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 C: Gene Names n001L through n849R

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

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

Xiao Li
University of Massachusetts–Lowell

Tamara Feldblyum
The Institute for Genomic Research, Rockville, MD

James Hartigan
Agencourt Bioscience Corporation, Beverly, MA

See next page for additional authors

Published in *Virology* 358:2 (February 20, 2007), pp. 459–471. doi:10.1016/j.virol.2006.08.034 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 C: Gene Names n001L through n849R

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
n001L	560--45	172	20,086	12.51		No Hit Found									No Hit Found							
n002R	94--345	84	9,447	11.09	1	cd00170	SEC14, Sec14p-like lipid-binding domain. Found in secretory proteins, such as S. cerevisiae phosphatidylinositol transfer protein (Sec14p), and in lipid regulated proteins such as RhoGAPs, RhoGEFs and neurofibromin (NF1). SEC14 domain of Dbl is known to associate with G protein beta/gamma subunits	26.14	9.94E-01	28%	52%	57--82	90--115		No Hit Found							
n004R	821--1066	82	8,824	9.51		No Hit Found									No Hit Found							
n005R	1421--1624	68	7,493	10.59		No Hit Found									No Hit Found							
n006R	2167--2409	81	8,982	6.65		No Hit Found									No Hit Found							
n008L	2803--2597	69	7,971	11.67		No Hit Found									No Hit Found							
n009R	3127--3360	78	9,070	10.01		No Hit Found									No Hit Found							
n011R	3578--3952	125	14,071	8.04		No Hit Found									No Hit Found							
n013L	4762--4400	121	14,018	10.94	1	pfam06021	Glycine_acyl_tr, Aralkyl acyl-CoA:amino acid N-acyltransferase. This family consists of several mammalian specific aralkyl acyl-CoA:amino acid N-acyltransferase (glycine N-acyltransferase) proteins EC:2.3.1.13..	26.87	7.26E-01	47.83%	56.52%	95--118	173--196		No Hit Found							
n017L	5838--5437	134	15,130	4.72		No Hit Found								1	AAC96566	a198R	42.74	4.33E-03	37%	59%	69--126	6--64
n018L	6205--5948	86	9,578	8.41		No Hit Found								1	AAC96562	a194R	76.26	3.56E-13	50%	59%	1--79	1--79
n021R	8881--9231	117	12,381	11.31		No Hit Found									No Hit Found							
n022R	10331--10531	67	7,096	6.05		No Hit Found									No Hit Found							
n024R	11057--11593	179	20,792	10.40		No Hit Found									No Hit Found							
n025R	11445--11681	79	8,549	10.27		No Hit Found									No Hit Found							
n026L	11880--11650	77	8,826	12.85		No Hit Found									No Hit Found							
n027R	12992--13393	134	14,645	8.98		No Hit Found									No Hit Found							
n028R	13173--13406	78	8,279	10.51	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	26.70	7.37E-01	32%	48%	26--51	147--172		No Hit Found							
n032R	15112--15447	112	12,187	10.05		No Hit Found									No Hit Found							
n036L	19108--18788	107	11,214	9.50		No Hit Found									No Hit Found							
n037L	19415--19056	120	13,550	9.48		No Hit Found									No Hit Found							
n038L	19693--19373	107	11,221	6.49		No Hit Found									No Hit Found							
n039L	20050--19778	91	9,852	9.97		No Hit Found									No Hit Found							
n041L	20516--20244	91	9,852	4.97		No Hit Found									No Hit Found							
n042L	20870--20661	70	7,481	6.73	1	pfam05958	tRNA_U5-meth_tr, tRNA (Uracil-5-)-methyltransferase. This family consists of (Uracil-5-)-methyltransferases EC:2.1.1.35 from bacteria, archaea and eukaryotes. A 5-methyluridine (m(5)U) residue at position 54 is a conserved feature of bacterial and eukaryotic tRNAs. The methylation of U54 is catalysed by the tRNA(m5U54)methyltransferase, which in Saccharomyces cerevisiae is encoded by the nonessential TRM2 gene. It is thought that tRNA modification enzymes might have a role in tRNA maturation not necessarily linked to their known catalytic activity..	27.19	4.54E-01	37.04%	48.15%	35--61	47--74		No Hit Found							
n043L	21422--21159	88	9,424	10.25	1	COG5101	CRM1, Importin beta-related nuclear transport receptor [Nuclear structure / Intracellular trafficking and secretion].	30.00	6.72E-02	29.41%	43.14%	1--55	358--409		No Hit Found							
n044R	22773--22976	68	7,575	7.30		No Hit Found									No Hit Found							
n046L	23568--23209	120	13,021	4.76		No Hit Found									No Hit Found							
n047L	23925--23623	101	10,888	10.24	1	COG1368	MdoB, Phosphoglycerol transferase and related proteins, alkaline phosphatase superfamily [Cell envelope biogenesis, outer membrane].	29.24	1.20E-01	20.34%	38.98%	10--69	106--165		No Hit Found							
n048L	24339--23953	129	13,677	6.37	1	cd00704	MDH, malate dehydrogenases (MDH); member of the family of NAD-dependent 2-hydroxycarboxylate dehydrogenases. MDH is one of the key enzymes in the citric acid cycle, facilitating both the conversion of malate to oxaloacetate and replenishing levels of oxaloacetate by reductive carboxylation of pyruvate..	27.00	6.53E-01	31.15%	45.90%	68--122	23--84		No Hit Found							
n049L	25269--24520	250	26,993	4.53		No Hit Found								1	AAC39774	hepatitis A virus cellular receptor 1 long form	46.21	1.19E-03	26%	35%	46--216	123--288
														2	AAC39773	hepatitis A virus cellular receptor 1 short form	46.21	1.19E-03	26%	35%	46--216	118--283
														3	CAA66906	HAVcr-1 protein	58.54	2.32E-07	26%	34%	47--244	142--325
														4	AAC39772	hepatitis A virus cellular receptor 1 long form	45.82	1.55E-03	24%	33%	46--246	123--339
														5	AAC39771	hepatitis A virus cellular receptor 1 short form	45.82	1.55E-03	24%	33%	46--246	118--334
														6	BAA21556	hepatitis A virus receptor	53.91	5.71E-06	29%	38%	47--209	162--338
														7	XP_981286	PREDICTED: similar to Maltase-glucoamylase, intestinal	43.51	7.71E-03	28%	34%	53--247	1902--2100
														8	NP_002448	mucin 2	45.05	2.65E-03	23%	30%	26--248	1772--1712
														9	AAA59163	mucin	45.05	2.65E-03	23%	30%	26--248	1747--1087
														1	AAA59875	mucin	43.90	5.91E-03	23%	29%	26--248	142--370
n050L	25860--25486	125	13,445	5.69		No Hit Found									No Hit Found							
n051R	26134--26418	95	10,578	9.69		No Hit Found									No Hit Found							
n053L	27905--27609	99	10,414	3.98	1	pfam01512	Complex1_51K, Respiratory-chain NADH dehydrogenase 51 Kd subunit..	26.36	9.36E-01	25.37%	34.33%	3--66	162--229		No Hit Found							
n054L	28652--28023	210	22,679	4.64		No Hit Found									No Hit Found							
n055L	29132--28920	71	7,604	4.77		No Hit Found									No Hit Found							
n056L	29417--29199	73	7,780	4.21		No Hit Found									No Hit Found							
n057L	30116--29883	78	8,281	3.98		No Hit Found									No Hit Found							
n058L	31796--31536	87	9,223	9.37		No Hit Found									No Hit Found							
n060L	32471--32112	120	12,815	4.02		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
n061L	32828–32526	101	11,033	4.82		No Hit Found									No Hit Found							
n062L	33155–32856	100	10,480	6.42		No Hit Found									No Hit Found							
n063L	33917–33702	72	7,621	4.21		No Hit Found									No Hit Found							
n064L	34619–34386	78	8,142	4.04		No Hit Found									No Hit Found							
n065R	34397–34651	85	9,252	7.92		No Hit Found									No Hit Found							
n066L	35576–35322	85	9,206	9.92		No Hit Found									No Hit Found							
n068L	37063–36650	138	14,723	10.61		No Hit Found									No Hit Found							
n069L	37327–37070	86	9,154	10.20		No Hit Found									No Hit Found							
n070L	38965–38765	67	7,458	9.52		No Hit Found									No Hit Found							
n071R	40174–40392	73	8,004	6.76		No Hit Found									No Hit Found							
n072L	40492–40202	97	10,436	8.68		No Hit Found									No Hit Found							
n073L	40828–40505	108	11,893	4.84		No Hit Found									No Hit Found							
n075R	41174–41473	100	11,484	11.91		No Hit Found									No Hit Found							
n076L	41852–41589	88	9,469	4.14	1	pfam02779	Transket_pyr, Transketolase, pyridine binding domain. This family includes transketolase enzymes, pyruvate dehydrogenases, and branched chain alpha-keto acid decarboxylases..	27.45	4.54E-01	33.33%	46.67%	35–65	41–71		No Hit Found							
n077R	42662–43054	131	15,522	12.61		No Hit Found									No Hit Found							
n078L	43045–42662	128	14,467	10.92		No Hit Found									No Hit Found							
n080L	43239–42841	133	14,491	8.22		No Hit Found									No Hit Found							
n083L	44416–43925	164	18,299	10.79		No Hit Found									No Hit Found							
n085L	45215–44688	176	19,305	11.29		No Hit Found									No Hit Found							
n088L	46680–46462	73	8,142	11.84		No Hit Found									No Hit Found							
n090R	47203–47433	77	8,675	7.12		No Hit Found									No Hit Found							
n094L	49848–49507	114	11,764	8.81		No Hit Found									No Hit Found							
n096R	50300–50527	76	8,126	6.20	1	pfam00109	ketoacyl-synt, Beta-ketoacyl synthase, N-terminal domain. The structure of beta-ketoacyl synthase is similar to that of the thiolase family (Pfam::PF00108) and also chalcone sythase. The active site of beta-ketoacyl synthase is located between the N and C-terminal domains. The N-terminal domain contains most of the structures involved in dimer formation and also the active site cysteine..	26.40	8.67E-01	30.43%	56.52%	17–40	140–163		No Hit Found							
