

A Study of Functional Genomics for Unknown Proteins in *Chlamydomonas reinhardtii*

Manisha D. Giripunje

Department of Environmental
Science, Sevadal Mahila
Mahavidyalaya, Nagpur- 440015,
Maharashtra, India.

Abhay B. Fulke

CSIR-National Environmental
Engineering Research Institute,
Nagpur-440020,
Maharashtra, India.

Pravin U. Meshram

Department of Environmental
Science, Sevadal Mahila
Mahavidyalaya, Nagpur- 440015,
Maharashtra, India.

ABSTRACT

Chlamydomonas reinhardtii is a unicellular green alga, which has been used as a reference organism for identifying proteins. Five hundred hypothetical proteins in *Chlamydomonas reinhardtii* have been sequenced for knowing functions of the proteins in their families. Functions of Five hundred hypothetical proteins in *Chlamydomonas reinhardtii* were predicted using bioinformatics web tools. The web tools like CDD-BLAST, INTERPROSCAN and PFAM were used for the prediction of functions in the proteins by searching protein databases for the presence of conserved domains. The current study was useful in better knowing the functions of hypothetical proteins in *Chlamydomonas reinhardtii* and the various regulatory mechanisms, which interact to yield a final output from the system.

Keywords

Chlamydomonas reinhardtii, Bioinformatics Web Tools, hypothetical proteins, Conserved Domains

1. INTRODUCTION

Chlamydomonas reinhardtii is the unicellular biflagellate green alga [1]. *Chlamydomonas reinhardtii* is known to us as model organism because of its physiological, metabolic and genetic characteristics that expresses the structure, function and regulation of the photosynthetic apparatus [2]. *C. reinhardtii* has been used for genome-wide studies of cell motility, photosynthesis [3], phototaxis, cell wall biogenesis and additional essential cellular actions. Bioinformatics analysis of *Chlamydomonas* has been highlighted the photosynthetic function, in which set of proteins indicated the “GreenCut” [4]. *Chlamydomonas reinhardtii* also develops in the dark on an organic carbon source while maintaining a functional photosynthetic apparatus. Many *Chlamydomonas* genes can be traced to the green plant or plant-animal common ancestor by comparative genomic analysis [5]. The *Chlamydomonas* nuclear genome includes seventeen linkage groups presumably corresponding to seventeen chromosomes, consistent with electron microscopy of meiotic synaptonemal complexes [6]. The *Chlamydomonas* genome has about uniform densities of genes, simple sequence repeats and transposable elements. Most *Chlamydomonas* proteins exhibit slightly more similarity to Arabidopsis than to human proteins [7]. Many hypothetical proteins in *Chlamydomonas reinhardtii* are not functionally annotated. The unknown proteins functions estimate with database and algorithm of bioinformatics tools. The early study of proteins functions in *Chlamydomonas reinhardtii* was annotated. However, the database was increased after the improvement in the gene

models. Therefore, about one-fourth portion of the protein coding gene models were correctly annotated [8].

The current study is an effort to know and assemble the information associated with functional annotations of unknown proteins in *Chlamydomonas reinhardtii*. Functions and classification of target protein in family can explore through homologous/orthologous sequence in biological sequence databases for the target sequences using computational tools like CDD-BLAST, INTERPROSCAN and PFAM [9, 10].

2. METHODOLOGY

2.1 Sequence analysis of *Chlamydomonas reinhardtii*

The original database of the genomic sequences of *Chlamydomonas reinhardtii* was generated by Joint Genome Institute (JGI). Kyoto Encyclopedia of Genes and Genomes (KEGG) maintained the database resource of genome. Sequences of *Chlamydomonas reinhardtii* hypothetical proteins were recollected from the KEGG database (<http://www.genome.jp/kegg/>) [11] (using the primary accession number T01039 and the entry name cre, CHLRE, 3055.)

2.2 Genomics analysis of *Chlamydomonas reinhardtii*

The workflow of genomics analysis was described and the unknown proteins from *Chlamydomonas reinhardtii* were screened for the use of bioinformatics tools. First, Conserved Domain Database CDD-BLAST tool search was executed for *Chlamydomonas reinhardtii* (<http://www.ncbi.nlm.nih.gov/BLAST/>) that presented as an annotation of protein domains [12, 13 and 14] NCBI-CDD is a protein annotation resource that consists of a collection of well-annotated multiple sequence alignment models for ancient domains and full-length proteins [15]. Then, search was performed by InterProScan tool, which combines protein signature recognition methods native to the InterPro member databases into one resource (<http://www.ebi.ac.uk/interpro>) [16]. The prediction of uncharacterized proteins was carried out by Pfam v27.0 (<http://www.pfam.sanger.ac.uk/>) which includes their annotations and multiple sequence alignments generated using hidden Markov models [17,18]. Unknown proteins studied with bioinformatics tools, resulted the conserved domains in hypothetical sequences. The domains in

hypothetical sequences are useful in advanced classification of proteins into their individual family. The present study focused on sequence genomics analysis of *Chlamydomonas reinhardtii* hypothetical proteins as summarized in Figure 1.

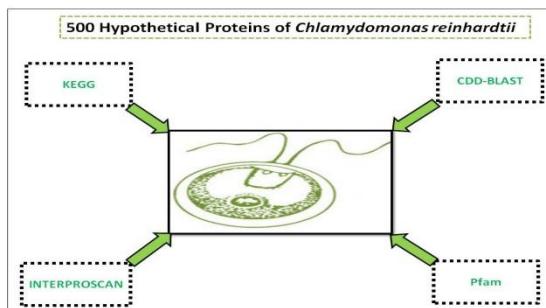


Table 1: Identified hypothetical proteins of *Chlamydomonas reinhardtii* using genomics.

