/Beta-glucan synthetase	105.53	2.54E-21	28%	42%	49--343	12--333
														10	BAC96166	conserved hypothetical protein	104.38	5.67E-21	28%	42%	49--343	23--344
M128R	60357--62039	561	64,127	7.38	1	COG1215	COG1215, Glycosyltransferases, probably involved in cell wall biogenesis [Cell envelope biogenesis, outer membrane]	91.53	8.00E-20	22%	42%	33--483	10--420	1	AAD26641	hyaluronan synthase	858.98	0.00E+00	70%	83%	1--561	8--568
					2	pfam03142	Chitin_synth_2, Chitin synthase. Members of this family are fungal chitin synthase EC:2.4.1.16 enzymes. They catalyse chitin synthesis as follows: UDP-N-acetyl-D-glucosamine + {(1,4)-(N-acetyl-beta-D-glucosaminyl)}(N) <=> UDP + {(1,4)-(N-acetyl-beta-D-glucosaminyl)}(N+1)	58.00	8.14E-10	23%	44%	186--345	193--373	2	NP_048446	PBCV-1 hyaluronic acid synthetase	857.06	0.00E+00	70%	82%	1--561	8--568
					3	pfam00535	Glycos_transf_2, Glycosyl transferase. Diverse family, transferring sugar from UDP-glucose, UDP-N-acetyl- galactosamine, GDP-mannose or CDP-abequose, to a range of substrates including cellulose, dolichol phosphate and teichoic acids..	44.83	9.01E-06	24%	44%	86--283	2--167	3	AAD26643	hyaluronan synthase	855.13	0.00E+00	70%	82%	1--561	8--568
														4	NP_037285	hyaluronan synthase 2	224.17	9.43E-57	30%	49%	7--530	12--532
														5	NP_005319	hyaluronan synthase 2	223.79	1.23E-56	30%	49%	7--530	12--532
														6	BAC37733	unnamed protein product	223.79	1.23E-56	30%	49%	7--530	12--532
														7	NP_032242	hyaluronan synthase 2	223.79	1.23E-56	30%	49%	7--530	12--532
														8	XP_528222	PREDICTED: similar to hyaluronan synthase 2	223.79	1.23E-56	30%	49%	7--530	24--544
														9	NP_776504	hyaluronan synthase 2	222.63	2.74E-56	30%	49%	7--530	12--532
														10	BAB63264	hyaluronic acid synthase 2	222.25	3.58E-56	29%	49%	7--530	12--532
M133R	62232--63188	319	35,900	8.34	1	pfam01331	mRNA_cap_enzyme, mRNA capping enzyme, catalytic domain. This family represents the ATP binding catalytic domain of the mRNA capping enzyme..	113.84	1.26E-26	27%	44%	50--223	1--192	1	NP_048451	PBCV-1 mRNA guanylyltransferase	331.26	2.50E-89	53%	69%	2--314	12--324
					2	COG5226	CEG1, mRNA capping enzyme, guanylyltransferase (alpha) subunit [RNA processing and modification].	90.48	1.33E-19	22%	37%	42--317	39--362	2	1CKN_B	Chain B, Structure Of Guanylylated Mrna Capping Enzyme Complexed With Gto	328.95	1.24E-88	52%	69%	2--314	12--324
					3	pfam03919	mRNA_cap_C, mRNA capping enzyme, C-terminal domain..	47.66	1.08E-06	26%	43%	228--314	3--109	3	NP_612286	putative RNA guanylytransferase	74.71	4.21E-12	27%	43%	44--311	221--479
					4	COG1793	CDC9, ATP-dependent DNA ligase [DNA replication, recombination, and repair].	35.75	0.004237	24%	41%	121--257	215--355	4	AAX82373	mRNA capping enzyme	72.79	1.60E-11	26%	42%	44--311	220--478
														5	AAT68133	mRNA capping enzyme	72.40	2.09E-11	23%	45%	8--317	234--566
														6	CAG09212	unnamed protein product	72.40	2.09E-11	24%	43%	9--317	234--560
														7	NP_974263	mRNA guanylyltransferase/ phosphoprotein phosphatase/ protein tyrosine/serine/threonine phosphatase	68.94	2.31E-10	22%	42%	33--318	337--650
														8	P78587	mRNA capping enzyme alpha subunit (mRNA guanylyltransferase) (GTP-RNA guanylyltransferase) (GTase)	67.78	5.15E-10	25%	39%	47--316	41--373
														9	1P16_B	Chain B, Structure Of An Mrna Capping Enzyme Bound To The Phosphorylated Carboxyl-Terminal Domain Of Rna Polymerase Ii	66.63	1.15E-09	25%	38%	47--316	41--373
														10	XP_397436	PREDICTED: similar to ENSANGP00000009960	66.24	1.50E-09	24%	42%	47--317	276--571
M136L	63572--63216	119	13,803	8.08		No Hit Found								1	NP_049001	A645R	80.88	1.22E-14	33%	57%	1--117	1--121
M137L	64469--63633	279	31,701	5.23	1	pfam00443	UCH, Ubiquitin carboxyl-terminal hydrolase..	107.79	9.38E-25	20%	40%	4--275	6--312	1	NP_048453	contains ubquitin carboxy-terminal hydrolase active sites; similar to human ubiquitin carboxy-terminal hydrolase, coresponds to Swiss-Prot Accession Number Q09879	261.54	1.96E-68	44%	67%	1--277	1--280
					2	cd02257	Peptidase_C19, Peptidase C19 contains ubiquitinyl hydrolases. They are intracellular peptidases that remove ubiquitin molecules from polyubiquinated peptides by cleavage of isopeptide bonds. They hydrolyse bonds involving the carboxyl group of the C-terminal Gly residue of ubiquitin The purpose of the de-ubiquitination is thought to be editing of the ubiquitin conjugates, which could rescue them from degradation, as well as recycling of the ubiquitin. The ubiquitin/proteasome system is responsible for most protein turnover in the mammalian cell, and with over 50 members, family C19 is one of the largest families of peptidases in the human genome.	78.81	5.43E-16	17%	33%	4--276	2--320	2	EAL49353	ubiquitin carboxyl-terminal hydrolase, putative	60.08	8.66E-08	23%	41%	4--278	68--347
M139L	65483--64548	312	34,925	7.50	1	COG1405	SUA7, Transcription initiation factor TFIIIB, Brf1 subunit/Transcription initiation factor TFIIB [Transcription].	79.58	3.15E-16	22%	42%	44--311	7--282	1	NP_048455	similar to Pyrococcus woesei factor TFIIB homolog, corresponds to GenBank Accession Number X70668	204.53	3.39E-51	35%	57%	30--312	1--290
														2	AAT93251	YPR086W	51.22	4.80E-05	21%	35%	44--304	27--316
														3	AAB68135	Sua7p: Transcription initiation factor IIB (Swiss Prot. accession number P29055)	51.22	4.80E-05	21%	35%	44--304	27--316
														4	CAG61530	unnamed protein product	50.45	8.19E-05	20%	36%	44--304	28--317
M140L	65990--65487	168	19,353	9.14		No Hit Found								1	NP_048457	A109L	128.64	6.63E-29	58%	73%	66--168	1--103
M143L	67855--66041	605	67,784	10.88	1	smart00220	S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-specific kinase subfamily. .	47.90	1.00E-06	33%	48%	63--257	4--151	1	NP_048636	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055	483.80	7.23E-135	44%	59%	1--605	15--568
					2	pfam00069	Pkinase, Protein kinase domain..	46.43	2.48E-06	29%	50%	63--257	4--151	2	NP_048632	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662	459.91	1.12E-127	42%	57%	1--539	1--608
					3	cd00180	S_TKc, Serine/Threonine protein kinases, catalytic domain. Phosphotransferases of the serine or threonine-specific kinase subfamily. The enzymatic activity of these protein kinases is controlled by phosphorylation of specific residues in the activation segment of the catalytic domain, sometimes combined with reversible conformational changes in the C-terminal autoregulatory tail.	44.43	1.02E-05	32%	48%	63--257	5--152	3	NP_048970	RPQT-like (9x)	311.23	6.43E-83	46%	64%	11--344	7--336
					4	pfam01636	APH, Phosphotransferase enzyme family. This family consists of bacterial antibiotic resistance proteins, which confer resistance to various aminoglycosides they include:- aminoglycoside 3'-phosphotransferase or kanamycin kinase / neomycin-kanamycin phosphotransferase and streptomycin 3''-kinase or streptomycin 3''-phosphotransferase. The aminoglycoside phosphotransferases inactivate aminoglycoside antibiotics via phosphorylation. This family also includes homoserine kinase. This family is related to fructosamine kinase pfam03881.	40.52	0.000146	44%	67%	225--251	171--198	4	NP_048441	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563	83.96	1.67E-14	52%	69%	453--539	1--85
					5	COG2334	COG2334, Putative homoserine kinase type II (protein kinase fold) [General function prediction only]	39.56	0.000339	40%	57%	225--255	200--230	5	NP_049032	similar to Chlorella virus PBCV-1 ORF A282L, corresponds to GenBank Accession Number U42580	66.63	2.76E-09	79%	89%	566--604	1--39
M148L	68171--67884	96	10,594	8.41		No Hit Found									No Hit Found	No Hit Found						
M149R	68260--68730	157	17,621	5.03		No Hit Found								1	NP_049020	similar to Chlorella virus CVK2 unknown ORF, corresponds to GenBank Accession Number D16505	145.98	3.21E-34	51%	72%	16--153	1--139
														2	NP_049021	similar to Chlorella virus CVK2 unknown ORF, corresponds to GenBank Accession Number D16505	116.32	2.72E-25	39%	67%	3--152	22--170
M150R	68760--69071	104	12,366	10.95		No Hit Found								1	NP_048469	A121R	124.02	1.24E-27	73%	86%	33--104	25--96
														2	AAL73473	Tlr 6Fp protein	60.85	1.29E-08	44%	64%	32--90	32--88
														3	AAL73467	Tlr 6Fp protein	60.08	2.21E-08	42%	64%	32--90	32--88

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
														4	AAL73477	Tlr 6Fp protein	53.53	2.07E-06	38%	59%	32--90	32--88
M152R	69108--74000	1631	167,109	5.10		No Hit Found								1	NP_048470	PBCV-1 Vp260 protein	461.84	8.96E-128	36%	47%	2--970	3--962
														2	AAA86307	glycoprotein Vp260	362.46	7.40E-98	35%	46%	2--812	3--809
														3	NP_048471	contains a leucine zipper motif; similar to Bacillus subtilis phage P2A preneck appendage protein, corresponds to Swiss-Prot Accession Number P07537	184.50	2.76E-44	37%	53%	1408--1629	41--304
														4	BAB83471	Vp260 like protein	176.41	7.52E-42	22%	37%	28--787	32--861
														5	BAB83467	Vp260 like protein	169.09	1.20E-39	22%	37%	11--825	9--838
														6	BAB83469	Vp260 like protein	164.85	2.26E-38	24%	38%	8--637	152--798
														7	BAB83470	Vp260 like protein	162.16	1.47E-37	24%	39%	8--643	152--805
														8	BAB83468	Vp260 like protein	158.30	2.12E-36	23%	38%	8--637	152--798
														9	XP_637068	hypothetical protein DDB0215928	115.16	2.06E-23	19%	34%	35--1459	441--1786
														10	NP_862008	rb133	109.77	8.64E-22	39%	55%	1448--1590	309--454
M160R	74021--75307	429	48,922	10.90		No Hit Found								1	NP_049032	similar to Chlorella virus PBCV-1 ORF A282L, corresponds to GenBank Accession Number U42580	220.71	7.32E-56	36%	53%	52--423	1--368
														2	NP_048636	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055	162.93	1.81E-38	57%	72%	10--141	418--567
														3	NP_048632	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662	140.20	1.26E-31	51%	70%	6--141	462--608
														4	NP_048441	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563	89.35	2.55E-16	41%	54%	10--143	4--145
M163L	75853--75314	180	20,783	9.34	1	pfam01096	TFIIS, Transcription factor S-II (TFIIS)..	66.15	2.80E-12	68%	74%	141--179	1--39	1	BAA04187	transcription elongation factor SII	242.28	4.91E-63	62%	70%	1--180	1--180
					2	smart00440	ZnF_C2C2, C2C2 Zinc finger; Nucleic-acid-binding motif in transcriptional elongation factor TFIIS and RNA polymerases. .	63.38	2.17E-11	56%	77%	141--180	1--40	2	NP_048472	contains a zinc ribbon domain; similar to Chlorella virus CVU1 TFIIS-like transcription factor, corresponds to GenBank Accession Number D29631	240.35	1.87E-62	62%	71%	1--180	1--180
					3	COG1594	RPB9, DNA-directed RNA polymerase, subunit M/Transcription elongation factor TFIIS [Transcription].	59.66	3.12E-10	56%	75%	143--179	75--111	3	S47662	transcription elongation factor TFIIS homolog - Chlorella virus CV-U1	238.81	5.43E-62	61%	70%	1--180	1--180
														4	BAA04186	transcription elongation factor SII	204.53	1.13E-51	57%	68%	1--164	1--164
														5	CAG98214	unnamed protein product	74.33	1.77E-12	33%	45%	59--179	174--290
														6	AAS54500	AGR011Wp	74.33	1.77E-12	35%	44%	66--179	189--302
														7	XP_762035	hypothetical protein UM05888.1	72.40	6.73E-12	30%	42%	14--179	145--313
														8	CAG58585	unnamed protein product	71.63	1.15E-11	36%	45%	82--179	209--304
														9	CAG86394	unnamed protein product	71.63	1.15E-11	31%	43%	52--179	161--288
														10	CAC19733	tfs1	71.25	1.50E-11	70%	82%	140--179	252--291
M166R	75958--76617	220	24,359	10.28		No Hit Found								1	NP_048475	A127R	291.58	1.13E-77	63%	76%	8--217	30--245
M167L	77379--76624	252	28,463	8.45		No Hit Found								1	NP_048357	A9R	203.37	5.28E-51	53%	77%	87--252	8--173
														2	NP_049005	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	101.68	2.16E-20	29%	45%	8--251	13--253
														3	NP_048807	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	91.66	2.24E-17	25%	46%	8--251	7--246
														4	NP_048629	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	88.97	1.45E-16	24%	44%	9--250	11--248
														5	NP_048525	A177R	80.11	6.74E-14	26%	45%	2--235	4--226
														6	NP_048427	A79R	77.03	5.70E-13	23%	45%	9--230	8--218
														7	AAU06304	hypothetical protein A275R	76.26	9.73E-13	25%	51%	87--250	4--167
														8	AAU06301	hypothetical protein A275R	75.10	2.17E-12	25%	51%	87--250	4--167
														9	AAU06302	hypothetical protein A275R	64.70	2.93E-09	26%	53%	119--251	1--132
M170R	77524--77889	122	13,843	8.14		No Hit Found								1	NP_048396	A48R	93.59	1.80E-18	38%	61%	8--121	5--122
M171L	78147--77896	84	8,672	3.51		No Hit Found									No Hit Found	No Hit Found						
M173R	78230--78679	150	17,359	4.76		No Hit Found								1	AAZ21386	hypothetical protein SAR11_0565	94.36	1.07E-18	39%	61%	1--117	1--115
M174L	80034--78661	458	51,959	5.86	1	COG0637	COG0637, Predicted phosphatase/phosphohexomutase [General function prediction only].	121.37	7.78E-29	31%	47%	9--198	2--190	1	CAH09375	putative nucleotidyltransferase	206.45	1.56E-51	48%	64%	220--450	2--236
					2	COG0546	Gph, Predicted phosphatases [General function prediction only].	101.37	7.60E-23	29%	46%	7--196	2--191	2	ZP_00368786	lipopolysaccharide biosynthesis protein, putative	200.29	1.12E-49	47%	63%	220--450	2--238
					3	pfam00702	Hydrolase, haloacid dehalogenase-like hydrolase. This family are structurally different from the alpha/ beta hydrolase family (pfam00561). This family includes L-2-haloacid dehalogenase, epoxide hydrolases and phosphatases. The structure of the family consists of two domains. One is an inserted four helix bundle, which is the least well conserved region of the alignment. The rest of the fold is composed of the core alpha/beta domain	90.43	1.42E-19	25%	39%	10--190	1--191	3	ZP_00371454	lipopolysaccharide biosynthesis protein, putative	196.82	1.23E-48	49%	63%	220--450	1--236
					4	COG1011	COG1011, Predicted hydrolase (HAD superfamily) [General function prediction only].	77.64	1.07E-15	25%	37%	7--203	2--208	4	AAT91796	putative nucleotidyl transferase family protein	196.05	2.10E-48	44%	63%	220--450	2--240
					5	COG1209	RfbA, dTDP-glucose pyrophosphorylase [Cell envelope biogenesis, outer membrane].	62.88	2.90E-11	25%	42%	223--446	4--228	5	BAE48960	dTDP-glucose pyrophosphorylase	182.19	3.14E-44	44%	59%	217--450	2--240
					6	COG1208	GCD1, Nucleoside-diphosphate-sugar pyrophosphorylase involved in lipopolysaccharide biosynthesis/translation initiation factor 2B, gamma/epsilon subunits (eIF-2Bgamma/eIF-2Bepsilon) [Cell envelope biogenesis, outer membrane / Translation, ribosomal structure and biogenesis]	58.05	8.08E-10	25%	42%	222--447	4--228	6	NP_790553	lipopolysaccharide biosynthesis protein, putative	173.33	1.46E-41	40%	57%	220--450	2--239
					7	pfam00483	NTP_transferase, Nucleotidyl transferase. This family includes a wide range of enzymes which transfer nucleotides onto phosphosugars	56.37	3.20E-09	20%	38%	223--392	3--184	7	ZP_00207801	COG1209: dTDP-glucose pyrophosphorylase	172.17	3.25E-41	42%	59%	226--450	1--229
					8	COG0647	NagD, Predicted sugar phosphatases of the HAD superfamily [Carbohydrate transport and metabolism].	43.29	2.54E-05	23%	36%	10--187	9--228	8	YP_233720	Nucleotidyl transferase	168.32	4.70E-40	38%	58%	220--450	2--239
					9	COG4750	LicC, CTP:phosphocholine cytidylyltransferase involved in choline phosphorylation for cell surface LPS epitopes [Cell envelope biogenesis, outer membrane].	41.05	0.00012	45%	68%	220--258	1--39	9	AAZ34091	lipopolysaccharide biosynthesis protein, putative	164.47	6.78E-39	38%	55%	220--450	2--239
					10	pfam02348	CTP_transf_3, Cytidylyltransferase. This family consists of two main Cytidylyltransferase activities: 1) 3-deoxy-manno-octulosonate cytidylyltransferase, EC:2.7.7.38 catalysing the reaction:- CTP + 3-deoxy-D-manno-octulosonate <=> diphosphate + CMP-3-deoxy-D-manno-octulosonate, 2) acylneuraminate cytidylyltransferase EC:2.7.7.43, catalysing the reaction:- CTP + N-acylneuraminate <=> diphosphate + CMP-N-acylneuraminate. NeuAc cytydilyltransferase of Mannheimia haemolytica has been characterised describing kinetics and regulation by substrate charge, energetic charge and amino-sugar demand	40.31	0.000193	19%	36%	222--392	2--175	10	BAE48961	Nucleoside-diphosphate-sugar pyrophosphorylase	148.29	5.03E-34	35%	54%	220--456	2--243
M177L	81483--80074	470	53,951	4.96		No Hit Found								1	ZP_00520036	Methyltransferase FkbM	67.40	1.17E-09	29%	45%	2--200	402--595
M178L	82188--81529	220	25,947	7.30		No Hit Found								1	YP_163203	hypothetical protein ZMO1468	51.60	1.97E-05	33%	54%	69--168	428--533
M179L	82706--82224	161	18,288	11.05		No Hit Found								1	NP_049018	A662L	210.69	1.17E-53	64%	79%	6--161	16--171
														2	NP_565983	unknown protein	63.93	1.77E-09	24%	49%	27--160	98--232
														3	BAD87006	unknown protein	62.77	3.94E-09	25%	50%	34--160	112--239
														4	XP_328621	hypothetical protein	50.06	2.64E-05	23%	44%	34--156	129--255
														5	AAM62733	contains similarity to 22 kDa peroxisomal membrane protein	49.29	4.51E-05	25%	51%	34--135	122--225
														6	BAB08278	unnamed protein product	49.29	4.51E-05	25%	51%	34--135	121--224
														7	NP_568621	unknown protein	49.29	4.51E-05	25%	51%	34--135	121--224
M180L	83379--82906	158	18,513	10.34		No Hit Found								1	NP_048482	similar to bacteriophage T4 intron-associated endonuclease, corresponds to Swiss-Prot Accession Number P13299	143.67	1.63E-33	52%	64%	1--148	1--148
														2	YP_293795	putative endonuclease	51.99	6.47E-06	34%	52%	8--92	2--88
														3	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	49.29	4.19E-05	34%	53%	9--93	2--85
														4	NP_048671	A315L	48.14	9.34E-05	26%	50%	9--134	2--128
M183R	83430--83714	95	10,293	7.78		No Hit Found								1	ABA40764	potassium ion channel	189.89	1.91E-47	100%	100%	1--95	1--95
														2	AAQ16135	potassium channel protein	83.19	2.52E-15	47%	64%	6--83	16--92
														3	AAQ16140	potassium channel protein	82.80	3.29E-15	46%	64%	6--83	16--92
														4	AAQ16137	potassium channel protein	82.80	3.29E-15	46%	64%	6--83	16--92
														5	NP_048599	PBCV-1 K+ ion channel protein	82.80	3.29E-15	48%	64%	6--83	16--92

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
														6	AAQ16141	potassium channel protein	82.42	4.29E-15	44%	65%	6--83	16--92
														7	AAQ16142	potassium channel protein	82.03	5.61E-15	46%	64%	6--83	16--92
														8	AAQ16138	potassium channel protein	81.65	7.32E-15	44%	64%	6--83	16--92
M184L	84274--83936	113	12,957	10.01		No Hit Found								1	NP_048487	A139L	113.24	2.20E-24	55%	79%	27--107	27--107
M186R	84369--85928	520	58,807	8.65	1	COG1215	COG1215, Glycosyltransferases, probably involved in cell wall biogenesis [Cell envelope biogenesis, outer membrane].	133.91	1.37E-32	20%	35%	58--460	1--412	1	NP_048829	similar to Acetobacter cellulose synthase, corresponds to Swiss-Prot Accession Number P21877	802.74	0.00E+00	72%	84%	1--520	1--517
					2	pfam00535	Glycos_transf_2, Glycosyl transferase. Diverse family, transferring sugar from UDP-glucose, UDP-N-acetyl- galactosamine, GDP-mannose or CDP-abequose, to a range of substrates including cellulose, dolichol phosphate and teichoic acids..	64.09	1.19E-11	18%	37%	118--280	1--164	2	XP_328932	hypothetical protein	436.80	8.37E-121	44%	60%	21--516	168--670
					3	pfam03552	Cellulose_synt, Cellulose synthase. Cellulose, an aggregate of unbranched polymers of beta-1,4-linked glucose residues, is the major component of wood and thus paper, and is synthesised by plants, most algae, some bacteria and fungi, and even some animals. The genes that synthesise cellulose in higher plants differ greatly from the well-characterized genes found in Acetobacter and Agrobacterium sp. More correctly designated as 'cellulose synthase catalytic subunits', plant cellulose synthase (CesA) proteins are integral membrane proteins, approximately 1,000 amino acids in length. There are a number of highly conserved residues, including several motifs shown to be necessary for processive glycosyltransferase activity	43.36	2.30E-05	30%	46%	292--387	517--618	3	ZP_00545426	Cellulose synthase (UDP-forming)	276.56	1.45E-72	33%	52%	23--497	60--550
					4	COG1216	COG1216, Predicted glycosyltransferases [General function prediction only].	42.81	3.38E-05	21%	34%	190--312	82--213	4	NP_790863	cellulose synthase, catalytic subunit	158.30	5.74E-37	34%	51%	111--367	155--417
														5	AAQ60348	cellulose synthase, subunit A	157.53	9.80E-37	34%	51%	115--385	264--540
														6	AAL71842	WssB	154.84	6.35E-36	34%	50%	113--367	157--417
														7	AAM38361	celullose synthase	154.45	8.30E-36	32%	49%	115--405	153--451
														8	NP_643825	celullose synthase	154.45	8.30E-36	32%	49%	115--405	155--453
														9	CAJ25375	cellulose synthase catalytic subunit	154.07	1.08E-35	32%	49%	115--405	155--453
														10	CAI87218	Cellulose synthase catalytic subunit	153.30	1.85E-35	30%	47%	115--449	294--620
M190R	85998--89204	1069	113,278	10.81	1	pfam05887	Trypan_PARP, Procyclic acidic repetitive protein (PARP). This family consists of several Trypanosoma brucei procyclic acidic repetitive protein (PARP) like sequences. The procyclic acidic repetitive protein (parp) genes of Trypanosoma brucei encode a small family of abundant surface proteins whose expression is restricted to the procyclic form of the parasite. They are found at two unlinked loci, parpA and parpB; transcription of both loci is developmentally regulated. .	55.37	6.06E-09	32%	56%	951--1010	64--123	1	BAE02830	surface protein	1631.31	0.00E+00	89%	89%	1--949	1--953
					2	pfam05616	Neisseria_TspB, Neisseria meningitidis TspB protein. This family consists of several Neisseria meningitidis TspB virulence factor proteins..	50.13	2.28E-07	38%	39%	953--1009	329--385	2	NP_048488	PBCV-1 surface protein	737.26	0.00E+00	46%	59%	1--865	1--888
					3	pfam04625	DEC-1_N, DEC-1 protein, N terminal region. The defective chorion-1 gene (dec-1) in Drosophila encodes follicle cell proteins necessary for proper eggshell assembly. Multiple products of the dec-1 gene are formed by alternative RNA splicing and proteolytic processing. Cleavage products include S80 (80 kDa) which is incorporated into the eggshell, and further proteolysis of S80 gives S60 (60 kDa)..	44.49	9.77E-06	41%	41%	951--1002	94--145	3	BAD22850	surface protein	734.18	0.00E+00	45%	59%	1--865	1--888
					4	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	42.82	3.27E-05	36%	43%	931--1007	40--117	4	BAD12236	surface protein	729.55	0.00E+00	45%	58%	1--865	1--888
					5	COG3147	DedD, Uncharacterized protein conserved in bacteria [Function unknown]	41.55	8.14E-05	33%	36%	946--1024	88--166	5	T17636	proline-rich protein A145R - Chlorella virus PBCV-1	78.57	1.35E-12	28%	43%	469--698	2--236
					6	pfam01213	CAP, CAP protein..	39.99	0.00021	30%	37%	867--987	219--326	6	NP_048741	Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	54.30	2.73E-05	43%	58%	890--947	206--263
M200L	89586--89278	103	11,555	5.19		No Hit Found								1	NP_048498	A150L	75.49	5.10E-13	38%	60%	1--100	5--105
M201R	89691--91052	454	50,601	8.41	1	COG1061	SSL2, DNA or RNA helicases of superfamily II [Transcription / DNA replication, recombination, and repair].	140.23	1.66E-34	26%	40%	54--397	7--379	1	NP_048501	similar to phage T5 helicase, corresponds to Swiss-Prot Accession Number P11107	612.45	9.30E-174	65%	81%	3--453	6--459
					2	smart00487	DEXDc, DEAD-like helicases superfamily; .	63.71	1.91E-11	23%	37%	94--233	13--173	2	NP_077551	EsV-1-66	182.19	3.10E-44	32%	50%	21--426	17--443
					3	cd00269	DEXHc, DEXH-box helicases. A diverse family of proteins involved in ATP-dependent DNA or RNA unwinding, needed in a variety of cellular processes. The name derives from the sequence of the Walker B motif (motif III). This domain contains the ATP-binding region..	55.47	5.62E-09	23%	34%	110--230	2--143	3	YP_142750	VV A18 helicase	170.24	1.22E-40	29%	48%	4--430	9--453
					4	cd00046	DEXDc, DEAD-like helicases superfamily. A diverse family of proteins involved in ATP-dependent RNA or DNA unwinding. This domain contains the ATP-binding region..	53.55	1.96E-08	25%	37%	110--230	2--144	4	ZP_00754006	COG1061: DNA or RNA helicases of superfamily II	144.44	7.17E-33	30%	46%	85--430	419--776
					5	COG4096	HsdR, Type I site-specific restriction-modification system, R (restriction) subunit and related helicases [Defense mechanisms].	44.19	1.23E-05	26%	40%	113--231	190--321	5	YP_437404	DNA or RNA helicase of superfamily II	141.35	6.07E-32	29%	46%	87--430	421--776
					6	pfam00176	SNF2_N, SNF2 family N-terminal domain. This domain is found in proteins involved in a variety of processes including transcription regulation (e.g., SNF2, STH1, brahma, MOT1) , DNA repair (e.g., ERCC6, RAD16, RAD5), DNA recombination (e.g., RAD54), and chromatin unwinding (e.g., ISWI) as well as a variety of other proteins with little functional information (e.g., lodestar, ETL1).	43.34	2.39E-05	22%	39%	110--230	18--156	6	BAB34728	hypothetical protein	135.96	2.55E-30	30%	49%	81--430	417--777
					7	pfam00270	DEAD, DEAD/DEAH box helicase. Members of this family include the DEAD and DEAH box helicases. Helicases are involved in unwinding nucleic acids. The DEAD box helicases are involved in various aspects of RNA metabolism, including nuclear transcription, pre mRNA splicing, ribosome biogenesis, nucleocytoplasmic transport, translation, RNA decay and organellar gene expression.	43.10	2.71E-05	18%	32%	102--233	19--175	7	NP_287072	putative helicase	135.96	2.55E-30	30%	49%	81--430	425--785
					8	COG1197	Mfd, Transcription-repair coupling factor (superfamily II helicase) [DNA replication, recombination, and repair / Transcription].	42.17	5.73E-05	26%	43%	89--230	607--752	8	ZP_00665630	Type III restriction enzyme, res subunit:DEAD/DEAH box helicase, N-terminal	131.34	6.28E-29	29%	45%	81--430	481--849
					9	cd00079	HELICc, Helicase superfamily c-terminal domain; associated with DEXDc-, DEAD-, and DEAH-box proteins, yeast initiation factor 4A, Ski2p, and Hepatitis C virus NS3 helicases; this domain is found in a wide variety of helicases and helicase related proteins; may not be an autonomously folding unit, but an integral part of the helicase; 4 helicase superfamilies at present according to the organization of their signature motifs; all helicases share the ability to unwind nucleic acid duplexes with a distinct directional polarity; they utilize the free energy from nucleoside triphosphate hydrolysis to fuel their translocation along DNA, unwinding the duplex in the process	40.68	0.000164	23%	40%	302--395	17--123	9	EAM63429	Type III restriction enzyme, res subunit:DEAD/DEAH box helicase, N-terminal	129.80	1.83E-28	30%	47%	51--430	394--790
					10	COG0610	COG0610, Type I site-specific restriction-modification system, R (restriction) subunit and related helicases [Defense mechanisms].	39.66	0.000293	23%	36%	95--338	258--517	10	ZP_00679441	DEAD/DEAH box helicase:Helicase, C-terminal:Type III restriction enzyme, res subunit	117.86	7.19E-25	29%	46%	82--430	427--786
M204L	91463--91113	117	12,749	3.86		No Hit Found								1	NP_048505	A157L	95.13	6.32E-19	44%	56%	1--117	1--109
M206L	91741--91523	73	8,893	3.28		No Hit Found									No Hit Found	No Hit Found						
M208L	92392--91994	133	14,349	10.12		No Hit Found									No Hit Found	No Hit Found						
M210L	92872--92588	95	10,895	10.03		No Hit Found								1	NP_048513	A165L	73.17	2.60E-12	39%	54%	5--95	171--279
M213L	93392--92937	152	17,762	10.22		No Hit Found								1	NP_048513	A165L	88.20	7.63E-17	37%	60%	5--118	20--134
M215R	93462--94298	279	32,168	5.13		No Hit Found								1	NP_048514	PBCV-1 exonuclease	298.90	1.11E-79	59%	73%	1--245	1--246
														2	NP_077549	EsV-1-64	93.97	5.41E-18	29%	50%	13--213	1--191
														3	AAG28903	F12A21.19	78.95	1.80E-13	30%	46%	28--229	95--294
														4	NP_176934	unknown protein	78.95	1.80E-13	30%	46%	28--229	116--315
														5	CAF27185	Exonuclease	77.41	5.24E-13	30%	45%	28--216	3--200
														6	CAF27488	Exonuclease	77.41	5.24E-13	30%	45%	28--216	3--200
														7	AAR26918	FirrV-1-B43 precursor	77.03	6.85E-13	30%	46%	11--211	3--196
														8	XP_480865	unknown protein	71.63	2.88E-11	28%	46%	27--208	134--317
														9	ZP_00620487	hypothetical protein RoseDRAFT_2901	71.63	2.88E-11	29%	46%	28--213	3--190
														10	XP_472944	OSJNBa0081L15.20	70.86	4.91E-11	28%	46%	27--208	141--324
M218R	94325--94807	161	17,492	4.72		No Hit Found								1	NP_048516	A168R	127.49	1.30E-28	51%	64%	26--157	24--161