n097L	50982–50647	112	12,444	12.24		No Hit Found									No Hit Found							
n098R	50885–51130	82	8,592	8.92		No Hit Found									No Hit Found							
n100L	51612–51319	98	10,903	8.46		No Hit Found									No Hit Found							
n104L	53606–53253	118	12,972	8.57		No Hit Found									No Hit Found							
n105L	54124–53747	126	13,943	8.46		No Hit Found									No Hit Found							
n108R	54820–55038	73	7,934	10.23		No Hit Found									No Hit Found							
n111L	55776–55576	67	7,599	10.22		No Hit Found									No Hit Found							
n112L	56138–55770	123	13,689	10.77		No Hit Found									No Hit Found							
n113L	57207–57004	68	7,656	10.78		No Hit Found									No Hit Found							
n114R	57461–57706	82	8,873	9.59		No Hit Found									No Hit Found							
n118L	58870–58646	75	8,152	7.46		No Hit Found									No Hit Found							
n120R	58910–59119	70	7,610	9.23	1	pfam00955	HCO3_cotransp, HCO3- transporter family. This family contains Band 3 anion exchange proteins that exchange CL-/HCO3-. This family also includes cotransporters of Na+/HCO3-..	26.78	7.11E-01	37.04%	62.96%	16–43	567–594		No Hit Found							
n121L	59273–59070	68	7,664	9.57		No Hit Found									No Hit Found							
n122R	59548–59922	125	14,427	10.70	1	pfam03926	DUF335, Putative metallopeptidase (SprT family). This family of uncharacterised proteins may be zinc metallopeptidases..	27.18	5.89E-01	23.08%	39.74%	29–104	59–137	1	AAC96464	a96R	42.36	5.68E-03	40%	49%	1–57	1–65
n123R	59696–59899	68	6,982	11.55		No Hit Found									No Hit Found							
n125L	60599–60138	154	17,288	7.00		No Hit Found									No Hit Found							
n126L	61171–60941	77	8,806	11.17	1	pfam00735	GTP_CDC, Cell division protein. Members of this family include CDC3, CDC10, CDC11 and CDC12/Septin. Members of this family bind GTP..	27.10	5.00E-01	37.04%	66.67%	15–42	103–130		No Hit Found							
n127R	61069–61395	109	11,538	9.89		No Hit Found									No Hit Found							
n129L	62792–62205	196	21,660	8.56		No Hit Found								1	AAC96472	a104L	50.45	3.77E-05	29%	54%	60–150	5–96
n130L	63178–62915	88	9,304	11.50	1	pfam03571	Peptidase M49, Peptidase family M49..	27.70	3.81E-01	27.94%	45.59%	12–71	460–528		No Hit Found							
n133R	63702–63965	88	9,920	5.14		No Hit Found									No Hit Found							
n134L	64306–64088	73	7,923	8.73		No Hit Found									No Hit Found							
n136R	64300–64515	72	7,805	11.75		No Hit Found									No Hit Found							
n138R	65239–65478	80	8,665	11.88	1	COG3919	COG3919, Predicted ATP-grasp enzyme [General function prediction only].	26.91	6.24E-01	22.22%	35.56%	19–64	9–54		No Hit Found							
n139R	65247–65486	80	8,967	7.62		No Hit Found									No Hit Found							
n141R	65780–66073	98	11,467	8.83		No Hit Found									No Hit Found							
n142R	66123–66416	98	9,977	4.24	1	cd00203	ZnMc, Zinc-dependent metalloprotease; Neutral zinc metallopeptidases. This alignment represents a subset of known subfamilies. The HExxH zinc-binding site/active site is best conserved..	27.06	5.29E-01	22.39%	31.34%	5–70	35–102		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
n143R	66169–66375	69	7,391	4.34		No Hit Found									No Hit Found								
n144R	67308–67514	69	7,479	11.92		No Hit Found									No Hit Found								
n145L	68436–67597	280	31,356	8.66		No Hit Found									No Hit Found								
n147R	67845–68171	109	12,561	11.37		No Hit Found									No Hit Found								
n148R	68271–68564	98	10,927	11.50		No Hit Found									No Hit Found								
n149L	68536–68273	88	9,795	6.67	1	cd03290	ABC_SUR2_N, The SUR domain 1. The sulfonylurea receptor SUR is an ATP transporter of the ABCC/MRP family with tandem ATPase binding domains. Unlike other ABC proteins, it has no intrinsic transport function, neither active nor passive, but associates with the potassium channel proteins Kir6.1 or Kir6.2 to form the ATP-sensitive potassium (K(ATP)) channel. Within the channel complex, SUR serves as a regulatory subunit that fine-tunes the gating of Kir6.x in response to alterations in cellular metabolism. It constitutes a major pharmaceutical target as it binds numerous drugs, K(ATP) channel openers and blockers, capable of up- or down-regulating channel activity. We here review current knowledge on the molecular basis of the interaction of classical K(ATP) channel openers (cromakalim, pinacidil, diazoxide) with SUR	27.37	4.78E-01	48.15%	66.67%	55–82	162–189		No Hit Found								
n153L	69803–69597	69	7,921	9.64		No Hit Found									No Hit Found								
n154L	69943–69740	68	7,716	5.70		No Hit Found									No Hit Found								
n156L	72077–69924	718	78,713	4.64	1	COG1038	PycA, Pyruvate carboxylase [Energy production and conversion].	30.61	7.89E-01	39.47%	55.26%	265–303	129–167	1	BAB03062	unnamed protein product	98.21	1.20E-18	26%	35%	2–646	57–1396
														2	XP_499823	hypothetical protein	102.83	4.86E-20	29%	45%	298–669	16–404
														3	CAE74002	Hypothetical protein CBG21638	93.97	2.26E-17	21%	39%	142–667	12–495
														4	NP_730262	CG13731-PA	100.91	1.85E-19	28%	32%	21–653	239–849
														5	XP_931991	PREDICTED: similar to CG13731-PA	82.42	6.79E-14	24%	37%	87–667	44–667
														6	NP_180307	ATEPR1	49.68	4.88E-04	26%	33%	350–691	52–363
														7	XP_501710	hypothetical protein	58.15	1.37E-06	22%	35%	10–524	395–876
														8	XP_941694	PREDICTED: similar to mucin 17	63.93	2.50E-08	23%	37%	40–689	70–2761
														9	ZP_00567828	Hypothetical protein Franean1DRAFT_6526	51.60	1.28E-04	26%	35%	17–362	266–553
n157R	70406–70615	70	7,370	8.37		No Hit Found								1	AAL04416	zonadhesin	78.95	7.51E-13	22%	35%	21–652	52–1071
n158L	72950–72117	278	30,538	4.44	1	COG2114	CyaA, Adenylate cyclase, family 3 (some proteins contain HAMP domain) [Signal transduction mechanisms].	30.14	3.41E-01	13.33%	40%	190–235	18–63	No Hit Found									
n159R	72509–72718	70	7,423	8.37		No Hit Found								1	XP_499823	hypothetical protein	50.45	7.68E-05	32%	44%	1–111	311–440
														2	AAC39772	hepatitis A virus cellular receptor 1 long form	43.90	7.19E-03	28%	36%	2–165	174–328
														3	AAC39771	hepatitis A virus cellular receptor 1 short form	43.90	7.19E-03	28%	36%	2–165	169–323
														4	BAA21556	hepatitis A virus receptor	53.14	1.19E-05	33%	41%	2–158	175–329
														5	AAC39774	hepatitis A virus cellular receptor 1 long form	48.14	3.81E-04	27%	36%	84–274	154–307
														6	AAC39773	hepatitis A virus cellular receptor 1 short form	48.14	3.81E-04	27%	36%	84–274	149–302
														7	CAA66906	HAVcr-1 protein	44.67	4.22E-03	27%	36%	70–260	127–294
														8	BAB03062	unnamed protein product	47.75	4.98E-04	24%	34%	8–274	262–560
														9	EAL42150	ENSANGP00000025869	51.22	4.50E-05	20%	35%	18–243	64–308
														1	AAA28405	calcium-binding protein	46.21	1.45E-03	24%	37%	37–270	153–383
n160L	73568–73359	70	7,587	7.46		No Hit Found									No Hit Found								
n161R	73742–74029	96	11,008	11.14		No Hit Found									No Hit Found								
n162L	74636–74268	123	13,266	7.86	1	COG3365	COG3365, Uncharacterized protein conserved in archaea [Function unknown].	27.56	3.99E-01	37.93%	58.62%	77–106	54–83		No Hit Found								
n163L	75493–74735	253	28,019	9.85		No Hit Found								1	AAC96649	a281R	45.05	2.72E-03	49%	54%	99–149	130–180
n166R	76095–76376	94	10,406	10.36	1	pfam05028	PARG_cat, Poly (ADP-ribose) glycohydrolase (PARG). Poly(ADP-ribose) glycohydrolase (PARG), is a ubiquitously expressed exo- and endoglycohydrolase which mediates oxidative and excitotoxic neuronal death..	27.60	3.22E-01	40%	45.71%	45–82	87–122	1	AAC96494	a126R	51.60	9.13E-06	50%	61%	28–71	23–66
n167R	76204–76410	69	7,447	10.52	1	COG3967	DltE, Short-chain dehydrogenase involved in D-alanine esterification of lipoteichoic acid and wall teichoic acid (D-alanine transfer protein) [Cell envelope biogenesis, outer membrane].	26.82	5.41E-01	36.67%	60%	22–58	113–143		No Hit Found								
n171L	78947–78672	92	10,651	11.06		No Hit Found									No Hit Found								
n172L	79387–79124	88	10,400	8.64		No Hit Found									No Hit Found								
n175R	79652–79978	109	11,824	12.02		No Hit Found									No Hit Found								
n178R	80489–80713	75	8,288	10.14		No Hit Found									No Hit Found								
n179R	81023–81250	76	8,533	5.86		No Hit Found									No Hit Found								
n181L	82395–82162	78	9,071	9.64	1	smart00351	PAX, Paired Box domain; .	28.31	2.43E-01	34.78%	60.87%	22–69	18–64		No Hit Found								
n182R	82193–82555	121	14,113	7.58		No Hit Found									No Hit Found								
n184R	83375–83581	69	7,525	5.47		No Hit Found									No Hit Found								
n188L	85666–85298	123	13,444	12.46		No Hit Found									No Hit Found								