KEGG	CDD-BLAST	INTERPROSCAN	Pfam
cre_10008	Leucine-rich repeats (LRRs), ribonuclease inhibitor (RNI-like subfamily)	No	Leucine-rich repeats
cre_100494	Helicase superfamily c-terminal domain	Signal-peptide	DEAD/DEAH box helicase, Helicase conserved C-terminal
cre_100535	cGMP-specific phosphodiesterases, adenylyl and guanylyl cyclases, phytochromes, FhlA and NifA	Signal-peptide	GAF domain
cre_100583	Catalytic domain of Protein Kinases; Protein Kinases (PKs)	Signal-peptide	Protein kinase domain
cre_100620	The Ady2 protein in is required for acetate in <i>Saccharomyces cerevisiae</i> and is probably an acetate transporter.	No	GPR1/FUN34/yaaH family
cre_100676	Catalytic domain of Protein Kinases	Signal-peptide	Protein kinase domain
cre_100718	Catalytic domain of Cyclin-Dependent protein Kinase Like Serine/Threonine Kinases	Signal-peptide	Protein kinase domain
cre_100723	Catalytic domain of Cyclin-Dependent protein Kinase Like Serine/Threonine Kinases	Signal-peptide	Protein kinase domain
cre_100734	Thiols, C-terminal domain	Signal-peptide	Protein kinase domain
cre_100734	Serine/threonine phosphatases, catalytic domain	Signal-peptide	Protein phosphatase 2C
cre_100930	ATP-binding cassette domain of a mitochondrial protein MTABC3 and related proteins	Signal-peptide	ABC transporter
cre_101034	N227-like protein	No	N227-like protein
cre_101073	Catalytic domain of Cyclin-Dependent protein Kinase Like Serine/Threonine Kinases; Serine/Threonine Kinases (STKs)	No	Protein kinase domain
cre_101226	Oxidoreductase family, NAD-binding Rossmann fold	Signal-peptide	Oxidoreductase family, NAD-binding Rossmann fold
cre_101248	SpoU tRNA Methylase family; This family of proteins probably use S-AdoMet.	No	SpoU tRNA Methylase family
cre_101262	Endonuclease	Signal-peptide	Endonuclease
cre_101276	Endonuclease_V	Signal-peptide	Endonuclease_V
cre_101285	Iron-sulfur cluster assembly accessory protein	No	Iron-sulfur cluster biosynthesis
cre_101353	Catalytic domain of the Protein Serine/Threonine Kinase	Signal-peptide	Protein kinase domain
cre_101516	Catalytic domain of the Serine/Threonine Kinase, Mitogen-Activated Protein Kinase	Signal-peptide	Protein kinase domain
cre_101561	Catalytic domain of Protein Kinases	No	Protein kinase domain
cre_101592	Complex I protein (CYTB family)	Signal-peptide	Complex I protein (CYTB family)
cre_101593	ATP-binding cassette domain of elongation factor 3.	Signal-peptide	ABC transporter
cre_101629	Fe/S cluster binding domain of Protein Kinases	Signal-peptide	Protein kinase domain
cre_101710	The Perl_NiH superfamily, cobW C-terminal domain, Putative GTPases	IGFBP_N_1	No
cre_101710	Ferritin-like superfamily	Signal-peptide	Protein of unknown function
cre_101843	Helicase superfamily c-terminal domain	No	Helicase conserved C-terminal domain
cre_101874	Cytochrome b5-like Heme/Steroid binding domain	No	Cytochrome b5-like Heme/Steroid binding domain
cre_101882	E2F-associated phosphoprotein	Signal-peptide	E2F-associated phosphoprotein
cre_101941	DEDD8 3'-5' exonuclease domain of <i>Caenorhabditis elegans</i> ERI-1, human 3' exonuclease	Signal-peptide	Exonuclease
cre_102026	Tynovirus 45/70Kd protein	AGO1/1_I	No
cre_102084	ankyrin repeats; S-adenosylmethionine-dependent methyltransferases	IGFBP_N_1	Ankyrin repeats (3 copies)
cre_102339	SacI homology domain	Signal-peptide	SacI homology domain
cre_102417	No	Signal-peptide	No
cre_102527	Squalene/phytoene synthase	Signal-peptide	Squalene/phytoene synthase
cre_102648	Catalytic domain of Cyclin-Dependent protein Kinase Like Serine/Threonine Kinases.	Signal-peptide	Protein kinase domain
cre_102723	CESA_CelA	Signal-peptide	Glycosyl transferase family group 2
cre_102759	Catalytic domain of the Protein Serine/Threonine Kinase	No	Protein kinase domain
cre_102784	Cinnamyl alcohol dehydrogenases (CAD)	Signal-peptide	Alcohol dehydrogenase
cre_102892	DHHC palmitoyltransferase	Signal-peptide	DHHC palmitoyltransferase
cre_102984	allophane hydrolase	Signal-peptide	No
cre_103049	Helicase superfamily c-terminal domain	No	DEAD/DEAH box helicase, Helicase conserved C-terminal domain
cre_103211	Second domain of the pleiotropic drug resistance-like (PDR) subfamily G of ATP-binding cassette transporters; ABC-2 type transporter	IGFBP_N_1;AGO1/1_Transmembrane regions	ABC transporter;ABC-2-type transporter;Plant PDR ABC transporter
cre_103256	Catalytic domain of the Protein Serine/Threonine Kinase, Phosphotransferases	Signal-peptide	Protein kinase domain
cre_103375	N-terminal domain of Pfg (formamidopyrimidine-DNA glycosylase, MutM). Nei (endonuclease VIII)	Signal-peptide;Transmembrane regions	Formamidopyrimidine-DNA glycosylase N-terminal domain, H2TH domain
cre_103389	B0A-like protein	No	B0A-like protein
cre_103408	Tynovirus 45/70Kd protein	Signal-peptide;Transmembrane regions	No
cre_103549	CHC domain	Signal-peptide	CHC domain
cre_103617	Metal dependent phosphohydrolases with conserved HD motif	Signal-peptide	3'-cyclic nucleotide phosphodiesterase
cre_103623	Sec14p-like lipid-binding domain; CRAL/TRIO_N-terminal domain	Signal-peptide	CRAL/TRIO domain
cre_103793	SWI3, ADA2, N-CuR and TFIIB DNA-binding domains	No	Myb-like DNA-binding domain;Myb-like DNA-binding domain
cre_103881	Carbonyl reductase sniffer-like, classical (c) SDRs; Sniffer is an NADPH-dependent carbonyl reductase of the classical SDR family.	No	No
cre_103983	The AAA+ (ATPases Associated with a wide variety of cellular Activities)	Signal-peptide	ATPase family associated with various cellular activities (AAA)
cre_104002	Cysteine hydrolases; Isomerase family	Signal-peptide	Isomerase family
cre_104003	ATPases involved in the bleomycin resistance protein, glyoxalase I, and type I ring-cleaving dioxygenases	Signal-peptide; Transmembrane regions	Carboxylate-like binding protein
cre_104082	Intercalating factor (IHF) and HU are small heterodimeric members of the DNABII protein family	Signal-peptide	EF-hand domain;EF-hand domain
cre_104097	EF-hand calcium binding motif	EGCGHEHL-Signal-peptide	DEAD/DEAH box helicase; Helicase conserved C-terminal domain
cre_104111	DEAD-box helicases	INTERGRIN_BETA, IGFBP_N_1	Domain of unknown function (DUF1768)
cre_104136	Domain of unknown function (DUF1768)	Signal-peptide	Protein of unknown function
cre_104300	Protein of unknown function (DUF1237)	Signal-peptide	RibD C-terminal domain
cre_104322	Dihydrofolate reductase (DHFR)	EGGSHELL_Signal-peptide	No
cre_104339	Dimerization-Docking domain of Cyclic GMP-dependent Protein Kinase I	Signal-peptide	Putative RNA binding domain
cre_104389	Catalytic domain of AGC Family Protein Serine/Threonine Kinases	Signal-peptide	Protein kinase domain
cre_104431	Members of the ICL/PEPM enzyme family catalyze either P-C or C-C bond formation/cleavage.	Signal-peptide	Phosphoenolpyruvate phosphomutase
cre_104441	RING-variant domain	Signal-peptide	No
cre_104448	Replication termination factor 2, Protein of unknown function (DUF3348)	Signal-peptide	Rif2 RING-finger
cre_104450	Methyltransferase domain	Signal-peptide; Transmembrane regions	Methyltransferase domain
cre_104519	RING-finger (Really Interesting New Gene) domain	EGGSHELL	Zinc finger, C3H4 type (RING finger)
cre_104582	4-hydroxy-2-oxoalerate/4-hydroxy-2-oxopentanoic acid aldolase	Signal-peptide	Protein of unknown function (DUF971)
cre_104611	cwf21 domain	Signal-peptide	cwf21 domain
cre_104719	Prion/Doppel alpha-helical domain	EGGSHELL_Signal-peptide;	No
cre_104732	tRNA-binding-domain-containing EMAP2-like proteins	Signal-peptide	Putative tRNA binding domain
cre_104774	Nudix hydrolase	EGGSHELL_Signal-peptide; Transmembrane regions	NUDIX domain
cre_104778	RNA recognition motif in yeast cell wall integrity protein scw1	Signal-peptide	RNA recognition motif (a.k.a. RRM, RBD, or RNP domain)
cre_104788	Catalytic domain of the Protein Serine/Threonine Kinase	MTECHINOIDEA	Protein kinase domain
cre_104798	S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase), class I	IGFBP_N_1; Signal-peptide	Methyltransferase domain