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
M219L	95631--94810	274	30,278	9.89	1	pfam01734	Patatin, Patatin-like phospholipase. This family consists of various patatin glycoproteins from plants. The patatin protein accounts for up to 40% of the total soluble protein in potato tubers. Patatin is a storage protein but it also has the enzymatic activity of lipid acyl hydrolase, catalysing the cleavage of fatty acids from membrane lipids. Members of this family have been found also in vertebrates	116.20	2.61E-27	30%	50%	16--189	1--179	1	NP_048521	similar to E. coli hypothetical protein, corresponds to Swiss-Prot Accession Number P39407	323.55	4.06E-87	57%	77%	5--274	14--288
					2	COG1752	RssA, Predicted esterase of the alpha-beta hydrolase superfamily [General function prediction only]	101.31	7.43E-23	31%	50%	16--196	14--193	2	ZP_00240206	Patatin-like phospholipase family	107.84	3.50E-22	34%	53%	15--185	7--190
					3	COG4667	COG4667, Predicted esterase of the alpha-beta hydrolase superfamily [General function prediction only]	51.03	1.04E-07	26%	42%	16--208	14--202	3	XP_800775	PREDICTED: hypothetical protein XP_795682, partial	98.98	1.63E-19	35%	50%	14--186	76--265
														4	CAG23338	hypothetical protein	89.74	9.88E-17	28%	48%	14--256	6--261
														5	NP_149926	463L	88.20	2.87E-16	30%	56%	14--186	25--204
														6	AAQ66865	conserved hypothetical protein	88.20	2.87E-16	32%	52%	14--185	6--194
														7	AAS94389	phospholipase, patatin family	87.81	3.75E-16	31%	49%	15--186	4--194
														8	XP_789091	PREDICTED: hypothetical protein XP_783998	87.81	3.75E-16	32%	51%	14--186	86--275
														9	ABB37620	esterase of the alpha-beta hydrolase superfamily-like	87.04	6.40E-16	31%	50%	15--185	4--193
														10	YP_142800	patatin-like phospholipase (463L)	82.80	1.21E-14	27%	49%	15--238	57--285
M221L	96553--95654	300	34,092	4.51	1	cd00180	S_TKc, Serine/Threonine protein kinases, catalytic domain. Phosphotransferases of the serine or threonine-specific kinase subfamily. The enzymatic activity of these protein kinases is controlled by phosphorylation of specific residues in the activation segment of the catalytic domain, sometimes combined with reversible conformational changes in the C-terminal autoregulatory tail	176.55	1.66E-45	32%	54%	36--295	1--256	1	AAU06282	protein kinase A248R	193.74	5.66E-48	38%	59%	21--295	32--305
					2	smart00220	S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-specific kinase subfamily. .	176.56	1.90E-45	31%	52%	37--295	1--256	2	AAU06280	protein kinase A248R	192.97	9.65E-48	38%	59%	21--295	32--305
					3	pfam00069	Pkinase, Protein kinase domain..	165.46	3.71E-42	28%	51%	37--295	1--258	3	AAU06275	protein kinase A248R	192.97	9.65E-48	38%	59%	21--295	11--284
					4	COG0515	SPS1, Serine/threonine protein kinase [General function prediction only / Signal transduction mechanisms / Transcription / DNA replication, recombination. and repair]	117.18	1.39E-27	28%	47%	36--296	1--279	4	NP_048597	PBCV-1 protein kinase	189.89	8.17E-47	38%	58%	24--295	32--304
					5	smart00219	TyrKc, Tyrosine kinase, catalytic domain; Phosphotransferases. Tyrosine-specific kinase subfamily. .	94.09	1.32E-20	28%	48%	39--227	3--192	5	AAA87065	serine/threonine protein kinase	187.58	4.05E-46	37%	58%	24--295	28--300
					6	cd00192	TyrKc, Tyrosine kinase, catalytic domain. Phosphotransferases; tyrosine-specific kinase subfamily. Enzymes with TyrKc domains belong to an extensive family of proteins which share a conserved catalytic core common to both serine/threonine and tyrosine protein kinases. Enzymatic activity of tyrosine protein kinases is controlled by phosphorylation of specific tyrosine residues in the activation segment of the catalytic domain or a C-terminal tyrosine (tail) residue with reversible conformational changes	91.40	7.55E-20	27%	46%	36--227	8--203	6	AAU06274	protein kinase A248R	162.54	1.40E-38	38%	60%	79--295	22--238
					7	COG3642	COG3642, Mn2+-dependent serine/threonine protein kinase [Signal transduction mechanisms]	53.31	2.26E-08	28%	41%	112--217	66--186	7	AAU06285	protein kinase A248R	157.92	3.44E-37	37%	59%	81--295	2--217
					8	COG4248	COG4248, Uncharacterized protein with protein kinase and helix-hairpin-helix DNA-binding domains [General function prediction only]	38.54	0.000635	20%	37%	127--214	114--208	8	NP_048643	similar to PBCV-1 serine/threonine protein kinase, corresponds to GenBank Accession Number U14660	127.87	3.81E-28	31%	53%	41--295	22--279
					9	pfam06293	Kdo, Lipopolysaccharide kinase (Kdo/WaaP) family. These lipopolysaccharide kinases are related to protein kinases pfam00069. This family includes waaP (rfaP) gene product is required for the addition of phosphate to O-4 of the first heptose residue of the lipopolysaccharide (LPS) inner core region. It has previously been shown that WaaP is necessary for resistance to hydrophobic and polycationic antimicrobials in E. coli and that it is required for virulence in invasive strains of S. enterica	35.63	0.004574	21%	36%	13--167	1--155	9	AAU06286	protein kinase A248R	127.10	6.50E-28	47%	63%	146--295	2--152
					10	smart00750	KIND, kinase non-catalytic C-lobe domain; It is an interaction domain identified as being similar to the C-terminal protein kinase catalytic fold (C lobe). Its presence at the N terminus of signalling proteins and the absence of the active-site residues in the catalytic and activation loops suggest that it folds independently and is likely to be non-catalytic. The occurrence of KIND only in metazoa implies that it has evolved from the catalytic protein kinase domain into an interaction domain possibly by keeping the substrate-binding features	35.35	0.006041	22%	41%	135--297	22--175	10	NP_048631	similar to PBCV-1 serine/threonine protein kinase, corresponds to GenBank Accession Number U14660	123.25	9.39E-27	31%	51%	31--292	13--273
M225L	98791--96647	715	80,084	7.18	1	COG4581	COG4581, Superfamily II RNA helicase [DNA replication, recombination, and repair]	379.70	1.26E-106	40%	55%	11--412	119--568	1	NP_048589	contains ATP-GTP binding motif; similar to Saccharomyces cervisiae antiviral protein SKI2, corresponds to Swiss-Prot Accession Number P35207	708.75	0.00E+00	49%	67%	3--715	2--725
					2	COG1204	COG1204, Superfamily II helicase [General function prediction only]	206.78	1.49E-54	32%	49%	25--403	46--430	2	EAL46725	DEAD/DEAH box helicase, putative	318.55	4.90E-85	38%	57%	7--454	91--567
					3	COG1202	COG1202, Superfamily II helicase, archaea-specific [General function prediction only]	130.08	1.62E-31	31%	52%	25--381	231--552	3	XP_624031	PREDICTED: similar to ENSANGP00000020973	311.23	7.83E-83	38%	58%	12--431	123--578
					4	pfam00270	DEAD, DEAD/DEAH box helicase. Members of this family include the DEAD and DEAH box helicases. Helicases are involved in unwinding nucleic acids. The DEAD box helicases are involved in various aspects of RNA metabolism, including nuclear transcription, pre mRNA splicing, ribosome biogenesis, nucleocytoplasmic transport, translation, RNA decay and organellar gene expression..	113.21	2.17E-26	21%	40%	10--193	9--206	4	CAG30992	hypothetical protein	308.53	5.07E-82	37%	56%	12--450	127--598
					5	smart00487	DEXDc, DEAD-like helicases superfamily; .	101.07	1.09E-22	26%	43%	4--178	1--190	5	EAL33229	GA17990-PA	308.15	6.63E-82	37%	55%	12--431	57--513
					6	COG1205	COG1205, Distinct helicase family with a unique C-terminal domain including a metal-binding cysteine cluster [General function prediction only]	93.54	1.90E-20	25%	41%	16--368	75--410	6	EAL27780	GA10159-PA	306.99	1.48E-81	35%	51%	12--541	261--846
					7	cd00046	DEXDc, DEAD-like helicases superfamily. A diverse family of proteins involved in ATP-dependent RNA or DNA unwinding. This domain contains the ATP-binding region..	81.29	1.00E-16	25%	44%	27--157	1--143	7	AAF53481	CG4152-PA	306.61	1.93E-81	37%	55%	12--431	152--608
					8	COG1201	Lhr, Lhr-like helicases [General function prediction only]	79.14	3.43E-16	25%	45%	16--377	27--358	8	EAA14456	ENSANGP00000020973	305.83	3.29E-81	36%	55%	12--456	50--531
					9	cd00269	DEXHc, DEXH-box helicases. A diverse family of proteins involved in ATP-dependent DNA or RNA unwinding, needed in a variety of cellular processes. The name derives from the sequence of the Walker B motif (motif II). This domain contains the ATP-binding region..	72.42	3.94E-14	28%	45%	27--157	1--142	9	CAE45877	hypothetical protein	305.45	4.29E-81	36%	56%	12--450	140--611
					10	COG0513	SrmB, Superfamily II DNA and RNA helicases [DNA replication, recombination, and repair / Transcription / Translation, ribosomal structure and biogenesis]	69.42	3.51E-13	20%	41%	16--427	56--425	10	AAH65258	SKIV2L2 protein	305.45	4.29E-81	36%	56%	12--450	141--612
M232R	98893--99204	104	12,386	10.56		No Hit Found									No Hit Found	No Hit Found						
M233L	100741--99236	502	56,422	9.74	1	pfam06408	Homo_sperm_syn, Homospermidine synthase. This family consists of several homospermidine synthase proteins (EC:2.5.1.44). Homospermidine synthase (HSS) catalyses the synthesis of the polyamine homospermidine from 2 mol putrescine in an NAD(+)-dependent reaction	531.45	2.42E-152	38%	53%	27--496	1--470	1	NP_048585	PBCV-1 homospermidine synthase	729.55	0.00E+00	69%	83%	21--502	36--518
					2	pfam03435	Saccharop_dh, Saccharopine dehydrogenase. This family comprised of three structural domains that can not be separated in the linear sequence. In some organisms this enzyme is found as a bifunctional polypeptide with lysine ketoglutarate reductase. The saccharopine dehydrogenase can also function as a saccharopine reductase	46.01	3.46E-06	22%	47%	30--185	1--137	2	ZP_00589757	Homospermidine synthase	334.72	4.30E-90	38%	57%	27--474	5--452
					3	COG1748	LYS9, Saccharopine dehydrogenase and related proteins [Amino acid transport and metabolism]	46.05	3.71E-06	24%	49%	28--269	2--184	3	AAM05046	homospermidine synthase	305.83	2.14E-81	37%	57%	27--474	13--459
					4	cd01483	E1_enzyme_family, Superfamily of activating enzymes (E1) of the ubiquitin-like proteins. This family includes classical ubiquitin-activating enzymes E1, ubiquitin-like (ubl) activating enzymes and other mechanistic homologes, like MoeB, ThiF1 and others. The common reaction mechanism catalyzed by MoeB, ThiF and the E1 enzymes begins with a nucleophilic attack of the C-terminal carboxylate of MoaD, ThiS and ubiquitin, respectively, on the alpha-phosphate of an ATP molecule bound at the active site of the activating enzymes, leading to the formation of a high-energy acyladenylate intermediate and subsequently to the formation of a thiocarboxylate at the C termini of MoaD and ThiS	35.02	0.007331	20%	37%	30--130	2--121	4	AAM29862	homospermidine synthase	301.60	4.04E-80	36%	57%	27--474	6--452
														5	ZP_00543002	Homospermidine synthase	289.27	2.07E-76	36%	55%	27--474	6--452
														6	ZP_00584810	Homospermidine synthase	252.68	2.15E-65	33%	52%	30--474	14--446
														7	CAH16656	hypothetical protein	249.60	1.82E-64	34%	54%	26--474	13--446
														8	AAU28555	homospermidine synthase	248.83	3.10E-64	34%	54%	26--474	13--446
														9	BAE51070	Homospermidine synthase	234.57	6.06E-60	32%	52%	33--474	34--462
														10	NP_791691	homospermidine synthase	228.41	4.34E-58	34%	52%	77--478	56--448
M237L	101090--100764	109	12,662	5.20		No Hit Found									No Hit Found	No Hit Found						

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
M238R	101294--101650	119	13,221	10.74		No Hit Found								1	NP_048582	A234L	93.20	2.38E-18	43%	57%	1--118	1--107
M240L	101915--101646	90	11,406	10.51		No Hit Found								1	NP_048581	A233R	85.50	4.98E-16	47%	70%	5--86	11--91
M241L	102200--101940	87	10,153	4.84	1	COG0695	GrxC, Glutaredoxin and related proteins [Posttranslational modification, protein turnover. chaperones].	43.03	3.20E-05	31%	53%	3--72	4--72	1	CAE58431	Hypothetical protein CBG01566	48.91	5.22E-05	30%	52%	4--83	31--114
M242L	102804--102235	190	20,890	10.14		No Hit Found								1	NP_048578	A230R	157.15	2.38E-37	40%	65%	2--190	3--194
M243R	102829--103062	78	8,683	6.49		No Hit Found									No Hit Found	No Hit Found						
M246L	103535--103296	80	9,502	7.85		No Hit Found									No Hit Found	No Hit Found						
M247R	103586--104335	250	28,319	8.21		No Hit Found								1	NP_048357	A9R	238.42	1.46E-61	64%	79%	79--249	3--173
														2	NP_049005	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	103.61	5.60E-21	32%	47%	6--248	11--253
														3	NP_048807	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	99.75	8.09E-20	30%	46%	6--248	5--246
														4	NP_048427	A79R	85.89	1.21E-15	31%	44%	9--226	8--217
														5	NP_048629	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	84.73	2.69E-15	27%	46%	9--247	11--248
														6	NP_048525	A177R	71.25	3.08E-11	26%	44%	9--233	11--227
														7	AAU06304	hypothetical protein A275R	58.15	2.70E-07	25%	44%	81--247	1--167
														8	AAU06301	hypothetical protein A275R	57.00	6.01E-07	25%	44%	81--247	1--167
M249R	104467--104880	138	15,412	9.27		No Hit Found								1	NP_048923	A567L	77.41	1.34E-13	38%	52%	12--127	10--123
M253L	106765--104888	626	67,332	8.25	1	COG0508	AceF, Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component, and related enzymes [Energy production and conversion].	41.13	0.000118	20%	34%	336--438	77--176	1	NP_048921	A565R	261.54	6.10E-68	41%	59%	25--336	26--336
					2	COG3883	COG3883, Uncharacterized protein conserved in bacteria [Function unknown].	40.32	0.000214	21%	47%	12--176	5--170	2	BAA11343	DNA binding protein	259.23	3.03E-67	41%	59%	25--336	26--336
					3	pfam03276	Gag_spuma, Spumavirus gag protein..	39.68	0.000275	27%	40%	333--433	178--273	3	NP_048741	Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	63.93	1.87E-08	45%	53%	251--312	29--99
M258R	106836--108620	595	64,077	10.10	1	COG3064	TolA, Membrane protein involved in colicin uptake [Cell envelope biogenesis. outer membrane].	44.39	1.17E-05	24%	37%	111--305	56--253	1	NP_048917	similar to Chlorella virus CVK2 DNA binding protein, corresponds to GenBank Accession Number D78305	208.38	5.74E-52	45%	63%	378--594	417--647
					2	pfam06519	TolA, TolA protein. This family consists of several bacterial TolA proteins as well as two eukaryotic proteins of unknown function. Tol proteins are involved in the translocation of group A colicins. Colicins are bacterial protein toxins, which are active against Escherichia coli and other related species (See pfam01024). TolA is anchored to the cytoplasmic membrane by a single membrane spanning segment near the N-terminus, leaving most of the protein exposed to the periplasm..	43.96	1.46E-05	24%	40%	145--339	85--269	2	BAA11342	DNA binding protein	204.91	6.35E-51	44%	62%	378--594	416--646
					3	pfam01576	Myosin_tail_1, Myosin tail. The myosin molecule is a multi-subunit complex made up of two heavy chains and four light chains it is a fundamental contractile protein found in all eukaryote cell types. This family consists of the coiled-coil myosin heavy chain tail region. The coiled-coil is composed of the tail from two molecules of myosin. These can then assemble into the macromolecular thick filament. The coiled-coil region provides the structural backbone the thick filament. .	42.99	2.99E-05	24%	42%	88--273	379--550	3	BAA83789	alginate lyase	95.90	4.15E-18	36%	51%	410--574	153--317
					4	COG0845	AcrA, Membrane-fusion protein [Cell envelope biogenesis, outer membrane].	42.03	5.35E-05	16%	32%	37--257	2--216	4	NP_048562	PBCV-1 alginate lyase	94.36	1.21E-17	35%	50%	410--574	141--305
					5	COG4372	COG4372, Uncharacterized protein conserved in bacteria with the myosin-like domain [Function unknown]	39.73	0.000274	19%	35%	102--297	157--345	5	BAB19127	vAL-1	93.20	2.69E-17	35%	51%	410--574	169--333
					6	pfam01213	CAP, CAP protein..	39.21	0.000407	25%	39%	252--383	192--330	6	BAE57035	unnamed protein product	55.07	8.12E-06	30%	42%	81--277	567--776
M263R	108657--109310	218	24,292	10.15		No Hit Found								1	NP_048915	A559L	73.94	3.64E-12	27%	43%	1--176	1--185
M264L	109732--109313	140	14,802	4.77	1	pfam00692	dUTPase, dUTPase. dUTPase hydrolyses dUTP to dUMP and pyrophosphate	131.51	6.05E-32	52%	69%	11--139	2--129	1	AAW51452	deoxyuridine triphosphatase	218.39	4.99E-56	76%	88%	2--140	3--141
					2	COG0717	Dcd, Deoxycytidine deaminase [Nucleotide transport and metabolism].	54.12	1.33E-08	29%	48%	32--117	74--158	2	NP_048907	similar to tomato dUTP pyrophosphatase, corresponds to GenBank Accesssion Number S40549	218.01	6.51E-56	75%	89%	2--140	3--141
														3	AAW51453	deoxyuridine triphosphatase	215.70	3.23E-55	75%	87%	2--140	3--141
														4	EAL88001	dUTPase	172.94	2.40E-42	63%	77%	4--140	114--250
														5	XP_657875	hypothetical protein AN0271.2	172.17	4.10E-42	63%	78%	4--139	71--206
														6	NP_190278	dUTP diphosphatase/ hydrolase	169.86	2.03E-41	62%	76%	3--140	29--166
														7	AAB22611	deoxyuridine triphosphatase; dUTPase; P18	167.93	7.73E-41	62%	73%	3--140	32--169
														8	XP_629169	dUTP diphosphatase	166.78	1.72E-40	59%	74%	2--140	37--175
														9	XP_469212	putative deoxyuridine triphosphatase	166.78	1.72E-40	61%	75%	5--140	90--225
														10	BAE55800	unnamed protein product	166.39	2.25E-40	61%	76%	4--139	58--193
M266R	109853--110671	273	30,421	8.38	1	pfam00352	TBP, Transcription factor TFIID (or TATA-binding protein, TBP)..	37.03	0.002008	24%	48%	142--237	3--85	1	NP_048908	similar to Sulfolobus TATA-binding protein, corresponds to GenBank Accession Number S55311'	179.49	9.41E-44	41%	59%	25--268	13--268
														2	YP_142807	TATA-box binding protein (TBP)	50.06	8.61E-05	25%	41%	75--252	138--346
M269L	111880--110681	400	44,555	5.69	1	pfam04451	Capsid_Iridovir, Iridovirus major capsid protein. This family includes the major capsid protein of iridoviruses, chlorella virus and Spodoptera ascovirus, which are all dsDNA viruses with no RNA stage. This is the most abundant structural protein and can account for up to 45% of virion protein. In Chlorella virus MT325 the major capsid protein is a glycoprotein	363.48	9.75E-102	40%	57%	1--396	2--443	1	NP_048914	similar to Chlorella virus PBCV-1 ORF A11L, corresponds to GenBank Accession Number U42580	510.38	4.21E-143	60%	76%	1--400	1--400
														2	NP_048359	contains aminoacyl-tRNA synthetase class-II signature	397.13	5.19E-109	49%	67%	2--400	3--403
														3	NP_048358	similar to PBCV-1 major capsid protein, corresponds to Swiss-Prot Accession Number P30328	322.01	2.12E-86	40%	62%	2--398	4--399
														4	AAC27492	major capsid protein Vp49	276.17	1.34E-72	38%	54%	2--400	3--432
														5	BAA76601	major capsid protein MCP1	271.17	4.30E-71	39%	54%	2--400	3--437
														6	NP_048787	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	270.78	5.61E-71	38%	54%	2--400	3--437
														7	BAA76600	major capsid protein	267.70	4.75E-70	37%	54%	2--400	3--436
														8	BAA22198	major capsid protein Vp54	261.92	2.61E-68	37%	53%	2--400	3--437
														9	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	240.74	6.22E-62	37%	53%	32--400	9--413
														10	1M3Y_D	Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid Containing, Dna Virus	240.74	6.22E-62	37%	53%	32--400	9--413
M272L	113304--111928	459	51,475	8.82	1	COG0553	HepA, Superfamily II DNA/RNA helicases, SNF2 family [Transcription / DNA replication. recombination. and repair].	158.77	4.66E-40	27%	43%	8--438	339--851	1	NP_048904	similar to Caenorhabditis transcription activator, corresponds to Swiss-Prot Accession Number P41877	451.44	2.79E-125	51%	71%	8--457	7--458
					2	pfam00176	SNF2_N, SNF2 family N-terminal domain. This domain is found in proteins involved in a variety of processes including transcription regulation (e.g., SNF2, STH1, brahma, MOT1) , DNA repair (e.g., ERCC6, RAD16, RAD5), DNA recombination (e.g., RAD54), and chromatin unwinding (e.g., ISWI) as well as a variety of other proteins with little functional information (e.g., lodestar, ETL1)	126.16	2.44E-30	24%	44%	11--262	1--287	2	BAC15031	helicase (swi/snf family)	168.32	4.71E-40	30%	51%	8--422	606--1027
					3	COG1061	SSL2, DNA or RNA helicases of superfamily II [Transcription / DNA replication. recombination. and repair].	81.30	8.96E-17	22%	38%	7--393	36--375	3	AAP98807	swf/snf helicase	162.16	3.38E-38	29%	47%	4--437	698--1154
					4	smart00487	DEXDc, DEAD-like helicases superfamily; .	78.73	5.89E-16	19%	36%	8--193	9--199	4	CAG35724	probable helicase	161.38	5.76E-38	28%	47%	2--450	938--1397
					5	cd00269	DEXHc, DEXH-box helicases. A diverse family of proteins involved in ATP-dependent DNA or RNA unwinding, needed in a variety of cellular processes. The name derives from the sequence of the Walker B motif (motif II). This domain contains the ATP-binding region..	73.58	2.00E-14	25%	49%	29--164	1--143	5	CAC96916	lin1685	161.00	7.52E-38	29%	47%	4--443	612--1058