n190R	86951–87160	70	7,678	9.01	1	pfam05519	MSP4, Merozoite surface protein 4/5 (MSP4/5). This family consists of Merozoite surface proteins 4 and 5 (MSP4/5). MSP4 is a protein with apparent molecular mass of 40 kDa that is synthesised by mature stage parasites and anchored to the merozoite membrane by a glycosylphosphatidylinositol moiety. MSP4 is immunogenic in both laboratory animals and during natural infection. Antibodies raised to this protein can inhibit parasite growth in vitro. Its homologue in the rodent malaria species Plasmodium yoelii, PyMSP4/5, is capable of conferring significant protection against lethal challenge in mice. All of these suggest that MSP4 is a candidate for inclusion in an effective asexual-stage malaria vaccine	26.56	6.86E-01	22.45%	46.94%	1–49	199–248		No Hit Found								
n192L	87290–87048	81	9,064	10.42		No Hit Found									No Hit Found								
n193R	88095–88358	88	9,710	7.33	1	COG2192	COG2192, Predicted carbamoyl transferase, NodU family [Posttranslational modification, protein turnover, chaperones]	28.72	1.86E-01	29.27%	46.34%	25–68	107–148		No Hit Found								
n194L	88896–88543	118	12,355	9.37		No Hit Found									No Hit Found								
n196R	88716–88916	67	7,382	12.25		No Hit Found									No Hit Found								
n197L	89632–89297	112	12,154	8.22		No Hit Found									No Hit Found								
n198R	89598–89873	92	10,322	12.10		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
n199L	90265--89906	120	13,144	6.89		No Hit Found									No Hit Found								
n200L	90562--90326	79	8,543	6.72	1	cd01977	Nitrogenase_VFe_alpha, Nitrogenase_VFe_alpha -like: Nitrogenase VFe protein, alpha subunit like. This group contains proteins similar to the alpha subunits of, the VFe protein of the vanadium-dependent (V-) nitrogenase and the FeFe protein of the iron-on	26.41	7.86E-01	25.86%	43.10%	5--66	165--223		No Hit Found								
n201L	91429--91070	120	13,336	8.05		No Hit Found									No Hit Found								
n202L	91827--91363	155	14,902	4.42		No Hit Found									No Hit Found								
n203R	91443--91682	80	9,100	12.98		No Hit Found									No Hit Found								
n204L	91759--91490	90	10,739	11.55		No Hit Found									No Hit Found								
n207L	92915--92640	92	10,518	10.82		No Hit Found									No Hit Found								
n208L	92971--92717	85	9,179	8.81		No Hit Found									No Hit Found								
n212L	95134--94607	176	19,072	9.52		No Hit Found									No Hit Found								
n213R	94661--94966	102	10,901	11.01		No Hit Found									No Hit Found								
n215R	95195--95596	134	15,344	10.19		No Hit Found									No Hit Found								
n217L	95940--95722	73	8,048	12.08		No Hit Found									No Hit Found								
n219L	96775--96566	70	7,670	11.01	1	pfam05677	DUF818, Chlamydia CHLPS protein (DUF818). This family consists of several Chlamydia CHLPS proteins, the function of which are unknown..	26.84	5.74E-01	27.12%	47.46%	10--68	124--183		No Hit Found								
n220L	97306--97034	91	9,138	8.37	1	COG2808	PaiB, Transcriptional regulator [Transcription].	28.77	1.43E-01	35.48%	48.39%	31--62	22--53		No Hit Found								
n224R	98258--98932	225	25,955	11.19	1	COG0514	RecQ, Superfamily II DNA helicase [DNA replication, recombination, and repair].	34.94	8.22E-03	24.07%	36.11%	96--202	146--254		No Hit Found								
n225R	98484--98699	72	7,578	12.10		No Hit Found									No Hit Found								
n226L	99231--98947	95	10,241	10.36		No Hit Found									No Hit Found								
n228R	99249--100028	260	28,714	10.89		No Hit Found									No Hit Found								
n229R	100584--100934	117	12,773	10.00	1	COG4109	COG4109, Predicted transcriptional regulator containing CBS domains [Transcription].	26.77	7.88E-01	20.45%	39.77%	9--79	208--296		No Hit Found								
n230R	100850--101176	109	12,606	8.82		No Hit Found									No Hit Found								
n233R	102076--102285	70	7,603	10.85	1	COG1139	COG1139, Uncharacterized conserved protein containing a ferredoxin-like domain [Energy production and conversion].	26.78	5.81E-01	34.29%	60%	8--43	224--259	1	AAC96606	a238L	88.58	6.78E-17	68%	73%	2--68	48--114
n234R	102418--102768	117	12,888	8.83		No Hit Found									No Hit Found								
n235R	102939--103367	143	16,249	7.62		No Hit Found									No Hit Found								
n236R	103225--103707	161	18,831	11.76		No Hit Found									No Hit Found								
n238R	103388--103624	79	8,414	12.38		No Hit Found									No Hit Found								
n245L	106022--105795	76	8,386	11.17		No Hit Found								1	AAC96596	a228R	72.40	5.12E-12	54%	64%	1--74	1--73
n247L	106601--106326	92	10,495	10.47		No Hit Found									No Hit Found								
n248L	107102--106842	87	9,476	11.45		No Hit Found									No Hit Found								
n249R	107555--107965	137	14,626	11.35		No Hit Found								1	XP_782809	PREDICTED: similar to ankyrin repeat domain 28	41.59	9.54E-03	26%	46%	4--135	396--542
n250R	108397--108615	73	8,295	7.77		No Hit Found									No Hit Found								
n251R	108599--108934	112	11,696	4.88	1	COG5316	COG5316, Uncharacterized conserved protein [Function unknown].	26.14	9.68E-01	20%	34.12%	7--97	137--222		No Hit Found								
n252R	108667--108870	68	7,462	7.81		No Hit Found									No Hit Found								
n256L	112190--111915	92	9,839	11.84	1	cd01375	KISc_KIF9_like, Kinesin motor domain, KIF9-like subgroup; might play a role in cell shape remodeling. This catalytic (head) domain has ATPase activity and belongs to the larger group of P-loop NTPases. Kinesins are microtubule-dependent molecular motors	26.06	9.18E-01	60%	70%	22--42	174--194		No Hit Found								
n257R	112109--112522	138	15,260	8.43	1	COG3434	COG3434, Predicted signal transduction protein containing EAL and modified HD-GYP domains [Signal transduction mechanisms].	26.78	8.84E-01	33.33%	55.56%	31--58	67--94		No Hit Found								
n258L	112392--112189	68	7,555	12.26		No Hit Found									No Hit Found								
n260R	112732--113007	92	10,518	8.36		No Hit Found									No Hit Found								
n261L	113373--113155	73	7,579	5.80		No Hit Found									No Hit Found								
n264R	115226--115576	117	13,442	10.04	1	pfam06665	DUF1172, Protein of unknown function (DUF1172). This family represents a conserved region of unknown function within NAC1 and a number of hypothetical proteins whose sequences bear resemblance to it. NAC1 is a constitutively-expressed POZ/BTB transcription factor found in mammalian neurones that can regulate behaviours associated with cocaine use. All family members contain the pfam00651 domain	26.52	7.72E-01	29.63%	40.74%	35--62	9--36		No Hit Found								
n265R	115468--115761	98	10,869	10.32		No Hit Found									No Hit Found								
n268R	117593--117958	122	13,691	10.53	1	COG1982	LdcC, Arginine/lysine/ornithine decarboxylases [Amino acid transport and metabolism]	26.82	8.38E-01	23.68%	52.63%	31--69	45--83		No Hit Found								
n270L	118375--118124	84	9,249	10.32	1	pfam05934	MCLC, Mid-1-related chloride channel (MCLC). This family consists of several mid-1-related chloride channels. mid-1-related chloride channel (MCLC) proteins function as a chloride channel when incorporated in the planar lipid bilayer..	28.57	1.66E-01	39.29%	57.14%	17--45	506--534		No Hit Found								
n272L	118635--118417	73	7,953	4.56		No Hit Found									No Hit Found								
n273R	118821--119039	73	8,512	11.89		No Hit Found									No Hit Found								
n275R	119944--120168	75	8,625	10.99		No Hit Found									No Hit Found								
n276R	119952--120152	67	7,318	7.29		No Hit Found									No Hit Found								
n278R	120244--120495	84	9,438	9.04		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
n279R	120547--120777	77	8,259	8.67		No Hit Found									No Hit Found							
n280L	121105--120845	87	10,070	12.31		No Hit Found									No Hit Found							
n281R	120919--121206	96	10,784	11.93		No Hit Found									No Hit Found							
n282R	121200--121403	68	7,530	8.53		No Hit Found									No Hit Found							
n283L	122244--121672	191	21,348	7.67		No Hit Found									No Hit Found							
n285R	121898--122221	108	12,060	12.51		No Hit Found									No Hit Found							
n286R	122243--122467	75	8,504	12.60	1	pfam01237	Oxysterol BP, Oxysterol-binding protein..	26.12	9.63E-01	37.50%	45.83%	5--29	177--201		No Hit Found							
n290L	123665--123297	123	14,414	10.52	1	pfam02298	Cu_bind_like, Plastocyanin-like domain. This family represents a domain found in flowering plants related to the copper binding protein plastocyanin. Some members of this family may not bind copper due to the lack of key residues..	27.20	6.05E-01	18.87%	39.62%	70--115	27--80		No Hit Found							
n291L	123646--123377	90	9,883	11.95	1	COG4993	Gcd, Glucose dehydrogenase [Carbohydrate transport and metabolism].	28.41	2.17E-01	31.43%	45.71%	47--82	100--135		No Hit Found							
n292L	124238--123954	95	10,513	9.97	1	cd02013	TPP_Xsc_like, Thiamine pyrophosphate (TPP) family, Xsc-like subfamily, TPP-binding module; composed of proteins similar to Alcaligenes defragrans sulfoacetaldehyde acetyltransferase (Xsc). Xsc plays a key role in the degradation of taurine, catalyzing the desulfonation of 2-sulfoacetaldehyde into sulfite and acetyl phosphate. This enzyme requires TPP and divalent metal ions for activity	26.44	7.96E-01	32%	48%	3--28	154--179		No Hit Found							