Fig.1: Schematic diagram of hypothetical proteins of *Chlamydomonas reinhardtii*

cre: 104804	Tesmin/TSOI-like CXC domain	Signal-peptide	Tesmin/TSOI-like CXC domain, cysteine-rich domain
cre: 104846	ABC1 family	Signal-peptide	ABC1 family
cre: 104871	Forkhead associated domain (FOF): FOG	Signal-peptide; Transmembrane regions	FHA domain
cre: 104900	Catalytic domain of Protein Tyrosine Kinases	Signal-peptide	Protein kinase domain
cre: 104999	N-terminal region of Choroin	Signal-peptide	N-terminal region of Choroin, a TM vesicle-mediated sorter
cre: 105035	Catalytic domain of the Protein Serine/Threonine Kinase	No	Protein kinase domain
cre: 105117	MORN repeat	Signal-peptide	MORN repeat
cre: 105305	Sec14p-like lipid-binding domain	INTEGRIN BETA; Signal-peptide	CRAL TRIO domain
cre: 105319	EamA-like transporter family	Signal-peptide	Magnesium transporter NIPA
cre: 105357	S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase), class I.	Signal-peptide	Methyltransferase domain
cre: 105419	Hypothetical protein; Provisional	Signal-peptide	Uncharacterized BCRA_YnfA/UPF0060 family
cre: 105431	tRNA Methyltransferase family	Signal-peptide	tRNA tRNA Methyltransferase family
cre: 105488	RBM1CTR (NUC064) family, sic protein; Cornifin (SPRR) family	Signal-peptide	No
cre: 105567	cobW C-terminal domain	INTEGRIN BETA; Signal-peptide	Cobalamin synthesis protein cobW C-terminal domain
cre: 105644	Tesmin/TSOI-like CXC domain	Signal-peptide	Tesmin/TSOI-like CXC domain, cysteine-rich domain
cre: 105828	Biotinophilic RFP-TM chloride channel	Signal-peptide; Transmembrane regions	Biotinophilic RFP-TM chloride channel
cre: 105835	DEAD-box helicases	IGFBP_N_1; Signal-peptide	DEAD/DEAH box helicase; Helicase conserved C-terminal domain
cre: 105845	Calcium/calmodulin dependent protein kinase II Association; Smn1-like domain	Signal-peptide	Calcium/calmodulin dependent protein kinase II Association
cre: 105873	Uncharacterized protein family UPF0016	No	Uncharacterized protein family UPF0016
cre: 105896	Macro domain	Signal-peptide; Transmembrane regions	No
cre: 105899	Zinc-binding domain of Paramyxoviride V protein.	No	No
cre: 105918	Catalytic domain of Protein Tyrosine Kinases	INTEGRIN BETA isomerase; AGOUTL_1; Signal-peptide	ACT domain; Protein tyrosine k
cre: 105941	Regulator of chromosome condensation (RCC1) repeat.	Signal-peptide	Regulator of chromosome condensation (RCC1) repeat
cre: 105965	High molecular weight glutenin subunit	Signal-peptide	No
cre: 105992	Usp: Universal stress protein family; ATP-sulfurylase	No	Universal stress protein family
cre: 106083	Scavenger receptor_Cys-rich	EGGSHELL; Signal-peptide	Scavenger receptor cysteine-rich domain
cre: 106207	Male specific sperm protein	Signal-peptide	Bacterial transferase hexapeptide (six repeats)
cre: 106211	Hypothetical protein	Signal-peptide; Transmembrane regions	Protein of unknown function (DUF2498)
cre: 106306	Serine/threonine phosphatases, family 2C	Signal-peptide; Transmembrane regions	Protein phosphatase 2C
cre: 106339	Posttranslational modification, protein turnover, chaperones; SelR domain;	Signal-peptide	SelR domain
cre: 106402	CobW/HypB/UreG_nucleotide-binding domain; Putative GTPases (G3E family)	Signal-peptide	CobW/HypB/UreG_nucleotide-binding domain; Cobalamin synthesis protein
cre: 106436	RNA recognition motif 2 found in U2 large nuclear ribonucleoprotein auxiliary factor U2AF 65 kDa subunit (U2AF65)	No	RNA recognition motif, (a.k.a. RRM, RBD, or RNP domain)
cre: 106461	EF-hand, calcium binding motif	Signal-peptide	EF-hand domain pair
cre: 106469	cyclase homology	Signal-peptide	Adenylate and Guanylate cyclase catalytic domain
cre: 106548	Prolyl oligopeptidase family; Alpha/beta hydrolase family	Signal-peptide	Alpha/beta hydrolase family
cre: 106555	Helicase superfamily c-terminal domain; associated with DEXDC-, DEAD-, and DEAH-box proteins, yeast initiation factor 4A, Sk1p	Signal-peptide	Helicase conserved C-terminal domain
cre: 106655	SWI3, ADA2, N-Cor and TFIIB DNA-binding domains.	Signal-peptide	Myb-like DNA-binding domain
cre: 106711	Dial domain or 1-domain.	EGGSHELL; Signal-peptide	Dial domain
cre: 106717	ATP-binding cassette transporter nucleotide-binding domain	signal_peptide	MutS domain V
cre: 106748	The Fer1_NiH superfamily	EGGSHELL; Signal-peptide	CobW/HypB/UreG_nucleotide-binding domain; Cobalamin synthesis protein
cre: 106820	error-prone DNA polymerase	Signal-peptide; Transmembrane regions	No
cre: 106821	Tumor necrosis factor receptor (TNFR) domain	Signal-peptide; Transmembrane regions	No
cre: 106822	Histidine kinase-like ATPases	Signal-peptide	Histidine kinase, DNA gyrase B-, and HSP90-like ATPase
cre: 106870	C1 Peptidase family	Signal-peptide; Transmembrane regions	Papain family cysteine protease
cre: 106887	Histidine kinase-like ATPases	Signal-peptide	Histidine kinase, DNA gyrase B-, and HSP90-like ATPase
cre: 106896	Granulin, Metallothionein	Signal-peptide	No
cre: 106901	No	AGOUTL_1	No
cre: 106906	Kinesin motor domain, catalytic domain	EGGSHELL	Kinesin motor domain
cre: 106912	Scavenger receptor_Cys-rich	Signal-peptide	Scavenger receptor cysteine-rich domain
cre: 107085	AAA+ superfamily, ABC, RecA-like, VirD4-like, PilT-like, and SF1/2 helicases.	Signal-peptide	ATPase family associated with various cellular activities (AAA)
cre: 107111	AAA+ superfamily	Signal-peptide	ATPase family associated with various cellular activities (AAA)
cre: 107159	RING variant domain	Signal-peptide	RING variant domain
cre: 107183	Sec-independent protein translocase protein (TatC)	INTEGRIN BETA; Signal-peptide	Sodium/hydrogen exchanger family
cre: 107200	RING-finger (Really Interesting New Gene) domain	No	Ring finger domain
cre: 107202	Translation initiation factor IF-3, C-terminal domain; Translation initiation factor IF-3, N-terminal domain	Signal-peptide	Translation initiation factor IF-3, N-terminal domain; Translation initiation
cre: 107266	flavonoid reductase (FR), extended (e) SDRs	Signal-peptide	NAD dependent epimerase/dehydratase family
cre: 107285	No	Signal-peptide	No
cre: 107313	Macro domain, Appr-1'-pase like family.	Signal-peptide	Macro domain
cre: 107359	GST_N family, Saccharomyces cerevisiae GTT1-like subfamily	Signal-peptide	Glutathione S-transferase, N-terminal domain
cre: 107381	Leucine rich repeat	AGOUTL_1; Signal-peptide; Transmembrane regions	Leucine rich repeat
cre: 107387	Radical SAM superfamily	Signal-peptide; Transmembrane regions	Radical SAM superfamily
cre: 107417	CAP-Gly domain	Signal-peptide	CAP-Gly domain
cre: 107425	NADH dehydrogenase subunit 2; V-type ATP synthase subunit E; GldM N-terminal domain	Signal-peptide	No
cre: 107441	Short repeat of unknown function (DUF308)	Signal-peptide	No
cre: 107546	mttA/HCf106 family	AGOUTL_1; Signal-peptide; Transmembrane regions	No
cre: 107572	The MutT pyrophosphorylase	Signal-peptide	NUDIX domain
cre: 107624	HIT family	Signal-peptide	HIT domain
cre: 107663	RING-finger (Really Interesting New Gene) domain	Signal-peptide; Transmembrane regions	Ring finger domain
cre: 107819	Mitotic-spindle organizing gamma-tubulin ring associated	Signal-peptide	Mitotic-spindle organizing gamma-tubulin ring associated
cre: 107828	Uncharacterized conserved protein (DUF2358)	EGGSHELL; Signal-peptide	Uncharacterized conserved protein (DUF2358)
cre: 107862	Aldo-keto reductases (AKRs)	INTEGRIN BETA; Signal-peptide	Aldo/keto reductase family
cre: 108035	The AAA+ (ATPases Associated with a wide variety of cellular Activities)	IGFBP_N_1; Signal-peptide	ATPase family associated with various cellular activities (AAA)
cre: 108061	MtrB38-like domain	Transmembrane regions; Signal-peptide	Protein of unknown function (DUF498/DUF598)
cre: 108112	PAP2, subfamily similar to human phosphatidic acid phosphatase type 2 domain containing 1	Signal-peptide	PAP2 superfamily
cre: 108125	ankyrin repeats	Signal-peptide	Ankyrin repeats
cre: 108123	Fe4S4 single cluster domain	Signal-peptide; Transmembrane regions	Fe4S4 single cluster domain
cre: 108137	Catalytic domain of the Serine/Threonine Kinase, Cyclin-Dependent protein Kinase 7	Signal-peptide	Protein kinase domain
cre: 108218	Recombinant Multi-domain subfamily RCC1_2; large tegument protein UL36	Signal-peptide	Recombinant cl Chromosome condensation (RCC1) repeat
cre: 108412	Uncharacterized conserved protein	Signal-peptide	Uncharacterized conserved protein
cre: 108443	ATP-binding cassette domain 2 of multidrug resistance-associated protein	Signal-peptide; Transmembrane regions	ABC transporter
cre: 108464	bleomycin resistance protein, glyoxalase I, and type I farnesylation dioxygenases	J_ACTX; Signal-peptide	Glyoxalase-like domain
cre: 108536	Carbohydrate kinase and ribulose-phosphate 3-epimerase fusion protein	Signal-peptide	FGGY family of carbohydrate kinases, N-terminal domain
cre: 108537	The AAA+ (ATPases Associated with a wide variety of cellular Activities)	Signal-peptide	MCM2/5 family
cre: 108564	YTF21-B-like domain	Signal-peptide	YTF21-B-like domain
cre: 108580	Ankyrin repeats; Atrophin-1 family	Signal-peptide	Ankyrin repeats
cre: 108616	Haemolytic domain	IGFBP_N_1	No
cre: 108648	putative FAD-binding dehydrogenase	EGGSHELL; Signal-peptide	FAD binding domain
cre: 108674	RNA recognition motif 1 in heterogeneous nuclear ribonucleoprotein hnRNP A and hnRNP D	Signal-peptide	RNA recognition motif, (a.k.a. RRM, RBD, or RNP domain)
cre: 108707	ankyrin repeats	Signal-peptide	Ankyrin repeats
cre: 108741	NO	Signal-peptide	No
cre: 108766	NruU protein	Signal-peptide	NruU protein
cre: 108818	Revorius minor core protein Mu-2	Signal-peptide	No
cre: 108839	DEAD-like helicases superfamily.	Signal-peptide	Type II/IV secretion system protein
cre: 108925	cyclase homology domain	Signal-peptide	Adenylate and Guanylate cyclase catalytic domain
cre: 108934	cell division protein DamX	Signal-peptide; Transmembrane regions	No
cre: 108940	RBMLCTR (NUC064) family	Signal-peptide	No
cre: 108983	Herpesvirus UL46 protein; phosphoribosyl-dephospho-CoA transferase	Signal-peptide; Transmembrane regions	No
cre: 108984	Herpesvirus UL46 protein; Formin Homology Region 1; UL43 envelope protein	MTECHINOIDEA; Signal-peptide	No
cre: 109028	NO	Signal-peptide; Transmembrane_regions	No
cre: 109029	NO	Signal-peptide; Transmembrane_regions	No
cre: 109119	NO	Signal-peptide; Transmembrane_regions	No
cre: 109222	DNA polymerase III subunits gamma and tau	INTEGRIN BETA; Signal-peptide; Transmembrane regions	No
cre: 109250	Methyltransferase domain	Signal-peptide	Methyltransferase domain
cre: 109252	Ribosomal protein L24eL24	Signal-peptide	Ribosomal protein L24e
cre: 109269	Herpesvirus helicase-prime complex component	Signal-peptide	No
cre: 109352	Tub family	Signal-peptide	No
cre: 109358	Tymovirus 45/70Kd protein	AGOUTL_1; Signal-peptide; Transmembrane regions	No
cre: 109422	G1 family of glycosyltransferases.	Signal-peptide	Glycosyl transferases group 1
cre: 109425	Mallophilophagase superfamily, metallophosphatase domain; E4 protein	Signal-peptide	No
cre: 109509	No	AGOUTL_1; Signal-peptide	No
cre: 109540	Ribosomal protein L6; UDP galactose-4-epimerase	Signal-peptide	No
cre: 109545	Histidine phosphotransferase domain	Signal-peptide	Histidine phosphotransferase superfamily (branch 1)
cre: 109602	C-terminal domain of Asparagine Synthase B.	Signal-peptide	Asparagine synthase
cre: 109818	Pyridine nucleotide-disulphide oxidoreductase; NAD(P)-binding Rossmann-like domain	Signal-peptide	Pyridine nucleotide-disulphide oxidoreductase
cre: 109828	Protein of unknown function (DUF1005)	Signal-peptide; Transmembrane regions	No
cre: 109976	Variable surface protein Vz28; Tumor necrosis factor receptor (TNFR) domain	Signal-peptide; Transmembrane_regions	No
cre: 110032	Leucine-rich repeats (LRRs); ribonuclease inhibitor (RI)	IGFBP_N_2; Signal-peptide	Leucine rich repeat
cre: 110145	NADH dehydrogenase subunit 2; V-type ATP synthase subunit E; GldM N-terminal domain	Signal-peptide	No
cre: 110216	Enoylacylglycerol C; EBNA-3B; Streptogramin lyase	INTEGRIN BETA; Signal-peptide; Transmembrane regions	No
cre: 110256	Extensin-like region	Signal-peptide; Transmembrane regions	No
cre: 110360	Ctp amino terminal domain	Signal-peptide	Ctp amino terminal domain
cre: 110955	Cystathione beta-lyase family protein	Signal-peptide	Methionine gamma-lyase
cre: 110956	No	Signal-peptide	No
cre: 110991	No	MTMOLLUSC; Signal-peptide; Transmembrane_regions	No
cre: 110992	No	Signal-peptide; Transmembrane_regions	No
cre: 111008	High molecular weight glutenin subunit	Signal-peptide	No
cre: 111044	Uncharacterized protein conserved in bacteria (DUF2065)	Signal-peptide	No
cre: 111066	AAA+ (ATPases Associated with a wide variety of cellular Activities)	Signal-peptide	Sigma-54 interaction domain; Bacterial regulatory protein, Fis family
cre: 111068	alphaA-amylose/limit dextrinase	Signal-peptide	No
cre: 111086	YigYER057c/UK114-like proteins present in bacteria, archaea, and eukaryotes	Signal-peptide	No
cre: 111092	AAA+	Signal-peptide	ATPase family associated with various cellular activities (AAA)
cre: 111122	Rat sarcoma (Ras)	Signal-peptide	Elongation factor Tu GTP binding domain
cre: 111160	Eukaryotic long-chain fatty acid CoA synthetase (LC-FACS)	Signal-peptide	No
cre: 111185	Glutaryl-CoA dehydrogenase	Signal-peptide	Acyl-CoA dehydrogenase, C-terminal domain
cre: 111255	Prolyl 4-hydroxylase alpha subunit homologues	No	2OG-Fe(II) oxygenase superfamily; Shk domain-like
cre: 111258	HECT domain	Signal-peptide	HECT-domain (ubiquitin-transfere