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					6	cd00079	HELICc, Helicase superfamily c-terminal domain; associated with DEXDc-, DEAD-, and DEAH-box proteins, yeast initiation factor 4A, Ski2p, and Hepatitis C virus NS3 helicases; this domain is found in a wide variety of helicases and helicase related proteins; may not be an autonomously folding unit, but an integral part of the helicase; 4 helicase superfamilies at present according to the organization of their signature motifs; all helicases share the ability to unwind nucleic acid duplexes with a distinct directional polarity; they utilize the free energy from nucleoside triphosphate hydrolysis to fuel their translocation along DNA, unwinding the duplex in the process	73.42	2.22E-14	26%	44%	275--402	3--130	6	NP_300906	SWI/SNF family helicase_2	159.84	1.68E-37	29%	46%	4--437	698--1154
					7	smart00490	HELICc, helicase superfamily c-terminal domain; .	60.63	1.56E-10	24%	47%	314--395	3--82	7	CAE79539	putative helicase/SNF2 family domain protein	159.84	1.68E-37	29%	48%	3--422	856--1294
					8	cd00046	DEXDc, DEAD-like helicases superfamily. A diverse family of proteins involved in ATP-dependent RNA or DNA unwinding. This domain contains the ATP-binding region..	59.71	3.06E-10	18%	36%	30--164	2--144	8	AAT04439	helicase, Snf2 family	159.07	2.86E-37	28%	47%	4--437	612--1053
					9	pfam00271	Helicase_C, Helicase conserved C-terminal domain. TThis domain family is found in a wide variety of helicases and helicase related proteins. It may be that this is not an autonomously folding unit, but an integral part of the helicase..	59.46	3.23E-10	26%	47%	316--395	1--78	9	NP_296465	helicase, Snf2 family	159.07	2.86E-37	28%	47%	2--437	711--1169
					10	pfam00270	DEAD, DEAD/DEAH box helicase. Members of this family include the DEAD and DEAH box helicases. Helicases are involved in unwinding nucleic acids. The DEAD box helicases are involved in various aspects of RNA metabolism, including nuclear transcription, pre mRNA splicing, ribosome biogenesis, nucleocytoplasmic transport, translation, RNA decay and organellar gene expression	45.03	6.79E-06	20%	39%	8--138	11--142	10	ZP_00231434	helicase, Snf2 family	159.07	2.86E-37	28%	47%	4--437	612--1053
M278R	113482--114282	267	29,095	4.11		No Hit Found									No Hit Found	No Hit Found						
M282R	114658--114915	86	9,836	10.85		No Hit Found								1	NP_048875	A519L	98.60	5.76E-20	60%	72%	7--80	4--77
M283R	114950--116107	386	44,286	9.76	1	COG1231	COG1231, Monoamine oxidase [Amino acid transport and metabolism].	65.77	4.64E-12	18%	30%	1--384	1--439	1	BAA83788	orf1	545.43	1.12E-153	68%	80%	8--386	19--397
					2	COG1233	COG1233, Phytoene dehydrogenase and related proteins [Secondary metabolites biosynthesis. transport. and catabolism].	47.39	1.45E-06	38%	48%	11--63	7--59	2	NP_048564	similar to bovine monoamine oxidase, corresponds to Swiss-Prot Accession Number P21398	534.64	1.98E-150	67%	79%	8--386	16--394
M288L	117169--116114	352	40,610	10.08		No Hit Found									No Hit Found	No Hit Found						
M289R	117311--118414	368	40,590	10.22		No Hit Found								1	NP_048562	PBCV-1 alginate lyase	412.15	1.39E-113	64%	71%	61--366	5--320
														2	BAB19127	vAL-1	410.99	3.09E-113	57%	66%	1--366	1--348
														3	BAA83789	alginate lyase	407.91	2.62E-112	56%	65%	1--366	1--332
														4	BAE48156	hypothetical methionyl-tRNA synthetase	89.74	1.58E-16	29%	47%	125--361	5--247
														5	NP_048917	similar to Chlorella virus CVK2 DNA binding protein, corresponds to GenBank Accession Number D78305	73.94	8.95E-12	24%	42%	133--367	409--648
														6	BAA11342	DNA binding protein	72.02	3.40E-11	25%	42%	166--367	430--647
M294R	118438--119088	217	23,981	8.96		No Hit Found									No Hit Found	No Hit Found						
M298R	119184--119654	157	18,097	5.79		No Hit Found								1	NP_048561	A214L	130.57	1.40E-29	45%	66%	22--154	3--135
M299R	119736--120182	149	16,648	4.54		No Hit Found								1	NP_048560	A213L	137.89	8.51E-32	52%	79%	1--126	1--125
M302L	121198--120233	322	36,434	3.97	1	COG5271	MDN1, AAA ATPase containing von Willebrand factor type A (vWA) domain [General function prediction only].	49.71	2.61E-07	29%	51%	141--313	340--4007		No Hit Found	No Hit Found						
M307L	122464--121349	372	41,792	5.47	1	pfam02784	Orn_Arg_deC_N, Pyridoxal-dependent decarboxylase, pyridoxal binding domain. These pyridoxal-dependent decarboxylases acting on ornithine, lysine, arginine and related substrates This domain has a TIM barrel fold..	196.63	1.77E-51	38%	57%	25--260	4--246	1	NP_048554	PBCV-1 arginine decarboxylase	492.66	8.14E-138	63%	78%	1--372	1--372
					2	COG0019	LysA, Diaminopimelate decarboxylase [Amino acid transport and metabolism].	192.42	3.35E-50	26%	45%	19--372	29--394	2	CAE02644	ornithine decarboxylase	269.63	1.12E-70	40%	59%	1--364	43--407
					3	pfam00278	Orn_DAP_Arg_deC, Pyridoxal-dependent decarboxylase, C-terminal sheet domain. These pyridoxal-dependent decarboxylases act on ornithine, lysine, arginine and related substrates. .	91.03	1.17E-19	30%	46%	263--365	1--103	3	CAA71498	ornithine decarboxylase	264.23	4.71E-69	38%	57%	1--364	48--418
					4	COG1166	SpeA, Arginine decarboxylase (spermidine biosynthesis) [Amino acid transport and metabolism]	54.10	1.47E-08	23%	40%	45--270	118--374	4	BAA83427	ornithine decarboxylase	263.08	1.05E-68	39%	57%	1--364	47--417
														5	AAL87201	ornithine decarboxylase	262.69	1.37E-68	39%	58%	1--364	50--420
														6	AAL83709	putative ornithine decarboxylase 1	261.15	3.99E-68	39%	58%	1--364	50--420
														7	AAK13622	ornithine decarboxylase	260.77	5.20E-68	38%	57%	1--364	48--418
														8	NP_571876	ornithine decarboxylase 1	260.38	6.80E-68	37%	55%	2--372	21--409
														9	AAQ14852	ornithine decarboxylase	258.84	1.98E-67	38%	56%	1--364	48--418
														10	AAG45222	ornithine decarboxylase	258.84	1.98E-67	39%	57%	1--364	47--417
M308L	123252--122548	235	25,935	11.57		No Hit Found								1	NP_048552	A205R	178.72	1.22E-43	47%	60%	1--228	1--205
M310L	123954--123295	220	24,037	4.83		No Hit Found								1	NP_048550	A203R	269.63	4.61E-71	63%	74%	1--220	1--213
M312R	124001--124336	112	12,026	4.52		No Hit Found								1	NP_048549	A202L	117.09	1.53E-25	52%	71%	3--112	2--112
M313L	124897--124343	185	20,826	10.87	1	cd00127	DSPc, Dual specificity phosphatases (DSP); Ser/Thr and Tyr protein phosphatases. Structurally similar to tyrosine-specific phosphatases but with a shallower active site cleft and a distinctive active site signature motif. HCxxGxxR. Characterized as VHR- or Cdc25-like..	132.67	2.68E-32	35%	53%	26--159	4--139	1	NP_048659	similar to human protein Tyr-phosphatase, corresponds to GenBank Accession Number U27193	242.66	4.08E-63	66%	85%	6--168	7--169
					2	pfam00782	DSPc, Dual specificity phosphatase, catalytic domain. Ser/Thr and Tyr protein phosphatases. The enzyme's tertiary fold is highly similar to that of tyrosine-specific phosphatases, except for a "recognition" region..	125.37	4.49E-30	35%	55%	24--161	2--139	2	NP_998405	dual specificity phosphatase 16	87.43	2.19E-16	34%	52%	24--164	159--300
					3	smart00195	DSPc, Dual specificity phosphatase, catalytic domain; .	122.69	2.95E-29	34%	56%	24--161	2--139	3	XP_543810	PREDICTED: similar to Dual specificity protein phosphatase 16 (Mitogen-activated protein kinase phosphatase 7) (MAP kinase phosphatase 7) (MKP-7)	84.73	1.42E-15	34%	53%	24--163	159--299
					4	COG2453	CDC14, Predicted protein-tyrosine phosphatase [Signal transduction mechanisms]	52.40	3.94E-08	23%	44%	58--181	65--180	4	AAH42101	DUSP16 protein	83.57	3.17E-15	34%	53%	24--163	159--299
					5	smart00404	PTPc_motif, Protein tyrosine phosphatase, catalytic domain motif; .	36.17	0.003842	22%	48%	99--147	36--90	5	AAH31643	Unknown (protein for IMAGE:5176724)	83.57	3.17E-15	34%	53%	24--163	110--250
														6	AAI09236	Dual specificity phosphatase 16	83.57	3.17E-15	34%	53%	24--163	159--299
														7	AAI09235	Dual specificity phosphatase 16	83.57	3.17E-15	34%	53%	24--163	159--299
														8	XP_520751	PREDICTED: similar to KIAA1700 protein	83.57	3.17E-15	34%	53%	24--163	268--408
														9	BAB21791	KIAA1700 protein	83.57	3.17E-15	34%	53%	24--163	184--324
														10	AAH59232	Dusp16 protein	82.42	7.05E-15	35%	52%	24--162	159--298
M314L	125260--125006	85	10,248	3.73		No Hit Found								1	NP_048660	A306L	105.92	3.62E-22	84%	96%	34--85	35--86
M315L	125720--125385	112	13,007	4.87		No Hit Found								1	NP_048663	A308L	48.14	8.73E-05	38%	68%	43--98	20--79
M317L	126306--125803	168	18,118	8.10		No Hit Found								1	NP_048665	A310L	182.57	3.87E-45	53%	69%	1--168	1--170
M319L	127075--126353	241	26,796	5.32		No Hit Found								1	NP_048667	PBCV-1 33kDa translation peptide	340.50	2.56E-92	71%	82%	1--235	1--230
														2	NP_077561	EsV-1-76	84.34	3.30E-15	28%	43%	5--236	10--259
														3	AAR26966	FirrV-1-I1	57.77	3.31E-07	26%	40%	9--213	10--222
														4	AAR26885	FirrV-1-B10	55.84	1.26E-06	25%	39%	1--222	13--247
M322L	127493--127242	84	9,374	8.18		No Hit Found									No Hit Found	No Hit Found						
M323R	127579--127815	79	9,151	5.01		No Hit Found								1	NP_048670	A314R	132.88	2.72E-30	83%	87%	2--79	5--80
M324R	127875--128990	372	41,246	4.66		No Hit Found								1	CAI82895	hypothetical membrane protein	91.28	5.46E-17	26%	47%	129--364	18--256
														2	AAW39962	membrane protein, putative	90.89	7.13E-17	25%	46%	129--364	18--256
														3	AAL80842	hypothetical protein	78.57	3.66E-13	24%	52%	125--357	2--231
														4	AAB85442	unknown	75.10	4.05E-12	25%	47%	118--365	14--270
M326R	129017--129367	117	12,851	10.70		No Hit Found								1	NP_048676	A320R	88.97	4.53E-17	37%	59%	1--115	24--138

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
M328R	129391--129729	113	11,846	4.87		No Hit Found								1	NP_048677	A321R	62.00	5.82E-09	35%	56%	4--95	11--101
M329L	130346--129849	166	19,230	5.50		No Hit Found								1	NP_048678	A322L	122.09	5.95E-27	38%	57%	1--159	1--170
M332L	131644--130379	422	46,792	5.07	1	pfam01298	Lipoprotein_5, Transferrin binding protein-like solute binding protein. This family of proteins are distantly related to other families of solute binding proteins	34.97	0.006974	19%	37%	2--116	47--167	1	NP_048680	A324L	396.74	7.28E-109	52%	62%	11--422	13--453
														2	AAR26897	FirrV-1-B22	88.58	4.24E-16	29%	47%	46--254	58--270
														3	NP_077588	EsV-1-103	76.64	1.67E-12	28%	50%	74--256	192--365
														4	YP_142803	unknown	56.23	2.33E-06	30%	52%	140--250	225--333
M334R	131667--132263	199	23,258	10.51		No Hit Found									No Hit Found	No Hit Found						
M337L	133396--132341	352	40,349	9.76		No Hit Found								1	NP_048684	A328L	266.54	8.78E-70	38%	60%	1--352	1--350
M341R	133727--134449	241	26,780	5.31		No Hit Found								1	NP_048667	PBCV-1 33kDa translation peptide	342.04	8.79E-93	71%	82%	1--235	1--230
														2	NP_077561	EsV-1-76	84.34	3.30E-15	28%	43%	5--236	10--259
														3	AAR26966	FirrV-1-I1	58.54	1.94E-07	26%	40%	9--213	10--222
														4	AAR26885	FirrV-1-B10	57.00	5.65E-07	26%	40%	1--222	13--247
tRNAs																						
Ile	134513--134585	73 bs					anticodon TAT															
Leu	134610--134693	84 bs					anticodon TAA															
Phe	134718--134790	73 bs					anticodon GAA															
Arg	134814--134886	73 bs					anticodon TCT															
Gly	134890--134960	71 bs					anticodon TCC															
Asn	134984--135057	74 bs					anticodon GTT															
Asn	135080--135151	72 bs					anticodon GTT															
Tyr	135174--135259	86 bs					anticodon GTA															
Intron	135211--135223	13 bs					Intron (135211-135223)															
Lys	135262--135334	73 bs					anticodon CTT															
Thr	135496--135567	72 bs					anticodon CGT															
M344L	136141--135620	174	20,192	6.76		No Hit Found								1	NP_048693	A337L	66.63	3.39E-10	48%	59%	100--173	5--78
M346L	136605--136189	139	15,952	4.56		No Hit Found									No Hit Found	No Hit Found						
M348L	138420--136744	559	60,603	9.29		No Hit Found								1	NP_048699	A342L	734.95	0.00E+00	68%	78%	16--550	37--561
M354R	138534--139847	438	48,898	4.48	1	COG3534	AbfA, Alpha-L-arabinofuranosidase [Carbohydrate transport and metabolism].	41.05	0.00011	26%	41%	57--247	66--256	1	CAD86595	cellulase precursor	54.68	7.16E-06	25%	36%	54--290	282--516
M357L	140281--139901	127	14,450	10.21		No Hit Found								1	BAA22200	URF14.2	148.29	6.26E-35	66%	78%	2--108	3--109
														2	BAA22202	URF14.2	147.90	8.18E-35	66%	79%	2--108	3--109
														3	BAA22201	URF14.2	147.52	1.07E-34	65%	78%	2--108	3--109
														4	BAA22199	URF14.2	146.36	2.38E-34	64%	78%	2--108	3--109
														5	NP_048706	A349L	112.46	3.81E-24	73%	85%	1--69	73--141
M358R	140365--140835	157	18,719	9.88		No Hit Found								1	NP_048707	a350R	153.68	1.54E-36	50%	76%	1--120	1--120
M359L	140726--141079	118	38,475	8.32	1	cd00315	Cyt_C5_DNA_methylase, Cytosine-C5 specific DNA methylases; Methyl transfer reactions play an important role in many aspects of biology. Cytosine-specific DNA methylases are found both in prokaryotes and eukaryotes. DNA methylation, or the covalent addition of a methyl group to cytosine within the context of the CpG dinucleotide, has profound effects on the mammalian genome. These effects include transcriptional repression via inhibition of transcription factor binding or the recruitment of methyl-binding proteins and their associated chromatin remodeling factors, X chromosome inactivation, imprinting and the suppression of parasitic DNA sequences. DNA methylation is also essential for proper embryonic development and is an important player in both DNA repair and genome stability	160.08	1.74E-40	29%	43%	5--259	4--271	1	NP_048873	M.CviAII cytosine DNA methyltransferase	450.28	4.09E-125	62%	74%	1--340	1--342
					2	pfam00145	DNA methylase, C-5 cytosine-specific DNA methylase..	156.24	2.29E-39	28%	43%	5--286	4--289	2	NP_048886	M.CviAIV cytosine DNA methyltransferase	440.27	4.24E-122	61%	76%	1--328	2--332
					3	COG0270	Dcm, Site-specific DNA methylase [DNA replication, recombination, and repair].	109.01	4.34E-25	27%	44%	1--191	3--197	3	AAV84097	CviPII m5C DNA methyltransferase	308.15	2.51E-82	47%	61%	2--340	16--357
														4	AAC64006	cytosine methyltransferase	288.12	2.69E-76	44%	58%	2--342	3--358
														5	AAC55063	cytosine methyltransferase	266.16	1.09E-69	40%	56%	5--342	6--362
														6	NP_049039	nonfunctional M.CviAV cytosine DNA methyltransferase	265.00	2.44E-69	40%	56%	5--342	6--362
														7	NP_945098	gp67	100.14	1.04E-19	32%	48%	2--162	6--192
														8	NP_818425	gp127	93.59	9.73E-18	35%	51%	5--155	6--165
														9	CAC48079	putative DNA methylase	92.05	2.83E-17	33%	44%	5--222	4--226
														10	ZP_00517640	C-5 cytosine-specific DNA methylase	90.89	6.30E-17	31%	48%	2--180	72--249
M361R	141459--141671	71	7,643	11.29		No Hit Found								1	NP_048874	a518R	112.46	3.76E-24	76%	81%	1--71	1--71
M362L	142574--141936	213	23,843	3.72		No Hit Found								1	NP_048709	Asp/Glu rich; DAEDDDIYxxET (2X) negative charge cluster	231.49	1.29E-59	58%	66%	1--213	1--207
														2	YP_142843	unknown	51.22	2.39E-05	32%	58%	134--213	140--221
M363L	143695--142673	341	38,820	3.99	1	COG5271	MDN1, AAA ATPase containing von Willebrand factor type A (vWA) domain [General function prediction only].	47.01	1.67E-06	24%	50%	17--152	950--4085	1	NP_048714	A357L	161.77	2.89E-38	51%	71%	160--320	86--249
M367L	144396--143740	219	25,122	4.82		No Hit Found								1	NP_048479	A131L	66.63	5.87E-10	36%	50%	88--218	11--135
M370R	144446--145246	267	30,020	10.18	1	cd00283	GIY-YIG_Cterm, GIYX(10-11)YIG family of class I homing endonucleases C-terminus (GIY-YIG_Cterm). Homing endonucleases promote the mobility of intron or intein by recognizing and cleaving a homologous allele that lacks the sequence. They catalyze a double-strand break in the DNA near the insertion site of that element to facilitate homing at that site. Class I homing endonucleases are sorted into four families based on the presence of these motifs in their respective N-terminii: LAGLIDADG, His-Cys box, HNH, and GIY-YIG. This CD contains several but not all members of the GIY-YIG family. The C-terminus of GIY-YIG is a DNA-binding domain which is separated from the N-terminus by a long, flexible linker. The DNA-binding domain consists of a minor-groove binding alpha-helix, and a helix-turn-helix. Some also contain a zinc finger (i.e. I-TevI) which is not required for DNA binding or catalysis, but is a component of the linker and directs the catalytic domain to cleave the homing site at a fixed distance from the intron insertion site	67.33	1.43E-12	48%	63%	134--267	1--113	1	NP_048671	A315L	244.97	1.75E-63	47%	64%	1--267	1--240
					2	smart00497	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished).	45.51	5.54E-06	40%	60%	216--267	1--51	2	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	226.87	4.93E-58	47%	62%	1--267	1--223
					3	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	41.98	5.99E-05	34%	55%	1--91	1--83	3	NP_048641	PBCV-1 33kd peptide	197.59	3.20E-49	42%	60%	5--267	7--247
					4	pfam07453	NUMOD1, NUMOD1 domain..	36.17	0.003294	41%	69%	216--248	1--33	4	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	164.47	3.00E-39	38%	54%	1--258	1--212
														5	NP_899393	SegD	80.49	5.71E-14	31%	47%	1--213	1--198
														6	AAK09365	intron encoded BmoI	73.94	5.34E-12	29%	43%	4--265	5--266
														7	AAC49248	ORF211	72.40	1.55E-11	33%	46%	96--252	77--193
														8	AAC49244	ORF301	72.40	1.55E-11	30%	43%	9--192	80--240
														9	YP_293795	putative endonuclease	72.02	2.03E-11	36%	55%	2--109	3--111
														10	CAC51107	putative GIY-YIG endonuclease	71.63	2.65E-11	35%	54%	148--251	105--218
M372R	145289--148681	1131	124,388	10.87	1	pfam04484	DUF566, Family of unknown function (DUF566). Family of related proteins that is plant specific..	44.31	1.30E-05	23%	37%	1008--1130	44--169	1	NP_048720	similar to chicken vitellogenin II, corresponds to Swiss-Prot Accession Number P02845	836.25	0.00E+00	61%	75%	353--1029	1--651
					2	pfam03999	MAP65_ASE1, Microtubule associated protein (MAP65/ASE1 family)..	38.86	0.000503	24%	41%	980--1131	410--559	2	NP_048717	A360R	251.91	9.52E-65	59%	71%	14--226	27--239
					3	smart00490	HELICc, helicase superfamily c-terminal domain; .	37.52	0.001249	31%	57%	827--891	19--80	3	NP_048718	A361R	133.27	4.94E-29	72%	84%	249--332	2--85

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
M375R	148717--149253	179	21,201	4.78		No Hit Found								1	A48779	translation elongation factor EF-3 homolog - Chlorella virus CVK2	61.62	1.17E-08	39%	58%	75--168	84--171
														2	NP_049002	A646L	54.68	1.43E-06	41%	58%	85--162	54--129
M376R	149243--149830	196	22,945	10.92		No Hit Found									No Hit Found	No Hit Found						
M377L	150512--149901	204	23,518	10.92	1	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	35.44	0.005785	21%	38%	76--172	1--83	1	NP_048736	A379L	165.24	1.04E-39	44%	61%	12--203	5--203
M379R	150719--151156	146	17,153	4.36		No Hit Found								1	NP_048842	A486L	94.74	8.11E-19	37%	51%	1--146	1--152
M381R	151222--152637	472	53,700	6.38	1	pfam04451	Capsid_Iridovir, Iridovirus major capsid protein. This family includes the major capsid protein of iridoviruses, chlorella virus and Spodoptera ascovirus, which are all dsDNA viruses with no RNA stage. This is the most abundant structural protein and can account for up to 45% of virion protein. In Chlorella virus MT325 the major capsid protein is a glycoprotein..	178.20	6.16E-46	29%	45%	81--470	3--422	1	NP_048740	similar to PBCV-1 major capsid protein, corresponds to Swiss-Prot Accession Number P30328	291.97	2.94E-77	53%	72%	148--405	1--257
														2	BAE06835	hypothetical major capsid protein	163.31	1.58E-38	30%	48%	81--446	3--401
														3	BAA76601	major capsid protein MCP1	139.81	1.87E-31	30%	45%	81--438	3--380
														4	BAA22198	major capsid protein Vp54	137.89	7.09E-31	29%	42%	81--469	3--410
														5	NP_048787	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	137.89	7.09E-31	29%	42%	81--469	3--410
														6	BAA76600	major capsid protein	137.50	9.26E-31	29%	43%	81--469	3--409
														7	AAC27492	major capsid protein Vp49	135.96	2.69E-30	30%	44%	81--438	3--371
														8	1M3Y_D	Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid Containing, Dna Virus	115.93	2.89E-24	28%	40%	103--469	1--386
														9	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	115.55	3.77E-24	28%	40%	103--469	1--386
														10	NP_048741	Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	95.52	4.04E-18	62%	75%	401--472	385--456
M385L	154232--152652	527	59,190	11.06		No Hit Found								1	NP_048441	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563	150.21	1.59E-34	30%	47%	61--391	3--296
														2	NP_048636	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055	117.09	1.49E-24	46%	55%	4--161	414--568
														3	NP_048439	a91L	114.39	9.69E-24	48%	65%	403--527	1--126
														4	NP_048632	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662	95.52	4.66E-18	41%	52%	7--161	465--609
														5	NP_049032	similar to Chlorella virus PBCV-1 ORF A282L, corresponds to GenBank Accession Number U42580	66.24	3.02E-09	72%	84%	58--101	2--44
M388R	154334--155038	235	27,144	4.58		No Hit Found								1	AAX46836	hypothetical protein PSSM4_035	55.45	1.56E-06	26%	42%	6--205	306--511
M389R	155125--156621	499	57,656	8.49	1	pfam01171	ATP_bind_3, PP-loop family. This family of proteins belongs to the PP-loop superfamily	124.59	7.26E-30	27%	47%	196--399	1--204	1	NP_048910	similar to MesJ cell cycle protein	525.78	1.30E-147	52%	67%	1--494	1--497
					2	cd01992	PP-ATPase, N-terminal domain of predicted ATPase of the PP-loop fainly implicated in cell cycle control [Cell division and chromosome partitioning]. This is a subfamily of Adenine nucleotide alpha hydrolases superfamily.Adeninosine nucleotide alpha hydrolases superfamily includes N type ATP PPases and ATP sulphurylases. It forms a apha/beta/apha fold which binds to Adenosine group. This domain has a strongly conserved motif SGGXD at the N terminus..	123.03	2.62E-29	29%	48%	196--381	1--185	2	T18059	hypothetical protein A557L - Chlorella virus PBCV-1	104.38	9.38E-21	51%	65%	26--122	14--112
					3	COG0037	MesJ, Predicted ATPase of the PP-loop superfamily implicated in cell cycle control [Cell division and chromosome partitioning].	104.54	8.67E-24	25%	44%	181--389	6--218	3	CAG39532	conserved hypothetical protein	87.04	1.55E-15	23%	47%	197--495	14--320
					4	cd01993	Alpha_ANH_like_II, This is a subfamily of Adenine nucleotide alpha hydrolases superfamily.Adeninosine nucleotide alpha hydrolases superfamily includes N type ATP PPases and ATP sulphurylases. It forms a apha/beta/apha fold which binds to Adenosine group. This subfamily of proteins is predicted to bind ATP. This domainhas a strongly conserved motif SGGKD at the N terminus.	66.04	3.30E-12	23%	41%	196--367	1--180	4	CAI80146	conserved hypothetical protein	86.27	2.64E-15	23%	46%	197--495	14--320
					5	pfam06041	DUF924, Bacterial protein of unknown function (DUF924). This family consists of several hypothetical bacterial proteins of unknown function..	46.44	2.53E-06	23%	40%	2--105	4--117	5	BAB56671	conserved hypothetical protein	85.89	3.45E-15	23%	46%	197--495	14--320
					6	cd01712	ThiI, ThiI is required for thiazole synthesis in the thiamine biosynthesis pathway. It belongs to the Adenosine Nucleotide Hydrolysis supperfamily and predicted to bind to Adenosine nucleotide	36.73	0.002192	23%	41%	201--356	6--153	6	AAW37665	tRNA(Ile)-lysidine synthetase	85.89	3.45E-15	23%	46%	197--495	14--320
					7	COG0175	CysH, 3'-phosphoadenosine 5'-phosphosulfate sulfotransferase (PAPS reductase)/FAD synthetase and related enzymes [Amino acid transport and metabolism / Coenzyme metabolism].	35.02	0.008227	20%	38%	178--364	23--199	7	YP_193201	putative cell cycle	84.73	7.69E-15	25%	45%	196--440	23--272
					8	COG0301	ThiI, Thiamine biosynthesis ATP pyrophosphatase [Coenzyme metabolism].	34.87	0.008691	24%	46%	196--356	177--329	8	EAN10266	PP-loop	84.34	1.00E-14	26%	47%	192--419	28--260
														9	CAG42241	conserved hypothetical protein	83.96	1.31E-14	22%	46%	197--495	14--320
														10	ZP_00576297	PP-loop	80.88	1.11E-13	27%	51%	196--394	20--217
M395L	156892--156632	87	9,642	10.44		No Hit Found									No Hit Found	No Hit Found						
M397R	157055--157816	254	28,999	8.32		No Hit Found								1	NP_048357	A9R	215.31	1.36E-54	57%	78%	89--254	8--173
														2	NP_048807	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	98.98	1.42E-19	25%	45%	2--253	1--246
														3	NP_049005	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	95.13	2.06E-18	27%	44%	5--253	10--253
														4	NP_048629	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	93.59	5.98E-18	25%	44%	2--252	4--248
														5	NP_048525	A177R	92.82	1.02E-17	26%	47%	2--241	4--230
														6	NP_048427	A79R	83.19	8.09E-15	25%	46%	2--232	1--218
														7	AAU06304	hypothetical protein A275R	79.72	8.94E-14	30%	50%	89--252	4--167
														8	AAU06301	hypothetical protein A275R	78.57	1.99E-13	30%	50%	89--252	4--167
														9	AAU06302	hypothetical protein A275R	66.63	7.83E-10	32%	52%	121--253	1--132
M399L	158394--157846	183	20,781	7.72	1	pfam02940	mRNA_triPase, mRNA capping enzyme, beta chain. The beta chain of mRNA capping enzyme has triphosphatase activity. The function of the capping enzyme also depends on the guanylyltransferase activity conferred by the alpha chain (see pfam01331).	41.15	9.92E-05	34%	55%	94--156	191--256	1	NP_048806	PBCV-1 RNA triphosphatase	184.11	1.67E-45	54%	69%	5--183	11--193
M401R	158497--159168	224	25,101	7.96	1	smart00702	P4Hc, Prolyl 4-hydroxylase alpha subunit homologues. Mammalian enzymes catalyse hydroxylation of collagen, for example. Prokaryotic enzymes might catalyse hydroxylation of antibiotic peptides. These are 2-oxoglutarate-dependent dioxygenases, requiring 2-oxoglutarate and dioxygen as cosubstrates and ferrous iron as a cofactor	116.71	2.03E-27	29%	44%	34--222	1--178	1	NP_048433	PBCV-1 prolyl 4-hydroxylase	227.64	2.09E-58	55%	71%	32--224	48--241
					2	pfam03171	2OG-FeII_Oxy, 2OG-Fe(II) oxygenase superfamily. This family contains members of the 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily. This family includes the C-terminal of prolyl 4-hydroxylase alpha subunit. The holoenzyme has the activity EC:1.14.11.2 catalysing the reaction: Procollagen L-proline + 2-oxoglutarate + O2 <=> procollagen trans- 4-hydroxy-L-proline + succinate + CO2. The full enzyme consists of a alpha2 beta2 complex with the alpha subunit contributing most of the parts of the active site. The family also includes lysyl hydrolases, isopenicillin synthases and AlkB..	42.02	5.20E-05	27%	39%	119--223	2--96	2	AAZ62310	Procollagen-proline,2-oxoglutarate-4-dioxygenase	93.20	6.16E-18	34%	49%	34--222	90--274
														3	CAD16521	HYPOTHETICAL PROTEIN	90.51	3.99E-17	33%	47%	34--222	97--281
														4	ZP_00943867	Prolyl 4-hydroxylase alpha subunit	89.74	6.81E-17	33%	47%	34--222	97--281
														5	ZP_00599099	Procollagen-proline,2-oxoglutarate-4-dioxygenase	87.43	3.38E-16	32%	46%	32--222	99--285
														6	ZP_00984285	hypothetical protein BdolA_01003928	85.89	9.83E-16	30%	46%	34--222	73--257
														7	ZP_00508297	Procollagen-proline,2-oxoglutarate-4-dioxygenase	83.57	4.88E-15	31%	46%	34--222	55--237
														8	YP_142947	prolyl 4-hydroxylase	83.19	6.37E-15	29%	44%	30--222	56--237
														9	AAF08583	unknown protein	78.57	1.57E-13	32%	43%	34--221	45--223
														10	ZP_00245340	hypothetical protein Rgel02000555	78.57	1.57E-13	29%	40%	34--222	87--269
M403R	159213--159530	106	11,844	10.09	1	pfam00085	Thioredoxin, Thioredoxin. Thioredoxins are small enzymes that participate in redox reactions, via the reversible oxidation of an active centre disulfide bond. Some members with only the active site are not separated from the noise..	61.78	6.44E-11	29%	48%	29--105	33--108	1	NP_048805	contains cytochrome C family heme-binding site signature; similar to maize protein disulphide isomerase, correspond to Swiss-Prot Accession Number P52588	125.56	4.40E-28	56%	72%	4--99	6--101
					2	COG3118	COG3118, Thioredoxin domain-containing protein [Posttranslational modification, protein turnover, chaperones].	38.35	0.000665	20%	40%	14--99	39--125	2	AAQ23042	transglutaminase	68.17	8.34E-11	35%	58%	12--105	8--106
														3	AAF78087	protein disulfide isomerase ER-60	67.40	1.42E-10	36%	57%	23--105	43--123