n294R	124439--124639	67	7,468	8.46		No Hit Found									No Hit Found							
n295L	124850--124539	104	11,367	4.64	1	pfam00218	IGPS, Indole-3-glycerol phosphate synthase..	26.72	6.03E-01	34.29%	54.29%	58--93	134--169		No Hit Found							
n296L	125097--124711	129	14,233	7.19		No Hit Found									No Hit Found							
n297L	125408--125157	84	9,033	10.16	1	pfam05021	NPL4, NPL4 family. The HRD4 gene was identical to NPL4, a gene previously implicated in nuclear transport. Using a diverse set of substrates and direct ubiquitination assays, analysis revealed that HRD4/NPL4 is required for a poorly characterised step in ER-associated degradation after ubiquitination of target proteins but before their recognition by the 26S proteasome. Npl4p physically associates with Cdc48p via Ufd1p to form a Cdc48p-Ufd1p-Npl4p complex. The Cdc48-Ufd1-Npl4 complex functions in the recognition of several polyubiquitin-tagged proteins and facilitates their presentation to the 26S proteasome for processive degradation or even more specific processing	27.67	3.52E-01	28.57%	42.86%	31--73	193--235		No Hit Found							
n298L	125725--125474	84	9,155	7.31		No Hit Found									No Hit Found							
n300L	126167--125544	208	23,536	10.49		No Hit Found									No Hit Found							
n301L	126354--126067	96	10,539	8.47		No Hit Found									No Hit Found							
n302L	126490--126251	80	8,629	5.33		No Hit Found									No Hit Found							
n305L	127297--126986	104	11,861	9.48		No Hit Found									No Hit Found							
n306L	127535--127308	76	9,261	13.26		No Hit Found									No Hit Found							
n308R	127550--128293	248	25,403	12.71		No Hit Found									No Hit Found							
n309L	127850--127626	75	8,728	12.93		No Hit Found									No Hit Found							
n310R	127957--128319	121	12,931	7.41		No Hit Found									No Hit Found							
n311R	128395--128844	150	16,562	5.43	1	COG4133	CcmA, ABC-type transport system involved in cytochrome c biogenesis, ATPase component [Posttranslational modification, protein turnover, chaperones].	28.32	3.63E-01	25.86%	37.93%	75--127	98--156		No Hit Found							
n314R	129876--130133	86	9,200	10.96		No Hit Found									No Hit Found							
n316R	130620--130850	77	7,719	10.79		No Hit Found								1	AAC96572	a204L	84.34	1.30E-15	65%	67%	2--74	6--78
n319R	131912--132160	83	9,099	10.46		No Hit Found									No Hit Found							
n322R	132394--132690	99	10,913	9.42		No Hit Found									No Hit Found							
n324R	132761--133066	102	11,781	9.95		No Hit Found									No Hit Found							
n326R	133663--134079	139	14,728	11.07		No Hit Found									No Hit Found							
n328R	134151--134564	138	14,076	4.67		No Hit Found									No Hit Found							
n332L	136570--136370	67	7,306	9.45		No Hit Found									No Hit Found							
n334L	136974--136567	136	15,422	11.35		No Hit Found									No Hit Found							
n337R	137318--137695	126	13,577	6.23		No Hit Found									No Hit Found							
n338R	137670--137873	68	8,061	8.12		No Hit Found									No Hit Found							
n340R	138043--138702	220	24,154	5.88		No Hit Found									No Hit Found							
n341L	138845--138489	119	12,958	12.11		No Hit Found									No Hit Found							
n343R	139348--139554	69	7,808	8.53	1	pfam00724	Oxidored_FMN, NADH:flavin oxidoreductase / NADH oxidase famil..	26.37	7.84E-01	34.78%	39.13%	3--26	265--288		No Hit Found							
n344R	139379--139624	82	9,263	11.08		No Hit Found									No Hit Found							
n345R	139672--140145	158	18,643	11.21		No Hit Found									No Hit Found							
n347R	140323--140631	103	11,992	8.17		No Hit Found									No Hit Found							
n348R	140390--140764	125	14,123	10.61		No Hit Found									No Hit Found							
n349R	140877--141122	82	8,771	11.48		No Hit Found									No Hit Found							
n350L	141349--140951	133	13,658	3.84	1	pfam07040	DUF1326, Protein of unknown function (DUF1326). This family consists of several hypothetical bacterial proteins which seem to be found exclusively in Rhizobium and Ralstonia species. Members of this family are typically around 210 residues in length and contain 5 highly conserved cysteine residues at their N-terminus. The function of this family is unknown..	29.10	2.10E-01	37.84%	51.35%	13--52	30--67		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
n352L	141798--141421	126	13,155	11.25		No Hit Found									No Hit Found							
n353R	141943--142152	70	7,763	12.17		No Hit Found									No Hit Found							
n355L	143585--143385	67	7,482	9.44		No Hit Found									No Hit Found							
n356L	144138--143425	238	26,844	7.80		No Hit Found									No Hit Found							
n358R	143922--144278	119	12,732	9.04	1	COG5558	COG5558, Transposase [DNA replication, recombination, and repair].	26.91	6.19E-01	39.39%	51.52%	75--108	187--220		No Hit Found							
n360L	144546--144178	123	13,268	12.35		No Hit Found									No Hit Found							
n361R	144323--144607	95	10,186	8.22		No Hit Found									No Hit Found							
n362L	145161--144688	158	16,906	10.50		No Hit Found									No Hit Found							
n363L	146019--145669	117	13,428	9.08		No Hit Found									No Hit Found							
n365R	145868--146104	79	8,676	11.92		No Hit Found									No Hit Found							
n366R	146231--146440	70	7,515	12.04		No Hit Found									No Hit Found							
n368R	147662--148138	159	18,784	10.04	1	COG4341	COG4341, Predicted HD phosphohydrolase [General function prediction only].	30.36	9.96E-02	18.75%	41.67%	60--106	70--118	1	AAC96718	a350R	152.53	3.84E-36	50%	75%	1--120	1--120
														2	XP_636443	hypothetical protein DDB0188233	47.75	1.33E-04	22%	44%	3--120	53--205
														3	XP_363704	hypothetical protein MG01630.4	43.90	1.92E-03	22%	45%	3--122	49--201
														4	XP_380701	hypothetical protein FG00525.1	41.97	7.31E-03	23%	42%	3--120	50--199
n370R	148023--148289	89	9,969	12.10		No Hit Found									No Hit Found							
n371L	148918--148700	73	7,799	8.95		No Hit Found									No Hit Found							
n374R	150145--150498	118	12,895	11.47		No Hit Found								1	AAP09598	Cell surface protein	42.36	5.64E-03	39%	69%	53--105	56--4111
														2	AAT59485	conserved hypothetical protein	41.59	9.62E-03	39%	69%	53--105	56--4111
														3	AAT30727	conserved repeat domain protein	41.59	9.62E-03	39%	69%	53--105	56--4111
														4	EAR77175	Protein of unknown function DUF11	41.59	9.62E-03	37%	67%	53--105	20--3775
n375R	150465--150956	164	19,332	6.08		No Hit Found									No Hit Found							
n376L	150859--150467	131	16,696	13.52		No Hit Found									No Hit Found							
n378R	151106--151543	146	16,389	8.64		No Hit Found									No Hit Found							
n379L	151605--151372	78	8,446	8.73		No Hit Found									No Hit Found							
n381L	152106--151894	71	7,727	6.19		No Hit Found									No Hit Found							
n382L	152317--151967	117	14,139	7.82		No Hit Found									No Hit Found							
n384L	154673--154326	116	12,771	11.00		No Hit Found									No Hit Found							
n385L	155004--154801	68	7,230	4.30		No Hit Found									No Hit Found							
n386R	154967--155173	69	7,840	8.85		No Hit Found									No Hit Found							
n388L	156434--156123	104	11,653	10.78	1	COG5245	DYN1, Dynein, heavy chain [Cytoskeleton].	28.10	2.42E-01	21.05%	29.82%	16--73	360--2617		No Hit Found							
n390L	157181--156954	76	8,100	8.17	1	pfam06537	DUF1111, Protein of unknown function (DUF1111). This family consists of several hypothetical bacterial proteins of unknown function..	27.98	2.57E-01	41.03%	53.85%	16--49	320--359		No Hit Found							
n392R	157779--158039	87	9,900	12.46		No Hit Found									No Hit Found							
n393L	158410--158183	76	8,245	11.57		No Hit Found									No Hit Found							
n394L	158615--158364	84	9,144	10.37	1	COG0786	GltS, Na+/glutamate symporter [Amino acid transport and metabolism]	26.67	6.28E-01	25%	37.50%	24--64	67--107		No Hit Found							
n396L	158975--158733	81	9,394	11.68		No Hit Found									No Hit Found							
n397L	159171--158956	72	8,275	6.37		No Hit Found									No Hit Found							
n398L	159302--159099	68	6,958	8.52		No Hit Found									No Hit Found							
n399L	159761--159558	68	7,627	8.07	1	pfam06388	DUF1075, Protein of unknown function (DUF1075). This family consists of several eukaryotic proteins of unknown function..	27.34	4.70E-01	30.95%	50%	24--64	94--136		No Hit Found							
n401R	160012--160305	98	10,862	7.64	1	pfam01820	Dala_Dala_ligas, D-ala D-ala ligase. This family contains D-alanine -D-alanine ligase enzymes EC:6.3.2.4..	26.78	6.78E-01	40%	55%	27--47	241--261		No Hit Found							
n402R	160600--161598	333	36,802	4.97		No Hit Found								1	AAC96649	a281R	76.64	1.35E-12	41%	50%	66--202	8--164