cre_111278	No	Signal-peptide	No
cre_111290	Anticodon-binding domain	Signal-peptide	Anticodon-binding domain
cre_111291	Catalytic domain of Cyclin-Depended protein Kinase	Signal-peptide	Protein kinase domain
cre_111296	Catalytic domain of Cyclin-Depended protein Kinase	Signal-peptide	No
cre_111308	Uncharacterized protein family (UPF0041)	IGFBP_N_1; Signal-peptide	No
cre_111403	Histidine phosphotransferase domain	Signal-peptide	Uncharacterised protein family (UPF0041)
cre_111413	Catalytic domain of Protein Tyrosine Kinases	Signal-peptide	Histidine phosphatase superfamily (branch 1)
cre_111421	Formyltetrahydrofolate deformylase (Formyl-FH4 hydrolase), C-terminal hydrolase domain	Signal-peptide	Protein kinase domain
cre_111463	Catalytic domain of Protein Kinases	Signal-peptide	ACT domain; Formyl transferase
cre_111483	Alde-keto reductases (AKR)s	Signal-peptide	Protein kinase domain
cre_111544	RNA-recognition motif; RNA-binding proteins (RRM domain)	Signal-peptide	Aldo/keto reductase family
cre_111555	Prolyl 4-hydroxylase alpha/beta; ShK toxin domain	Signal-peptide	RNA recognition motifs (a.k.a. RRM, RBD, or RNP domain)
cre_111581	Tetratricopeptide repeat domain	Signal-peptide	2OG-FoII) oxygenase superfamily
cre_111593	Per1	Signal-peptide	TPR repeat; Tetratricopeptide repeat
cre_111637	Catalytic domain of Protein Kinases	Signal-peptide	Per1-like
cre_111701	Oba GTase	Signal-peptide	Protein kinase domain
cre_111714	Nucleoporin autoproteidase; UDP-3-O-[3-hydroxyvinylstoyl] glucosamine N-acetyltransferase	Signal-peptide	GTP1/ORG-505 ribosome-binding GTPase
cre_111779	YGGT family	Signal-peptide	Nucleoporin autoproteidase
cre_111793	RNA recognition motif in RNA-binding protein 42 (RBM42)	Signal-peptide	YGGT family
cre_111796	MARCKS family; Tymovirus 45/70Kd protein	Signal-peptide	No
cre_111806	Sulfatase; Arylsulfatase A	Signal-peptide	Sulfatase
cre_111844	myb-like DNA-binding domain; SHAQKYF class	Signal-peptide	Myb-like DNA-binding domain
cre_111927	Esterases and lipases (includes fungal lipases, cholinesterases, etc.); Carboxylesterase family	J_ACTX; IGFBP_N_1; Signal-peptide	Carboxylesterase family
cre_111935	Protein of unknown function (DUF788); Keratin, high sulfur B2 protein	Signal-peptide	Protein of unknown function (DUF788)
cre_111940	NLI interacting factor-like phosphatase	Signal-peptide	NLI interacting factor-like phosphatase
cre_111943	Hypothetical protein	Signal-peptide	Uncharacterized ACR, YdU/UPF0061 family
cre_111945	EF-hand, calcium binding motif	Signal-peptide	EF-hand domain pair
cre_111968	Phosphatidylinositol 3- and 4-kinase	Signal-peptide; Transmembrane regions	Phosphatidylinositol 3- and 4-kinase
cre_111970	Non-haem dixygenase in morphine synthesis N-terminal	IGFBP_N_1; Signal-peptide	No
cre_111971	No	Signal-peptide	No
cre_111975	WD40 domain	IGFBP_N_1; Signal-peptide	WD domain, G-beta repeat
cre_111979	Sec14p-like lipid-binding domain.	Signal-peptide	CRAL/TRIO_N-terminal domain; CRAL/TRIO domain
cre_111999	AAA+	Signal-peptide	ATPase family associated with various cellular activities (AAA)
cre_112011	Serine/Threonine Kinases (STKs); Cyclin-dependent protein kinase (CDKL), catalytic (c) domain.	Signal-peptide	Protein kinase domain
cre_112030	Nodulin-21 and CCC1-related protein family	Signal-peptide	VIT family
cre_112046	WD40 domain	EGGSHELL; Signal-peptide	WD domain, G-beta repeat
cre_112052	Catalytic domain of Cyclin-Dependent protein Kinase Like Serine/Threonine Kinases	Signal-peptide	Protein kinase domain
cre_112057	START/THYR-alpha; C_PIT1; C_CoxG_Calc (SRPBCC) ligand-binding domain superfamily	Signal-peptide	Polyketide cyclase / dehydrase and lipid transport
cre_112232	Serine/Threonine Kinases (STKs); Cyclin-dependent protein kinase (CDKL), catalytic (c) domain.	IGFBP_N_1; Signal-peptide	Protein kinase domain
cre_112237	SW13_ADA2_N-CuR; TET2-B DNA-binding domains.	IGFBP_N_1; Signal-peptide	Myb-like DNA-binding domain
cre_112246	AAA domain (dynamin-related subfamily)	Signal-peptide	AAA domain (dynamin-related subfamily)
cre_112349	GST_N family	Signal-peptide	Glutathione S-transferase_N-terminal domain (shorten)
cre_112370	Tryptophan-rich protein (DUF2389)	Signal-peptide	Tryptophan-rich protein (DUF2389)
cre_112628	SW13_ADA2_N-CuR and THH-B DNA-binding domains	Transmembrane_regions; Signal-peptide	Myb-like DNA-binding domain
cre_112723	ATP-binding cassette domain of elongation factor 3, subfamily F	IGFBP_N_1; Signal-peptide	ABC transporter
cre_112727	OTU-like cysteine protease	Transmembrane_regions	OTU-like cysteine protease
cre_112842	Formiminotransferease domain, N-terminal subdomain	INTEGRIN_BETA; Signal-peptide	Formiminotransferease domain, N-terminal subdomain
cre_112949	Zinc-binding domain	Signal-peptide	Zinc-binding domain
cre_112975	Domain of unknown function (DUF3594)	Signal-peptide	Domain of unknown function (DUF3594)
cre_112991	Dipeptidyl aminopeptidases/acylaminoacyl-peptidases	MTTECHINOIDEA; Signal-peptide; Transmembrane_regions	No
cre_113013	Mitochondrial ATPase inhibitor, IATP	Signal-peptide	No
cre_113074	Nuclear transport factor 2 (NTF2) domain	Signal-peptide	Nuclear transport factor 2 (NTF2) domain
cre_113076	Armadillo/beta-catenin-like repeats.	Signal-peptide	No
cre_113086	Diacylglycerol kinase catalytic domain	No	Diacylglycerol kinase catalytic domain
cre_113093	Protein of unknown function (DUF1295); Keratin, high sulfur B2 protein;	No	Diacylglycerol kinase catalytic domain
cre_113099	Catalytic domain of Protein Kinases; Protein Kinases (PKs), catalytic (c) domain.	Signal-peptide	Protein tyrosine kinase
cre_113103	Chalcone and stilbene synthases	INTEGRIN_BETA; Signal-peptide	FAE1/Type III polyketide synthase; Chalcone and stilbene synthases, C-terminal domain
cre_113105	Glutathione-dependent formaldehyde dehydrogenase related proteins	No	Alcohol dehydrogenase GroES-like domain
cre_113124	Serine carboxypeptidase	Signal-peptide	Serine carboxypeptidase
cre_113145	Protein of unknown function (DUF679)	Signal-peptide	Protein of unknown function (DUF679)
cre_113182	phosphoribosylpyrophosphate synthetase	Signal-peptide	No
cre_113197	Prolyl 4-hydroxylase alpha subunit homologues; Mammalian enzymes catalyse hydroxylation of collagen	Signal-peptide	2OG-FoII) oxygenase superfamily
cre_113206	Catalytic domain of Protein Tyrosine Kinases; Protein Tyrosine Kinase (PTK) family, catalytic domain.	Signal-peptide	Protein tyrosine kinase
cre_113303	Protein of unknown function (DUF172)	Signal-peptide	Protein of unknown function (DUF172)
cre_113308	WD40 domain	Signal-peptide	WD domain, G-beta repeat
cre_113366	Maltose O-acetyltransferase (MAT) and Galactoside O-acetyltransferase (GAT	Signal-peptide	Maltose acetyltransferase; Bacterial transferase hexapeptide
cre_113377	Cyclin box fold.	Signal-peptide	Cyclin_N-terminal domain