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
														4	AAH67155	Hypothetical protein zqc:77086	66.24	3.17E-10	36%	56%	23--105	45--125
														5	XP_419952	PREDICTED: similar to Protein disulfide isomerase A6 precursor (Protein disulfide isomerase P5) (Thioredoxin domain containing protein 7)	66.24	3.17E-10	45%	62%	29--102	60--134
														6	CAH65062	hypothetical protein	66.24	3.17E-10	45%	62%	29--102	60--134
														7	AAC24752	transglutaminase precursor	64.70	9.22E-10	35%	56%	12--105	37--135
														8	CAG03659	unnamed protein product	64.70	9.22E-10	45%	62%	29--97	55--124
														9	AAH44524	Sb:cb825 protein	63.54	2.05E-09	37%	56%	23--105	43--123
														10	AAM93973	protein disulfide isomerase 1	62.77	3.50E-09	36%	58%	23--99	46--120
M404R	159584--160972	463	52,177	6.01	1	COG0661	AarF, Predicted unusual protein kinase [General function prediction onlv].	175.14	5.17E-45	27%	43%	24--385	37--423	1	NP_048802	similar to Clostridium pasteurianum ORF, corresponds to GenBank Accession Number Z28353	620.16	4.58E-176	64%	84%	13--463	11--462
					2	pfam03109	ABC1, ABC1 family. This family includes ABC1 from yeast and AarF from E. coli. These proteins have a nuclear or mitochondrial subcellular location in eukaryotes. The exact molecular functions of these proteins is not clear, however yeast ABC1 suppresses a cytochrome b mRNA translation defect and is essential for the electron transfer in the bc 1 complex and E. coli AarF is required for ubiquinone production. It has been suggested that members of the ABC1 family are novel chaperonins. These proteins are unrelated to the ABC transporter proteins..	114.54	9.83E-27	41%	58%	96--209	7--119	2	ZP_00673531	ABC-1	138.27	5.28E-31	24%	46%	49--448	70--501
														3	BAD78862	hypothetical protein	134.42	7.63E-30	24%	46%	49--448	83--515
														4	ABA22988	ABC-1	134.04	9.97E-30	24%	46%	37--448	71--516
														5	BAB66733	488aa long conserved hypothetical protein	134.04	9.97E-30	24%	46%	12--457	9--459
														6	BAB76214	alr4515	133.27	1.70E-29	23%	46%	37--448	82--527
														7	BAC92119	gl4178	130.18	1.44E-28	25%	46%	54--394	102--470
														8	ZP_01006074	Predicted protein kinase	128.64	4.19E-28	26%	48%	36--392	48--428
														9	AAY80974	universally conserved protein	127.87	7.14E-28	26%	45%	12--384	9--399
														10	ZP_00112431	COG0661: Predicted unusual protein kinase	127.49	9.33E-28	23%	45%	49--448	78--511
M407R	161042--161380	113	13,229	5.00		No Hit Found								1	NP_048801	A444L	71.63	7.35E-12	35%	56%	4--112	2--99
M409R	161649--163055	469	50,358	10.79	1	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis. outer membrane].	60.16	2.18E-10	32%	36%	66--194	36--161		No Hit Found	No Hit Found						
M411L	163494--163072	141	15,822	4.42		No Hit Found								1	NP_048798	A441L	125.95	3.35E-28	46%	61%	6--141	2--137
M413L	164611--163613	333	36,902	6.59	1	cd00204	ANK, ankyrin repeats; ankyrin repeats mediate protein-protein interactions in very diverse families of proteins. The number of ANK repeats in a protein can range from 2 to over 20 (ankyrins, for example). ANK repeats may occur in combinations with other types of domains. The structural repeat unit contains two antiparallel helices and a beta-hairpin, repeats are stacked in a superhelical arrangement; this alignment contains 4 consecutive repeats	135.60	4.30E-33	59%	79%	62--188	1--126	1	EAL29245	GA14074-PA	197.59	4.62E-49	41%	59%	9--270	170--434
					2	COG0666	Arp, FOG: Ankyrin repeat [General function prediction onlv].	91.49	7.70E-20	34%	50%	46--233	17--212	2	AAM11327	GH01626p	196.05	1.35E-48	42%	59%	9--270	25--289
					3	pfam00023	Ank, Ankyrin repeat. There's no clear separation between noise and signal on the HMM search Ankyrin repeats generally consist of a beta, alpha, alpha, beta order of secondary structures. The repeats associate to form a higher order structure..	51.22	1.13E-07	56%	81%	67--99	1--33	3	XP_681288	hypothetical protein AN8019.2	196.05	1.35E-48	43%	60%	10--270	801--1064
					4	smart00248	ANK, ankyrin repeats; Ankyrin repeats are about 33 amino acids long and occur in at least four consecutive copies. They are involved in protein protein interactions. The core of the repeat seems to be an helix-loop-helix structure. .	42.64	4.08E-05	55%	76%	67--96	1--30	4	AAN12046	CG7462-PC, isoform C	196.05	1.35E-48	42%	59%	9--270	175--439
														5	AAF50525	CG7462-PB, isoform B	196.05	1.35E-48	42%	59%	9--270	175--439
														6	ABA81818	RE55168p	196.05	1.35E-48	42%	59%	9--270	325--589
														7	EAA03765	ENSANGP00000006233	192.97	1.14E-47	42%	59%	9--270	175--439
														8	XP_658734	hypothetical protein AN1130.2	190.66	5.65E-47	43%	56%	13--273	577--840
														9	ZP_00374082	ankyrin 1, erythrocyte splice form 1	189.89	9.64E-47	42%	59%	25--274	81--333
														10	XP_395788	PREDICTED: similar to CG7462-PB, isoform B	187.96	3.66E-46	40%	59%	9--270	180--444
M416R	164684--165577	298	33,175	5.33		No Hit Found								1	NP_048800	A443R	182.19	1.68E-44	33%	54%	6--297	8--306
M417L	167392--165593	600	64,526	8.50	1	pfam05887	Trypan_PARP, Procyclic acidic repetitive protein (PARP). This family consists of several Trypanosoma brucei procyclic acidic repetitive protein (PARP) like sequences. The procyclic acidic repetitive protein (parp) genes of Trypanosoma brucei encode a small family of abundant surface proteins whose expression is restricted to the procyclic form of the parasite. They are found at two unlinked loci, parpA and parpB; transcription of both loci is developmentally regulated	59.22	4.27E-10	50%	52%	53--115	60--122	1	BAE02830	surface protein	65.08	7.94E-09	45%	56%	181--255	997--1071
M421L	167746--167417	110	12,636	6.50		No Hit Found								1	NP_048796	A439R	124.79	7.39E-28	61%	82%	19--103	21--105
M423R	167767--168024	86	9,971	10.79	1	COG0695	GrxC, Glutaredoxin and related proteins [Posttranslational modification, protein turnover. chaperones].	51.89	5.60E-08	34%	55%	3--72	5--72	1	NP_048795	similar to E. coli glutaredoxin, corresponds to Swiss-Prot Accession Number P37687	91.28	9.20E-18	55%	81%	3--72	5--74
					2	pfam00462	Glutaredoxin, Glutaredoxin..	38.70	0.000524	25%	49%	1--72	2--73	2	AAQ59708	probable peroxiredoxin/glutaredoxin family protein	50.83	1.38E-05	38%	57%	1--74	173--241
														3	AAL51366	GLUTAREDOXIN	50.45	1.80E-05	32%	57%	1--70	8--73
														4	AAN30771	glutaredoxin 3	50.45	1.80E-05	32%	57%	1--70	4--69
														5	ZP_00399122	Glutaredoxin	49.29	4.01E-05	31%	59%	3--71	8--72
M425R	168038--168595	186	21,497	4.90	1	cd01673	dNK, Deoxyribonucleoside kinase (dNK) catalyzes the phosphorylation of deoxyribonucleosides to yield corresponding monophosphates (dNMPs). This family consists of various deoxynucleoside kinases including deoxyribo- cytidine (EC 2.7.1.74), guanosine (EC 2.7.1.113), adenosine (EC 2.7.1.76), and thymidine (EC 2.7.1.21) kinases. They are key enzymes in the salvage of deoxyribonucleosides originating from extra- or intracellular breakdown of DNA	101.08	8.98E-23	31%	52%	3--165	1--178	1	NP_048773	contains ATP/GTP-binding site motif A; similar to Bacillus subtilus 24.1 kDa protein, corresponds to Swiss-Prot Accession Number P37530	190.27	2.44E-47	50%	72%	3--178	2--180
					2	COG1428	COG1428, Deoxynucleoside kinases [Nucleotide transport and metabolism].	91.09	9.10E-20	31%	51%	1--181	4--210	2	NP_149606	143R	88.20	1.31E-16	31%	51%	1--152	1--157
					3	pfam01712	dNK, Deoxynucleoside kinase. This family consists of various 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 from a deoxynucleoside. Using ATP and yielding ADP in the process..	77.23	1.29E-15	32%	49%	53--180	1--138	3	YP_073536	deoxynucleoside kinases	78.57	1.03E-13	31%	56%	3--152	4--161
					4	COG0125	Tmk, Thymidylate kinase [Nucleotide transport and metabolism].	54.12	1.48E-08	22%	46%	4--184	6--204	4	CAG42289	putative deoxyadenosine kinase protein	72.79	5.68E-12	29%	47%	3--181	11--211
					5	cd01672	TMPK, Thymidine monophosphate kinase (TMPK), also known as thymidylate kinase, catalyzes the phosphorylation of thymidine monophosphate (TMP) to thymidine diphosphate (TDP) utilizing ATP as its preferred phophoryl donor. TMPK represents the rate-limiting step in either de novo or salvage biosynthesis of thymidine triphosphate (TTP)..	49.50	3.19E-07	19%	43%	3--181	2--198	5	CAI80195	deoxypurine kinase subunit	72.79	5.68E-12	29%	48%	3--181	11--211
					6	cd02019	NK, Nucleoside/nucleotide kinase (NK) is a protein superfamily consisting of multiple families of enzymes that share structural similarity and are functionally related to the catalysis of the reversible phosphate group transfer from nucleoside triphosphates to nucleosides/nucleotides, nucleoside monophosphates, or sugars. Members of this family play a wide variety of essential roles in nucleotide metabolism, the biosynthesis of coenzymes and aromatic compounds, as well as the metabolism of sugar and sulfate..	41.73	7.78E-05	21%	39%	3--159	1--165	6	AAW37713	deoxynucleoside kinase family protein	72.02	9.68E-12	32%	50%	3--153	11--172
					7	pfam02223	Thymidylate_kin, Thymidylate kinase..	41.42	9.97E-05	20%	38%	6--177	1--186	7	XP_770504	deoxypurine kinase subunit	71.63	1.26E-11	33%	51%	4--155	33--192
					8	cd02030	NDUO42, NADH:Ubiquinone oxioreductase, 42 kDa (NDUO42) is a family of proteins that are highly similar to deoxyribonucleoside kinases (dNK). Members of this family have been identified as one of the subunits of NADH:Ubiquinone oxioreductase (complex I), a multi-protein complex located in the inner mitochondrial membrane. The main function of the complex is to transport electrons from NADH to ubiquinone, which is accompanied by the translocation of protons from the mitochondrial matrix to the inter membrane space	39.85	0.000245	25%	45%	3--153	1--184	8	CAC84481	thymidine kinase	71.63	1.26E-11	25%	46%	4--182	13--208
					9	COG0572	Udk, Uridine kinase [Nucleotide transport and metabolism].	36.38	0.002876	22%	44%	3--159	10--179	9	YP_187798	deoxynucleoside kinase family protein	71.63	1.26E-11	30%	48%	3--177	11--197
														10	NP_078725	Deoxynucleoside kinase	69.71	4.81E-11	29%	52%	3--152	4--161

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
M427R	168662--169717	352	39,681	8.71		No Hit Found								1	NP_048477	similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903	272.71	1.22E-71	43%	61%	9--352	8--356
														2	NP_048502	A154L	261.92	2.16E-68	40%	59%	1--352	1--347
														3	NP_048920	similar to Chlorella virus PBCV-1 ORF A154L, corresponds to GenBank Accession Number U42580	254.22	4.51E-66	39%	58%	10--352	9--350
														4	NP_077492	EsV-1-7	146.36	1.32E-33	35%	47%	4--199	5--200
														5	NP_077600	EsV-1-115	84.34	6.18E-15	30%	42%	4--166	146--303
														6	BAD61685	loricrin-like	56.23	1.80E-06	27%	36%	4--212	109--277
														7	XP_464880	loricrin-like	50.45	9.89E-05	26%	35%	4--194	319--490
M430L	170803--169724	360	40,028	8.34	1	COG1910	COG1910, Periplasmic molybdate-binding protein/domain [Inorganic ion transport and metabolism].	35.25	0.006527	35%	49%	40--83	87--130	1	NP_048774	A417L	212.62	1.56E-53	37%	57%	1--348	1--372
M435L	171179--170838	114	13,067	10.50		No Hit Found								1	NP_048777	A420L	70.09	2.20E-11	48%	65%	45--114	1--70
M437R	171162--171413	84	10,358	11.31		No Hit Found									No Hit Found	No Hit Found						
M438L	171644--171402	81	9,009	7.94		No Hit Found									No Hit Found	No Hit Found						
M440R	171671--172138	156	17,496	4.92		No Hit Found								1	NP_048780	A423R	82.42	4.24E-15	36%	61%	36--154	33--155
M442R	172162--172521	120	13,655	11.03		No Hit Found									No Hit Found	No Hit Found						
M444R	172564--172911	116	13,307	4.84		No Hit Found								1	NP_048783	A426R	95.90	3.72E-19	42%	64%	6--116	4--114
M445L	173265--172918	116	13,195	5.80	1	pfam00085	Thioredoxin, Thioredoxin. Thioredoxins are small enzymes that participate in redox reactions, via the reversible oxidation of an active centre disulfide bond. Some members with only the active site are not separated from the noise..	43.29	2.50E-05	25%	49%	25--93	25--96	1	NP_048784	contains thioredoxin active site-like sequence; similar to Synechocystis thioredoxin-like protein, corresponds to Swiss-Prot Accession Number P52232	103.22	2.33E-21	44%	65%	2--115	7--119
														2	ZP_00769720	COG0526: Thiol-disulfide isomerase and thioredoxins	53.14	2.77E-06	31%	54%	17--105	21--112
														3	AAC71342	thioredoxin (trxA)	51.60	8.05E-06	29%	51%	4--103	2--102
														4	AAP04832	thioredoxin	51.60	8.05E-06	29%	55%	4--102	1--99
														5	CAH63539	putative thioredoxin	50.83	1.37E-05	30%	53%	4--102	1--99
														6	CAH59450	thioredoxin 1	50.45	1.79E-05	28%	51%	3--95	12--106
														7	BAE56042	unnamed protein product	50.45	1.79E-05	30%	51%	2--90	1--92
														8	AAH84527	Hypothetical protein LOC496541	50.45	1.79E-05	28%	55%	4--90	5--91
														9	CAG25528	thioredoxin	49.29	4.00E-05	28%	53%	10--83	11--86
														10	NP_395913	TrxA1_2	49.29	4.00E-05	29%	51%	11--101	23--114
M448L	173669--173334	112	12,692	5.76	1	COG3118	COG3118, Thioredoxin domain-containing protein [Posttranslational modification, protein turnover, chaperones].	40.66	0.00015	38%	56%	55--89	81--115	1	NP_048784	contains thioredoxin active site-like sequence; similar to Synechocystis thioredoxin-like protein, corresponds to Swiss-Prot Accession Number P52232	53.91	1.59E-06	28%	53%	1--105	7--110
M449L	174017--173706	104	11,732	8.10		No Hit Found								1	NP_048784	contains thioredoxin active site-like sequence; similar to Synechocystis thioredoxin-like protein, corresponds to Swiss-Prot Accession Number P52232	52.37	4.60E-06	32%	52%	2--95	9--106
M451L	175368--174043	442	51,066	5.24	1	cd00204	ANK, ankyrin repeats; ankyrin repeats mediate protein-protein interactions in very diverse families of proteins. The number of ANK repeats in a protein can range from 2 to over 20 (ankyrins, for example). ANK repeats may occur in combinations with other types of domains. The structural repeat unit contains two antiparallel helices and a beta-hairpin, repeats are stacked in a superhelical arrangement; this alignment contains 4 consecutive repeats.	44.69	8.95E-06	23%	41%	157--287	1--126	1	NP_048786	A429L	217.62	6.47E-55	33%	51%	7--437	34--466
														2	NP_200815	protein binding	55.84	3.26E-06	25%	44%	51--237	61--250
														3	BAB03143	ankyrin-like protein	53.53	1.62E-05	23%	43%	49--253	599--805
														4	NP_187842	protein binding	51.60	6.14E-05	22%	43%	49--237	114--305
M454R	175405--175728	108	12,899	9.58		No Hit Found									No Hit Found	No Hit Found						
M455R	175778--176056	93	10,779	10.74		No Hit Found								1	NP_048957	A601R	91.28	8.97E-18	52%	65%	11--91	12--101
M457R	176074--177201	376	42,427	10.07		No Hit Found								1	NP_048579	contains ATP/GTP-binding motif A	493.81	3.71E-138	61%	76%	2--376	3--383
														2	NP_048983	similar to Chlorella virus PBCV-1 ORF A231L, corresponds to GenBank Accession Number U42580	305.83	1.43E-81	44%	62%	10--336	104--428
M460R	177267--178019	251	28,857	10.08	1	cd00283	GIY-YIG_Cterm, GIYX(10-11)YIG family of class I homing endonucleases C-terminus (GIY-YIG_Cterm). Homing endonucleases promote the mobility of intron or intein by recognizing and cleaving a homologous allele that lacks the sequence. They catalyze a double-strand break in the DNA near the insertion site of that element to facilitate homing at that site. Class I homing endonucleases are sorted into four families based on the presence of these motifs in their respective N-termini: LAGLIDADG, His-Cys box, HNH, and GIY-YIG. This CD contains several but not all members of the GIY-YIG family. The C-terminus of GIY-YIG is a DNA-binding domain which is separated from the N-terminus by a long, flexible linker. The DNA-binding domain consists of a minor-groove binding alpha-helix, and a helix-turn-helix. Some also contain a zinc finger (i.e. I-TevI) which is not required for DNA binding or catalysis, but is a component of the linker and directs the catalytic domain to cleave the homing site at a fixed distance from the intron inserting site.	39.98	0.000211	44%	57%	124--173	11--65	1	CAC51107	putative GIY-YIG endonuclease	61.62	2.46E-08	28%	45%	39--212	14--187
					2	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	36.98	0.001988	23%	43%	26--112	1--78	2	AAU16837	GIY-YIG catalytic domain containing protein; possible intron encoded endonuclease	57.00	6.06E-07	30%	51%	26--149	2--136
					3	pfam01541	GIY-YIG, GIY-YIG catalytic domain. This domain called GIY-YIG is found in the amino terminal region of excinuclease abc subunit c (uvrC), bacteriophage T4 endonucleases segA, segB, segC, segD and segE; it is also found in putative endonucleases encoded by group I introns of fungi and phage. The structure of I-TevI a GIY-YIG endonuclease, reveals a novel alpha/beta-fold with a central three-stranded antiparallel beta-sheet flanked by three helices. The most conserved and putative catalytic residues are located on a shallow, concave surface and include a metal coordination site. .	36.68	0.002641	26%	43%	29--115	4--85	3	CAA38813	GIY ND1 i4 grp IB protein b	51.22	3.33E-05	30%	45%	14--167	54--220
														4	NP_074961	orf261	51.22	3.33E-05	30%	45%	14--167	53--219
M463L	179395--178097	433	47,651	6.92	1	pfam04451	Capsid_Iridovir, Iridovirus major capsid protein. This family includes the major capsid protein of iridoviruses, chlorella virus and Spodoptera ascovirus, which are all dsDNA viruses with no RNA stage. This is the most abundant structural protein and can account for up to 45% of virion protein. In Chlorella virus MT325 the major capsid protein is a glycoprotein.	478.65	1.87E-136	54%	65%	1--429	1--443	1	AAC27492	major capsid protein Vp49	763.45	0.00E+00	88%	89%	1--433	1--432
														2	NP_048787	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	650.97	0.00E+00	75%	80%	1--433	1--437
														3	BAA22198	major capsid protein Vp54	643.65	0.00E+00	74%	80%	1--433	1--437
														4	BAA76601	major capsid protein MCP1	643.27	0.00E+00	74%	80%	1--433	1--437
														5	BAA76600	major capsid protein	637.88	0.00E+00	73%	79%	1--433	1--436
														6	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	603.98	3.12E-171	73%	79%	25--433	1--413
														7	1M3Y_D	Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid Containing, Dna Virus	603.98	3.12E-171	73%	79%	25--433	1--413
														8	BAE06835	hypothetical major capsid protein	316.62	9.94E-85	41%	54%	1--433	1--440
														9	NP_048359	contains aminoacyl-tRNA synthetase class-II signature	248.05	4.34E-64	37%	52%	1--433	1--403
														10	NP_048914	similar to Chlorella virus PBCV-1 ORF A11L, corresponds to GenBank Accession Number U42580	246.13	1.65E-63	35%	51%	3--433	2--400

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
M465L	180410--179550	287	32,625	8.99	1	cd00283	GIY-YIG_Cterm, GIYX(10-11)YIG family of class I homing endonucleases C-terminus (GIY-YIG_Cterm). Homing endonucleases promote the mobility of intron or intein by recognizing and cleaving a homologous allele that lacks the sequence. They catalyze a double-strand break in the DNA near the insertion site of that element to facilitate homing at that site. Class I homing endonucleases are sorted into four families based on the presence of these motifs in their respective N-termini: LAGLIDADG, His-Cys box, HNH, and GIY-YIG. This CD contains several but not all members of the GIY-YIG family. The C-terminus of GIY-YIG is a DNA-binding domain which is separated from the N-terminus by a long, flexible linker. The DNA-binding domain consists of a minor-groove binding alpha-helix, and a helix-turn-helix. Some also contain a zinc finger (i.e. I-TevI) which is not required for DNA binding or catalysis, but is a component of the linker and directs the catalytic domain to cleave the homing site at a fixed distance from the intron insertion site.	68.10	7.64E-13	57%	73%	116--179	1--64	1	NP_048671	A315L	237.65	3.16E-61	46%	60%	1--283	1--240
					2	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	43.52	1.96E-05	32%	54%	1--90	1--83	2	NP_048641	PBCV-1 33kd peptide	221.86	1.79E-56	44%	61%	3--285	6--249
					3	smart00497	IENR1, Intron encoded nuclease repeat motif. Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished). .	41.66	8.38E-05	37%	58%	232--285	1--53	3	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	192.20	1.52E-47	40%	52%	1--286	1--226
					4	pfam01541	GIY-YIG, GIY-YIG catalytic domain. This domain called GIY-YIG is found in the amino terminal region of excinuclease abc subunit c (uvrC), bacteriophage T4 endonucleases segA, segB, segC, segD and segE; it is also found in putative endonucleases encoded by group I introns of fungi and phage. The structure of I-TevI a GIY-YIG endonuclease, reveals a novel alpha/beta-fold with a central three-stranded antiparallel beta-sheet flanked by three helices. The most conserved and putative catalytic residues are located on a shallow, concave surface and include a metal coordinating site.	37.06	0.001784	30%	56%	1--59	1--64	4	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	137.50	4.45E-31	35%	48%	1--286	1--211
														5	AAK09365	intron encoded BmoI	82.42	1.70E-14	31%	45%	4--210	5--217
														6	NP_899393	SegD	79.34	1.44E-13	30%	47%	1--219	1--221
														7	AAC49244	ORF301	75.87	1.59E-12	31%	46%	4--173	76--239
														8	AAT53588	group I intron GIY-YIG endonuclease	73.17	1.03E-11	25%	45%	4--254	5--215
														9	YP_293795	putative endonuclease	71.63	3.00E-11	36%	54%	2--112	3--115
														10	CAA73995	unnamed protein product	71.25	3.91E-11	30%	45%	13--183	124--277
M467R	180530--183076	849	95,588	6.42	1	pfam00535	Glycos_transf_2, Glycosyl transferase. Diverse family, transferring sugar from UDP-glucose, UDP-N-acetyl- galactosamine, GDP-mannose or CDP-abequose, to a range of substrates including cellulose, dolichol phosphate and teichoic acids..	49.45	3.41E-07	30%	43%	257--384	2--124	1	NP_048462	A114R	640.96	0.00E+00	64%	77%	374--848	5--477
					2	COG0463	WcaA, Glycosyltransferases involved in cell wall biogenesis [Cell envelope biogenesis, outer membrane].	47.44	1.35E-06	18%	35%	252--526	2--258	2	NP_048459	A111R	495.35	3.58E-138	58%	76%	1--373	1--379
					3	COG1215	COG1215, Glycosyltransferases, probably involved in cell wall biogenesis [Cell envelope biogenesis, outer membrane].	44.15	1.51E-05	25%	38%	252--368	53--167	3	AAP95345	possible glycosyltransferase	90.51	2.66E-16	32%	46%	255--427	6--189
					4	COG1216	COG1216, Predicted glycosyltransferases [General function prediction only]	38.96	0.000491	26%	46%	253--370	3--116	4	CAG34747	hypothetical protein	89.74	4.53E-16	30%	45%	1--226	1--225
					5	pfam03016	Exostosin, Exostosin family. The EXT family is a family of tumour suppressor genes. Mutations of EXT1 on 8q24.1, EXT2 on 11p11-13, and EXT3 on 19p have been associated with the autosomal dominant disorder known as hereditary multiple exostoses (HME). This is the most common known skeletal dysplasia. The chromosomal locations of other EXT genes suggest association with other forms of neoplasia. EXT1 and EXT2 have both been shown to encode a heparan sulphate polymerase with both D-glucuronyl (GlcA) and N-acetyl-D-glucosaminoglycan (GlcNAC) transferase activities. The nature of the defect in heparan sulphate biosynthesis in HME is unclear.	35.03	0.007743	23%	43%	658--819	177--325	5	ZP_00202013	COG0463: Glycosyltransferases involved in cell wall biogenesis	86.27	5.01E-15	32%	49%	255--427	48--231
														6	AAU37201	unknown	81.65	1.23E-13	26%	40%	255--496	6--259
														7	ZP_00154867	COG1216: Predicted glycosyltransferases	79.34	6.13E-13	29%	44%	255--428	29--215
														8	AAU03775	conserved hypothetical protein	58.15	1.46E-06	26%	50%	59--207	103--250
														9	ZP_00340129	hypothetical protein RakaH01000503	57.00	3.26E-06	30%	51%	59--207	107--254
														10	ZP_00153455	COG0457: FOG: TPR repeat	55.45	9.48E-06	29%	50%	59--207	107--254
M472R	183130--183585	152	16,607	8.18		No Hit Found								1	NP_048789	Lys-rich	114.01	1.30E-24	46%	67%	1--109	1--109
M475R	183617--183964	116	12,729	11.93		No Hit Found								1	NP_048927	A571R	157.15	1.36E-37	79%	89%	1--101	1--101
														2	NP_048792	Arg-rich	56.61	2.50E-07	61%	70%	1--44	4--47
M476L	184825--184220	202	22,832	10.53	1	pfam05854	MC1, Non-histone chromosomal protein MC1. This family consists of archaeal chromosomal protein MC1 sequences which protect DNA against thermal denaturation..	48.51	6.22E-07	37%	53%	106--169	2--70	1	NP_048794	similar to Methanothrix chromosomal protein MC1A, corresponds to Swiss-Prot Accession Number P15251	141.35	1.57E-32	70%	80%	101--200	1--100
														2	P15249	Chromosomal protein MC1a	54.68	1.93E-06	42%	50%	108--191	4--84
														3	P15250	Chromosomal protein MC1b	51.22	2.13E-05	45%	60%	108--167	4--62
M478L	185179--184862	106	12,215	4.88		No Hit Found								1	NP_048767	A410L	113.62	1.73E-24	48%	74%	4--105	8--110
														2	NP_048765	A408L	78.57	6.17E-14	33%	61%	3--102	40--139
														3	NP_077527	EsV-1-42	50.83	1.38E-05	30%	50%	3--101	6--110
														4	AAR26867	FirrV-1-A43	49.68	3.07E-05	33%	51%	3--103	5--110
M479R	185275--185766	164	18,674	5.08		No Hit Found								1	NP_048768	A411R	79.34	4.24E-14	37%	56%	1--131	1--133
M482R	185802--186338	179	20,331	7.46		No Hit Found								1	NP_048769	A412R	178.72	6.55E-44	49%	65%	1--178	1--179
M484L	187082--186345	246	26,921	9.59		No Hit Found								1	NP_048770	Gln-rich, QQQQM(4x); similar to human transcription factor TFIID, corresponds to Swiss-Prot Accession Number P20226	150.21	5.05E-35	37%	50%	2--245	3--244
M485R	187170--187862	231	25,816	4.81		No Hit Found								1	NP_048765	A408L	209.92	4.76E-53	51%	75%	1--187	37--233
														2	NP_077527	EsV-1-42	64.31	3.23E-09	26%	49%	5--143	7--157
														3	NP_048767	A410L	59.69	7.96E-08	29%	53%	4--97	7--108
														4	AAR26867	FirrV-1-A43	58.54	1.77E-07	24%	53%	4--130	5--143
M488L	188632--187886	249	26,091	12.88		No Hit Found								1	ZP_00283849	COG4991: Uncharacterized protein with a bacterial SH3 domain homologue	98.21	2.33E-19	36%	61%	125--235	153--263
														2	ZP_00840459	collagen triple helix repeat:antifreeze protein, type I	85.89	1.20E-15	37%	47%	125--235	8--124
														3	AAK14819	hsp70-like protein	85.11	2.04E-15	52%	52%	138--225	257--344
														4	CAC51030	procollagen type I alpha 2 chain	84.34	3.49E-15	38%	45%	126--235	878--990
														5	AAB96638	precollagen D	83.57	5.95E-15	37%	45%	125--235	354--464
														6	AAM34601	precollagen-D	83.19	7.77E-15	37%	45%	125--235	347--457
														7	CAE29034	Collagen triple helix repeat:Antifreeze protein, type I	83.19	7.77E-15	36%	45%	123--235	6--121
														8	NP_031761	procollagen, type IV, alpha 4	82.80	1.01E-14	40%	44%	126--235	457--573
														9	NP_892013	collagen, type I, alpha 2	82.80	1.01E-14	38%	44%	126--235	878--990
														10	NP_703618	hsp70 interacting protein, putative	82.42	1.33E-14	51%	51%	134--225	312--395
M491R	188838--189983	382	43,963	8.00		No Hit Found									No Hit Found	No Hit Found						
M495R	190020--190646	209	23,349	8.61		No Hit Found								1	NP_048764	A407L	309.30	4.72E-83	70%	82%	1--209	1--209
M496L	191314--190700	205	22,858	8.17		No Hit Found								1	NP_048877	contains Gln-rich, neutral zinc metallopeptidase, zinc binding region signature	220.71	2.11E-56	63%	79%	1--167	1--165
														2	YP_142679	metal-dependent hydrolase	61.62	1.64E-08	29%	47%	4--181	3--191
M497R	191377--191910	178	19,621	8.62		No Hit Found								1	NP_048879	A523R	230.72	1.43E-59	74%	83%	27--175	20--168
M501R	191937--192380	148	16,447	8.39		No Hit Found								1	NP_048882	A526R	130.95	1.04E-29	52%	63%	18--145	1--127