n405L	162949--162665	95	10,760	5.65	1	cd03457	intradiol_dioxygenase_like, Intradiol dioxygenase supgroup. Intradiol dioxygenases catalyze the critical ring-cleavage step in the conversion of catecholate derivatives to citric acid cycle intermediates. They break the catechol C1-C2 bond and utilize Fe3+, as opposed to the extradiol-cleaving enzymes which break the C2-C3 or C1-C6 bond and utilize Fe2+ and Mn+. The family contains catechol 1,2-dioxygenases and protocatechuate 3,4-dioxygenases. The specific function of this subgroup is unknown..	26.37	8.59E-01	45%	65%	30--50	76--96		No Hit Found							
n406L	163360--163055	102	11,363	4.99		No Hit Found									No Hit Found							
n407L	163503--163267	79	8,749	9.47		No Hit Found									No Hit Found							
n408R	163287--163511	75	8,628	10.76		No Hit Found									No Hit Found							
n409L	163950--163636	105	11,612	7.34		No Hit Found									No Hit Found							
n411R	164005--164289	95	10,540	11.36		No Hit Found									No Hit Found							
n414L	165443--165186	86	9,066	3.14		No Hit Found									No Hit Found							
n415R	165338--165607	90	10,077	11.88		No Hit Found									No Hit Found							
n417L	166338--166054	95	9,771	10.22	1	COG2187	COG2187, Uncharacterized protein conserved in bacteria [Function unknown]	26.04	9.67E-01	28%	48%	4--29	205--230		No Hit Found							
n419L	166920--166654	89	9,679	8.81		No Hit Found									No Hit Found							
n422R	168121--168324	68	7,328	11.34	1	pfam05858	BIV_Env, Bovine immunodeficiency virus surface envelope protein (ENV). The bovine lentivirus also known as the bovine immunodeficiency-like virus (BIV) has conserved and hypervariable regions in the surface envelope gene..	26.97	6.14E-01	41.86%	55.81%	3--42	352--395		No Hit Found							
n423L	168760--168440	107	11,743	11.79		No Hit Found								1	AAC96814	a446R	53.91	1.89E-06	58%	64%	55--105	2--52

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
n424L	168809--168558	84	9,526	8.10		No Hit Found									No Hit Found							
n425L	169450--169097	118	12,850	11.31		No Hit Found									No Hit Found							
n427L	170433--169624	270	26,518	9.10		No Hit Found									No Hit Found							
n429L	171346--171125	74	7,900	9.64		No Hit Found									No Hit Found							
n431R	171213--171431	73	7,867	9.38		No Hit Found									No Hit Found							
n432L	171979--171617	121	13,122	11.05		No Hit Found									No Hit Found							
n435L	173276--172875	134	15,394	10.81		No Hit Found									No Hit Found							
n437L	173224--172985	80	8,827	12.34		No Hit Found									No Hit Found							
n439L	173630--173403	76	8,155	5.61	1	COG3537	COG3537, Putative alpha-1,2-mannosidase [Carbohydrate transport and metabolism].	26.49	8.44E-01	12.31%	40%	3--71	687--752		No Hit Found							
n440R	173774--174076	101	11,850	10.35	1	cd00891	PI3Kc, Phosphoinositide 3-kinase (PI3K), catalytic domain; PI3Ks phosphorylate the 3-position in the inositol ring of inositol phospholipids. PI3Ks play an important role in a variety of fundamental cellular processes, including cell motility, the Ras pathway, vesicle trafficking and secretion, and apoptosis. They can be divided into 3 main classes, defined by their substrate specificity and domain structure	28.25	2.45E-01	42.86%	65.71%	32--73	267--302		No Hit Found							
n442R	173871--174119	83	9,456	11.16		No Hit Found									No Hit Found							
n443R	174210--174413	68	7,234	9.54		No Hit Found									No Hit Found							
n444R	174774--175097	108	12,166	5.57		No Hit Found									No Hit Found							
n448L	176031--175657	125	13,983	8.91		No Hit Found									No Hit Found							
n450L	176682--176227	152	18,022	10.20		No Hit Found									No Hit Found							
n454R	176997--177284	96	10,580	11.53		No Hit Found									No Hit Found							
n455R	177263--177604	114	12,348	9.41		No Hit Found									No Hit Found							
n456R	177321--177596	92	10,687	8.15		No Hit Found									No Hit Found							
n459R	177773--178042	90	9,387	10.46		No Hit Found									No Hit Found							
n461R	179071--179289	73	8,436	8.33		No Hit Found									No Hit Found							
n463L	180136--179663	158	18,723	8.79		No Hit Found									No Hit Found							
n466L	180531--180133	133	15,472	11.21		No Hit Found									No Hit Found							
n467L	180867--180595	91	10,093	10.85		No Hit Found									No Hit Found							
n468L	180983--180705	93	10,367	7.81		No Hit Found									No Hit Found							
n473L	184863--184639	75	8,736	9.49		No Hit Found									No Hit Found							
n474L	185152--184790	121	13,213	8.67		No Hit Found								1	AAC96481	a113L	47.75	1.33E-04	40%	63%	4--58	38--92
n475L	185653--185447	69	7,581	10.71		No Hit Found								1	AAC96483	a115L	41.59	9.54E-03	31%	57%	1--66	1--66
n476L	186801--186568	78	8,093	11.66		No Hit Found									No Hit Found							
n478L	187243--187010	78	8,279	7.86	1	COG0056	AtpA, F0F1-type ATP synthase, alpha subunit [Energy production and conversion].	28.60	1.86E-01	68.75%	68.75%	17--33	227--243		No Hit Found							
n479L	187749--187330	140	15,768	7.18		No Hit Found								1	AAC96804	A435R	43.51	2.48E-03	52%	59%	91--134	27--70
n481L	188396--187947	150	16,689	9.51		No Hit Found									No Hit Found							
n483R	189144--189455	104	10,754	4.77		No Hit Found									No Hit Found							
n486L	189865--189602	88	9,859	8.80	1	COG1152	CdhA, CO dehydrogenase/acetyl-CoA synthase alpha subunit [Energy production and conversion].	28.03	2.96E-01	34.88%	41.86%	46--86	87--130		No Hit Found							
n489L	190736--190500	79	8,425	10.64		No Hit Found									No Hit Found							
n490L	190935--190693	81	9,053	11.36		No Hit Found									No Hit Found							
n492R	191220--191567	116	13,281	11.72		No Hit Found									No Hit Found							
n495L	192686--192390	99	10,647	4.68		No Hit Found									No Hit Found							
n497R	192903--193514	204	21,489	9.04	1	COG5180	PBP1, Protein interacting with poly(A)-binding protein [RNA processing and modification].	28.25	7.62E-01	31.68%	37.62%	10--117	517--618	1	AAM62141	ESX1L	45.44	1.33E-03	34%	45%	45--126	239--305
														2	AAH42633	Extraembryonic, spermatogenesis, homeobox 1-like	45.44	1.33E-03	34%	45%	45--126	239--305
														3	YP_559513	Hypothetical proline rich protein	43.90	3.86E-03	40%	48%	5--81	242--307
														4	NP_054012	collagen-like protein	55.45	1.28E-06	27%	52%	43--126	29--112
														5	XP_780848	PREDICTED: similar to Aryl hydrocarbon receptor precursor (Ah receptor) (AhR)	55.45	1.28E-06	36%	45%	4--108	414--517
														6	ZP_01147863	Initiation factor 2:Small GTP-binding protein domain	71.63	1.73E-11	39%	44%	2--126	201--325
														7	ZP_00675930	hypothetical protein TerzDRAFT_0036	62.39	1.05E-08	32%	38%	9--126	15--132
														8	XP_421396	PREDICTED: similar to adenylosuccinate synthetase:ISOTYPE=muscle	51.22	2.42E-05	45%	55%	4--60	538--597
														9	XP_728216	circumsporozoite protein precursor	48.14	2.05E-04	36%	36%	9--98	207--284
														1	XP_523024	PREDICTED: MAGE-like protein 2	53.53	4.87E-06	36%	40%	12--126	68--184
n498L	193635--193420	72	8,430	11.87		No Hit Found									No Hit Found							
n499L	193921--193679	81	9,051	9.93		No Hit Found									No Hit Found							
n501L	194208--193915	98	11,308	12.59		No Hit Found									No Hit Found							
n502R	194364--194603	80	8,283	11.02		No Hit Found									No Hit Found							
n503L	194581--194372	70	7,641	9.46		No Hit Found									No Hit Found							
n505R	195525--195764	80	8,839	7.64		No Hit Found									No Hit Found							
n508L	196771--196460	104	12,479	9.79		No Hit Found									No Hit Found							
n509R	196583--196789	69	7,232	7.09		No Hit Found									No Hit Found							
n510L	196911--196711	67	6,940	8.70		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
n511L	197126--196905	74	7,847	8.38		No Hit Found									No Hit Found							
n517L	199271--198972	100	10,467	4.33		No Hit Found									No Hit Found							
n519R	200068--200277	70	7,122	8.43		No Hit Found									No Hit Found							
n522R	200943--201335	131	14,081	8.04		No Hit Found									No Hit Found							
n523R	201304--201654	117	12,801	7.50	1	pfam05770	Ins134_P3_kin, Inositol 1, 3, 4-trisphosphate 5/6-kinase. This family consists of several inositol 1, 3, 4-trisphosphate 5/6-kinase proteins. Inositol 1,3,4-trisphosphate is at a branch point in inositol phosphate metabolism. It is dephosphorylated by specific phosphatases to either inositol 3,4-bisphosphate or inositol 1,3-bisphosphate. Alternatively, it is phosphorylated to inositol 1,3,4,6-tetrakisphosphate or inositol 1,3,4,5-tetrakisphosphate by inositol trisphosphate 5/6-kinase	26.81	6.11E-01	31.25%	47.92%	49--104	167--215		No Hit Found							
n526L	202168--201950	73	8,027	12.02		No Hit Found									No Hit Found							
n527R	203872--204213	114	12,777	9.86		No Hit Found								1	AAF75279	bvssal protein Dpfp1 precursor	46.21	3.83E-04	28%	46%	12--112	104--196
n527R														2	AAC39038	foot protein 1 precursor	43.90	1.90E-03	28%	47%	12--107	85--172
n528R	204370--204576	69	7,073	4.47		No Hit Found									No Hit Found							
n529R	205045--205278	78	8,608	4.17	1	COG5565	COG5565, Bacteriophage terminase large (ATPase) subunit and inactivated derivatives [General function prediction only]	26.65	7.33E-01	35.42%	50%	6--54	25--73		No Hit Found							