cre_113394	AAA+ superfamily	Signal-peptide	ATPase family associated with various cellular activities (AAA)
cre_113457	NO	EGGSHELL; Signal-peptide; Transmembrane_regions	PAMP1_N-terminal
cre_113524	Metal dependent phosphohydrolases with conserved HD' motif	Signal-peptide	3'-cyclic nucleotide phosphodiesterase
cre_113566	PAS domain	Signal-peptide	PAS domain
cre_113573	The AAA+ (ATPases Associated with a wide variety of cellular Activities)	Signal-peptide	AAA+ family associated with various cellular activities (AAA)
cre_113576	Protein, high sulfur B2 protein, CG2 binding protein (Siva)	Signal-peptide	No
cre_113594	quenine tRNA-ribosyltransferase; Quenine tRNA-ribosyltransferase	Signal-peptide; Transmembrane_regions	Quenine tRNA-ribosyltransferase
cre_113642	Catalytic domain of Protein Kinases; Protein Kinases (PKs), catalytic (c) domain.	Signal-peptide	Protein kinase domain
cre_113656	NAD(P)-binding Rossmann-like domain	Signal-peptide	Flavin containing amine oxidoreductase
cre_113668	Protein of unknown function (DUF1538)	IGFBP_N_1	Protein of unknown function (DUF1538)
cre_113676	NTPase-HAM1	Signal-peptide	Ham1 family
cre_113763	Malonyl-CoA synthetase (MCS)	Signal-peptide	AMP-binding enzyme C-terminal domain
cre_113785	N-terminal domain	Signal-peptide	PP-loop family
cre_113806	Hevein or type I chitin binding domain; YppG-like protein	Signal-peptide	Chitin recognition protein
cre_113811	2OG-FoII) oxygenase superfamily	INTEGRIN_BETA; Signal-peptide	No
cre_113818	Circularly permuted YifA-related GTPases.	Signal-peptide; Transmembrane_regions	50S ribosome-binding GTPase
cre_113825	Myosin and Kinesin motor domain.	Signal-peptide	Kinesin motor domain
cre_113826	PAS domain	No	PAS domain
cre_113857	Dephosphi-CoA kinase	IGFBP_N_1; EGGSHELL	Dephosphi-CoA kinase
cre_113892	Methyltransferase domain	EGGSHELL; Signal-peptide	Methyltransferase domain
cre_113895	Peptidase M14	Signal-peptide	Zinc carboxypeptidase
cre_113900	RNA recognition motif (RRM) superfamily; RRM, also known as RBD or RNP	Signal-peptide	RNA recognition mot
cre_113924	PurB like adenylosuccinate (adenylsuccinate lyase, ASL)	Signal-peptide	No
cre_113936	PIP-5-kinases and protein kinases; phosphoribosyl-dephospho-CoA transferase	INTEGRIN_BETA; EGGSHELL; Signal-peptide	MORN repeat
cre_113957	Arabidopsis thaliana Bsu I phophatase and related proteins, C-terminal metallophosphatase domain	No	Calcineurin-like phosphoesterase
cre_113976	Phosphotryosyl phosphatase activator (PTPA)	INTEGRIN_BETA; Signal-peptide	Phosphotryosyl phosphate activator (PTPA) protein
cre_113992	Rieske domain	Signal-peptide	Rieske [2Fe-2S] domain
cre_114005	PAS domain	No	PAS domain
cre_114023	Leucine rich repeat N-terminal domain	Signal-peptide	Leucine rich repeat N-terminal domain; Leucine rich repeat
cre_114085	Kinesin motor domain	Signal-peptide; Transmembrane_regions	Kinesin motor domain
cre_114119	Type I glutamine amidotransferase (GATase)	Signal-peptide	N-terminal domain of DJ-I_PipI family; DJ-I_PipI family
cre_114154	EngA GTase	EGF_1	50S ribosome-binding GTase; KH-domain-like of EngA bacterial GTase enzymes, C-terminal
cre_114162	Protein of unknown function (DUF3755)	Signal-peptide	Protein of unknown function (DUF3755)
cre_114199	Ankyrin repeats; ty kinase domain	Signal-peptide	Ankyrin repeats; Protein kinase domain
cre_114209	First domain of ATP-binding cassette component of cobalt transport system	Signal-peptide	ABC transporter; Protein of unknown function (DUF1602)
cre_114331	Glutathione hydroxylase; The Jmfc domain, acyltransferase domain	Signal-peptide	JM domain, hydroxylase
cre_114381	TBC and LysM domain; Plasmid histidine-rich protein (HRPIII/III	Signal-peptide	TBC domain
cre_114371	Rot sarcomeric Rho	Signal-peptide; Transmembrane_regions	Root hair defective 3 GTP-binding protein (RHD3)
cre_114393	ORF080 virion core protein	Signal-peptide; Transmembrane_regions	Rieske [2Fe-2S] domain
cre_114479	C-terminal catalytic domain	Signal-peptide; Transmembrane_regions	Protein kinase domain
cre_114487	Catalytic domain of the Protein Serine/Threonine Kinase	Signal-peptide	Nucleotide-diphospho-sugar transferase
cre_114491	Nucleotide-diphospho-sugar transferase	Signal-peptide	His_Kinase_A-domain; Histidine kinase; DNA gyrase B; and HSP90-like ATPase
cre_114537	Histidine kinase-like ATPases	Signal-peptide	Yip1 domain
cre_114544	Rab-GTPase interacting factor; Golgi membrane protein [Intracellular trafficking and secretion]	Signal-peptide	Protein kinase domain
cre_114555	Catalytic domain of the Protein Serine/Threonine Kinase, cAMP-dependent protein kinase	Signal-peptide	PPR repeat family
cre_114572	Tetratricopeptide repeat domain	Signal-peptide	EamA-like transporter family
cre_114780	EamA-like transporter family	Signal-peptide	Glycosyl hydrolases family 32 N-terminal domain
cre_114834	Glycosyl hydrolase family 32_beta-fructosidases	Signal-peptide; Transmembrane_regions	Carbon-nitrogen hydrolase
cre_114854	Nit1, Nit2 proteins	IGFBP_N_1; Signal-peptide	No
cre_114874	DEAD-like helicases superfamily.	Signal-peptide	EF-hand
cre_114938	EF-hand, calcium binding motif	Signal-peptide; Transmembrane_regions	AMP-binding enzyme
cre_115017	Eukaryotic long-chain fatty acid CoA synthetase (LC-FACS)	EGGSHELL	Cupin domain
cre_115035	Cupin domain	Signal-peptide	Methionine gamma-lyase
cre_115040	Cystathione beta-lyase family protein	Signal-peptide	NADH(P)-binding
cre_115042	atypical (a) SDRs, subgroup 5	Signal-peptide	Protein kinase domain
cre_115049	Catalytic domain of AGC family Protein Serine/Threonine Kinases	Signal-peptide	Protein tyrosine kinase
cre_115082	Catalytic domain of Ephrin Receptor Protein Tyrosine Kinases	Signal-peptide	AGOUTI_I; Signal-peptide; Transmembrane_regions
cre_115086	Rieske domain	Signal-peptide	WD domain, G-beta repeat
cre_115105	WD40 domain	Signal-peptide	Rhodanese domain
cre_115108	Rhodanese Homology Domain (RHOD)	Signal-peptide	ABC transporter
cre_115132	ATP-binding cassette transporter nucleotide-binding domain	Signal-peptide	Tetratricopeptide repeat
cre_115145	Tetratricopeptide repeat domain	Signal-peptide	Adenylate cyclase associated (CAP) C terminal
cre_115172	Domain in CAPs (cyclase-associated proteins) and X-linked retinitis pigmentosa 2 gene product	Signal-peptide	Myb-like DNA-binding domain
cre_115247	myb-like DNA-binding domain	Signal-peptide	No
cre_115272	No	J_ACTX; Signal-peptide	Glyoxalase-like domain
cre_115316	Type I extradiol dioxygenases, glyoxalase I and a group of antibiotic resistance proteins	Signal-peptide	Glyoxalase-like domain
cre_115363	Halocidic dehalogenase	No	Halocidic dehalogenase-like hydrolase
cre_115434	Sec14p; CRAL TRIO, N-terminal domain	Signal-peptide	CRAL TRIO, N-terminal domain
cre_115450	Tetratricopeptide repeat domain	CTCK_1; Signal-peptide	Tetratricopeptide repeat