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
M502L	193151--192417	245	27,605	9.83	1	cd00283	GIY-YIG_Cterm, GIYX(10-11)YIG family of class I homing endonucleases C-terminus (GIY-YIG_Cterm). Homing endonucleases promote the mobility of intron or intein by recognizing and cleaving a homologous allele that lacks the sequence. They catalyze a double-strand break in the DNA near the insertion site of that element to facilitate homing at that site. Class I homing endonucleases are sorted into four families based on the presence of these motifs in their respective N-termini: LAGLIDADG, His-Cys box, HNH, and GIY-YIG. This CD contains several but not all members of the GIY-YIG family. The C-terminus of GIY-YIG is a DNA-binding domain which is separated from the N-terminus by a long, flexible linker. The DNA-binding domain consists of a minor-groove binding alpha-helix, and a helix-turn-helix. Some also contain a zinc finger (i.e. I-TevI) which is not required for DNA binding or catalysis, but is a component of the linker and directs the catalytic domain to cleave the homing site at a fixed distance from the intron insertion site	57.70	1.00E-09	43%	55%	140--241	16--113	1	NP_048671	A315L	245.74	8.76E-64	53%	69%	1--241	1--240
					2	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	44.68	9.16E-06	34%	51%	1--90	1--83	2	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	229.95	4.97E-59	49%	66%	1--244	1--226
					3	smart00497	IENR1, Intron encoded nuclease repeat motif; Repeat of unknown function, but possibly DNA-binding via helix-turn-helix motif (Ponting, unpublished)	40.50	0.000189	40%	58%	190--243	1--53	3	NP_048641	PBCV-1 33kd peptide	175.64	1.11E-42	42%	60%	3--241	6--247
					4	pfam01541	GIY-YIG, GIY-YIG catalytic domain. This domain called GIY-YIG is found in the amino terminal region of excinuclease abc subunit c (uvrC), bacteriophage T4 endonucleases segA, segB, segC, segD and segE; it is also found in putative endonucleases encoded by group I introns of fungi and phage. The structure of I-TevI a GIY-YIG endonuclease, reveals a novel alpha/beta-fold with a central three-stranded antiparallel beta-sheet flanked by three helices. The most conserved and putative catalytic residues are located on a shallow, concave surface and include a metal coordination site. .	40.15	0.00022	27%	45%	1--87	1--89	4	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	165.24	1.50E-39	45%	62%	1--196	1--179
														5	NP_899393	SegD	95.90	1.12E-18	36%	53%	1--190	1--184
														6	YP_293795	putative endonuclease	75.10	2.05E-12	38%	58%	2--107	3--111
														7	AAK09365	intron encoded BmoI	73.94	4.56E-12	31%	48%	4--216	5--242
														8	CAA38804	GIY COII i1 grp IB protein	68.94	1.47E-10	29%	47%	4--222	74--286
														9	AAC49244	ORF301	67.01	5.57E-10	28%	43%	4--179	76--236
														10	AAT53588	group I intron GIY-YIG endonuclease	65.47	1.62E-09	29%	47%	4--212	5--215
M504R	193217--193495	93	11,055	10.73		No Hit Found								1	NP_048883	A527R	76.64	2.29E-13	60%	74%	6--69	5--69
M506L	193756--193547	70	7,977	10.35		No Hit Found									No Hit Found	No Hit Found						
M507L	194020--193784	79	8,652	9.93		No Hit Found								1	NP_048888	A532L	95.13	6.29E-19	60%	77%	1--73	1--74
M508R	194269--195849	527	57,118	4.98		No Hit Found								1	NP_048889	A533R	429.10	1.78E-118	58%	72%	1--362	1--365
														2	AAA66400	unknown protein	427.17	6.75E-118	58%	71%	1--362	1--365
														3	NP_048890	a534R	152.53	3.21E-35	76%	87%	438--526	16--104
														4	NP_077576	EsV-1-91	52.76	3.46E-05	28%	47%	3--161	8--158
M511L	196067--195852	72	8,112	4.28		No Hit Found								1	NP_048891	A535L	62.77	3.41E-09	45%	66%	1--72	1--71
M512L	196357--196121	79	9,145	10.91		No Hit Found								1	NP_048892	A536L	62.77	3.46E-09	41%	55%	1--79	1--73
M515R	196755--197237	161	17,937	9.41		No Hit Found									No Hit Found	No Hit Found						
M518L	201296--197730	1189	120,260	4.65		No Hit Found								1	NP_048896	A540L	617.85	6.95E-175	34%	47%	82--1180	1--1157
														2	AAA66404	unknown protein	140.20	4.25E-31	41%	56%	1004--1180	70--273
														3	ZP_00950302	outer membrane protein	125.18	1.41E-26	28%	37%	383--917	196--765
														4	AAX44675	possible T4-like proximal tail fiber	122.09	1.20E-25	26%	38%	342--910	180--734
														5	ZP_00533161	Hep_Hag	116.70	5.03E-24	23%	35%	132--913	15--776
														6	ZP_00532602	Hep_Hag	100.91	2.86E-19	23%	36%	75--669	307--865
														7	ZP_00732297	hypothetical protein AsucDRAFT_1054	94.74	2.05E-17	21%	33%	16--903	282--1213
														8	AAM39974	outer membrane protein	94.36	2.67E-17	21%	33%	39--903	55--890
														9	CAH36064	putative membrane protein	92.82	7.78E-17	23%	34%	281--1098	81--736
														10	AAU49476	haemagluttinin family protein	92.82	7.78E-17	23%	34%	281--1098	81--736
M522L	203211--201367	615	68,418	8.61		No Hit Found								1	NP_048383	contains Pro-rich Px motif, PAPK (19X); similar to Arabidopsis anter-specific Pro-rich protein, corresponds to Swiss-Prot Accession Number P40602	61.23	1.18E-07	20%	35%	1--295	1--266
M525L	204063--203305	253	28,489	7.70		No Hit Found								1	NP_048357	A9R	219.16	9.37E-56	63%	79%	86--248	11--173
														2	NP_049005	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	126.33	8.26E-28	32%	49%	2--249	8--255
														3	NP_048807	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	112.85	9.45E-24	31%	48%	1--247	1--246
														4	NP_048629	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	103.61	5.74E-21	27%	48%	1--246	4--248
														5	NP_048427	A79R	97.06	5.37E-19	30%	46%	1--226	1--218
														6	NP_048525	A177R	91.28	2.95E-17	25%	45%	1--250	4--245
														7	NP_049003	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	63.54	6.58E-09	50%	68%	1--58	1--54
														8	AAU06302	hypothetical protein A275R	57.00	6.16E-07	28%	47%	115--249	1--134
														9	AAU06304	hypothetical protein A275R	57.00	6.16E-07	26%	46%	86--246	7--167
														10	AAU06301	hypothetical protein A275R	55.84	1.37E-06	26%	46%	86--246	7--167
M526L	207559--204266	1098	110,869	6.64		No Hit Found								1	NP_048896	A540L	584.72	5.99E-165	33%	47%	82--1097	1--1164
														2	ZP_00950302	outer membrane protein	147.13	3.19E-33	25%	39%	29--599	197--773
														3	AAA66404	unknown protein	136.35	5.64E-30	33%	50%	885--1097	32--280
														4	ZP_00533161	Hep_Hag	129.41	6.89E-28	23%	37%	12--710	54--755
														5	AAX44675	possible T4-like proximal tail fiber	125.56	9.95E-27	24%	37%	296--967	14--705
														6	CAH36064	putative membrane protein	122.48	8.42E-26	25%	40%	360--988	77--680
														7	ZP_00500391	COG5295: Autotransporter adhesin	122.48	8.42E-26	25%	40%	360--988	61--664
														8	ZP_00488067	COG5295: Autotransporter adhesin	121.32	1.88E-25	25%	40%	360--988	61--664
														9	AAU49476	haemagluttinin family protein	120.94	2.45E-25	25%	40%	360--988	77--680
														10	ZP_00445997	COG5295: Autotransporter adhesin	120.94	2.45E-25	25%	40%	360--988	61--664
M530L	208054--207623	144	16,139	5.38	1	cd01286	deoxycytidylate_deaminase, Deoxycytidylate deaminase domain. Deoxycytidylate deaminase catalyzes the deamination of dCMP to dUMP, providing the nucleotide substrate for thymidylate synthase. The enzyme binds Zn++, which is required for catalytic activity. The activity of the enzyme is allosterically regulated by the ratio of dCTP to dTTP not only in eukaryotic cells but also in T-even phage-infected Escherichia coli, with dCTP acting as an activator and dTTP as an inhibitor	130.40	1.48E-31	43%	57%	17--135	11--131	1	NP_048952	similar to Vibrio fischeri dCMP deaminase, corresponds to Swiss-Prot Accession Number P33968	205.68	3.29E-52	66%	82%	5--144	3--142
					2	COG2131	ComEB, Deoxycytidylate deaminase [Nucleotide transport and metabolism].	109.28	3.26E-25	34%	53%	2--141	5--148	2	YP_223954	deoxycytidylate deaminase	122.09	4.78E-27	50%	65%	9--139	110--237
					3	pfam00383	dCMP_cyt_deam, Cytidine and deoxycytidylate deaminase zinc-binding region..	87.73	1.05E-18	42%	57%	11--113	9--100	3	ZP_00052863	COG2131: Deoxycytidylate deaminase	110.15	1.88E-23	43%	63%	21--141	20--139
					4	cd00786	cytidine_deaminase-like, Cytidine and deoxycytidylate deaminase zinc-binding region. The family contains cytidine deaminases, nucleoside deaminases, deoxycytidylate deaminases and riboflavin deaminases. Also included are the apoBec family of mRNA editing enzymes. All members are Zn dependent. The zinc ion in the active site plays a central role in the proposed catalytic mechanism, activating a water molecule to form a hydroxide ion that performs a nucleophilic attack on the substrate..	62.72	3.26E-11	33%	52%	21--113	13--92	4	BAE51501	Deoxycytidylate deaminase	108.23	7.15E-23	42%	63%	21--141	20--139

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					5	cd01284	Riboflavin_deaminase-reductase, Riboflavin-specific deaminase. Riboflavin biosynthesis protein RibD (Diaminohydroxyphosphoribosylaminopyrimidine deaminase) catalyzes the deamination of 2,5-diamino-6-ribosylamino-4(3H)-pyrimidinone 5'-phosphate, which is an intermediate step in the biosynthesis of riboflavin.The ribG gene of Bacillus subtilis and the ribD gene of E. coli are bifunctional and contain this deaminase domain and a reductase domain which catalyzes the subsequent reduction of the ribosyl side chain	59.87	2.74E-10	38%	55%	23--113	16--93	5	CAE77204	dCMP deaminase	105.92	3.55E-22	40%	56%	1--142	5--147
					6	COG0117	RibD, Pyrimidine deaminase [Coenzyme metabolism].	56.02	3.75E-09	32%	54%	20--113	22--100	6	YP_424377	probable deoxycytidylate deaminase	103.61	1.76E-21	41%	57%	9--142	6--140
					7	cd01285	nucleoside_deaminase, Nucleoside deaminases include adenosine, guanine and cytosine deaminases. These enzymes are Zn dependent and catalyze the deamination of nucleosides. The zinc ion in the active site plays a central role in the proposed catalytic mechanism, activating a water molecule to form a hydroxide ion that performs a nucleophilic attack on the substrate. The functional enzyme is a homodimer. Cytosine deaminase catalyzes the deamination of cytosine to uracil and ammonia and is a member of the pyrimidine salvage pathway. Cytosine deaminase is found in bacteria and fungi but is not present in mammals; for this reason, the enzyme is currently of interest for antimicrobial drug design and gene therapy applications against tumors. Some members of this family are tRNA-specific adenosine deaminases that generate inosine at the first position of their anticodon (position 34) of specific tRNAs; this modification is thought to enlarge the codon recognition capacity during protein synthesis. Other members of the family are guanine deaminases which deaminate guanine to xanthine as part of the utilization of guanine	55.27	6.23E-09	33%	57%	28--120	19--100	7	IP_001016265	hypothetical protein LOC549019	102.83	3.00E-21	37%	57%	2--140	22--162
					8	COG0590	CumB, Cytosine/adenosine deaminases [Nucleotide transport and metabolism / Translation. ribosomal structure and biogenesis].	51.11	1.20E-07	41%	54%	28--113	30--104	8	ZP_00527356	Cytidine/deoxycytidylate deaminase, zinc-binding region	100.91	1.14E-20	46%	68%	21--113	22--115
														9	XP_781375	PREDICTED: similar to Deoxycytidylate deaminase (dCMP deaminase)	100.14	1.95E-20	38%	54%	1--139	60--200
														10	XP_849027	PREDICTED: similar to Deoxycytidylate deaminase (dCMP deaminase) isoform 1	99.75	2.54E-20	38%	54%	2--139	11--150
M531L	209217--208123	365	41,273	8.49		No Hit Found								1	NP_048502	A154L	295.05	2.44E-78	44%	59%	24--364	6--347
														2	NP_048920	similar to Chlorella virus PBCV-1 ORF A154L, corresponds to GenBank Accession Number U42580	264.23	4.61E-69	43%	56%	15--365	1--351
														3	NP_048477	similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903	260.38	6.66E-68	39%	57%	24--365	8--357
														4	NP_077492	EsV-1-7	105.15	3.58E-21	37%	47%	19--196	5--190
														5	EAA73437	hypothetical protein FG03969.1	55.84	2.49E-06	26%	36%	55--230	316--477
														6	NP_077600	EsV-1-115	53.53	1.24E-05	26%	34%	5--196	86--289
M535L	211856--209244	871	95,334	6.52	1	COG0474	MgtA, Cation transport ATPase [Inorganic ion transport and metabolism].	535.71	1.49E-153	30%	51%	20--849	33--898	1	XP_483341	putative calcium-transporting ATPase 8, plasma membrane-type	534.26	7.17E-150	36%	55%	34--849	172--1053
					2	COG2217	ZntA, Cation transport ATPase [Inorganic ion transport and metabolism].	232.81	2.00E-62	27%	48%	72--681	152--678	2	XP_636219	hypothetical protein DDB0188438	531.95	3.56E-149	35%	55%	36--854	64--977
					3	pfam00122	E1-E2_ATPase, E1-E2 ATPase..	158.74	4.77E-40	33%	55%	101--334	1--223	3	NP_194719	ACA10; calcium-transporting ATPase/ calmodulin binding	526.55	1.50E-147	36%	55%	35--850	159--1035
					4	COG2216	KdpB, High-affinity K+ transport system, ATPase chain B [Inorganic ion transport and metabolism]	158.44	5.43E-40	28%	49%	48--699	20--609	4	CAD67616	calcium-dependent ATPase	521.93	3.68E-146	37%	55%	25--849	140--1015
					5	pfam00702	Hydrolase, haloacid dehalogenase-like hydrolase. This family are structurally different from the alpha/ beta hydrolase family (pfam00561). This family includes L-2-haloacid dehalogenase, epoxide hydrolases and phosphatases. The structure of the family consists of two domains. One is an inserted four helix bundle, which is the least well conserved region of the alignment. The rest of the fold is composed of the core alpha/beta domain..	64.23	1.31E-11	20%	40%	449--631	31--197	5	CAD67615	putative P-type II calcium ATPase	512.30	2.92E-143	38%	56%	40--850	153--1017
					6	pfam00689	Cation_ATPase_C, Cation transporting ATPase, C-terminus. Members of this families are involved in Na+/K+, H+/K+, Ca++ and Mg++ transport	56.40	2.54E-09	21%	38%	726--850	1--147	6	XP_473800	OSJNBb0015N08.12	510.38	1.11E-142	35%	55%	31--848	158--1040
					7	pfam00690	Cation_ATPase_N, Cation transporter/ATPase, N-terminus. Members of this families are involved in Na+/K+, H+/K+, Ca++ and Mg++ transport..	47.13	1.67E-06	33%	52%	23--80	21--79	7	NP_851200	ACA8 (AUTOINHIBITED CA2+ -ATPASE, ISOFORM 8); calcium-transporting ATPase/ calmodulin binding	509.22	2.47E-142	36%	55%	35--849	159--1032
					8	COG4087	COG4087, Soluble P-type ATPase [General function prediction only].	46.50	2.67E-06	30%	53%	530--644	46--138	8	EAL90415	P-type calcium ATPase, putative	508.45	4.22E-142	35%	54%	35--849	275--1191
					9	COG0561	Cof, Predicted hydrolases of the HAD superfamily [General function prediction only].	45.87	3.81E-06	32%	47%	594--660	197--263	9	CAB43665	Ca2+-transporting ATPase-like protein	506.91	1.23E-141	35%	53%	35--850	159--1059
					10	COG0560	SerB, Phosphoserine phosphatase [Amino acid transport and metabolism].	40.30	0.000187	25%	40%	488--637	49--195	10	AAU44048	putative P-type ATPase	504.98	4.66E-141	35%	55%	30--859	121--994
M543R	212005--212820	272	30,497	8.76	1	cd00180	S_TKc, Serine/Threonine protein kinases, catalytic domain. Phosphotransferases of the serine or threonine-specific kinase subfamily. The enzymatic activity of these protein kinases is controlled by phosphorylation of specific residues in the activation segment of the catalytic domain, sometimes combined with reversible conformational changes in the C-terminal autoregulatory tail..	153.82	1.28E-38	29%	51%	15--265	1--256	1	NP_048643	similar to PBCV-1 serine/threonine protein kinase, corresponds to GenBank Accession Number U14660	134.81	2.64E-30	32%	49%	3--265	5--279
					2	smart00220	S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-specific kinase subfamily. .	151.52	5.62E-38	28%	49%	16--265	1--256	2	CAG59101	unnamed protein product	102.45	1.45E-20	29%	50%	20--267	49--309
					3	pfam00069	Pkinase, Protein kinase domain..	148.51	5.37E-37	29%	48%	16--265	1--258	3	XP_666308	calcium-dependent protein kinase	94.36	3.95E-18	30%	50%	22--270	191--445
					4	COG0515	SPS1, Serine/threonine protein kinase [General function prediction only / Signal transduction mechanisms / Transcription / DNA replication, recombination. and repair].	108.32	6.85E-25	25%	44%	16--270	2--283	4	AAS47705	calcium-dependent protein kinase 1	94.36	3.95E-18	30%	50%	22--270	191--445
					5	cd00192	TyrKc, Tyrosine kinase, catalytic domain. Phosphotransferases; tyrosine-specific kinase subfamily. Enzymes with TyrKc domains belong to an extensive family of proteins which share a conserved catalytic core common to both serine/threonine and tyrosine protein kinases. Enzymatic activity of tyrosine protein kinases is controlled by phosphorylation of specific tyrosine residues in the activation segment of the catalytic domain or a C-terminal tyrosine (tail) residue with reversible conformational changes	82.54	3.41E-17	27%	44%	13--205	6--207	5	XP_504770	hypothetical protein	94.36	3.95E-18	28%	48%	20--265	103--350
					6	smart00219	TyrKc, Tyrosine kinase, catalytic domain; Phosphotransferases. Tyrosine-specific kinase subfamily. .	78.68	4.74E-16	26%	47%	17--205	2--196	6	CAA99015	CMK2	93.59	6.75E-18	26%	49%	20--265	51--309
														7	ABA60892	calmodulin-like domain protein kinase isoform 3	93.59	6.75E-18	26%	46%	15--265	75--333
														8	EAA76933	hypothetical protein FG07121.1	93.20	8.81E-18	27%	46%	14--269	285--556
														9	CAA40281	calmodulin-dependent protein kinase type II	92.82	1.15E-17	26%	49%	20--265	51--309
														10	XP_667020	calmodulin-domain protein kinase 1	92.82	1.15E-17	24%	48%	16--270	54--315
M546R	212942--216115	1058	117,663	9.64	1	smart00433	TOP2c, TopoisomeraseII; Eukaryotic DNA topoisomerase II, GyrB, ParE	498.55	2.12E-142	33%	49%	50--610	1--589	1	AAU95770	topoisomerase II	1928.68	0.00E+00	91%	93%	1--1058	1--1058
					2	smart00434	TOP4c, DNA Topoisomerase IV; Bacterial DNA topoisomerase IV, GyrA, ParC .	423.06	1.28E-119	34%	50%	626--1051	1--457	2	NP_048939	PBCV-1 DNA topoisomerase II	1360.90	0.00E+00	63%	76%	3--1056	2--1061
					3	cd00187	TOP4c, DNA Topoisomerase, subtype IIA; domain A'; bacterial DNA topoisomerase IV (C subunit, ParC), bacterial DNA gyrases (A subunit, GyrA),mammalian DNA toposiomerases II. DNA topoisomerases are essential enzymes that regulate the conformational changes in DNA topology by catalysing the concerted breakage and rejoining of DNA strands during normal cellular growth	398.01	4.23E-112	33%	50%	644--1056	1--445	3	CAD25222	DNA TOPOISOMERASE II	893.65	0.00E+00	46%	62%	5--1053	8--1066
					4	COG0187	GyrB, Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), B subunit [DNA replication, recombination, and repair]	360.60	8.10E-101	28%	44%	2--626	8--635	4	AAW40881	DNA topoisomerase II, putative	866.69	0.00E+00	44%	59%	5--1057	112--1235
					5	pfam00521	DNA_topoisoIV, DNA gyrase/topoisomerase IV, subunit A..	300.93	6.09E-83	31%	46%	646--1053	1--434	5	BAD86854	DNA topoisomerase II	860.14	0.00E+00	42%	59%	5--1057	103--1249
					6	COG0188	GyrA, Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), A subunit [DNA replication, recombination, and repair].	267.90	5.37E-73	25%	44%	617--1051	2--466	6	NP_189031	TOPII (TOPOISOMERASE II); ATP binding / DNA binding / DNA topoisomerase (ATP-hydrolyzing)	859.75	0.00E+00	43%	60%	5--1056	35--1166
					7	pfam00204	DNA_gyraseB, DNA gyrase B. This family represents the second domain of DNA gyrase B which has a ribosomal S5 domain 2-like fold. This family is structurally related to PF01119..	72.97	3.02E-14	27%	45%	237--382	14--168	7	XP_467311	putative DNA topoisomerase II	854.74	0.00E+00	43%	58%	5--1054	32--1182
														8	BAE06274	topoisomerase II	853.20	0.00E+00	44%	58%	5--1054	35--1166
														9	AAN85208	DNA topoisomerase II	850.89	0.00E+00	43%	59%	5--1054	27--1148
														10	XP_759648	hypothetical protein UM03501.1	837.80	0.00E+00	42%	58%	5--1057	100--1235
M553R	216182--216589	136	15,899	11.01		No Hit Found								1	NP_048933	A577L	109.00	4.19E-23	65%	69%	26--111	2--85
M556R	216631--217152	174	19,111	9.91		No Hit Found								1	NP_048732	A375R	202.22	5.18E-51	58%	70%	4--171	6--172

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
M557L	218777--217155	541	56,442	10.43	1	pfam05616	Neisseria_TspB, Neisseria meningitidis TspB protein. This family consists of several Neisseria meningitidis TspB virulence factor proteins..	48.97	5.00E-07	33%	40%	123--196	322--395	1	NP_048762	Pro-, Lys-rich, PAPK (30x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	200.29	1.39E-49	37%	57%	266--541	214--496
					2	pfam05518	Totivirus_coat, Totivirus coat protein..	48.44	6.02E-07	24%	29%	16--145	630--751	2	BAE02830	surface protein	55.07	7.22E-06	44%	62%	201--258	572--629
M560L	219451--218903	183	20,996	3.20	1	pfam04931	DNA_pol_V, DNA polymerase V. This family includes the fifth essential DNA polymerase in yeast EC:2.7.7.7. Pol5p is localised exclusively to the nucleolus and binds near or at the enhancer region of rRNA-encoding DNA repeating units..	36.57	0.0024	24%	42%	55--154	644--740	1	NP_048761	Glu-, Asn-rich	78.95	7.56E-14	63%	77%	1--57	2--59
M563L	219757--219494	88	10,052	10.28		No Hit Found								1	NP_048760	A403R	95.13	6.33E-19	52%	69%	2--88	6--93
M564L	220484--219801	228	24,869	6.88		No Hit Found								1	NP_048759	A402R	231.88	1.14E-59	48%	72%	4--228	5--227
M567L	221450--220581	290	32,586	5.48		No Hit Found								1	NP_048758	A401R	337.42	2.98E-91	60%	75%	4--287	2--274
														2	BAC51116	bll5851	140.58	5.35E-32	34%	53%	34--284	7--245
														3	ZP_00863599	conserved hypothetical protein	130.57	5.54E-29	33%	53%	34--284	11--249
														4	YP 190685	hypothetical protein GOX0246	119.01	1.67E-25	32%	50%	28--287	5--253
														5	ZP_00577151	conserved hypothetical protein	69.32	1.52E-10	24%	41%	23--289	2--263
M569L	221845--221495	117	13,086	10.03		No Hit Found								1	NP_048757	similar to PBCV-1 ORF A214, corresponds to GenBank Accession Number U42580	84.34	1.12E-15	37%	57%	5--117	4--116
M570L	221876--222109	78	18,239	9.90	1	pfam00075	RnaseH, RNase H. RNase H digests the RNA strand of an RNA/DNA hybrid. Important enzyme in retroviral replication cycle, and often found as a domain associated with reverse transcriptases. Structure is a mixed alpha+beta fold with three a/b/a layers..	73.40	1.85E-14	30%	47%	16--136	2--124	1	NP_048756	A399R	147.90	9.70E-35	51%	64%	4--158	17--180
					2	COG0328	RnhA, Ribonuclease HI [DNA replication, recombination, and repair].	57.29	1.52E-09	23%	37%	17--147	3--153	2	XP_662792	hypothetical protein AN5188.2	53.53	2.49E-06	32%	44%	18--136	228--373
														3	CAA41925	unnamed protein product	52.76	4.25E-06	32%	51%	15--140	238--369
														4	S37765	hypothetical protein - fruit fly (Drosophila miranda) transposon TRIM	52.76	4.25E-06	32%	51%	15--140	143--274
														5	EAL40632	ENSANGP00000006819	50.83	1.61E-05	28%	46%	12--137	247--372
														6	1RDH_B	Chain B, Hiv-1 Reverse Transcriptase (Ribonuclease H Domain) (E.C.2.7.7.49)	50.45	2.11E-05	27%	42%	9--136	1--141
														7	EAL40633	ENSANGP00000020605	50.06	2.75E-05	28%	46%	12--137	272--397
														8	AAB65093	Lian-Aa1 retrotransposon protein	49.68	3.59E-05	31%	52%	15--147	910--1046
														9	EAL88427	ribonuclease h1	49.29	4.70E-05	31%	46%	18--136	239--383
														10	BAE61068	unnamed protein product	48.52	8.01E-05	31%	45%	18--136	236--380
M571R	222368--221877	164	9,065	8.81		No Hit Found									No Hit Found	No Hit Found						
M572R	222471--222824	118	12,885	10.39		No Hit Found								1	NP 048755	A398L	117.86	9.08E-26	49%	63%	1--118	1--118
M574L	223289--222831	153	17,499	11.73		No Hit Found								1	AAU06284	hypothetical protein	69.71	2.80E-11	32%	49%	7--143	23--141
M577L	223684--223316	123	14,265	10.12		No Hit Found								1	NP_048752	A395R	107.46	1.24E-22	66%	78%	43--122	1--80
M578R	223670--224293	208	23,718	8.20		No Hit Found								1	NP_048834	Lys-, Arg-rich; contains eukaryotic putative RNA-binding region RNP-1 signature; similar to PBCV-1 ORF A267L, corresponds to GenBank Accession Number U42580	166.01	6.39E-40	43%	58%	19--207	106--294
														2	NP_048846	Lys-, Glu-rich	159.46	5.98E-38	42%	59%	17--195	104--282
														3	YP 142777	unknown	98.60	1.25E-19	34%	49%	19--185	272--441
														4	NP_048621	A267L	93.97	3.08E-18	32%	49%	13--207	59--247
M580L	224861--224292	190	21,463	3.81		No Hit Found								1	NP_048751	A394R	92.43	7.21E-18	44%	63%	6--109	14--116
M584R	224914--225219	102	12,138	10.30		No Hit Found									No Hit Found	No Hit Found						
M586L	226085--225321	255	29,033	9.23	1	smart00382	AAA, ATPases associated with a variety of cellular activities; AAA - ATPases associated with a variety of cellular activities. This profile/alignment only detects a fraction of this vast family. The poorly conserved N-terminal helix is missing from the alignment..	40.89	0.000126	17%	37%	16--154	2--146	1	NP_048749	contains ATP/GTP-binding site motif A; similar to frog virus 3 ATPase, corresponds to GenBank Accession Number M80551	405.60	7.17E-112	78%	86%	1--252	1--251
					2	COG4615	PvdE, ABC-type siderophore export system, fused ATPase and permease components [Secondary metabolites biosynthesis, transport, and catabolism / Inorganic ion transport and metabolism].	37.62	0.001405	28%	40%	12--102	345--448	2	BAB69884	AGB-1	152.14	1.43E-35	35%	55%	1--245	1--253
					3	COG1123	COG1123, ATPase components of various ABC-type transport systems, contain duplicated ATPase [General function prediction only]	36.70	0.002422	32%	52%	12--98	31--97	3	AAR26836	FirrV-1-A12	147.90	2.69E-34	39%	57%	32--230	36--243
					4	cd01120	RecA-like_NTPases, RecA-like NTPases. This family includes the NTP binding domain of F1 and V1 H+ATPases, DnaB and related helicases as well as bacterial RecA and related eukaryotic and archaeal recombinases. This group also includes bacterial conjugation proteins and related DNA transfer proteins involved in type II and type IV secretion..	35.09	0.007282	23%	40%	18--105	1--95	4	NP_077511	EsV-1-26	135.19	1.81E-30	35%	57%	32--234	47--258
														5	AAT09674	AAA-ATPase	122.87	9.28E-27	33%	56%	32--235	43--248
														6	AAL77796	putative ATPase	122.87	9.28E-27	33%	56%	32--235	43--248
														7	AAM82544	putative ATPase	122.48	1.21E-26	33%	56%	32--235	43--248
														8	YP 003858	ATPase	121.71	2.07E-26	33%	56%	32--231	43--244
														9	AAS18149	ATPase	120.55	4.61E-26	31%	56%	32--235	46--251
														10	AAV91100	ATPase-like protein	120.55	4.61E-26	31%	56%	32--235	46--251
M591R	226168--227925	586	61,374	6.77	1	pfam05887	Trypan_PARP, Procyclic acidic repetitive protein (PARP). This family consists of several Trypanosoma brucei procyclic acidic repetitive protein (PARP) like sequences. The procyclic acidic repetitive protein (parp) genes of Trypanosoma brucei encode a small family of abundant surface proteins whose expression is restricted to the procyclic form of the parasite. They are found at two unlinked loci, parpA and parpB; transcription of both loci is developmentally regulated.	56.53	2.70E-09	31%	42%	460--534	50--124	1	NP_048741	Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472	177.56	1.06E-42	39%	50%	214--464	8--263
					2	pfam05616	Neisseria_TspB, Neisseria meningitidis TspB protein. This family consists of several Neisseria meningitidis TspB virulence factor proteins..	54.75	8.62E-09	39%	43%	478--532	323--377	2	NP_048921	A565R	115.93	3.80E-24	34%	41%	130--375	252--466
					3	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	49.76	2.99E-07	32%	39%	441--535	31--126	3	BAA11343	DNA binding protein	112.85	3.22E-23	34%	41%	130--375	252--458
					4	pfam04625	DEC-1_N, DEC-1 protein, N terminal region. The defective chorion-1 gene (dec-1) in Drosophila encodes follicle cell proteins necessary for proper eggshell assembly. Multiple products of the dec-1 gene are formed by alternative RNA splicing and proteolytic processing. Cleavage products include S80 (80 kDa) which is incorporated into the eggshell, and further proteolysis of S80 gives S60 (60 kDa)	46.03	4.01E-06	33%	34%	468--532	78--142	4	NP_048688	a332L	73.94	1.66E-11	61%	80%	544--585	2--43
					5	COG3147	DedD, Uncharacterized protein conserved in bacteria [Function unknown].	45.02	6.90E-06	35%	42%	460--535	79--150	5	NP_473058	hypothetical protein PFB0650w	70.86	1.40E-10	57%	76%	9--59	596--647
					6	pfam05518	Totivirus_coat, Totivirus coat protein..	44.97	7.71E-06	25%	29%	466--535	680--753	6	BAC69973	putative serine/threonine protein kinase	67.01	2.03E-09	67%	80%	14--59	557--602
					7	pfam06735	DUF1210, Protein of unknown function (DUF1210). This family represents a conserved region within plant proline-rich proteins	41.64	7.32E-05	42%	44%	464--532	136--202	7	NP_048917	similar to Chlorella virus CVK2 DNA binding protein, corresponds to GenBank Accession Number D78305	63.54	2.24E-08	39%	51%	284--375	306--403
					8	COG5373	COG5373, Predicted membrane protein [Function unknown].	40.73	0.000154	31%	34%	468--532	49--111	8	BAA11342	DNA binding protein	60.08	2.48E-07	38%	50%	284--375	304--401
					9	COG3898	COG3898, Uncharacterized membrane-bound protein [Function unknown].	40.43	0.00016	24%	34%	461--544	439--525	9	P12348	Period circadian protein	59.69	3.23E-07	52%	58%	9--59	744--794
					10	pfam07462	MSP1_C, Merozoite surface protein 1 (MSP1) C-terminus. This family represents the C-terminal region of merozoite surface protein 1 (MSP1) which are found in a number of Plasmodium species. MSP-1 is a 200-kDa protein expressed on the surface of the P. vivax merozoite. MSP-1 of Plasmodium species is synthesised as a high-molecular-weight precursor and then processed into several fragments. At the time of red cell invasion by the merozoite, only the 19-kDa C-terminal fragment (MSP-119), which contains two epidermal growth factor-like domains, remains on the surface. Antibodies against MSP-119 inhibit merozoite entry into red cells, and immunisation with MSP-119 protects monkeys from challenging infections. Hence, MSP-119 is considered a promising vaccine candidate.	40.05	0.000259	23%	27%	478--531	259--311	10	XP_500901	hypothetical protein	57.00	2.10E-06	37%	62%	13--60	327--374
M599R	228277--229137	287	30,575	6.51	1	COG3889	COG3889, Predicted solute binding protein [General function prediction only].	38.10	0.00079	25%	40%	153--228	764--841		No Hit Found	No Hit Found						