n530L	205606--205406	67	7,451	10.79		No Hit Found									No Hit Found							
n532R	207882--208118	79	9,212	9.09		No Hit Found									No Hit Found							
n534R	208478--208726	83	9,109	4.81		No Hit Found									No Hit Found							
n536R	209670--209873	68	7,405	8.21		No Hit Found									No Hit Found							
n537R	209848--210138	97	10,416	8.66		No Hit Found									No Hit Found							
n538R	210145--210363	73	7,599	9.57		No Hit Found									No Hit Found							
n539R	212254--212505	84	9,379	4.03		No Hit Found									No Hit Found							
n541R	213261--213467	69	7,977	8.05		No Hit Found									No Hit Found							
n543R	214156--214392	79	8,320	11.19		No Hit Found									No Hit Found							
n544R	214478--214693	72	8,064	7.50		No Hit Found									No Hit Found							
n545R	214639--215043	135	13,213	10.60		No Hit Found									No Hit Found							
n547L	215888--215643	82	9,324	11.44		No Hit Found									No Hit Found							
n550L	217089--216865	75	7,885	8.42		No Hit Found									No Hit Found							
n551R	217223--217438	72	8,041	9.65		No Hit Found									No Hit Found							
n552L	217464--217231	78	8,481	10.89		No Hit Found									No Hit Found							
n553L	218003--217602	134	14,506	12.51		No Hit Found									No Hit Found							
n554L	218703--218047	219	24,149	5.43		No Hit Found									No Hit Found							
n555L	218540--218247	98	10,868	10.35		No Hit Found								1	AAC97012	a586R	73.94	1.76E-12	60%	68%	1--64	12--75
n556R	218441--218740	100	11,054	12.51		No Hit Found									No Hit Found							
n557L	219540--219304	79	8,368	5.77		No Hit Found									No Hit Found							
n558L	219881--219645	79	9,441	11.07		No Hit Found									No Hit Found							
n560L	220389--220153	79	9,027	7.62		No Hit Found									No Hit Found							
n563L	221685--221371	105	11,894	9.32	1	COG2103	COG2103, Predicted sugar phosphate isomerase [General function prediction only]	27.91	3.25E-01	26.98%	49.21%	2--73	168--231		No Hit Found							
n566R	223073--223798	242	19,621	3.49		No Hit Found									No Hit Found							
n568R	223991--224263	91	10,568	12.42		No Hit Found									No Hit Found							
n571R	225336--225857	174	18,962	7.54		No Hit Found									No Hit Found							
n573R	225893--226252	120	12,570	5.42		No Hit Found									No Hit Found							
n576R	227024--227263	80	8,990	9.80	1	pfam06644	ATP11, ATP11 protein. This family consists of several eukaryotic ATP11 proteins. In Saccharomyces cerevisiae, expression of functional F1-ATPase requires two proteins encoded by the ATP11 and ATP12 genes..	27.69	3.83E-01	29.31%	44.83%	9--69	111--169		No Hit Found							
n577R	227464--227685	74	8,477	10.75		No Hit Found									No Hit Found							
n580R	228151--228354	68	7,658	7.82	1	pfam00871	Acetate_kinase, Acetokinase family. This family includes acetate kinase, butyrate kinase and 2-methylpropanoate kinase..	29.88	7.95E-02	31.11%	42.22%	3--48	54--99		No Hit Found							
n581R	228182--228445	88	9,606	8.74	1	smart00079	PBPe, Eukaryotic homologues of bacterial periplasmic substrate binding proteins. Prokaryotic homologues are represented by a separate alignment: PBPb .	26.90	5.42E-01	16.25%	32.50%	3--83	117--197		No Hit Found							
n583R	228518--228769	84	9,337	6.22		No Hit Found									No Hit Found							
n585L	229654--229418	79	8,938	8.37		No Hit Found									No Hit Found							
n587R	229496--229759	88	10,221	8.94		No Hit Found									No Hit Found							
n588R	229653--229970	106	12,570	9.92		No Hit Found									No Hit Found							
n590R	230476--230724	83	9,692	9.50		No Hit Found									No Hit Found							
n591R	230489--230740	84	9,080	10.35		No Hit Found									No Hit Found							
n592L	230726--230514	71	8,290	11.53		No Hit Found									No Hit Found							
n594L	231556--231338	73	8,068	12.05		No Hit Found									No Hit Found							
n595L	231720--231445	92	9,681	8.85	1	COG1928	PMT1, Dolichyl-phosphate-mannose--protein O-mannosyl transferase [Posttranslational modification, protein turnover, chaperones].	26.92	5.49E-01	22.73%	43.18%	12--56	603--647	1	AAC96757	a389R	43.13	3.27E-03	54%	58%	6--50	21--64

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
n596L	232266--231946	107	10,895	11.17		No Hit Found									No Hit Found							
n597L	233110--232892	73	8,177	6.37		No Hit Found									No Hit Found							
n599L	233832--233371	154	14,660	4.19	1	cd00755	YgdL_like, Family of activating enzymes (E1) of ubiquitin-like proteins related to the E.coli hypothetical protein ygdL. The common reaction mechanism catalyzed by E1-like enzymes begins with a nucleophilic attack of the C-terminal carboxylate of the ubiquitin-like substrate, 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 the substrate. The exact function of this family is unknown..	30.55	1.02E-01	57.14%	76.19%	43--64	15--36		No Hit Found							
n601R	234394--234609	72	7,848	11.26		No Hit Found									No Hit Found							
n602R	234606--234836	77	8,454	9.03		No Hit Found									No Hit Found							
n605L	235925--235695	77	7,950	11.07		No Hit Found									No Hit Found							
n607R	235999--236385	129	14,524	11.29	1	cd00228	eu-GS, Eukaryotic Glutathione Synthetase (eu-GS); catalyses the production of glutathione from gamma-glutamylcysteine and glycine in an ATP-dependent manner. Belongs to the ATP-grasp superfamily	29.09	1.88E-01	19.10%	33.71%	28--117	102--191		No Hit Found							
n608R	236189--236437	83	8,954	9.61		No Hit Found									No Hit Found							
n610L	236971--236579	131	14,448	12.15		No Hit Found									No Hit Found							
n611R	237127--237435	103	11,228	12.21		No Hit Found									No Hit Found							
n612L	238160--237669	164	18,422	10.62		No Hit Found									No Hit Found							
n615L	239103--238894	70	8,128	10.59	1	pfam01007	IRK, Inward rectifier potassium channel..	26.35	7.50E-01	26.09%	56.52%	21--63	18--64		No Hit Found							
n617R	239996--240343	116	13,011	10.01		No Hit Found									No Hit Found							
n620R	241636--242085	150	13,320	9.42		No Hit Found									No Hit Found							
n622L	242741--242520	74	8,039	9.13	1	pfam00342	PGI, Phosphoglucose isomerase. Phosphoglucose isomerase catalyses the interconversion of glucose-6-phosphate and fructose-6-phosphate..	28.40	2.33E-01	41.18%	70.59%	20--37	330--347		No Hit Found							
n623R	242710--243294	195	21,680	8.34		No Hit Found									No Hit Found							
n625R	243119--243343	75	7,589	12.50		No Hit Found									No Hit Found							
n626L	243665--243447	73	7,879	12.60		No Hit Found									No Hit Found							
n627R	243530--243745	72	7,962	10.53		No Hit Found									No Hit Found							
n630L	245168--244944	75	8,438	11.43		No Hit Found									No Hit Found							
n632R	245262--245510	83	9,098	10.92		No Hit Found									No Hit Found							
n634R	245562--245975	138	14,923	11.16		No Hit Found									No Hit Found							
n636R	246306--246581	92	11,304	11.39		No Hit Found									No Hit Found							
n637R	246723--246929	69	7,667	8.20	1	pfam06035	DUF920, Bacterial protein of unknown function (DUF920). This family consists of several hypothetical bacterial proteins of unknown function	26.45	8.80E-01	32.35%	55.88%	9--43	94--128		No Hit Found							
n638R	246841--247191	117	13,109	10.69		No Hit Found								1	AAC96649	a281R	46.60	3.00E-04	56%	58%	68--112	63--108
n641L	248204--247962	81	9,072	7.77		No Hit Found									No Hit Found							
n644R	249038--249238	67	7,657	6.61		No Hit Found									No Hit Found							
n645L	249532--249047	162	16,967	4.31		No Hit Found								1	XP_639322	hypothetical protein DDB0185331	45.05	9.32E-04	37%	49%	2--75	500--582
n646L	249760--249536	75	7,613	8.69		No Hit Found									No Hit Found							
n647L	250028--249546	161	18,759	7.87	1	pfam05295	Luciferase, Luciferase. This family consists of dinoflagellate luciferase and luciferin binding proteins. Luciferase is involved in catalysing the light emitting reaction in bioluminescence and luciferin binding protein (LBP) is known to bind to luciferin (the substrate for luciferase) to stop it reacting with the enzyme and therefore switching off the bioluminescence function. The expression of these two proteins is controlled by a circadian clock at the translational level, with synthesis and degradation occurring on a daily basis	29.35	2.35E-01	35%	47.50%	15--55	213--253		No Hit Found							
n648L	250309--250070	80	7,863	3.87		No Hit Found									No Hit Found							
n650L	251421--251008	138	15,341	10.07		No Hit Found									No Hit Found							
n652R	251602--252162	187	20,684	10.58		No Hit Found								1	XP_455758	unnamed protein product	94.74	1.55E-18	44%	53%	1--133	1--133
														2	AAL79317	unknown	52.37	8.85E-06	38%	51%	64--133	2--71
														3	NP_851581	CPXV051A protein	51.99	1.16E-05	55%	63%	82--128	31--77
														4	AAW23723	ribonucleoside-diphosphate reductase	50.06	4.39E-05	55%	65%	82--124	31--73