cre_115459	Prion/Doppel alpha-helical domain; Tymovirus 45/70Kd protein	EGGSHELL; Signal-peptide	No
cre_115491	GPR1/FUN34/yaaH family	No	GPR1/FUN34/yaaH family
cre_115526	Putative glycerate dehydrogenase and related proteins of the D-specific 2-hydroxy dehydrogenase family	No	NAD-binding of NADP-dependent 3-hydroxysobutyrate dehydrogenase
cre_115546	Protein of unknown function (DUF647)	No	Vitamin B6 photo-protection and homeostasis
cre_115555	N-Acyltransferase superfamily; Myristoyl-CoA	Signal-peptide; Transmembrane_regions	Acetyltransferase (GNAT) domain
cre_115597	YGGT family	Signal-peptide	YGGT family
cre_115645	Predicted NAD/FAD-binding protein	Signal-peptide; Transmembrane_regions; Signal-peptide;	NAD(P)-binding Rossmann-like domain
cre_115665	Kef-type K+ transport systems; TrkA-N domain; L-aspartate oxidase	Signal-peptide	Sodium/hydrogen exchanger family; TrkA-N domain
cre_115715	Peptidoglycane repeats	Signal-peptide	Peptidoglycane repeats
cre_115716	Uncharacterized conserved protein (DUF2042)	Signal-peptide	Uncharacterized conserved protein (DUF2042)
cre_115745	Metallo-beta-lactamase superfamily	Signal-peptide	Metallo-beta-lactamase superfamily; RNA-metabolising metallo-beta-lactamase
cre_115782	First domain of the ATP-binding cassette component of cobalt transport system	EGGSHELL; Signal-peptide	ABC transporter
cre_115803	NLI interacting factor-like phosphatase	No	Ubiquitin family; NLI interacting factor-like phosphatase
cre_115878	Catalytic domain of Protein Kinases; Protein Kinases (PKs)	No	Protein kinase domain
cre_115896	HYR domain	Signal-peptide	HYR domain
cre_115900	Catalytic domain of AGC family Protein Serine/Threonine Kinases	IGFBP_N_1	Protein kinase domain
cre_115915	WD40 domain	EGGSHELL; Signal-peptide	WD domain; G-beta repeat
cre_115925	AAA domain	Signal-peptide	AAA domain
cre_115938	ATP-binding cassette domain I of multidrug resistance-associated protein	Signal-peptide	ABC transporter; Protein of unknown function (DUF1602)
cre_115983	Catalytic domain of Cyclin-Dependent protein Kinase-like Serine/Threonine Kinases; Phosphotransferases.	Signal-peptide	Protein kinase domain
cre_115988	Thioredoxin (TRX)-like [2Fe-2S] Ferredoxin (Fd) family	Signal-peptide	No
cre_116001	WD40 domain	Signal-peptide	WD domain; G-beta repeat
cre_116006	17beta hydroxysteroid dehydrogenase	Signal-peptide	short chain dehydrogenase
cre_116033	Catalytic domain of the Protein Serine/Threonine Kinase	Signal-peptide	Protein kinase domain
cre_116056	SWI3, ADA2, N-CoR and TFIIB DNA-binding domains.	IGFBP_N_1	Myb-like DNA-binding domain
cre_116096	Possible lysine decarboxylase	No	Possible lysine decarboxylase
cre_116116	Catalytic domain of Protein Tyrosine Kinases	Signal-peptide; Transmembrane_regions	Possible lysine decarboxylase
cre_116132	Helicase superfamily c-terminal domain	Signal-peptide	Helicase conserved C-terminal domain
cre_116145	Catalytic domain of Protein Kinases; Protein Kinases (PKs), catalytic (c) domain.	Signal-peptide	Protein kinase domain
cre_116196	This is a subfamily of Adenine nucleotide alpha hydrolases superfamily.	No	ATP-binding region
cre_116235	Serine carboxypeptidase	Signal-peptide	Serine carboxypeptidase
cre_116256	RNA recognition motif (RRM) superfamily	Signal-peptide; Transmembrane_regions	RNA recognition motif, (a.k.a. RRM, RBD, or RNP domain)
cre_116259	effector domain of the CAP family of transcription factors	Signal-peptide	Cyclic nucleotide-binding domain
cre_116281	DEAD-like helicases superfamily.	Signal-peptide	DEAD/DEAH box helicase; Helicase conserved C-terminal domain
cre_116326	Catalytic domain of Protein Tyrosine Kinases; Protein Tyrosine Kinase (PTK) family, catalytic domain.	Signal-peptide	Protein tyrosine kinase
cre_116326	AAA+ superfamily	J_ACTX	AFG1-like ATPase
cre_116429	Zn-dependent protease with chaperone function	Signal-peptide	Peptidase family M48
cre_116435	ATP-grasp domain	No	No
cre_116444	Calmodulin-related protein	Signal-peptide	Calmodulin related protein
cre_116449	Lipase/protein ligase A: Aspartate aminotransferase (AAT) superfamily (fold type I) of pyridoxal phosphate (PLP)-dependent enzymes.	INTTEGRIN_BETA; Signal-peptide	Biotin-lipope_A-B protein ligase family
cre_116461	SWI3, ADA2, N-CoR and TFIIB DNA-binding domains.	Signal-peptide	Myb-like DNA-binding domain
cre_116475	Uncharacterized protein family_UPF0047	Signal-peptide; Transmembrane_regions	Uncharacterised protein family UPF0047
cre_116553	Helix-turn-helix domain	Signal-peptide	Helix-turn-helix domain
cre_116557	AAA+ Superfamily	IGFBP_N_1; Signal-peptide	ATPase family associated with various cellular activities (AAA)
cre_116572	family of mostly uncharacterized proteins similar to B subtilis YvcK	Signal-peptide; Transmembrane_regions	Uncharacterised protein family UPF0052
cre_116588	SWI3, ADA2, N-CoR and TFIIB DNA-binding domains	Signal-peptide	Myb-like DNA-binding domain
cre_116701	Protein of unknown function (DUF760)	Signal-peptide	Protein of unknown function (DUF760)
cre_116723	DEAD-like helicases superfamily.	Signal-peptide	SNF2 family N-terminal domain
cre_116773	EXS family	Signal-peptide	EXS family
cre_116793	Catalytic domain of Cyclin-Dependent protein Kinase Like Serine/Threonine Kinases	INTTEGRIN_BETA; IGFBP_N_1; Signal-peptide	Protein kinase domain
cre_116827	MutL, Trans: transducin domain	No	Histidine kinase DNA gyrase B, Hsp90-like ATPase; Topoisomerase VI B subunit, transducer
cre_116839	Scavenger receptor Cys-rich	Signal-peptide	Scavenger receptor cysteine-rich domain
cre_116959	Catalytic domain of Protein Kinases; Modified RING finger domain	Signal-peptide	Protein kinase domain; U-box domain
cre_116962	Catalytic domain of Protein Tyrosine Kinases.	No	Protein tyrosine kinase
cre_117116	catalytic core domain of valyltRNA synthetases; Valine amino-acyl tRNA synthetase (ValRS) catalytic core domain. T	Signal-peptide	tRNA synthetases class I (I, L, M and V)
cre_117171	Rossman-fold NAD(P)+-binding proteins	Signal-peptide	NAD(P)+-binding
cre_117208	methionine sulfoxide reductase B	Signal-peptide	SeC1R domain
cre_117219	Catalytic domain of the Protein Serine/Threonine Kinase, cAMP-dependent protein kinase; cAMP-dependent protein kinase (PKA)	Signal-peptide	Protein kinase domain
cre_117237	Vanadium-binding protein 2	Signal-peptide	No
cre_117250	effector domain of the CAP family of transcription factors	Signal-peptide	Cyclic nucleotide-binding domain
cre_117282	Halocid dehalogenase-like hydrolases	Signal-peptide	Halocid dehalogenase-like hydrolase
cre_117291	SWI3, ADA2, N-CoR and TFIIB DNA-binding domains.	No	Myb-like DNA-binding domain
cre_117345	Helicase superfamily c-terminal domain	Signal-peptide	SNF2 family N-terminal domain; Helicase conserved C-terminal domain
cre_117386	alpha/beta hydrolase fold	Signal-peptide	alpha/beta hydrolase fold
cre_117458	The Fer4_NiFH superfamily	Signal-peptide	ConW/HypB/UreG, nucleotide-binding domain
cre_117463	EBNA-3B	Signal-peptide	No
cre_117475	No	Signal-peptide	No
cre_117479	Predicted thio-disulfide oxidoreductase	Signal-peptide	Protein of unknown function, DUF393
cre_117521	Dynein heavy chain and region D6 of dynein motor	Signal-peptide	Dynein heavy chain and region D6 of dynein motor
cre_117533	Catalytic domain of Protein Tyrosine Kinases	No	Protein kinase domain
cre_117562	GAPC_G catalytic domain	IGFBP_N_1; Signal-peptide	GAPC/GatE catalytic domain; GatB domain
cre_117574	pyridine nucleotide-disulfide oxidoreductase family protein	EGGSHELL; Signal-peptide	No
cre_117578	3-oxo-5-alpha-steroid 4-dehydrogenase	Signal-peptide	Protein kinase domain
cre_117611	Catalytic domain of Cyclin-Dependent protein Kinase Like Serine/Threonine Kinases	Signal-peptide	WD domain; G-beta repeat
cre_117660	Catalytic domain of Protein Tyrosine Kinases	Signal-peptide; Transmembrane_regions	Protein tyrosine kinase
cre_117672	flavonoid reductase (FR), extended (e) SDRs	Signal-peptide	NAD(P)+-binding
cre_117686	Predicted ATPase or kinase	Signal-peptide	Uncharacterised P-loop hydrolase UPP0079
cre_117820	AN1-like Zinc finger; putative Helix-turn-helix domain of transposase IS666	INTTEGRIN_BETA	AN1-like Zinc finger
cre_117828	Catalytic domain of Protein Tyrosine Kinases receptor tyk kinases.	Signal-peptide	Protein tyrosine kinase
cre_117865	NAD/NADP octopine/nopaline dehydrogenase, alpha-helical domain	IGFBP_N_1; AGOULT_1; Signal-peptide	NAD-dependent glycerol-3-phosphate dehydrogenase N-terminus
cre_117912	Bola-like protein	Signal-peptide	No
cre_117918	Leucine-rich repeats (LRRs), ribonuclease inhibitor (RI)-like subfamily	Signal-peptide	Leucine rich repeat; Uncharacterised protein family UPP0047
cre_117942	Uncharacterized protein family_UPF0047	No	Protein tyrosine kinase
cre_117946	Catalytic domain of Protein Tyrosine Kinases	Signal-peptide	No
cre_117978	HNH endonuclease;	No	No
cre_117999	RING-finger (Really Interesting New Gene) domain	Signal-peptide	Ring finger domain
cre_118003	S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase), class I	Signal-peptide	Ribosomal protein L11 methyltransferase (PrmA)
cre_118056	PGAPI-like protein	Signal-peptide	Alpha/beta hydrolase family
cre_118057	S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase), class I; Methyltransferase domain	INTTEGRIN_BETA; Signal-peptide	DinB superfamily; Sulfatase-modifying factor enzyme 1; Methyltransferase domain
cre_118061	Like-Sm protein 14, N-terminal domain	Signal-peptide	Scd6-like Sm domain
cre_118081	Catalytic domain of Cyclin-Dependent protein Kinase-like Serine/Threonine Kinases	IGFBP_N_1; Signal-peptide	Protein kinase domain
cre_118118	RNA recognition motif in eukaryotic RNA-binding protein 25 and similar proteins	Signal-peptide	RNA recognition motif, (a.k.a. RRM, RBD, or RNP domain)
cre_118132	Myosin tail	Signal-peptide	No
cre_118138	Catalytic domain of Protein Kinases	Signal-peptide	Protein kinase domain
cre_118166	Armadillo-beta-catenin-like repeats.	IGFBP_N_1	Spinocerebellar ataxia type 10 protein domain
cre_118177	PWI domain in splicing factors; sensor protein KdpD; Provisional	Signal-peptide; Transmembrane_regions	PWI domain
cre_118179	Catalytic domain of Protein Kinases	Signal-peptide	Protein kinase domain
cre_118180	Oxidoglycolate hydrolase	Signal-peptide	No
cre_118195	HisRS Histidyl-anticodon binding domain.	EGGSHELL; Transmembrane_regions	Aminoacid binding domain
cre_118244	Metal dependent phosphonylhydrolases with conserved HD' motif	Signal-peptide; Transmembrane_regions	3'-cyclic nucleotide phosphodiesterase
cre_118289	Met Metal-dependent phosphohydrolases with conserved HD' motif	Signal-peptide	Iron
cre_118289	Met Metal-dependent phosphohydrolases with conserved HD' motif	Signal-peptide	3'-cyclic nucleotide phosphodiesterase
cre_118285	Catalytic domain of Protein Kinases	Signal-peptide; Transmembrane_regions	Protein kinase domain
cre_118391	Dihydroxyline synthase-like (DUS) like FMN-binding domain.	IGFBP_N_1; Signal-peptide	Dihydroxyline synthase (Dus)
cre_118399	Domain of unknown function (DUF3346)	No	Domain of unknown function (DUF4346)
cre_118402	Dullard-like phosphatase domain	Signal-peptide	NLI interacting factor-like phosphatase
cre_118405	Methyltransferase involved in Williams-Beuren syndrome	EGGSHELL; Signal-peptide; Transmembrane_regions	Methyltransferase domain; Methyltransferase involved in Williams-Beuren syndrome
cre_118437	Collagen triple helix	Signal-peptide; Transmembrane_regions	No
cre_118469	WD40 domain	Signal-peptide	WD domain; G-beta repeat
cre_118470	AAA domain	No	AAA domain
cre_118587	Quinone oxidoreductase (QOR)	Signal-peptide	Alcohol dehydrogenase GroES-like domain; Zinc-binding dehydrogenase
cre_118615	GT_2 like, a represents a glycosyltransferase family-2	Signal-peptide	Glycosyl transferase family 2
cre_118641	Protein of unknown function (DUF3735)	IGFBP_N_1	Protein of unknown function (DUF3735); Abscisic acid G-protein coupled receptor
cre_118675	R3H domain of a group of proteins with unknown function	IGFBP_N_2; Signal-peptide	ATPase family associated with various cellular activities (AAA)
cre_118677	EF-hand calcium binding motif	Signal-peptide	EF-hand domain
cre_118714	Domain of unknown function (DUF3411)	IGFBP_N_2; Signal-peptide; Transmembrane_regions	Domain of unknown function (DUF3411)
cre_118754	Rat sarcoma (Ras)	No	Guanylate-binding protein, N-terminal domain
cre_118915	LEM3 (ligand-effect modulator 3) family / CDC50 family	Signal-peptide	LEM3 (ligand-effect modulator 3) family / CDC50 family
cre_118927	GHMP kinases N terminal domain	Signal-peptide	GHMP kinases N terminal domain
cre_118928	ankyrin repeats	IGFBP_N_1; Transmembrane_regions; Signal-peptide	No
cre_118935	Domain of unknown function	No	Domain of unknown function (DUF4205)
cre_118937	Catalytic domain of AGC family Protein Serine/Threonine Kinases	Signal-peptide	Protein kinase domain
cre_119014	Non	Signal-peptide	Copine
cre_119111	Non Willebrand factor type A (vWA) domain	INTTEGRIN_BETA; Signal-peptide; Transmembrane_regions	No
cre_119114	The Major Facilitator Superfamily (MFS)	Signal-peptide; Transmembrane_regions	Major Facilitator Superfamily
cre_119116	classical (c) SDR, subgroup 1	Signal-peptide	short chain dehydrogenase
cre_119122	Halocid dehalogenase-like hydrolases.	Signal-peptide	Halocid dehalogenase-like hydrolase
cre_119140	envelope glycoprotein C; EBNA-3B; Streptogramin lyase	INTTEGRIN_BETA; Signal-peptide; Transmembrane_regions	No
cre_119156	IBR domain	Signal-peptide	IBR domain
cre_119168	Protein of unknown function (DUF1264)	Myb-like DNA-binding domain	Protein of unknown function (DUF1264)
cre_119194	SWI3, ADA2, N-CoR and TFIIB DNA-binding domains.	EGGSHELL	Protein of unknown function (DUF1264)
cre_119201	RING-type zinc-finger, Lsh dimerisation motif	Signal-peptide	CT1L/CRA-C-terminal to Lsh motif domain
cre_119209	zinc finger	Signal-peptide	Zinc finger C-x8-C-x3-H type (and similar)
cre_119217	No	Signal-peptide	Iron-containing redox enzyme