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
M601L	230203--229127	359	40,298	7.39	1	COG0076	GadB, Glutamate decarboxylase and related PLP-dependent proteins [Amino acid transport and metabolism]	147.45	1.02E-36	24%	46%	55--310	102--380	1	NP_048954	similar to tomato histidine decarboxylase, corresponds to Swiss-Prot Accession Number P54772	464.54	2.27E-129	61%	78%	4--353	11--362
					2	pfam00282	Pyridoxal_deC, Pyridoxal-dependent decarboxylase conserved domain..	118.87	3.90E-28	27%	45%	56--289	81--352	2	P28578	Histidine decarboxylase (HDC)	216.08	1.40E-54	35%	55%	14--347	26--369
					3	COG0520	CsdB, Selenocysteine lyase [Amino acid transport and metabolism].	48.40	6.36E-07	25%	41%	63--308	75--339	3	AAA25071	histidine decarboxylase	216.08	1.40E-54	35%	55%	14--347	26--369
					4	COG1104	NifS, Cysteine sulfinate desulfinase/cysteine desulfurase and related enzymes [Amino acid transport and metabolism]	48.63	6.49E-07	24%	44%	51--307	36--298	4	ZP_00510529	Pyridoxal-dependent decarboxylase	213.39	9.10E-54	35%	57%	14--347	47--390
					5	pfam00266	Aminotran_5, Aminotransferase class-V..	36.83	0.002041	24%	39%	77--231	55--206	5	ZP_00106716	COG0076: Glutamate decarboxylase and related PLP-dependent proteins	207.22	6.52E-52	36%	55%	15--349	25--365
					6	COG2008	GLY1, Threonine aldolase [Amino acid transport and metabolism].	34.84	0.00787	19%	35%	57--216	38--206	6	AAO92385	histidine decarboxylase	205.30	2.48E-51	33%	55%	14--347	26--369
														7	AAR12533	histidine decarboxylase	205.30	2.48E-51	33%	55%	14--347	40--383
														8	P05034	Histidine decarboxylase (HDC)	202.60	1.61E-50	32%	53%	14--355	26--377
														9	AAO65983	putative pyridoxal 5' phosphate-dependent histidine decarboxylase	198.36	3.03E-49	32%	54%	14--347	26--369
														10	BAA78331	serine decarboxylase	198.36	3.03E-49	32%	54%	11--347	112--458
M605L	230627--230226	134	15,194	4.86		No Hit Found								1	NP_048958	A602L	54.68	9.47E-07	47%	76%	8--53	68--113
M606R	230722--231036	105	11,996	6.00		No Hit Found								1	NP_048959	A603R	96.67	2.20E-19	43%	63%	1--103	1--104
M607L	231719--231219	167	18,477	11.04		No Hit Found									No Hit Found	No Hit Found						
M610L	232680--232162	173	20,120	6.04		No Hit Found								1	NP_048848	A492L	112.85	4.05E-24	37%	59%	7--173	18--189
														2	NP_048849	A493L	85.89	5.31E-16	39%	57%	28--165	29--174
M612R	232763--233833	357	42,324	7.99		No Hit Found								1	NP_048850	A494R	403.68	4.70E-111	55%	71%	7--357	4--360
														2	YP_142783	PBCV1-A494R-like protein	150.60	7.18E-35	34%	54%	92--357	166--431
														3	NP_077586	EsV-1-101	129.41	1.71E-28	36%	54%	166--356	121--318
														4	AAR26895	FirrV-1-B20	122.87	1.60E-26	32%	51%	148--356	86--300
														5	YP_294161	hypothetical protein EhV 403	112.85	1.66E-23	33%	52%	152--353	156--362
														6	NP 149745	282R	70.86	7.23E-11	22%	43%	96--354	101--380
														7	YP_142784	unknown	57.38	8.27E-07	27%	46%	236--351	131--247
M615R	233914--234339	142	14,986	10.14		No Hit Found									No Hit Found	No Hit Found						
M616L	235330--234347	328	35,318	4.75	1	pfam05887	Trypan_PARP, Procyclic acidic repetitive protein (PARP). This family consists of several Trypanosoma brucei procyclic acidic repetitive protein (PARP) like sequences. The procyclic acidic repetitive protein (parp) genes of Trypanosoma brucei encode a small family of abundant surface proteins whose expression is restricted to the procyclic form of the parasite. They are found at two unlinked loci, parpA and parpB; transcription of both loci is developmentally regulated	54.60	1.08E-08	40%	55%	129--196	60--127	1	NP_048857	A501L	63.54	1.02E-08	48%	67%	4--73	5--66
					2	pfam05616	Neisseria_TspB, Neisseria meningitidis TspB protein. This family consists of several Neisseria meningitidis TspB virulence factor proteins..	51.29	9.36E-08	32%	38%	121--193	323--397	2	NP_048856	Pro-, Ser-rich	60.46	8.61E-08	50%	67%	255--303	4--64
M618L	235642--235418	75	8,175	4.28		No Hit Found								1	NP 048847	A491R	63.16	2.68E-09	38%	56%	4--75	5--76
M620L	236651--235692	320	34,621	5.24		No Hit Found								1	NP_048844	Pro-,Gln-rich	358.61	1.47E-97	60%	72%	1--320	1--317
														2	YP_142771	unknown	58.54	3.14E-07	25%	39%	8--256	5--245
M622L	237725--236679	349	40,425	9.69	1	pfam01844	HNH, HNH endonuclease..	38.39	0.000787	16%	24%	237--286	1--51	1	NP_048779	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081	104.38	5.69E-21	30%	50%	64--316	59--313
														2	YP_025062	putative endonuclease	86.27	1.60E-15	37%	50%	193--346	15--164
														3	YP_142599	HNH endonuclease	83.96	7.96E-15	29%	41%	16--283	10--267
														4	YP_223875	putative HNH homing endonuclease	79.72	1.50E-13	33%	50%	197--346	20--179
														5	YP_142601	HNH endonuclease	73.56	1.08E-11	36%	49%	190--346	18--173
														6	CAD43927	putative endodeoxyribonuclease	70.48	9.11E-11	35%	50%	215--346	72--203
														7	YP 164778	orf143	68.55	3.46E-10	33%	46%	183--346	4--173
														8	AAT34477	prophage LambdaBa03, HNH endonuclease family protein	67.01	1.01E-09	33%	51%	180--311	9--146
														9	NP_047162	putative HNH homing endonuclease	66.63	1.32E-09	32%	47%	190--348	10--165
														10	YP_358806	putative HNH endonuclease	65.08	3.83E-09	30%	44%	185--346	5--184
M625R	237774--238019	82	9,216	6.80		No Hit Found									No Hit Found	No Hit Found						
M627L	238682--238251	144	16,779	10.18	1	pfam03013	Pyr_excise, Pyrimidine dimer DNA glycosylase. Pyrimidine dimer DNA glycosylases excise pyrimidine dimers by hydrolysis of the glycosylic bond of the 5' pyrimidine, followed by the intra-pyrimidine phosphodiester bond. Pyrimidine dimers are the major UV-lesions of DNA	160.21	1.43E-40	47%	60%	1--134	1--131	1	AAD33374	pyrimidine dimer-specific glycosylase	163.70	1.43E-39	59%	69%	1--134	1--133
														2	NP_048398	PBCV-1 pyrimidine dimer-specific glycosylase	163.31	1.87E-39	59%	69%	1--134	1--133
														3	AAD33369	pyrimidine dimer-specific glycosylase	162.93	2.45E-39	58%	70%	1--134	1--133
														4	AAD33379	pyrimidine dimer-specific glycosylase	162.16	4.17E-39	58%	69%	1--134	1--133
														5	AAD33391	pyrimidine dimer-specific glycosylase	161.77	5.45E-39	58%	71%	1--134	1--133
														6	AAD33381	pyrimidine dimer-specific glycosylase	161.77	5.45E-39	58%	71%	1--134	1--133
														7	AAD33380	pyrimidine dimer-specific glycosylase	161.77	5.45E-39	58%	71%	1--134	1--133
														8	AAD33377	pyrimidine dimer-specific glycosylase	161.38	7.11E-39	58%	70%	1--134	1--133
														9	AAD33352	pyrimidine dimer-specific glycosylase	161.38	7.11E-39	58%	70%	1--134	1--133
														10	AAD33375	pyrimidine dimer-specific glycosylase	161.00	9.29E-39	58%	71%	1--134	1--133
M628L	239312--239040	91	9,598	10.67		No Hit Found								1	NP_048836	Gly-rich	92.82	3.11E-18	59%	81%	3--68	4--69
M629L	240038--239343	232	26,198	4.58	1	COG5219	COG5219, Uncharacterized conserved protein, contains RING Zn-finger [General function prediction only]	39.74	0.000253	28%	41%	149--198	171--1525	1	NP_048837	A481L	208.38	1.40E-52	44%	65%	1--228	1--224
M631L	241199--240105	365	41,836	10.81		No Hit Found								1	NP_048441	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563	176.79	9.68E-43	39%	59%	1--229	46--296
														2	NP_048439	a91L	120.17	1.08E-25	49%	64%	241--364	1--126
														3	NP_049032	similar to Chlorella virus PBCV-1 ORF A282L, corresponds to GenBank Accession Number U42580	60.46	1.01E-07	56%	70%	3--61	2--60
														4	NP_048636	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055	57.00	1.12E-06	72%	82%	3--41	530--569
														5	NP_048632	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662	57.00	1.12E-06	72%	82%	3--41	571--610
M634R	241316--241588	91	10,827	4.50		No Hit Found								1	NP_048809	A452L	48.14	8.77E-05	37%	56%	11--91	3--79
M635R	241632--242276	215	24,419	9.28		No Hit Found								1	NP_048838	A482R	199.90	4.25E-50	44%	64%	3--214	1--213
														2	AAR26970	FirrV-1-I5	64.70	2.14E-09	32%	44%	33--130	31--146
														3	YP_294196	hypothetical protein EhV 438	63.54	4.76E-09	34%	50%	33--130	73--168
														4	AAR26889	FirrV-1-B14	63.16	6.21E-09	31%	46%	25--130	20--145
														5	NP_077581	EsV-1-96	55.07	1.69E-06	31%	46%	37--130	70--181
M637L	242741--242277	155	18,317	6.80		No Hit Found								1	NP_048840	A484L	181.03	8.81E-45	54%	75%	1--154	1--154
M638R	242798--244501	568	61,762	10.39	1	cd01820	PAF_acetylesterase_like, PAF_acetylhydrolase (PAF-AH)_like subfamily of SGNH-hydrolases. Platelet-activating factor (PAF) and PAF-AH are key players in inflammation and in atherosclerosis. PAF-AH is a calcium independent phospholipase A2 which exhibits strong substrate specificity towards PAF, hydrolyzing an acetyl ester at the sn-2 position. PAF-AH also degrades a family of oxidized PAF-like phospholipids with short sn-2 residues. In addition, PAF and PAF-AH are associated with neural migration and mammalian reproduction..	69.51	3.01E-13	26%	42%	254--412	35--210	1	NP_048488	PBCV-1 surface protein	86.27	3.10E-15	26%	41%	321--566	659--890

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					2	cd01828	sialate_O-acetylesterase_like2, sialate_O-acetylesterase_like subfamily of the SGNH-hydrolases, , a diverse family of lipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alpha/beta hydrolase family and unique among all known hydrolases; its active site closely resembles the Ser-His-Asp(Glu) triad found in other serine hydrolases	60.68	1.37E-10	24%	41%	254--393	2--155	2	BAD22850	surface protein	83.96	1.54E-14	33%	48%	415--566	753--890
					3	cd01833	XynB_like, SGNH_hydrolase subfamily, similar to Ruminococcus flavefaciens XynB. Most likely a secreted hydrolase with xylanase activity. SGNH hydrolases are a diverse family of lipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alpha/beta hydrolase family and unique among all known hydrolases; its active site closely resembles the Ser-His-Asp(Glu) triad found in other serine hydrolases	48.02	9.22E-07	20%	30%	254--399	3--190	3	BAD12236	surface protein	81.65	7.63E-14	30%	44%	396--566	734--890
					4	cd00229	SGNH_hydrolase, SGNH_hydrolase, or GDSL_hydrolase, is a diverse family of lipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alpha/beta hydrolase family and unique among all known hydrolases; its active site closely resembles the typical Ser-His-Asp(Glu) triad from other serine hydrolases, but may lack the carboxlic acid	43.29	2.12E-05	18%	30%	253--400	1--192	4	T17636	proline-rich protein A145R - Chlorella virus PBCV-1	74.71	9.33E-12	37%	50%	452--566	2--99
					5	cd01841	NnaC_like, NnaC (CMP-NeuNAc synthetase) _like subfamily of SGNH_hydrolases, a diverse family of lipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alpha/beta hydrolase family and unique among all known hydrolases; its active site closely resembles two of the three components of typical Ser-His-Asp(Glu) triad from other serine hydrolases. E. coli NnaC appears to be involved in polysaccharide synthesis	43.33	2.57E-05	25%	43%	253--400	1--160	5	BAE02830	surface protein	70.86	1.35E-10	30%	48%	418--564	504--629
					6	cd01838	Isoamyl_acetate_hydrolase_like, Isoamyl-acetate hydrolyzing esterase-like proteins. SGNH_hydrolase subfamily similar to the Saccharomyces cerevisiae IAH1. IAH1 may be the major esterase that hydrolyses isoamyl acetate in sake mash. The SGNH-family of hydrolases is a diverse family of lipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alpha/beta hydrolase family and unique among all known hydrolases; its active site closely resembles the Ser-His-Asp(Glu) triad found in other serine hydrolases.	40.24	0.000193	18%	34%	254--408	2--198	6	AAH59352	MGC69176 protein	54.30	1.30E-05	27%	44%	254--414	42--227
M647L	245060--244572	163	18,896	3.57	1	COG5406	COG5406, Nucleosome binding factor SPN, SPT16 subunit [Transcription / DNA replication, recombination, and repair / Chromatin structure and dynamics].	40.08	0.000241	29%	55%	96--163	929--995		No Hit Found	No Hit Found						
M649R	245243--246649	469	53,138	10.92		No Hit Found								1	NP_048441	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563	143.67	1.28E-32	31%	49%	23--331	31--296
														2	NP 048439	a91L	126.33	2.12E-27	52%	68%	343--469	1--126
														3	NP_048636	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055	58.15	7.08E-07	36%	42%	7--122	417--542
														4	NP_048632	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662	52.76	2.97E-05	45%	52%	7--60	516--590
M651R	246879--247628	250	28,263	8.34		No Hit Found								1	NP 048357	A9R	248.83	1.08E-64	67%	84%	79--249	3--173
														2	NP_048629	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	92.43	1.29E-17	28%	48%	3--247	5--248
														3	NP_048427	A79R	88.97	1.43E-16	27%	47%	5--226	4--217
														4	NP_048807	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	87.04	5.42E-16	28%	47%	5--248	4--246
														5	NP_049005	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	84.34	3.52E-15	29%	46%	5--248	10--253
														6	NP_048525	A177R	83.96	4.59E-15	27%	49%	3--233	5--227
														7	AAU06304	hypothetical protein A275R	62.77	1.10E-08	28%	49%	83--247	3--167
														8	AAU06301	hypothetical protein A275R	61.62	2.44E-08	27%	49%	83--247	3--167
														9	NP 048723	A366L	55.07	2.28E-06	26%	47%	3--144	10--155
														10	AAU06302	hypothetical protein A275R	50.83	4.31E-05	28%	50%	116--248	1--132
M653L	248608--247631	326	37,173	4.45	1	cd01049	RNRR2, Ribonucleotide Reductase, R2/beta subunit (RNRR2) is a member of a broad superfamily of ferritin-like diiron-carboxylate proteins. The RNR protein catalyzes the conversion of ribonucleotides to deoxyribonucleotides and is found in all eukaryotes, many prokaryotes, several viruses, and few archaea. The catalytically active form of RNR is a proposed alpha2-beta2 tetramer. The homodimeric alpha subunit (R1) contains the active site and redox active cysteines as well as the allosteric binding sites. The beta subunit (R2) contains a diiron cluster that, in its reduced state, reacts with dioxygen to form a stable tyrosyl radical and a diiron(III) cluster. This essential tyrosyl radical is proposed to generate a thiyl radical, located on a cysteine residue in the R1 active site that initiates ribonucleotide reduction. The beta subunit is composed of 10-13 helices, the 8 longest helices form an alpha-helical bundle; some have 2 addition beta strands. Yeast is unique in that it assembles	295.61	3.04E-81	49%	64%	21--293	4--284	1	NP_048832	contains ribonucleotide reductase (RR) signature; similar to tobacco RR small subunit, corresponds to Swiss-Prot Accession Number P49730	507.29	2.63E-142	75%	84%	5--326	3--324
					2	pfam00268	Ribonuc red sm, Ribonucleotide reductase, small chain..	287.15	9.98E-79	44%	63%	7--289	1--285	2	NP 189342	ribonucleoside-diphosphate reductase	369.01	1.12E-100	54%	72%	6--326	6--332
					3	COG0208	NrdF, Ribonucleotide reductase, beta subunit [Nucleotide transport and metabolism].	235.95	2.72E-63	32%	48%	21--326	30--348	3	AAO62422	ribonucleotide reductase small subunit	366.31	7.26E-100	54%	71%	6--326	6--333
					4	cd00657	Ferritin_like, Ferritin-like, diiron-carboxylate proteins participate in a range of functions including iron regulation, mono-oxygenation, and reactive radical production. These proteins are characterized by the fact that they catalyze dioxygen-dependent oxidation-hydroxylation reactions within diiron centers; one exception is manganese catalase, which catalyzes peroxide-dependent oxidation-reduction within a dimanganese center. Diiron-carboxylate proteins are further characterized by the presence of duplicate metal ligands, glutamates and histidines (ExxH) and two additional glutamates within a four-helix bundle. Outside of these conserved residues there is little obvious homology. Members include bacterioferritin, ferritin, rubrerythrin, aromatic and alkene monooxygenase hydroxylases (AAMH), ribonucleotide reductase R2 (RNRR2), acyl-ACP-desaturases (Acyl_ACP_Desat), manganese (Mn) catalases, demethoxyubiquinone hydroxylases (DMQH), DNA protecting proteins (DPS), and ubiquinol oxidases (AOX). Additional members include the Fe containing subunit of the aerobic cyclase system (ACSF), the ferritin-like c	52.41	4.31E-08	20%	30%	69--215	2--140	4	CAF24073	probable ribonucleoside-diphosphate reductase small chain	363.61	4.70E-99	56%	73%	7--326	5--323
														5	BAD46317	putative ribonucleotide reductase R2	357.07	4.40E-97	53%	70%	2--326	10--339
														6	XP_550581	putative ribonucleotide reductase R2	355.14	1.67E-96	53%	71%	6--326	20--345
														7	NP 910365	putative ribonucleoside-diphosphate reductase	355.14	1.67E-96	53%	71%	6--326	2--327
														8	CAA71741	ribonucleotide reductase (Class I)	354.76	2.19E-96	54%	70%	6--326	16--337
														9	EAN80229	ribonucleoside-diphosphate reductase small chain	353.21	6.36E-96	54%	70%	6--326	16--337
														10	O15910	Ribonucleoside-diphosphate reductase small chain (Ribonucleotide reductase small subunit) (Ribonucleotide reductase R2 subunit)	353.21	6.36E-96	54%	70%	6--326	16--337
M657L	249244--248729	172	20,087	9.95		No Hit Found								1	NP 048827	A471R	208.38	6.96E-53	52%	77%	2--172	3--173
														2	YP_142861	unknown	106.69	2.85E-22	37%	61%	16--155	43--191
														3	AAR26829	FirrV-1-A5	87.81	1.37E-16	31%	60%	2--145	14--158
														4	NP_077626	EsV-1-141	70.09	2.96E-11	31%	58%	3--119	20--137
M660L	249878--249285	198	22,211	4.64		No Hit Found								1	NP_048826	A470R	231.49	1.11E-59	58%	71%	1--197	1--202
														2	BAB69883	UKCH-2	92.43	8.06E-18	39%	58%	6--127	49--170
M664L	251214--249922	431	48,625	8.54		No Hit Found								1	NP_048824	A468R	431.41	2.75E-119	55%	72%	3--377	4--393
														2	AAR26870	FirrV-1-A46	66.63	1.78E-09	26%	48%	15--178	17--180
														3	NP_077530	EsV-1-45	62.77	2.57E-08	26%	47%	6--187	1--187
M667R	251310--252230	307	35,056	7.09		No Hit Found								1	NP 048823	A467L	398.28	1.57E-109	62%	80%	4--300	6--302
														2	NP 065022	Hypothetical protein	63.16	1.20E-08	28%	46%	48--221	17--218

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
M670R	252261--252608	116	13,165	10.47	1	pfam04777	Evr1_Alr, Erv1 / Alr family. Biogenesis of Fe/S clusters involves a number of essential mitochondrial proteins. Erv1p of Saccharomyces cerevisiae mitochondria is required for the maturation of Fe/S proteins in the cytosol. The ALR (augmenter of liver regeneration) represents a mammalian orthologue of yeast Erv1p. Both Erv1p and full-length ALR are located in the mitochondrial intermembrane an d it thought to operate downstream of the mitochondrial ABC transporter	78.03	8.91E-16	40%	52%	18--110	2--90	1	NP_048821	PBCV-1 thiol oxidoreductase	177.95	7.45E-44	67%	80%	2--116	4--118
					2	COG5054	ERV1, Mitochondrial sulfhydryl oxidase involved in the biogenesis of cytosolic Fe/S proteins [Posttranslational modification, protein turnover, chaperones].	55.83	4.41E-09	30%	45%	1--110	70--175	2	YP_142722	putative thiol oxidoreductase	67.40	1.42E-10	36%	51%	10--97	3--90
														3	XP_503294	hypothetical protein	62.00	5.96E-09	33%	50%	9--111	88--186
														4	NP_078699	Thiol oxidoreductase	62.00	5.96E-09	33%	55%	2--99	21--115
														5	CAH02199	unnamed protein product	60.08	2.26E-08	31%	50%	14--111	85--178
														6	CAA48192	ERV1	58.54	6.59E-08	31%	48%	14--110	19--111
														7	CAA97017	ERV1	58.54	6.59E-08	31%	48%	14--110	91--183
														8	NP_612265	ORF043L	56.61	2.50E-07	34%	52%	7--99	5--96
														9	AAX82354	thiol oxidoreductase	56.23	3.27E-07	34%	52%	7--99	5--96
														10	NP_149810	347L	56.23	3.27E-07	30%	50%	7--98	3--98
M672L	253412--252615	266	29,615	6.63	1	COG0571	Rnc, dsRNA-specific ribonuclease [Transcription].	172.76	2.47E-44	35%	51%	38--263	4--234	1	NP_048820	similar to Bacillus ribonuclease III, corresponds to Swiss-Prot Accession Number P51833	348.59	1.12E-94	66%	81%	16--265	25--274
					2	smart00535	RIBOc, Ribonuclease III family; .	130.75	1.10E-31	42%	58%	58--187	2--130	2	YP_445467	ribonuclease III	133.65	5.64E-30	35%	53%	33--258	9--241
					3	cd00593	RIBOc, RIBOc. Ribonuclease III C terminal domain. This group consists of eukaryotic, bacterial and archeal ribonuclease III (RNAse III) proteins. RNAse III is a double stranded RNA-specific endonuclease. Prokaryotic RNAse III is important in post-transcriptional control of mRNA stability and translational efficiency. It is involved in the processing of ribosomal RNA precursors. Prokaryotic RNAse III also plays a role in the maturation of tRNA precursors and in the processing of phage and plasmid transcripts. Eukaryotic RNase III's participate (through direct cleavage) in rRNA processing, in processing of small nucleolar RNAs (snoRNAs) and snRNA's (components of the spliceosome). In eukaryotes RNase III or RNaseIII like enzymes such as Dicer are involved in RNAi (RNA interference) and miRNA (micro-RNA) gene silencing	127.32	1.16E-30	48%	62%	70--181	16--127	3	ZP_00590199	Ribonuclease III	128.26	2.37E-28	36%	54%	45--255	47--265
					4	pfam00636	Ribonuclease 3, RNase3 domain..	123.30	1.78E-29	56%	70%	76--166	1--91	4	ZP_00532592	Ribonuclease III	127.49	4.04E-28	33%	53%	16--255	21--259
					5	cd00048	DSRM, Double-stranded RNA binding motif. Binding is not sequence specific but is highly specific for double stranded RNA. Found in a variety of proteins including dsRNA dependent protein kinase PKR, RNA helicases, Drosophila staufen protein, E. coli RNase III, RNases H1, and dsRNA dependent adenosine deaminases	58.84	5.27E-10	31%	52%	192--258	1--68	5	ABB23018	Ribonuclease III	122.87	9.95E-27	35%	55%	58--255	47--252
					6	COG1939	COG1939, Uncharacterized protein conserved in bacteria [Function unknown]	34.87	0.007537	29%	46%	78--163	18--112	6	ZP_00591208	Ribonuclease III	122.48	1.30E-26	34%	52%	45--256	44--266
					7	pfam05948	DUF880, Protein of unknown function (DUF880). This family consists of a number of hypothetical bacterial and plant proteins. The family also contains the C terminal region of a Cysteinyl-tRNA synthetase from Staphylococcus epidermidis. The function of this family is unknown..	35.17	0.007546	26%	43%	78--167	11--111	7	ZP_00528534	Ribonuclease III	122.48	1.30E-26	33%	54%	41--255	58--281
														8	ZP_00511103	Ribonuclease III	121.32	2.90E-26	34%	54%	45--257	42--264
														9	ZP_00661649	Ribonuclease III	121.32	2.90E-26	32%	50%	35--257	19--254
														10	AAM73335	ribonuclease III	119.78	8.43E-26	32%	52%	30--262	15--260
M674R	253807--255765	653	75,012	5.23	1	COG3378	COG3378, Predicted ATPase [General function prediction only].	70.01	2.38E-13	24%	37%	281--593	146--445	1	NP_048813	contains ATP/GTP-binding site motif A	870.54	0.00E+00	65%	81%	20--648	22--650
					2	pfam03288	Pox_D5, Poxvirus D5 protein-like. This family includes D5 from Poxviruses which is necessary for viral DNA replication, and is a nucleic acid independent nucleoside triphosphatase. Members of this family are also found outside of poxviruses..	63.73	1.55E-11	22%	39%	290--614	3--313	2	NP_077594	EsV-1-109	278.10	6.60E-73	34%	54%	190--607	141--565
														3	AAR26902	FirrV-1-B27	267.70	8.92E-70	29%	51%	35--591	6--547
														4	YP_294217	putative nucleic acid independent nucleoside triphosphatase	166.78	2.14E-39	25%	44%	187--641	168--672
														5	ZP_00123428	COG3378: Predicted ATPase	65.86	5.15E-09	28%	48%	338--523	310--495
														6	BAE05402	putative DNA primase-phage associated	61.62	9.71E-08	24%	42%	292--566	107--366
														7	AAF27348	phage phi-R73 primase-like protein	61.23	1.27E-07	24%	43%	265--523	213--458
														8	ZP_00503756	Phage/plasmid primase P4, C-terminal	60.46	2.16E-07	22%	42%	232--506	241--499
														9	ZP_00593388	Phage/plasmid primase P4, C-terminal	58.54	8.22E-07	23%	42%	321--613	307--592
														10	ZP_00111378	COG3378: Predicted ATPase	58.15	1.07E-06	30%	53%	376--517	439--581
M680R	255802--256674	291	31,259	4.95		No Hit Found								1	NP_048811	A454L	354.37	2.37E-96	64%	77%	20--291	20--289
M683L	257390--256689	234	26,931	10.80	1	cd00283	GIY-YIG_Cterm, GIYX(10-11)YIG family of class I homing endonucleases C-terminus (GIY-YIG_Cterm). Homing endonucleases promote the mobility of intron or intein by recognizing and cleaving a homologous allele that lacks the sequence. They catalyze a double-strand break in the DNA near the insertion site of that element to facilitate homing at that site. Class I homing endonucleases are sorted into four families based on the presence of these motifs in their respective N-termini: LAGLIDADG, His-Cys box, HNH, and GIY-YIG. This CD contains several but not all members of the GIY-YIG family. The C-terminus of GIY-YIG is a DNA-binding domain which is separated from the N-terminus by a long, flexible linker. The DNA-binding domain consists of a minor-groove binding alpha-helix, and a helix-turn-helix. Some also contain a zinc finger (i.e. I-TevI) which is not required for DNA binding or catalysis, but is a component of the linker and directs the catalytic domain to cleave the homing site at a fixed distance from the intron inserting site	36.90	0.002104	38%	55%	120--199	1--77	1	NP_048851	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440	150.98	2.70E-35	41%	57%	12--234	2--219
					2	smart00465	GIYc, GIY-YIG type nucleases (URI domain); .	36.21	0.003097	29%	49%	10--99	1--83	2	NP_049007	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580	99.37	9.33E-20	34%	51%	12--205	2--195
														3	NP_048671	A315L	96.67	6.05E-19	32%	50%	12--211	2--216
														4	NP_048641	PBCV-1 33kd peptide	73.94	4.20E-12	31%	51%	15--194	6--203
														5	NP_899393	SegD	64.31	3.32E-09	31%	50%	21--182	12--154
														6	AAT53588	group I intron GIY-YIG endonuclease	62.00	1.65E-08	29%	47%	11--186	2--190
														7	ZP_00391680	COG0532: Translation initiation factor 2 (IF-2; GTPase)	56.23	9.05E-07	28%	47%	35--186	2--165
														8	AAU16837	GIY-YIG catalytic domain containing protein; possible intron encoded endonuclease	55.84	1.18E-06	23%	41%	10--197	2--182
														9	ABA03239	putative homing endonuclease	54.68	2.63E-06	42%	64%	118--185	109--171
														10	CAC51107	putative GIY-YIG endonuclease	51.99	1.71E-05	24%	42%	16--208	7--206
M686L	258678--257446	411	44,516	4.13	1	pfam07415	Herpes_LMP2, Gammaherpesvirus latent membrane protein (LMP2) protein. This family consists of several Gammaherpesvirus latent membrane protein (LMP2) proteins. Epstein-Barr virus is a human Gammaherpesvirus that infects and establishes latency in B lymphocytes in vivo. The latent membrane protein 2 (LMP2) gene is expressed in latently infected B cells and encodes two protein isoforms, LMP2A and LMP2B, that are identical except for an additional N-terminal 119 aa cytoplasmic domain which is present in the LMP2A isoform. LMP2A is thought to play a key role in either the establishment or the maintenance of latency and/or the reactivation of productive infection from the latent state. The significance of LMP2B and its role in pathogenesis remain unclear	34.95	0.007536	19%	32%	266--331	31--94	1	NP_048415	contains Pro-rich Px motif EPSPEPxP (5X), and PEST sequence; similar to trypanosome procyclin precursor, corresponds to Swiss-Prot Accession Number P08469	52.37	3.24E-05	24%	48%	16--149	177--308
M690R	258822--259511	230	25,409	10.74		No Hit Found									No Hit Found	No Hit Found						
M693R	259590--259793	68	7,882	11.75		No Hit Found									No Hit Found	No Hit Found						
M694L	260573--260178	132	15,046	6.40		No Hit Found								1	NP_048926	A570L	177.18	1.27E-43	65%	76%	1--118	1--121
M695R	260638--261159	174	19,819	7.15		No Hit Found								1	NP_048928	A572R	243.82	1.55E-63	65%	85%	4--163	11--170
M697L	261989--261162	276	30,772	4.22	1	pfam02747	PCNA_C, Proliferating cell nuclear antigen, C-terminal domain. N-terminal and C-terminal domains of PCNA are topologically identical. Three PCNA molecules are tightly associated to form a closed ring encircling duplex DNA	64.96	7.73E-12	28%	49%	152--267	3--128	1	NP_048930	similar to Periwinkle PCNA, corresponds to GenBank Accession Number X55052	230.34	4.75E-59	44%	67%	21--276	4--264