														5	AAA48019	putative F ORF B	50.06	4.39E-05	55%	65%	82--124	31--73
														6	NP_477712	wsv190	44.67	1.84E-03	44%	49%	89--149	1--61
n653L	251970--251731	80	8,264	10.71	1	COG4861	COG4861, Uncharacterized protein conserved in bacteria [Function unknown].	26.95	6.22E-01	32%	46%	13--58	202--252		No Hit Found							
n654L	252163--251894	90	9,657	11.99		No Hit Found									No Hit Found							
n656R	252676--252882	69	7,605	11.17		No Hit Found									No Hit Found							
n657L	253344--253135	70	7,762	12.60		No Hit Found									No Hit Found							
n659R	253328--253657	110	11,457	6.90		No Hit Found									No Hit Found							
n660R	253464--253670	69	7,468	12.41		No Hit Found									No Hit Found							
n661L	254022--253768	85	9,086	10.34		No Hit Found									No Hit Found							
n663R	254151--254375	75	8,600	8.81		No Hit Found									No Hit Found							
n664R	254161--254403	81	8,720	4.65		No Hit Found									No Hit Found							
n666L	256054--255176	293	32,701	8.34		No Hit Found									No Hit Found							
n667L	256380--255958	141	15,902	10.11		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
n669L	256535--256254	94	10,574	12.28		No Hit Found								1	AAC96833	a466L	67.40	1.61E-10	52%	58%	24--91	19--86
n671R	256475--256789	105	11,380	9.09		No Hit Found									No Hit Found							
n673R	257404--257682	93	10,454	8.80	1	COG4989	COG4989, Predicted oxidoreductase [General function prediction only].	27.55	4.18E-01	36.59%	53.66%	31--77	159--200		No Hit Found							
n675L	258706--258173	178	19,923	9.96	1	COG5421	COG5421, Transposase [DNA replication, recombination, and repair].	28.36	5.56E-01	21.77%	38.71%	44--164	340--464	1	AAC96828	a460R	53.91	2.67E-06	30%	60%	74--149	2--77
n676L	258524--258321	68	7,214	6.32		No Hit Found									No Hit Found							
n677L	259395--258760	212	23,333	10.44		No Hit Found								1	AAC96827	a459R	105.92	9.12E-22	72%	78%	131--205	1--75
														2	AAC96825	a457R	75.87	1.01E-12	60%	64%	14--88	2--76
n678R	259642--259980	113	12,672	10.73	1	COG2265	TrmA, SAM-dependent methyltransferases related to tRNA (uracil-5-)-methyltransferase [Translation, ribosomal structure and biogenesis].	27.60	3.90E-01	27.27%	40.91%	66--104	35--79		No Hit Found							
n680L	260163--259708	152	15,632	8.04		No Hit Found								1	AAC96823	a455R	88.20	8.89E-17	47%	54%	27--150	1--122
n681L	260347--260117	77	8,329	7.43		No Hit Found									No Hit Found							
n683R	260700--260933	78	8,786	11.71		No Hit Found									No Hit Found							
n686L	262444--262172	91	10,234	12.61		No Hit Found									No Hit Found							
n688R	262594--262818	75	8,055	9.38		No Hit Found									No Hit Found							
n691R	263834--264088	85	10,032	11.52		No Hit Found									No Hit Found							
n692L	264522--264109	138	15,532	6.65	1	cd03285	ABC_MSH2_euk, MutS2 homolog in eukaryotes. The MutS protein initiates DNA mismatch repair by recognizing mispaired and unpaired bases embedded in duplex DNA and activating endo- and exonucleases to remove the mismatch. Members of the MutS family possess	29.49	1.36E-01	28.26%	45.65%	47--88	129--175		No Hit Found							
n694L	265622--265275	116	12,632	12.00		No Hit Found									No Hit Found							
n695L	266256--265405	284	31,371	9.44		No Hit Found								1	AAC96874	a507R	65.47	2.40E-09	34%	54%	1--132	53--186
n697L	265973--265692	94	10,302	10.78		No Hit Found									No Hit Found							
n698L	266264--266040	75	7,881	8.43		No Hit Found									No Hit Found							
n699R	266255--266467	71	7,863	12.20		No Hit Found									No Hit Found							
n700L	266811--266269	181	19,598	7.33		No Hit Found									No Hit Found							
n701L	266675--266289	129	14,361	12.12		No Hit Found								1	AAC96873	a506R	82.03	6.36E-15	60%	68%	38--107	1--70
n702R	266495--266899	135	14,873	11.08		No Hit Found									No Hit Found							
n704L	267554--267252	101	11,569	9.40	1	COG4297	COG4297, Uncharacterized protein containing double-stranded beta helix domain [Function unknown]	27.25	4.34E-01	34.78%	56.52%	5--30	40--63		No Hit Found							
n707L	268426--267974	151	16,868	9.54		No Hit Found									No Hit Found							
n708R	268582--268962	127	14,170	11.62	1	pfam04443	LuxE, Acyl-protein synthetase, LuxE. LuxE is an acyl-protein synthetase found in bioluminescent bacteria. LuxE catalyses the formation of an acyl-protein thioester from a fatty acid and a protein. This is the second step in the bioluminescent fatty acid reduction system, which converts tetradecanoic acid to the aldehyde substrate of the luciferase-catalysed bioluminescence reaction. A conserved cysteine found at position 364 in Photobacterium phosphoreum LuxE is thought to be acylated during the transfer of the acyl group from the synthetase subunit to the reductase. The carboxyl terminal of the synthetase is though to act as a flexible arm to transfer acyl groups between the sites of activation and reduction. This family also includes Vibrio cholerae RBFN protein, which is involved in the biosynthesis of the O-antigen component 3-deoxy-l-glycero-tetronic acid	29.15	1.72E-01	41.94%	61.29%	51--82	89--120		No Hit Found							
n709L	269107--268892	72	8,111	9.43		No Hit Found									No Hit Found							
n711R	269902--270114	71	7,739	9.95		No Hit Found									No Hit Found							
n713L	270756--270469	96	10,400	12.51		No Hit Found									No Hit Found							
n714R	271383--271748	122	13,997	8.35	1	COG3898	COG3898, Uncharacterized membrane-bound protein [Function unknown].	32.34	1.62E-02	29.27%	41.46%	75--116	343--384	1	AAC96593	a225L	49.29	4.54E-05	39%	67%	68--120	2--54
n716L	272020--271778	81	8,755	8.46		No Hit Found									No Hit Found							
n717R	272257--272487	77	8,125	11.61		No Hit Found								1	AAC97060	a223aL	61.23	1.18E-08	69%	76%	2--40	19--57
n718R	272484--272996	171	18,788	11.60		No Hit Found								1	T17711	hypothetical protein a221L - Chlorella virus PBCV-1	48.52	1.00E-04	37%	62%	66--124	1--57
n721R	273968--274384	139	15,141	3.06		No Hit Found								1	XP_521636	PREDICTED: similar to putative protein family member (XC187)	44.67	1.12E-03	33%	54%	38--109	169--240
														2	AAO81743	tape measure protein, putative	43.90	1.91E-03	22%	39%	6--108	665--765
														3	ZP_00382165	COG5280: Phage-related minor tail protein	42.74	4.25E-03	25%	37%	20--107	372--466
														4	NP_112708	TMP	44.67	1.12E-03	25%	38%	25--108	671--754
														5	ZP_01187196	TMP	41.97	7.24E-03	21%	43%	20--107	389--477
														6	NP_044960	unknown	43.51	2.49E-03	28%	47%	33--108	575--646
														7	NP_108725	unknown	43.90	1.91E-03	22%	33%	18--103	686--793
														8	NP_047129	putative tail tape measure protein	49.29	4.54E-05	28%	45%	20--125	570--673
														9	NP_690687	structural protein	42.74	4.25E-03	27%	39%	17--107	353--455
														1	AAU16851	possible prophage membrane protein	41.59	9.46E-03	23%	38%	25--105	175--266
n722R	274589--274804	72	7,908	12.29		No Hit Found									No Hit Found							
n723R	275117--275398	94	9,882	10.19		No Hit Found								1	AAC97027	a616R	51.99	6.99E-06	46%	50%	1--78	1--79
n723R														2	AAC96648	a280R	43.13	3.25E-03	42%	50%	1--56	10--64
n726R	276504--276869	122	13,118	10.72		No Hit Found									No Hit Found							
n728R	277092--277421	110	11,197	4.21		No Hit Found									No Hit Found							
n729L	277491--277222	90	10,724	12.36		No Hit Found									No Hit Found							
n730L	277775--277485	97	11,052	8.81	1	pfam05140	ResB, ResB-like family. This family includes both ResB and cytochrome c biogenesis proteins. Mutations in ResB indicate that they are essential for growth. ResB is predicted to be a transmembrane protein..	29.50	8.67E-02	44.44%	59.26%	30--66	54--81		No Hit Found							
n734L	278773--278516	86	9,240	11.75		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
n735L	278828--278544	95	10,729	5.59		No Hit Found									No Hit Found							
n736L	279490--279104	129	14,040	11.77		No Hit Found									No Hit Found							
n738R	281695--282087	131	14,276	8.71		No Hit Found									No Hit Found							
n739R	282710--282979	90	10,204	11.61		No Hit Found									No Hit Found							
n740R	282862--283173	104	11,383	6.99	1	cd03342	TCP1_zeta, TCP-1 (CTT or eukaryotic type II) chaperonin family, zeta subunit. Chaperonins are involved in productive folding of proteins. They share a common general morphology, a double toroid of 2 stacked rings. In contrast to bacterial group I chaperonins (GroEL), each ring of the eukaryotic cytosolic chaperonin (CTT) consists of eight different, but homologous subunits. Their common function is to sequester nonnative proteins inside their central cavity and promote folding by using energy derived from ATP hydrolysis. The best studied in vivo substrates of CTT are actin and tubulin	29.40	1.08E-01	29.03%	56.45%	11--83	204--266		No Hit Found							
n743R	283411--283962	184	20,550	11.40	1	pfam04450	BSP, Plant Basic Secretory Protein. These basic secretory proteins (BSPs) are believed to be part of the plants defence mechanism against pathogens	31.87	5.53E-02	38.10%	46.03%	49--107	40--103		No Hit Found							