cre: 119222	AMMECR1
cre: 119237	No
cre: 119262	Brown fat-inducible thioesterase (BFTT).
cre: 119263	SWI3, ADA2, N-Cor and TFIIB DNA-binding domains
cre: 119294	Polytopic alpha-helical domain
cre: 119311	ATP-binding cassette domain of the lipid transporters, subfamily A
cre: 119320	Catalytic domain of Protein Tyrosine Kinases
cre: 119338	Elongation Factor 1-alpha (EF1-alpha) protein family
cre: 119360	diphthamide biosynthesis enzyme Dph1/Dph2 domain
cre: 119366	Lysine-rich repeat (LRRs), ribonuclease inhibitor RD-like subfamily
cre: 119438	S-adenosylmethionine-dependent methyltransferases (SAM or AdoMet-MTase), class I
cre: 119447	Catalytic domain of Cyclin-Dependent protein Kinase Like Serine/Threonine Kinases
cre: 119467	Predicted ATPase
cre: 119482	Catalytic domain of the Protein Serine/Threonine Kinase

Signal-peptide	AMMECR1
Signal-peptide	No
Signal-peptide	Thioesterase superfamily
Signal-peptide	MvB-like DNA-binding domain
EGO5/HELL_Signal-peptide	No
EGO5/HELL_Signal-peptide	ABC transporter
Signal-peptide	Protein tyrosine kinase
J_ACTX_Signal-peptide; Transmembrane_regions	Elongation factor Tu GTP binding domain
IGFBP_N_I_Signal-peptide	Putative diphthamide synthesis protein
Signal-peptide	Lysine Rich repeats
Transmembrane_regions; Signal-peptide	Macrocyclic O-methyltransferase (TyfF)
Signal-peptide	Protein kinase domain
Signal-peptide	ATPG-I-like ATPase
Signal-peptide	Protein kinase domain