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					2	COG0592	DnaN, DNA polymerase sliding clamp subunit (PCNA homolog) [DNA replication, recombination, and repair].	59.16	3.81E-10	17%	37%	18--268	47--323	2	CAE67843	Hypothetical protein CBG13430	123.64	6.26E-27	29%	50%	32--274	2--260
					3	pfam00705	PCNA_N, Proliferating cell nuclear antigen, N-terminal domain. N-terminal and C-terminal domains of PCNA are topologically identical. Three PCNA molecules are tightly associated to form a closed ring encircling duplex DNA..	49.88	2.40E-07	25%	55%	42--129	12--100	3	XP_502661	hypothetical protein	115.93	1.30E-24	25%	52%	42--275	12--260
														4	CAA38636	proliferating cell nuclear antigen	115.16	2.22E-24	29%	49%	37--274	7--260
														5	AAC48257	Pcna (proliferating cell nuclear antigen) homolog protein 1	115.16	2.22E-24	29%	51%	65--273	1--226
														6	CAA38893	proliferating cell nuclear antigen	113.24	8.45E-24	30%	49%	42--275	12--262
														7	Q9MAY3	Proliferating cell nuclear antigen (PCNA)	113.24	8.45E-24	28%	50%	42--274	12--261
														8	AAG24908	proliferating cell nuclear antigen	112.85	1.10E-23	30%	49%	42--274	12--261
														9	XP_743255	proliferating cell nuclear antigen	112.85	1.10E-23	29%	51%	38--268	7--261
														10	AAB27811	PCNA	112.85	1.10E-23	28%	50%	42--275	12--262
M699L	262481--261993	163	18,190	10.90		No Hit Found									No Hit Found	No Hit Found						
M700L	263013--262513	167	19,586	10.14		No Hit Found								1	NP_048931	A575L	112.46	4.82E-24	35%	66%	31--166	30--167
M701R	263097--264128	344	37,917	8.48	1	pfam05887	Trypan_PARP, Procyclic acidic repetitive protein (PARP). This family consists of several Trypanosoma brucei procyclic acidic repetitive protein (PARP) like sequences. The procyclic acidic repetitive protein (parp) genes of Trypanosoma brucei encode a small family of abundant surface proteins whose expression is restricted to the procyclic form of the parasite. They are found at two unlinked loci, parpA and parpB; transcription of both loci is developmentally regulated.	53.06	3.06E-08	33%	52%	222--276	68--122	1	NP_048689	PLPRNLLL (4X), SPPPSKP (3X)	345.13	1.87E-93	70%	80%	7--220	1--213
					2	pfam05616	Neisseria_TspB, Neisseria meningitidis TspB protein. This family consists of several Neisseria meningitidis TspB virulence factor proteins..	50.13	1.91E-07	38%	54%	224--274	323--373	2	NP_048688	a332L	88.20	4.12E-16	74%	90%	284--326	1--43
					3	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	46.29	2.85E-06	51%	53%	222--278	68--125	3	EAA01148	ENSANGP00000018413	71.25	5.22E-11	27%	42%	25--217	4--202
					4	COG3147	DedD, Uncharacterized protein conserved in bacteria [Function unknown].	44.63	9.90E-06	28%	40%	222--275	97--150	4	EAA06469	ENSANGP00000012390	69.71	1.52E-10	27%	42%	33--218	1--196
					5	pfam04625	DEC-1_N, DEC-1 protein, N terminal region. The defective chorion-1 gene (dec-1) in Drosophila encodes follicle cell proteins necessary for proper eggshell assembly. Multiple products of the dec-1 gene are formed by alternative RNA splicing and proteolytic processing. Cleavage products include S80 (80 kDa) which is incorporated into the eggshell, and further proteolysis of S80 gives S60 (60 kDa).	39.48	0.0003	24%	45%	226--275	94--143	5	AAS15667	LP24064p	66.24	1.68E-09	28%	43%	18--216	8--210
					6	smart00494	ChtBD2, Chitin-binding domain type 2; .	39.34	0.0004	40%	50%	282--325	9--49	6	AAM50982	RE24790p	65.86	2.19E-09	29%	42%	20--218	38--249
					7	COG3397	COG3397, Uncharacterized protein conserved in bacteria [Function unknown].	39.27	0.000431	21%	31%	11--235	8--226	7	AAF46012	CG15786-PA	65.86	2.19E-09	29%	42%	20--218	19--230
					8	pfam07267	Nucleo_P87, Nucleopolyhedrovirus capsid protein P87. This family consists of several Nucleopolyhedrovirus capsid protein P87 sequences. P87 is expressed late in infection and concentrated in infected cell nuclei.	37.48	0.001331	19%	32%	215--275	319--378	8	EAL29006	GA18133-PA	64.31	6.38E-09	27%	42%	18--216	6--210
					9	pfam01607	CBM_14, Chitin binding Peritrophin-A domain. This domain is called the Peritrophin-A domain and is found in chitin binding proteins particularly peritrophic matrix proteins of insects and animal chitinases. Copies of the domain are also found in some baculoviruses. Relevant references that describe proteins with this domain include. It is an extracellular domain that contains six conserved cysteines that probably form three disulphide bridges. Chitin binding has been demonstrated for a protein containing only two of these domains.	37.34	0.001689	41%	62%	294--325	18--47	9	EAL32472	GA13958-PA	64.31	6.38E-09	27%	41%	30--218	11--209
					10	pfam03067	Chitin_bind_3, Chitin binding domain. This domain is found associated with a wide variety of cellulose binding domain. This domain however is a chitin binding domain. This domain is found in isolation in baculoviral spheroidins and spindolins, protein of unknown function..	37.28	0.001746	23%	30%	18--216	1--211	10	EAA00829	ENSANGP00000011567	60.08	1.20E-07	28%	39%	17--216	7--215
M704R	264221--264748	176	19,645	11.69		No Hit Found								1	NP_048877	contains Gln-rich, neutral zinc metallopeptidase, zinc binding region signature	200.68	1.56E-50	58%	71%	1--175	212--391
M706L	266241--264841	467	51,912	8.35		No Hit Found								1	NP_048863	a507R	83.96	1.20E-14	34%	52%	143--313	6--182
M714L	266984--266688	99	11,584	8.38		No Hit Found									No Hit Found	No Hit Found						
M715R	266983--267303	107	12,723	6.80		No Hit Found								1	NP_048858	A502L	103.99	1.37E-21	54%	65%	21--107	3--95
M717R	267359--268519	387	43,765	6.67	1	cd00204	ANK, ankyrin repeats; ankyrin repeats mediate protein-protein interactions in very diverse families of proteins. The number of ANK repeats in a protein can range from 2 to over 20 (ankyrins, for example). ANK repeats may occur in combinations with other types of domains. The structural repeat unit contains two antiparallel helices and a beta-hairpin, repeats are stacked in a superhelical arrangement; this alignment contains 4 consecutive repeats.	60.10	1.95E-10	30%	58%	65--170	8--112	1	NP_048963	contains 3 ankyrin repeat-like elements; similar to Drosophila ankyrin, corresponds to GenBank Accession Number L35601	196.05	1.67E-48	45%	63%	4--220	3--217
					2	COG0666	Arp, FOG: Ankyrin repeat [General function prediction only].	40.26	0.000187	17%	36%	63--219	72--234	2	NP_048964	A608R	96.67	1.38E-18	33%	54%	239--384	1--149
														3	XP_786577	PREDICTED: similar to CG6599-PA	70.86	8.11E-11	27%	44%	1--258	134--368
														4	AAO25691	ankyrin repeat protein E4 2	59.31	2.44E-07	31%	51%	46--170	29--152
														5	AAW23170	ankyrin domain protein	58.54	4.17E-07	26%	44%	28--231	144--328
														6	XP_475255	hypothetical protein	57.77	7.11E-07	28%	49%	42--170	202--332
														7	NP_039208	ORF FPV245 Ankyrin repeat gene family protein	56.61	1.58E-06	28%	47%	94--286	121--323
														8	XP_782539	PREDICTED: similar to ankyrin repeat domain 28	56.61	1.58E-06	24%	41%	33--265	188--415
														9	EAL45248	ankyrin repeat protein, putative	55.84	2.70E-06	27%	44%	42--300	192--430
														10	NP_048786	A429L	55.07	4.61E-06	22%	40%	26--308	57--357
M719L	269688--268522	389	43,206	5.23	1	COG1004	Ugd, Predicted UDP-glucose 6-dehydrogenase [Cell envelope biogenesis, outer membrane].	354.11	6.29E-99	36%	56%	4--375	2--392	1	NP_048965	PBCV-1 UDP-glucose dehydrogenase	466.08	8.73E-130	59%	77%	5--386	4--386
					2	COG0677	WecC, UDP-N-acetyl-D-mannosaminuronate dehydrogenase [Cell envelope biogenesis, outer membrane].	124.96	5.80E-30	28%	44%	1--364	8--393	2	AAK02860	unknown	422.94	8.48E-117	54%	74%	3--389	2--389
					3	pfam03721	UDPG_MGDP_dh_N, UDP-glucose/GDP-mannose dehydrogenase family, NAD binding domain. The UDP-glucose/GDP-mannose dehydrogenaseses are a small group of enzymes which possesses the ability to catalyse the NAD-dependent 2-fold oxidation of an alcohol to an acid without the release of an aldehyde intermediate.	122.65	3.33E-29	32%	49%	3--176	1--186	3	AAC67251	UDP-glucose dehydrogenase	421.78	1.89E-116	53%	74%	3--389	2--389
					4	pfam00984	UDPG_MGDP_dh, UDP-glucose/GDP-mannose dehydrogenase family, central domain. The UDP-glucose/GDP-mannose dehydrogenaseses are a small group of enzymes which possesses the ability to catalyse the NAD-dependent 2-fold oxidation of an alcohol to an acid without the release of an aldehyde intermediate.	91.02	9.51E-20	39%	55%	193--285	2--96	4	AAK17922	UDP-glucose dehydrogenase FcbC	414.46	3.02E-114	53%	73%	3--389	2--389
					5	COG1893	ApbA, Ketopantoate reductase [Coenzyme metabolism].	46.85	1.95E-06	16%	39%	4--249	2--233	5	AAK17904	UDP-glucose dehydrogenase DcbC	411.76	1.95E-113	53%	72%	3--389	2--389
					6	COG0240	GpsA, Glycerol-3-phosphate dehydrogenase [Energy production and conversion].	45.92	3.46E-06	32%	57%	3--87	2--83	6	CAG21035	putative UDP-glucose dehydrogenase	404.45	3.12E-111	51%	73%	5--389	3--388
					7	COG2084	MmsB, 3-hydroxyisobutyrate dehydrogenase and related beta-hydroxyacid dehydrogenases [Lipid metabolism].	43.29	2.72E-05	29%	45%	3--232	1--204	7	ZP_00851874	UDP-glucose 6-dehydrogenase	400.21	5.88E-110	51%	71%	5--389	3--388
					8	COG1250	FadB, 3-hydroxyacyl-CoA dehydrogenase [Lipid metabolism].	41.40	8.91E-05	25%	47%	1--88	2--95	8	ZP_00880860	UDP-glucose 6-dehydrogenase	396.36	8.50E-109	50%	71%	5--389	3--388
					9	pfam03720	UDPG_MGDP_dh_C, UDP-glucose/GDP-mannose dehydrogenase family, UDP binding domain. The UDP-glucose/GDP-mannose dehydrogenaseses are a small group of enzymes which possesses the ability to catalyse the NAD-dependent 2-fold oxidation of an alcohol to an acid without the release of an aldehyde intermediate.	37.96	0.00092	17%	38%	302--383	1--87	9	ZP_00579286	UDP-glucose 6-dehydrogenase	393.28	7.19E-108	48%	72%	5--389	11--396
					10	COG0654	UbiH, 2-polyprenyl-6-methoxyphenol hydroxylase and related FAD-dependent oxidoreductases [Coenzyme metabolism / Energy production and conversion].	37.36	0.001585	39%	68%	1--31	1--32	10	AAM38424	UDP-glucose dehydrogenase	392.89	9.39E-108	51%	70%	5--386	3--385
M721L	271560--269749	604	69,547	7.89	1	COG1215	COG1215, Glycosyltransferases, probably involved in cell wall biogenesis [Cell envelope biogenesis, outer membrane].	109.64	2.39E-25	19%	37%	56--447	14--402	1	NP_048569	similar to cellulose synthase catalytic subunit (UDP-forming)	898.66	0.00E+00	68%	81%	1--604	72--677

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					2	pfam03552	Cellulose_synt, Cellulose synthase. Cellulose, an aggregate of unbranched polymers of beta-1,4-linked glucose residues, is the major component of wood and thus paper, and is synthesised by plants, most algae, some bacteria and fungi, and even some animals. The genes that synthesise cellulose in higher plants differ greatly from the well-characterized genes found in Acetobacter and Agrobacterium sp. More correctly designated as 'cellulose synthase catalytic subunits', plant cellulose synthase (CesA) proteins are integral membrane proteins, approximately 1,000 amino acids in length. There are a number of highly conserved residues, including several motifs shown to be necessary for processive glycosyltransferase activity	40.66	0.00015	24%	43%	262--467	499--698	2	AAL44127	cellulose synthase	305.06	4.60E-81	41%	57%	82--474	91--478
					3	pfam00535	Glycos_transf_2, Glycosyl transferase. Diverse family, transferring sugar from UDP-glucose, UDP-N-acetyl- galactosamine, GDP-mannose or CDP-abequose, to a range of substrates including cellulose, dolichol phosphate and teichoic acids..	38.66	0.000678	15%	32%	106--273	10--168	3	BAB54246	mlr7873	288.12	5.81E-76	38%	53%	93--531	112--546
														4	ZP_00958843	cellulose synthase-like protein	284.65	6.43E-75	36%	53%	93--557	102--561
														5	CAC48842	putative cellulose synthase protein	283.88	1.10E-74	38%	53%	90--526	97--528
														6	AAQ87082	Cellulose synthase catalytic subunit	281.18	7.11E-74	38%	53%	93--535	102--539
														7	ZP_00913218	similar to cellulose synthase	280.41	1.21E-73	40%	57%	93--471	166--538
														8	YP 418867	glycosyl transferase, group 2 family protein	278.10	6.02E-73	37%	53%	90--523	93--516
														9	ABA79331	cellulose synthase-like protein	276.17	2.29E-72	35%	52%	93--559	110--568
														10	AAN33736	glycosyl transferase, group 2 family protein	276.17	2.29E-72	37%	53%	90--523	93--516
M727L	271992--271636	119	13,075	10.35	1	smart00317	SET, SET (Su(var)3-9, Enhancer-of-zeste, Trithorax) domain; Putative methyl transferase. based on outlier plant homologues .	58.87	5.39E-10	29%	38%	6--113	2--125	1	NP_048968	PBCV-1 histone H3-Lys 27 methyltransferase (vSET)	141.35	7.64E-33	57%	73%	2--114	1--113
					2	pfam00856	SET, SET domain. SET domains are protein lysine methyltransferase enzymes. SET domains appear to be protein-protein interaction domains. It has been demonstrated that SET domains mediate interactions with a family of proteins that display similarity with dual-specificity phosphatases (dsPTPases). A subset of SET domains have been called PR domains. These domains are divergent in sequence from other SET domains, but also appear to mediate protein-protein interaction..	48.14	8.37E-07	24%	39%	6--108	8--128	2	ZP_00588496	Nuclear protein SET	65.86	4.08E-10	34%	52%	7--114	39--150
					3	COG2940	COG2940, Proteins containing SET domain [General function prediction only].	42.79	3.73E-05	22%	35%	1--108	328--450	3	ZP_00661322	Nuclear protein SET	60.46	1.71E-08	35%	51%	6--117	37--152
														4	NP 701503	hypothetical protein PFL0690c	60.08	2.24E-08	31%	45%	3--111	29--174
														5	ZP_00511449	Nuclear protein SET	59.31	3.82E-08	34%	50%	7--114	38--149
														6	EAM63832	Nuclear protein SET	58.54	6.51E-08	31%	49%	7--116	39--152
														7	ZP_00543998	Nuclear protein SET	57.77	1.11E-07	31%	52%	7--106	3--109
														8	ABB23988	Nuclear protein SET	56.61	2.47E-07	34%	51%	6--114	38--150
														9	AAM32541	hypothetical protein MM 2845	55.45	5.51E-07	32%	55%	16--106	19--116
														10	AAM05086	conserved hypothetical protein	54.68	9.40E-07	32%	58%	16--106	13--110
M729L	273776--272028	583	65,559	11.57	1	cd00180	S_TKc, Serine/Threonine protein kinases, catalytic domain. Phosphotransferases of the serine or threonine-specific kinase subfamily. The enzymatic activity of these protein kinases is controlled by phosphorylation of specific residues in the activation segment of the catalytic domain, sometimes combined with reversible conformational changes in the C-terminal autoregulatory tail	43.66	2.15E-05	30%	46%	53--238	6--151	1	NP_048970	RPQT-like (9x)	478.02	3.78E-133	44%	62%	2--583	6--577
					2	COG0515	SPS1, Serine/threonine protein kinase [General function prediction only / Signal transduction mechanisms / Transcription / DNA replication, recombination, and repair].	41.30	9.80E-05	13%	28%	84--474	11--383	2	NP_048632	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662	300.06	1.41E-79	35%	55%	1--524	9--527
					3	smart00220	S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-specific kinase subfamily. .	40.58	0.000152	25%	48%	53--238	5--150	3	NP_048636	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055	293.12	1.73E-77	36%	54%	1--490	23--524
					4	pfam01636	APH, Phosphotransferase enzyme family. This family consists of bacterial antibiotic resistance proteins, which confer resistance to various aminoglycosides they include:- aminoglycoside 3'-phosphotransferase or kanamycin kinase / neomycin-kanamycin phosphotransferase and streptomycin 3''-kinase or streptomycin 3''-phosphotransferase. The aminoglycoside phosphotransferases inactivate aminoglycoside antibiotics via phosphorylation. This family also includes homoserine kinase. This family is related to fructosamine kinase pfam03881	40.52	0.000167	37%	70%	207--233	171--198	4	CAB10894	extensin	82.80	3.54E-14	33%	43%	371--502	104--235
					5	pfam00069	Pkinase, Protein kinase domain..	39.50	0.000343	21%	38%	53--321	5--257	5	AAK54495	neurofilament triplet H1-like protein	75.10	7.39E-12	36%	51%	306--506	42--228
					6	COG2334	COG2334, Putative homoserine kinase type II (protein kinase fold) [General function prediction only].	38.02	0.00086	36%	55%	207--240	200--233	6	XP_427855	PREDICTED: similar to p87, partial	73.94	1.65E-11	24%	46%	366--515	44--185
					7	COG5422	ROM1, RhoGEF, Guanine nucleotide exchange factor for Rho/Rac/Cdc42-like GTPases [Signal transduction mechanisms].	37.71	0.001145	18%	34%	330--498	19--197	7	XP_607274	PREDICTED: similar to neurofilament, heavy polypeptide 200kDa isoform 1	73.56	2.15E-11	32%	51%	327--499	582--762
					8	COG5032	TEL1, Phosphatidylinositol kinase and protein kinases of the PI-3 kinase family [Signal transduction mechanisms / Cell division and chromosome partitioning / Chromatin structure and dynamics / DNA replication, recombination, and repair / Intracellular trafficking and secretion].	36.21	0.002982	44%	50%	209--243	1947--1981	8	XP_870725	PREDICTED: similar to neurofilament, heavy polypeptide 200kDa isoform 2	73.17	2.81E-11	32%	51%	327--499	634--820
					9	COG0661	AarF, Predicted unusual protein kinase [General function prediction only].	36.08	0.003979	29%	57%	204--254	284--333	9	XP_881770	PREDICTED: similar to neurofilament, heavy polypeptide 200kDa isoform 3	73.17	2.81E-11	32%	51%	327--499	624--810
					10	pfam00454	PI3 PI4 kinase, Phosphatidylinositol 3- and 4-kinase..	35.72	0.004574	32%	40%	209--253	137--184	10	EAN84654	hypothetical protein, conserved	73.17	2.81E-11	28%	46%	306--499	141--351
M734L	274105--273866	80	9,277	9.60		No Hit Found									No Hit Found	No Hit Found						
M735L	275268--274810	153	17,289	3.97		No Hit Found								1	NP_048974	A618L	80.49	1.58E-14	55%	80%	82--151	63--125
M738L	275902--275231	224	25,232	4.23		No Hit Found									No Hit Found	No Hit Found						
M740L	276247--275957	97	11,140	8.64		No Hit Found								1	NP_048976	similar to Synechocystis orf 90, corresponds to GenBank Accession Number D90902	67.01	1.85E-10	34%	50%	1--95	1--81
														2	NP_048991	A635R	51.60	8.06E-06	30%	51%	1--91	1--82
M741L	276682--276335	116	12,705	8.65		No Hit Found								1	NP 048977	A621L	130.18	1.78E-29	52%	71%	4--116	5--117
M742R	276753--279455	901	99,481	6.19	1	COG0488	Uup, ATPase components of ABC transporters with duplicated ATPase domains [General function prediction only].	277.14	9.58E-76	27%	45%	300--848	2--530	1	NP_049022	Chlorella virus CVK2 translation elongation factor-3 homolog, refer to GenBank Accession Number D16505	1105.51	0.00E+00	65%	77%	37--901	55--918
					2	pfam00005	ABC_tran, ABC transporter. ABC transporters for a large family of proteins responsible for translocation of a variety of compounds across biological membranes. ABC transporters are the largest family of proteins in many completely sequenced bacteria. ABC transporters are composed of two copies of this domain and two copies of a transmembrane domain pfam00664. These four domains may belong to a single polypeptide, or belong in different polypeptide chains	117.70	1.07E-27	31%	48%	327--491	1--182	2	A48779	translation elongation factor EF-3 homolog - Chlorella virus CVK2	1010.36	0.00E+00	61%	74%	37--885	272--1120
					3	cd00267	ABC_ATPase, ABC (ATP-binding cassette) transporter nucleotide-binding domain; ABC transporters are a large family of proteins involved in the transport of a wide variety of different compounds, like sugars, ions, peptides and more complex organic molecules. The nucleotide binding domain shows the highest similarity between all members of the family. ABC transporters are a subset of nucleotide hydrolases that contain a signature motif, Q-loop, and H-loop/switch region in addition to the Walker A motif/P-loop andWalker B motif commonly found in a number of ATP- and GTP-binding and hydrolyzing proteins..	105.67	4.18E-24	30%	51%	306--476	5--194	3	CAG58023	unnamed protein product	761.91	0.00E+00	45%	63%	3--900	113--1043
					4	COG1121	ZnuC, ABC-type Mn/Zn transport systems, ATPase component [Inorganic ion transport and metabolism].	101.06	1.07E-22	31%	51%	306--493	9--216	4	BAA33959	translation elongation factor3	760.37	0.00E+00	45%	62%	3--900	113--1043
					5	COG1131	CcmA, ABC-type multidrug transport system, ATPase component [Defense mechanisms].	100.43	1.58E-22	28%	47%	312--487	16--208	5	XP_711404	translation elongation factor 3	758.83	0.00E+00	46%	62%	29--901	136--1050
					6	COG2274	SunT, ABC-type bacteriocin/lantibiotic exporters, contain an N-terminal double-glycine peptidase domain [Defense mechanisms].	94.52	1.02E-20	28%	46%	304--474	476--666	6	CAA78282	translation elongation factor 3	757.67	0.00E+00	45%	62%	29--901	136--1050
					7	COG1124	DppF, ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component [Amino acid transport and metabolism / Inorganic ion transport and metabolism].	93.76	1.44E-20	29%	50%	306--487	8--213	7	XP_711356	translation elongation factor 3	756.52	0.00E+00	46%	62%	29--901	136--1050
					8	COG4988	CydD, ABC-type transport system involved in cytochrome bd biosynthesis, ATPase and permease components [Energy production and conversion / Posttranslational modification, protein turnover, chaperones].	91.04	1.01E-19	29%	44%	262--475	279--514	8	CAA22654	SPCC417.08	752.67	0.00E+00	45%	63%	35--898	142--1046