n744L	284544--284281	88	9,347	9.33		No Hit Found									No Hit Found							
n749L	286722--286522	67	6,933	12.37		No Hit Found									No Hit Found							
n750R	286601--287098	166	18,473	8.98		No Hit Found								1	AAC97055	a684R	43.90	2.22E-03	29%	56%	34--122	10--95
n752L	287901--287650	84	10,042	6.23	1	pfam06309	Torsin, Torsin. This family consists of several eukaryotic torsin proteins. Torsion dystonia is an autosomal dominant movement disorder characterised by involuntary, repetitive muscle contractions and twisted postures. The most severe early-onset form of	26.37	7.56E-01	38.78%	48.98%	26--73	172--221		No Hit Found							
n753R	288315--288569	85	9,599	6.39		No Hit Found									No Hit Found							
n757R	289999--290352	118	12,204	8.44		No Hit Found									No Hit Found							
n758L	291108--290854	85	9,495	12.52		No Hit Found									No Hit Found							
n760R	291494--291940	149	16,960	10.75	1	COG5191	COG5191, Uncharacterized conserved protein, contains HAT (Half-A-TPR) repeat [General function prediction only]	28.49	3.26E-01	28.85%	48.08%	69--121	92--144		No Hit Found							
n761L	291747--291544	68	7,483	5.84		No Hit Found									No Hit Found							
n765L	292953--292696	86	9,988	9.06		No Hit Found									No Hit Found							
n767R	293214--294182	323	35,642	8.13	1	pfam07414	YbtU, Yersiniabactin synthetase thiazolinyl reductase component YbtU. This family represents the thiazolinyl reductase component YbtU (approximately 350 residues long) of the bacterial four-protein yersiniabactin synthetase complex. Yersiniabactin is a virulence factor secreted by Yersinia pestis in iron-deficient microenvironments, in order to scavenge ferric ions	36.07	5.73E-03	24.09%	42.34%	102--237	60--197		No Hit Found							
n768R	293626--293871	82	8,856	11.97		No Hit Found									No Hit Found							
n769L	294064--293777	96	10,625	12.84		No Hit Found									No Hit Found							
n770R	294312--295055	248	25,829	7.87		No Hit Found								1	AAC97030	a631L	53.91	5.61E-06	31%	56%	64--149	4--89
n771R	295036--295293	86	9,587	11.06		No Hit Found									No Hit Found							
n772R	295428--295910	161	17,960	10.36	1	cd01675	RNR_3, RNR, class III. 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	27.13	9.60E-01	24%	32%	104--156	244--294		No Hit Found							
n774L	295777--295526	84	9,286	11.71		No Hit Found									No Hit Found							
n775R	296482--296718	79	8,429	11.76		No Hit Found									No Hit Found							
n776L	296859--296518	114	12,681	10.68	1	pfam02283	CobU, Cobinamide kinase / cobinamide phosphate guanyltransferase. This family is composed of a group of bifunctional cobalamin biosynthesis enzymes which display cobinamide kinase and cobinamide phosphate guanyltransferase activity. The crystal structure of the enzyme reveals the molecule to be a trimer with a propeller-like shape	26.86	6.36E-01	46.67%	53.33%	94--109	47--62		No Hit Found							
n778R	297015--297251	79	9,160	10.62		No Hit Found									No Hit Found							
n780L	297744--297316	143	15,332	11.47		No Hit Found									No Hit Found							
n781L	298410--297994	139	14,852	7.67		No Hit Found									No Hit Found							
n782L	298692--298420	91	8,959	10.53		No Hit Found									No Hit Found							
n783R	299574--299813	80	9,184	9.02	1	pfam04670	Gtr1_RagA, Gtr1/RagA G protein conserved region. GTR1 was first identified in S. cerevisiae as a suppressor of a mutation in RCC1. Biochemical analysis revealed that Gtr1 is in fact a G protein of the Ras family. The RagA/B proteins are the human homologues of Gtr1. Included in this family is the human Rag C, a novel protein that has been shown to interact with RagA/B	28.76	1.56E-01	19.15%	34.04%	4--51	45--92		No Hit Found							
n785L	300510--299890	207	23,524	7.39	1	pfam01229	Glyco_hydro_39, Glycosyl hydrolases family 39..	29.20	3.97E-01	33.33%	43.59%	30--68	42--81		No Hit Found							
n786L	300503--300291	71	7,690	12.41		No Hit Found									No Hit Found							
n788L	300956--300618	113	12,652	8.44		No Hit Found									No Hit Found							
n789L	301718--301287	144	16,877	10.68	1	pfam05208	ALG3, ALG3 protein. The formation of N-glycosidic linkages of glycoproteins involves the ordered assembly of the common Glc3Man9GlcNAc2 core-oligosaccharide on the lipid carrier dolichyl pyrophosphate. Whereas early mannosylation steps occur on the cytoplasmic side of the endoplasmic reticulum with GDP-Man as donor, the final reactions from Man5GlcNAc2-PP-Dol to Man9GlcNAc2-PP-Dol on the lumenal side use Dol-P-Man. ALG3 gene encodes the Dol-P-Man:Man5GlcNAc2-PP-Dol mannosyltransferase	27.90	5.20E-01	24.14%	37.93%	73--126	258--316		No Hit Found							
n790R	301659--301874	72	7,882	12.30		No Hit Found									No Hit Found							
n793L	303314--303027	96	10,559	8.04		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
n796R	304082--304411	110	12,088	7.95	1	cd02930	DCR_FMN, 2,4-dienoyl-CoA reductase (DCR) FMN-binding domain. DCR in E. coli is an iron-sulfur flavoenzyme which contains FMN, FAD and a 4Fe-4S cluster. It is also a monomer, unlike that of its eukaryotic counterparts which form homotetramers and lack the	27.14	4.47E-01	42.59%	48.15%	13--70	210--264		No Hit Found						
n797L	304802--304476	109	12,559	11.81		No Hit Found									No Hit Found						
n798R	304979--305257	93	7,824	4.11		No Hit Found									No Hit Found						
n800L	306030--305800	77	8,307	10.79		No Hit Found									No Hit Found						
n801L	306314--306111	68	7,682	12.28		No Hit Found									No Hit Found						
n804R	306826--307278	151	17,467	9.95		No Hit Found									No Hit Found						
n807R	308573--308776	68	7,511	11.18		No Hit Found									No Hit Found						
n808R	308625--308858	78	8,261	11.42		No Hit Found									No Hit Found						
n810L	309495--309262	78	8,682	6.78		No Hit Found									No Hit Found						
n812R	309540--309782	81	9,002	11.92		No Hit Found									No Hit Found						
n813L	309880--309647	78	9,038	9.01		No Hit Found									No Hit Found						
n816R	310399--310767	123	13,960	7.88		No Hit Found									No Hit Found						
n817L	310778--310551	76	8,395	11.93		No Hit Found									No Hit Found						
n818R	310906--311193	96	11,014	7.30		No Hit Found									No Hit Found						
n821L	312773--312327	149	13,536	8.84		No Hit Found									No Hit Found						
n822L	313077--312655	141	16,333	8.64		No Hit Found									No Hit Found						
n823L	313901--313593	103	11,353	10.38		No Hit Found								1	AAC96687	a319L	63.16	3.04E-09	46%	52%	29--101	1--82
n824L	314259--313996	88	9,920	4.52	1	COG1273	COG1273, Uncharacterized conserved protein [Function unknown].	28.66	1.73E-01	29.87%	45.45%	10--86	94--171		No Hit Found						
n825R	314115--314369	85	9,254	12.08		No Hit Found									No Hit Found						
n826L	314782--314213	190	20,410	7.54		No Hit Found									No Hit Found						
n828L	314637--314419	73	8,174	11.30		No Hit Found									No Hit Found						
n829R	314505--314741	79	8,781	12.98		No Hit Found									No Hit Found						
n831L	315345--315076	90	9,770	8.25		No Hit Found									No Hit Found						
n832R	315194--315499	102	11,639	11.14		No Hit Found									No Hit Found						
n833R	315286--315552	89	10,148	8.66		No Hit Found									No Hit Found						
n834R	315701--316210	170	18,628	11.47		No Hit Found									No Hit Found						
n835L	316211--315870	114	12,722	8.67		No Hit Found									No Hit Found						
n837R	316303--316662	120	12,809	6.20		No Hit Found									No Hit Found						
n838R	316527--316988	154	17,866	12.28		No Hit Found									No Hit Found						
n839L	316756--316544	71	7,746	12.21		No Hit Found									No Hit Found						
n841L	317348--316923	142	15,562	9.63		No Hit Found									No Hit Found						
n842L	317356--317156	67	7,917	11.92		No Hit Found									No Hit Found						
n843R	317588--317944	119	13,938	12.62		No Hit Found									No Hit Found						
n844L	317895--317644	84	9,447	11.09	1	cd00170	SEC14, Sec14p-like lipid-binding domain. Found in secretory proteins, such as S. cerevisiae phosphatidylinositol transfer protein (Sec14p), and in lipid regulated proteins such as RhoGAPs, RhoGEFs and neurofibromin (NF1). SEC14 domain of Dbl is known to associate with G protein beta/gamma subunits	26.14	9.94E-01	28%	52%	57--82	90--115		No Hit Found						
n846L	318140--317892	83	8,232	12.53		No Hit Found									No Hit Found						
n848L	320123--319902	74	8,373	5.91		No Hit Found									No Hit Found						
n849R	320241--320765	175	20,253	9.57	1	cd03296	ABC_CysA_sulfate_importer, Part of the ABC transporter complex cysAWTP involved in sulfate import. Responsible for energy coupling to the transport system. The complex is composed of two ATP-binding proteins (cysA), two transmembrane proteins (cysT and cysW) and a solute-binding protein (cysP). 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..	28.69	3.66E-01	50%	65%	41--61	68--88		No Hit Found						