4. CONCLUSION

Authors have applied genomics approach to recommend the potential functions for hypothetical proteins of *Chlamydomonas reinhardtii*. The prediction of functions of proteins can be analyzed by using accurate annotations. Thus, the functions of proteins can help in better knowing about various regulatory mechanisms that interact to yield a final output from the system. Current study can offer activities associated with nitrogen fixation, acid biosynthesis, CO₂ fixation, ATP formation, electron transport reactions, photosynthesis function, identified proteins crucial for the harvesting and use of excitation energy.

5. REFERENCES

- [1] Boesger, J., Wagner, V., Weisheit, W., Mittag, M. 2009. Analysis of Flagellar Phosphoproteins from *Chlamydomonas reinhardtii*. *Eukaryotic cell.* 8(7), 922-932.
- [2] Harris, E. 2001. Chlamydomonas as a model organism. *Annu Rev Plant Physiol Plant Mol Biol.* 52, 363-406.
- [3] Rochaix, J., 1995. *Chlamydomonas reinhardtii* as the photosynthetic yeast. *Annu. Rev. Genet.* 29, 209-230.
- [4] Merchant, S., Prochnik, S., Vallon, O., Harris, E., Karpowicz, S., Witman, G., et al. 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 318, 245-250.
- [5] Li, J., 2004. *Cell.* 117, 541.
- [6] Storms, R., Hastings, P. 1977. *Exp Cell Res.* 104, 39.
- [7] Merchant, S., Allen, M., Kropat, J., Moseley, J., Long, J., Tottey, S., Terauchi, A., 2006. Between a rock and a hard place: trace element nutrition in *Chlamydomonas*. *Biochim Biophys Acta* 1763, 578-594.
- [8] Grossman, A. R., Karpowicz, S. J., Heinrichs, M., Dewez, D., Hamel, B., Dent, R., Niyogi, K. K., Johnson, X., Alric, J., Wollman, F. A., Li, H., Merchant, S. S. 2010. Phylogenomic analysis of the *Chlamydomonas* genome unmasks proteins potentially involved in photosynthetic function and regulation. *Photosynth Res.* 106, 3-17.
- [9] Gore, D. 2009. In silico Prediction of Structure and Enzymatic Activity for Hypothetical Proteins of *Shigella flexneri*. *Biofrontiers.* 1(2), 1-10.
- [10] Edward, E., Gary, L. G., Osnat, H., John, M., John, O., Roberto, J. P., Linda, B., Delwood, R., Andrew, J. H. 2000. Biological function made crystal clear- annotation of hypothetical proteins via structural genomics. *Current Opinion in Biotechnology.* 11, 25-30.
- [11] Choi, E. J., Bae, S. M., Ahn, W. S. 2008. Antiproliferative effects of quercetin through cell cycle arrest and apoptosis in human breast cancer MDA-MB-453 cells. *Arch Pharm Res.*
- [12] Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: "a new generation of protein database search programs". *Nucleic Acids Res.* 25 (17), 3389-402.
- [13] Schaffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. I., Koonin, E. V., Altschul, S. F. 2001. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* 29(14), 2994-3005.
- [14] Aron, M., Bauer, John, B. A., Myra, K. D., Carol, D. S., Noreen, R. G., Marc, G., Luning, H., Siqian, H., David, I. H., John, D.J., Zhaoxi, K., Dmitri, K., Christopher, J. L., Cynthia, A. L., Chunlei, L., Fu, L., Shennan, L., Gabriele, H. M., Mikhail, M., James, S. S., Narmada, T., Roxanne, A. Y., Jodie, J. Y., Dachuan, Z., Stephen, H. B. 2006. CDD: "a conserved domain database for interactive domain family analysis. " *Nucleic Acids Research.* 35, 237-240.
- [15] Marchler-Bauer, A. 2011. *Nucleic Acids Res.* 39, 225.
- [16] Zdobnov, E. M., Rolf, A. 2001. Interproscan- an integration platform for the signatures recognition methods in InterPro. *Bioinformatics.* 17, 847-848.
- [17] Alex, B., Lachlan, C., Richard, D., Robert, D. F., Volker, H., Sam, G. J., Ajay, K., Mhairi, M., Simon, M., Erik, L. L. S., David, J. S., Corin, Y., Sean, R. E. 2004. The Pfam families' database. *Nucleic Acids Research.* 32, 138-141.
- [18] Finn, R. D. 2010. *Nucleic Acids Res.* 38, 211.
- [19] Grossman, A. R., Steven, J., Karpowicz, M., Heinrichs, D., Dewez, B., Hamel, R., Dent, T., Niyogi, K. K. 2010. Phylogenomic analysis of the *Chlamydomonas* genome unmasks proteins potentially involved in photosynthetic function and regulation. *Photosynthesis research* 106, 3-17.
- [20] Misra, N., Panda P. K., Parida B. K., Mishra B. K. 2012. Phylogenomic Study of Lipid Genes Involved in Microalgal Biofuel Production—Candidate Gene Mining and Metabolic Pathway Analyses. *Evolutionary bioinformatics online* 8, 545.