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					9	COG1122	CbiO, ABC-type cobalt transport system, ATPase component [Inorganic ion transoort and metabolism].	86.87	1.88E-18	29%	47%	305--492	7--215	9	CAA77567	elongation factor 3	751.90	0.00E+00	45%	62%	29--901	136--1049
					10	COG1127	Ttg2A, ABC-type transport system involved in resistance to organic solvents, ATPase component [Secondary metabolites biosynthesis, transoort. and catabolism].	86.76	2.09E-18	26%	46%	306--487	13--217	10	CAG98340	unnamed protein product	751.51	0.00E+00	46%	62%	35--901	137--1044
M748L	281025--279463	521	57,929	5.26	1	pfam04451	Capsid_Iridovir, Iridovirus major capsid protein. This family includes the major capsid protein of iridoviruses, chlorella virus and Spodoptera ascovirus, which are all dsDNA viruses with no RNA stage. This is the most abundant structural protein and can account for up to 45% of virion protein. In Chlorella virus MT325 the major capsid protein is a glycoprotein..	365.79	1.79E-102	41%	53%	172--517	75--443	1	AAC27493	putative capsid protein	1039.25	0.00E+00	97%	97%	1--521	1--521
														2	NP_048978	similar to Simulium iridescent virus capsid protein, corresponds to Swiss-Prot Accession Number P22166	756.90	0.00E+00	70%	80%	1--521	1--520
														3	BAE06835	hypothetical major capsid protein	246.13	2.10E-63	35%	52%	126--521	34--440
														4	BAA76600	major capsid protein	197.98	6.56E-49	34%	51%	196--521	92--436
														5	AAC27492	major capsid protein Vp49	195.67	3.26E-48	32%	50%	196--521	89--432
														6	BAA76601	major capsid protein MCP1	194.13	9.47E-48	34%	51%	196--521	92--437
														7	NP_048787	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052	194.13	9.47E-48	34%	51%	196--521	92--437
														8	1M3Y_D	Chain D, The Structure Of Major Capsid Protein Of A Large, Lipid Containing, Dna Virus	194.13	9.47E-48	34%	51%	196--521	68--413
														9	1M4X_C	Chain C, Pbcv-1 Virus Capsid, Quasi-Atomic Model	193.74	1.24E-47	34%	51%	196--521	68--413
														10	BAA22198	major capsid protein Vp54	190.66	1.05E-46	34%	51%	196--521	92--437
M752R	281297--281653	119	13,504	9.75	1	COG4852	COG4852, Predicted membrane protein [Function unknown].	37.20	0.001517	24%	36%	8--117	10--124	1	NP_048980	A624R	130.95	1.03E-29	52%	72%	5--117	8--120
														2	ZP_00843145	conserved hypothetical protein	53.91	1.60E-06	30%	49%	6--108	9--115
														3	ZP_00653117	conserved hypothetical protein	50.45	1.77E-05	31%	46%	6--108	8--117
														4	ZP_00234461	conserved hypothetical protein	50.45	1.77E-05	24%	43%	6--119	9--127
														5	AAZ18596	conserved hypothetical protein	49.29	3.95E-05	27%	44%	6--108	8--117
														6	CAC98668	lmo0589	48.52	6.74E-05	24%	42%	6--119	9--127
M753L	282564--281662	301	33,934	10.83		No Hit Found								1	NP_048439	a91L	64.70	3.99E-09	34%	53%	177--291	2--112
M756R	282608--283156	183	20,183	3.48		No Hit Found								1	NP_049015	A659L	80.49	2.60E-14	38%	49%	1--139	1--125
M757R	283302--283733	144	16,320	6.80		No Hit Found								1	NP_049012	A656L	65.08	6.94E-10	38%	56%	9--92	8--100
M758R	283779--284369	197	21,905	5.09	1	pfam00583	Acetyltransf_1, Acetyltransferase (GNAT) family. This family contains proteins with N-acetyltransferase functions	45.24	6.81E-06	31%	52%	113--163	30--82	1	NP_049010	A654L	209.15	5.83E-53	51%	68%	1--197	1--197
					2	COG3153	COG3153, Predicted acetyltransferase [General function prediction only].	37.17	0.001578	40%	62%	113--162	81--126	2	ZP_00663133	GCN5-related N-acetyltransferase	65.08	1.36E-09	27%	47%	2--184	11--205
					3	COG1247	COG1247, Sortase and related acyltransferases [Cell envelope biogenesis. outer membrane].	36.40	0.003077	35%	49%	115--162	89--138	3	ZP_00675288	hypothetical protein TeryDRAFT_0585	62.39	8.81E-09	24%	44%	10--186	15--204
					4	COG0456	RimI, Acetyltransferases [General function prediction only].	35.37	0.005968	25%	41%	113--182	97--168	4	XP_392876	PREDICTED: similar to arylalkylamine N-acetyltransferase	56.23	6.32E-07	27%	44%	2--173	51--237
														5	BAC87874	arylalkylamine N-acetyltransferase	52.76	6.98E-06	21%	41%	7--173	41--225
M762L	285067--284549	173	19,326	9.68		No Hit Found								1	NP_049000	A644R	117.09	2.15E-25	72%	79%	89--172	2--84
M763L	286537--285098	480	54,551	11.33	1	COG4942	COG4942, Membrane-bound metallopeptidase [Cell division and chromosome partitioning]	42.26	4.72E-05	16%	32%	269--428	175--334	1	NP_048999	Gln-rich; KQQ (6X)	224.94	4.51E-57	47%	58%	196--480	1--269
					2	COG3064	TolA, Membrane protein involved in colicin uptake [Cell envelope biogenesis. outer membrane].	38.23	0.00091	19%	40%	233--380	52--196	2	NP_048998	A642R	60.46	1.47E-07	68%	80%	19--53	21--55
M766L	287657--286569	363	41,125	5.49	1	pfam04371	PAD_porph, Porphyromonas-type peptidyl-arginine deiminase. Peptidyl-arginine deiminase (PAD) enzymes catalyse the deimination of the guanidino group from carboxy-terminal arginine residues of various peptides to produce ammonia. PAD from Porphyromonas gingivalis (PPAD) appears to be evolutionarily unrelated to mammalian PAD (pfam03068), which is a metalloenzyme. PPAD is thought to belong to the same superfamily as aminotransferase and arginine deiminase, and to form an alpha/beta propeller structure. This family has previously been named PPADH (Porphyromonas peptidyl-arginine deiminase homologues). The predicted catalytic residues in PPAD are Asp130, Asp187, His236, Asp238 and Cys351. These are absolutely conserved with the exception of Asp187 which is absent in two family members. PPAD is also able to catalyse the deimination of free L-arginine, but has primarily peptidyl-arginine specificity. It may have a FMN cofactor..	435.03	2.64E-123	54%	73%	14--361	1--329	1	NP_048994	PBCV-1 Agmatine iminohydrolase	493.43	4.63E-138	62%	78%	7--361	1--358
					2	COG2957	COG2957, Peptidylarginine deiminase and related enzymes [Amino acid transoort and metabolism].	388.49	2.64E-109	46%	65%	9--361	10--344	2	CAC98253	lmo0038	384.03	3.96E-105	52%	69%	11--361	10--362
														3	ZP_00229955	peptidyl-arginine deiminase-like protein	382.10	1.50E-104	51%	68%	11--361	10--362
														4	ZP_00232718	peptidyl-arginine deiminase-like protein	381.72	1.97E-104	51%	68%	11--361	10--362
														5	ZP_00322658	COG2957: Peptidylarginine deiminase and related enzymes	378.25	2.17E-103	52%	68%	9--361	8--362
														6	ZP_00640290	Porphyromonas-type peptidyl-arginine deiminase	377.48	3.71E-103	51%	67%	9--362	15--368
														7	ZP_00229957	conserved hypothetical protein	377.48	3.71E-103	51%	66%	9--361	8--362
														8	AAT02837	conserved hypothetical protein	376.71	6.32E-103	51%	66%	9--361	8--362
														9	AAG03681	agmatine deiminase	374.02	4.10E-102	51%	67%	9--362	8--364
														10	ZP_00849916	conserved hypothetical protein	373.24	6.99E-102	51%	67%	9--362	15--368
M770R	287716--288117	134	15,340	9.58		No Hit Found								1	NP_048990	A634L	166.78	1.71E-40	58%	77%	2--129	4--131
M773L	288496--288116	127	14,776	8.63		No Hit Found									No Hit Found	No Hit Found						
M777L	291250--288956	765	85,431	7.89	1	cd01679	RNR_1, RNR, class I. Ribonucleotide reductase (RNR) catalyzes the reductive synthesis of deoxyribonucleotides from their corresponding ribonucleotides. It provides the precursors necessary for DNA synthesis. RNRs are separated into three classes based on their metallocofactor usage. Class I RNRs, found in eukaryotes, bacteria, and many viruses, use a diiron-tyrosyl radical, Class II RNRs, found in bacteria, and bacteriophages, use coenzyme B12 (adenosylcobalamin, AdoCbl). Class III RNRs, found in anaerobic bacteria, bacteriophages, and archaea, use an FeS cluster and S-adenosylmethionine to generate a glycyl radical. Many organisms have more than one class of RNR present in their genomes. All three RNRs have a ten-stranded alpha-beta barrel domain that is structurally similar to the domain of PFL (pyruvate formate lyase). Class I RNR is oxygen-dependent and can be subdivided into classes Ia (eukaryotes, prokaryotes, viruses and phages) and Ib (which is found in prokaryotes only). It is a tetrameric enzyme of two alpha and two beta subunits: this model covers the major part of the alpha or large subunit, as	733.93	0	51%	68%	176--746	1--578	1	NP_048985	similar to Schizosaccharomyces ribonucleotide reductase M1 chain, corresponds to Swiss-Prot Accession Number P36602	1215.29	0.00E+00	78%	88%	9--765	14--771
					2	pfam02867	Ribonuc_red_lgC, Ribonucleotide reductase, barrel domain..	634.64	0	47%	66%	224--748	1--532	2	IP_001026008	ribonucleoside-diphosphate reductase M1 chain	764.61	0.00E+00	50%	69%	9--759	1--754
					3	COG0209	NrdA, Ribonucleotide reductase, alpha subunit [Nucleotide transport and metabolism]	533.82	5.62E-153	37%	53%	65--764	7--696	3	AAH46846	RRM1 protein	758.44	0.00E+00	50%	69%	9--759	1--754
					4	cd02888	RNR_1_like, RNR, class I_like family. Ribonucleotide reductase (RNR) catalyzes the reductive synthesis of deoxyribonucleotides from their corresponding ribonucleotides. It provides the precursors necessary for DNA synthesis. RNRs are separated into three classes based on their metallocofactor usage. Class I RNRs, found in eukaryotes, bacteria, and many viruses, use a diiron-tyrosyl radical, Class II RNRs, found in bacteria, and bacteriophages, use coenzyme B12 (adenosylcobalamin, AdoCbl). Class III RNRs, found in anaerobic bacteria, bacteriophages, and archaea, use an FeS cluster and S-adenosylmethionine to generate a glycyl radical. Many organisms have more than one class of RNR present in their genomes. All three RNRs have a ten-stranded alpha-beta barrel domain that is structurally similar to the domain of PFL (pyruvate formate lyase). This family appears similar to class I RNRs, as judged by sequence similarity and the predicted active site..	303.28	1.50E-83	34%	51%	196--743	2--521	4	AAH74185	RRM1 protein	756.90	0.00E+00	50%	69%	9--759	1--754
					5	pfam00317	Ribonuc_red_lgN, Ribonucleotide reductase, all-alpha domain..	77.59	1.24E-15	40%	56%	150--222	1--78	5	CAB98233	ribonucleoside-diphosphate reductase large chain (un-24)	751.13	0.00E+00	50%	67%	9--759	1--755

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
					6	cd00576	RNR_PFL, RNR_PFL. Ribonucleotide reductase (RNR) and pyruvate formate lyase (PFL) have a structurally similar ten-stranded alpha-beta barrel active site domain and are believed to have diverged from a common ancestor. RNRs are found in all organisms and provide the only mechanism by which nucleotides are converted to deoxynucleotides, while PFL, an essential enzyme in anaerobic bacteria, catalyzes the conversion of pyruvate and CoA to acteylCoA and formate. Both RNR and PFL are glycyl radical enzymes	55.88	3.78E-09	22%	35%	249--678	70--461	6	AAD49743	ribonucleotide reductase large subunit	751.13	0.00E+00	50%	67%	9--759	1--755
					7	pfam03477	ATP-cone, ATP cone domain..	49.59	2.95E-07	24%	40%	9--110	1--89	7	AAH85906	Ribonucleotide reductase M1	747.66	0.00E+00	49%	68%	9--759	1--754
														8	AAH16450	Ribonucleotide reductase M1	745.35	0.00E+00	49%	68%	9--759	1--754
														9	NP_001024	ribonucleoside-diphosphate reductase M1 chain	743.81	0.00E+00	49%	68%	9--759	1--754
														10	AAA40061	ribonucleotide reductase subunit M1	743.42	0.00E+00	49%	68%	9--759	1--754
M785L	292819--291521	433	48,690	10.98		No Hit Found								1	NP_048983	similar to Chlorella virus PBCV-1 ORF A231L, corresponds to GenBank Accession Number U42580	472.24	1.41E-131	54%	73%	25--433	31--436
														2	NP_048579	contains ATP/GTP-binding motif A	300.06	9.62E-80	44%	62%	94--423	4--337
M789L	293169--292840	110	12,094	9.86	1	COG1293	COG1293, Predicted RNA-binding protein homologous to eukaryotic snRNP [Transcription]	78.49	6.81E-16	37%	57%	4--110	445--554	1	NP_048528	similar to Streptococcus pyogenes fibronectin protein, corresponds to GenBank Accession Number L28919	117.86	9.04E-26	52%	71%	1--106	1--105
					2	pfam05670	DUF814, Domain of unknown function (DUF814). This domain occurs in proteins that have been annotated as Fibronectin/fibrinogen binding protein by similarity. This annotation comes from a sequence, where the N-terminal region is involved in this activity. Hence the activity of this C-terminal domain is unknown. This domain contains a conserved motif D/E X-W/Y-X-H that may be functionally important..	72.55	4.05E-14	39%	60%	4--87	3--90	2	ZP_00504390	Protein of unknown function DUF814:Fibronectin-binding A, N-terminal	63.54	2.02E-09	34%	50%	5--109	469--576
														3	AAD35450	fibronectin-binding protein, putative	60.46	1.71E-08	32%	53%	6--106	440--542
														4	CAA08863	putative fibronectin-binding protein	59.69	2.92E-08	32%	50%	5--106	439--542
														5	CAB13438	yloA	58.54	6.51E-08	30%	50%	5--110	457--563
														6	YP_004218	fibronectin/fibrinogen-binding protein	57.77	1.11E-07	29%	51%	8--106	405--505
														7	BAD70431	probable RNA-binding protein	57.77	1.11E-07	29%	51%	8--106	405--505
														8	NP_782213	fibronectin/fibrinogen-binding protein	57.38	1.45E-07	34%	48%	5--106	463--566
														9	AAK80069	Fibronectin-binding protein	57.38	1.45E-07	32%	50%	5--104	460--561
														10	ZP_00886600	RNA-binding protein homologous to eukaryotic snRNP	57.38	1.45E-07	33%	49%	5--106	468--572
M791R	293278--294837	520	56,961	7.33	1	smart00637	CBD_II, CBD_II domain..	66.89	1.85E-12	30%	43%	12--93	4--85	1	BAA78554	vChti-1	736.87	0.00E+00	69%	76%	1--513	1--533
					2	pfam00704	Glyco_hydro_18, Glycosyl hydrolases family 18..	38.19	0.000916	18%	31%	117--301	8--187	2	NP_048529	PBCV-1 chitinase	729.17	0.00E+00	68%	75%	1--513	1--537
					3	COG3469	COG3469, Chitinase [Carbohydrate transport and metabolism].	34.63	0.009514	33%	47%	179--259	90--169	3	ZP_00570566	Cellulose-binding, bacterial type	225.71	2.93E-57	40%	57%	111--412	225--538
														4	BAC72964	putative sugar hydrolase	212.62	2.57E-53	32%	46%	14--412	46--487
														5	CAC10108	putative secreted sugar hydrolase	212.62	2.57E-53	31%	46%	14--407	45--483
														6	ZP_00687414	Chitinase	183.73	1.28E-44	36%	50%	118--422	26--350
														7	ZP_00464807	Carbohydrate-binding domain, family V/XII	176.41	2.04E-42	36%	49%	118--422	26--350
														8	ABB08413	Chitinase	175.25	4.54E-42	35%	49%	118--422	26--350
														9	CAH35762	putative exported chitinase	174.87	5.93E-42	37%	51%	125--409	37--339
														10	ZP_00479579	COG3979: Uncharacterized protein contain chitin-binding domain type 3	174.10	1.01E-41	37%	51%	125--409	33--335
M794R	294905--295714	270	30,668	6.52	1	cd00180	S_TKc, Serine/Threonine protein kinases, catalytic domain. Phosphotransferases of the serine or threonine-specific kinase subfamily. The enzymatic activity of these protein kinases is controlled by phosphorylation of specific residues in the activation segment of the catalytic domain, sometimes combined with reversible conformational changes in the C-terminal autoregulatory tail	190.03	1.42E-49	29%	53%	10--268	1--256	1	AAU06280	protein kinase A248R	240.74	3.37E-62	48%	66%	4--267	41--304
					2	smart00220	S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases. Serine or threonine-specific kinase subfamily. .	187.34	9.57E-49	31%	49%	11--268	1--256	2	AAU06275	protein kinase A248R	240.74	3.37E-62	48%	66%	4--267	20--283
					3	pfam00069	Pkinase, Protein kinase domain..	183.56	1.48E-47	32%	53%	11--268	1--258	3	AAU06282	protein kinase A248R	236.88	4.87E-61	46%	65%	4--267	41--304
					4	COG0515	SPS1, Serine/threonine protein kinase [General function prediction only / Signal transduction mechanisms / Transcription / DNA replication, recombination, and repair].	116.80	2.00E-27	27%	47%	10--268	1--278	4	NP_048597	PBCV-1 protein kinase	226.48	6.58E-58	43%	64%	4--267	38--303
					5	smart00219	TyrKc, Tyrosine kinase, catalytic domain; Phosphotransferases. Tyrosine-specific kinase subfamily. .	111.04	9.29E-26	26%	44%	12--264	2--253	5	AAA87065	serine/threonine protein kinase	222.63	9.50E-57	43%	64%	4--267	34--299
					6	cd00192	TyrKc, Tyrosine kinase, catalytic domain. Phosphotransferases; tyrosine-specific kinase subfamily. Enzymes with TyrKc domains belong to an extensive family of proteins which share a conserved catalytic core common to both serine/threonine and tyrosine protein kinases. Enzymatic activity of tyrosine protein kinases is controlled by phosphorylation of specific tyrosine residues in the activation segment of the catalytic domain or a C-terminal tyrosine (tail) residue with reversible conformational changes	103.35	1.82E-23	25%	43%	9--263	7--263	6	AAU06274	protein kinase A248R	201.83	1.73E-50	49%	66%	52--267	22--237
					7	COG3642	COG3642, Mn2+-dependent serine/threonine protein kinase [Signal transduction mechanisms].	35.98	0.003518	28%	44%	86--152	67--139	7	AAU06285	protein kinase A248R	192.97	8.06E-48	47%	64%	54--267	2--216
														8	AAU06286	protein kinase A248R	154.84	2.43E-36	48%	70%	119--267	1--151
														9	AAU06270	protein kinase A248R	139.43	1.06E-31	45%	67%	124--267	1--146
														10	NP_048643	similar to PBCV-1 serine/threonine protein kinase, corresponds to GenBank Accession Number U14660	132.88	9.90E-30	31%	53%	9--267	16--278
M797R	295768--296901	378	39,916	8.96		No Hit Found								1	NP_048735	A378L	133.65	9.83E-30	55%	69%	178--281	69--173
														2	CAA64974	QI74 protein	60.08	1.38E-07	22%	33%	130--378	344--580
M801R	297080--297832	251	28,635	7.58		No Hit Found								1	NP_048357	A9R	239.20	8.61E-62	63%	81%	80--248	3--171
														2	NP_049005	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	120.55	4.46E-26	31%	48%	1--250	4--254
														3	NP_048807	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	115.55	1.44E-24	29%	48%	4--250	1--247
														4	NP_048525	A177R	107.46	3.91E-22	29%	48%	1--234	1--227
														5	NP_048629	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055	106.30	8.71E-22	27%	47%	1--248	1--248
														6	NP_048427	A79R	98.60	1.82E-19	30%	45%	4--227	1--217
														7	NP_049003	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	62.77	1.10E-08	47%	65%	4--64	1--58
														8	AAU06304	hypothetical protein A275R	62.39	1.44E-08	25%	46%	82--248	1--167
														9	AAU06301	hypothetical protein A275R	61.23	3.21E-08	24%	46%	82--248	1--167
														10	AAU06302	hypothetical protein A275R	51.22	3.33E-05	24%	48%	117--250	1--133
M803L	299338--297839	500	53,361	10.75	1	pfam05887	Trypan_PARP, Procyclic acidic repetitive protein (PARP). This family consists of several Trypanosoma brucei procyclic acidic repetitive protein (PARP) like sequences. The procyclic acidic repetitive protein (parp) genes of Trypanosoma brucei encode a small family of abundant surface proteins whose expression is restricted to the procyclic form of the parasite. They are found at two unlinked loci, parpA and parpB; transcription of both loci is developmentally regulated	52.67	3.85E-08	34%	48%	123--185	60--122	1	NP_048389	contains Pro-rich Px motif, PAPK (8X); similar to Thermoproteus virus protein TPX, corresponds to Swiss-Prot Accession Number P19275	427.56	4.85E-118	65%	80%	186--499	99--410
					2	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	48.22	7.13E-07	27%	34%	98--196	39--138	2	NP_048519	similar to PBCV-1 ORF A41R, corresponds to GenBank Accession Number U17055	410.22	8.01E-113	63%	79%	189--499	77--385
					3	pfam05616	Neisseria_TspB, Neisseria meningitidis TspB protein. This family consists of several Neisseria meningitidis TspB virulence factor proteins..	47.43	1.30E-06	31%	35%	117--191	329--403	3	BAD86968	hypothetical protein	67.01	1.66E-09	23%	41%	194--476	95--375
					4	COG3147	DedD, Uncharacterized protein conserved in bacteria [Function unknown].	43.09	2.99E-05	33%	39%	99--187	63--151	4	NP_916095	P0481E12.18	67.01	1.66E-09	23%	41%	194--476	72--352
					5	pfam04625	DEC-1_N, DEC-1 protein, N terminal region. The defective chorion-1 gene (dec-1) in Drosophila encodes follicle cell proteins necessary for proper eggshell assembly. Multiple products of the dec-1 gene are formed by alternative RNA splicing and proteolytic processing. Cleavage products include S80 (80 kDa) which is incorporated into the eggshell, and further proteolysis of S80 gives S60 (60 kDa)	42.95	3.28E-05	33%	39%	108--175	75--142	5	BAE02830	surface protein	64.70	8.26E-09	50%	68%	56--113	572--629
					6	pfam06735	DUF1210, Protein of unknown function (DUF1210). This family represents a conserved region within plant proline-rich proteins..	42.41	4.83E-05	33%	37%	75--183	100--202	6	AAM63817	unknown	59.69	2.66E-07	21%	40%	195--477	63--341
					7	pfam02993	MCPVI, Minor capsid protein VI. This minor capsid protein may act as a link between the external capsid and the internal DNA-protein core. The C-terminal 11 residues may function as a protease cofactor leading to enzyme activation	41.17	9.40E-05	32%	38%	114--185	122--196	7	NP_191439	unknown protein	59.69	2.66E-07	21%	40%	195--477	63--341
					8	pfam05518	Totivirus_coat, Totivirus coat protein..	40.74	0.000133	18%	23%	103--183	670--750	8	XP_464783	unknown protein	51.22	9.45E-05	22%	41%	214--477	78--346

Gene Name	Genome Position	A.A. length	Peptide Mw	pI	CDD Hit Number	COGs	COG Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to	BLASTp Hit Number	Hit Accession	BLASTp Definition	Bit Score	E-value	% Identity	% Positive	Query from-to	Hit from-to
M807R	299385--299849	155	17,753	4.64	1	COG5201	SKP1, SCF ubiquitin ligase, SKP1 component [Posttranslational modification, protein turnover, chaperones].	93.20	2.25E-20	32%	55%	3--143	4--155	1	NP_048387	contains ATP/GTP-binding motif A; similar to Dictyostelium FP21 glycoprotein, corresponds to Swiss-Prot Accession Number P52285	109.00	4.25E-23	36%	61%	1--142	1--142
					2	smart00512	Skp1, Found in Skp1 protein family; Family of Skp1 (kinetochore protein required for cell cycle progression) and elongin C (subunit of RNA polymerase II transcription factor SIII) homologues. .	79.92	2.14E-16	36%	59%	1--95	2--107	2	NP_567959	ASK11 (ARABIDOPSIS SKP1-LIKE 11); ubiquitin-protein ligase	109.00	4.25E-23	38%	62%	3--144	6--151
					3	pfam01466	Skp1, Skp1 family, dimerisation domain..	64.48	1.03E-11	36%	65%	70--144	3--77	3	NP_567967	ASK12; protein binding / ubiquitin-protein ligase	107.07	1.62E-22	36%	61%	3--144	6--151
					4	pfam03931	Skp1_POZ, Skp1 family, tetramerisation domain..	59.53	3.50E-10	40%	66%	1--61	1--66	4	1P22_B	Chain B, Structure Of A Beta-Trcp1-Skp1-Beta-Catenin Complex: Destruction Motif Binding And Lysine Specificity On The Scfbeta-Trcp1 Ubiquitin Ligase	106.69	2.11E-22	38%	61%	2--142	3--141
														5	AAX83944	Skp1	106.30	2.76E-22	36%	57%	3--144	7--155
														6	BAB85607	kinetochore protein	104.76	8.02E-22	36%	58%	3--144	6--159
														7	CAH92499	hypothetical protein	103.99	1.37E-21	34%	57%	2--142	3--159
														8	AAO85510	SKP1	103.99	1.37E-21	37%	59%	3--144	4--152
														9	BAB85606	kinetochore protein	103.99	1.37E-21	35%	58%	3--144	6--160
														10	AAI02436	Hypothetical protein LOC615427	103.99	1.37E-21	34%	57%	2--142	3--159
M810R	299894--300142	83	9,480	10.61		No Hit Found									No Hit Found	No Hit Found						
M811L	300904--300149	252	29,824	5.27		No Hit Found									No Hit Found	No Hit Found						
M813R	301197--301880	228	25,821	4.93		No Hit Found									No Hit Found	No Hit Found						
M814L	302784--302212	191	21,820	10.62		No Hit Found									No Hit Found	No Hit Found						
M817R	302866--303330	155	18,715	6.79		No Hit Found									No Hit Found	No Hit Found						
M818R	303386--303766	127	14,321	4.20		No Hit Found									No Hit Found	No Hit Found						
M819R	303786--304223	146	16,768	10.18		No Hit Found									No Hit Found	No Hit Found						
M820R	304499--304954	152	18,365	6.28		No Hit Found									No Hit Found	No Hit Found						
M822R	305057--306853	599	65,968	10.89	1	cd01828	sialate_O-acetylesterase_like2, sialate_O-acetylesterase_like subfamily of the SGNH-hydrolases, , a diverse family of lipases and esterases. The tertiary fold of the enzyme is substantially different from that of the alpha/beta hydrolase family and unique among all known hydrolases; its active site closely resembles the Ser-His-Asp(Glu) triad found in other serine hydrolases	79.17	3.73E-16	25%	38%	120--275	1--167	1	NP_048674	A318R	243.43	1.62E-62	56%	71%	403--599	6--211
					2	COG0810	TonB, Periplasmic protein TonB, links inner and outer membranes [Cell envelope biogenesis, outer membrane].	60.54	1.47E-10	35%	42%	17--117	28--127	2	NP_048672	PAPK (17X); similar to PBCV-1 ORF A41R, corresponds to Genbank Accession Number U17055	80.11	2.38E-13	60%	69%	299--360	130--194
M828R	307542--308210	223	25,745	7.69		No Hit Found								1	NP_048436	A88R	73.56	5.00E-12	28%	42%	1--205	1--247
M829L	309223--308348	292	32,821	6.70	1	cd00204	ANK, ankyrin repeats; ankyrin repeats mediate protein-protein interactions in very diverse families of proteins. The number of ANK repeats in a protein can range from 2 to over 20 (ankyrins, for example). ANK repeats may occur in combinations with other types of domains. The structural repeat unit contains two antiparallel helices and a beta-hairpin, repeats are stacked in a superhelical arrangement; this alignment contains 4 consecutive repeats	98.23	6.93E-22	41%	68%	29--150	2--125	1	XP_681288	hypothetical protein AN8019.2	150.98	4.01E-35	40%	57%	5--233	867--1100
					2	COG0666	Arp, FOG: Ankyrin repeat [General function prediction only].	62.60	3.44E-11	30%	42%	5--148	77--224	2	CAE64680	Hypothetical protein CBG09456	136.35	1.02E-30	36%	54%	5--230	463--693
					3	pfam00023	Ank, Ankyrin repeat. There's no clear separation between noise and signal on the HMM search Ankyrin repeats generally consist of a beta, alpha, alpha, beta order of secondary structures. The repeats associate to form a higher order structure..	45.05	6.98E-06	59%	78%	193--225	1--33	3	XP_392578	PREDICTED: similar to CG7462-PB, isoform B	135.19	2.28E-30	36%	54%	5--230	441--671
					4	smart00248	ANK, ankyrin repeats; Ankyrin repeats are about 33 amino acids long and occur in at least four consecutive copies. They are involved in protein protein interactions. The core of the repeat seems to be an helix-loop-helix structure. .	38.02	0.000936	59%	78%	193--220	1--28	4	AAF61702	ankyrin 1	132.88	1.13E-29	39%	54%	5--230	506--736
														5	XP_581734	PREDICTED: similar to Ankyrin-1 (Erythrocyte ankyrin), partial	132.88	1.13E-29	39%	54%	5--230	679--909
														6	AAM11327	GH01626p	131.34	3.29E-29	36%	54%	5--230	224--454
														7	AAN12046	CG7462-PC, isoform C	131.34	3.29E-29	36%	54%	5--230	374--604
														8	AAF50525	CG7462-PB, isoform B	131.34	3.29E-29	36%	54%	5--230	374--604
														9	EAL87814	NACHT domain protein, putative	130.95	4.29E-29	36%	57%	20--240	1270--1489
														10	AAQ91911	Uncoordinated protein 44, isoform g	130.95	4.29E-29	35%	54%	5--227	462--689
M833L	310061--309258	268	28,814	5.83	1	cd00204	ANK, ankyrin repeats; ankyrin repeats mediate protein-protein interactions in very diverse families of proteins. The number of ANK repeats in a protein can range from 2 to over 20 (ankyrins, for example). ANK repeats may occur in combinations with other types of domains. The structural repeat unit contains two antiparallel helices and a beta-hairpin, repeats are stacked in a superhelical arrangement; this alignment contains 4 consecutive repeats	126.74	1.66E-30	46%	59%	38--156	8--126	1	NP_049038	contains 10 ankyrin-like repeats; similar to human ankyrin, corresponds to Swiss-Prot Accession Number P16157	167.93	2.74E-40	37%	54%	8--268	109--368
					2	COG0666	Arp, FOG: Ankyrin repeat [General function prediction only].	93.41	1.76E-20	38%	53%	21--159	57--203	2	NP_048353	contains 4 ankyrin repeats; similar to reticulocyte ankyrin, corresponds to Swiss-Prot Accession Number P16157	164.85	2.32E-39	37%	52%	1--252	1--251
					3	pfam00023	Ank, Ankyrin repeat. There's no clear separation between noise and signal on the HMM search Ankyrin repeats generally consist of a beta, alpha, alpha, beta order of secondary structures. The repeats associate to form a higher order structure..	44.28	1.17E-05	57%	67%	104--134	3--33	3	XP_794653	PREDICTED: similar to Ankyrin-1 (Erythrocyte ankyrin)	141.35	2.74E-32	44%	55%	8--189	234--415
					4	smart00248	ANK, ankyrin repeats; Ankyrin repeats are about 33 amino acids long and occur in at least four consecutive copies. They are involved in protein protein interactions. The core of the repeat seems to be an helix-loop-helix structure. .	39.17	0.000447	56%	70%	104--131	3--30	4	XP_796302	PREDICTED: similar to Ankyrin repeat domain protein 28, partial	140.97	3.58E-32	41%	53%	8--220	744--954
														5	EAL84954	ankyrin repeat protein	134.42	3.35E-30	40%	56%	8--215	430--634
														6	XP_782299	PREDICTED: similar to Ankyrin-3 (ANK-3) (Ankyrin G)	130.95	3.71E-29	41%	52%	7--189	559--741
														7	XP_782809	PREDICTED: similar to ankyrin repeat domain 28	130.95	3.71E-29	38%	55%	8--189	618--799
														8	XP_782887	PREDICTED: similar to Ankyrin-1 (Erythrocyte ankyrin) (Ankyrin R)	129.80	8.26E-29	36%	57%	3--189	374--560
														9	AAQ09555	inv-like protein	128.64	1.84E-28	40%	53%	8--189	411--594
														10	XP_788092	PREDICTED: similar to Ankyrin-2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid), partial	128.26	2.40E-28	41%	59%	7--189	875--1057
M838L	310686--310315	124	13,324	8.64		No Hit Found									No Hit Found	No Hit Found						
M842R	311500--312168	223	26,253	6.70		No Hit Found								1	ZP_00575949	RepA / Rep+ protein KID	55.45	1.41E-06	25%	59%	14--113	61--160
														2	NP_705165	hypothetical malaria antigen	51.60	2.03E-05	24%	51%	1--136	1993--2129
														3	YP_161395	hypothetical protein BGP110	49.68	7.73E-05	27%	54%	27--137	91--194
M844R	312868--313596	243	27,062	8.50		No Hit Found								1	NP_048723	A366L	153.30	5.82E-36	37%	56%	6--237	14--251
														2	NP_049005	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	59.31	1.14E-07	28%	40%	2--167	8--180
														3	NP_049003	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580	55.07	2.16E-06	41%	60%	1--60	1--59
														4	NP_048427	A79R	53.53	6.27E-06	44%	65%	1--47	1--47
														5	NP_048807	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580	53.14	8.19E-06	48%	63%	1--47	1--47
														6	NP_048525	A177R	51.60	2.38E-05	40%	63%	1--47	4--50
M845L	313927--313712	72	8,515	6.34		No Hit Found									No Hit Found	No Hit Found						