International Research Journal of Biological Sciences _____ ISSN 2278-3202
Vol. **2(12),** 50-59, December **(2013)** *Int. Res. J. Biological Sci.*

Comparative study of Glycerate Kinase (GK): Bioinformatical Approach

Mondal Sunil Kanti*, Neogi Chandra and Choudhury Kuntal
Department of Biotechnology, The University of Burdwan, Golapbag, Burdwan, 713104, West Bengal, INDIA

Available online at: www.isca.in, www.isca.me
Received 1st August 2013, revised 12th September 2013, accepted 7th October 2013

Abstract

There are three classes of Glycerate kinase (GK) which are class I GK, class II GK and class III GK. Class I and class II GKs produce glycerate 2-phosphate whereas class III GK (GLYK) only can produce glycerate 3-phosphate. Phylogenetic analysis on 16S ribosomal RNA sequences reveals the strong evolutionary relationship between cyanobacteria and plants. Phylogeny using GK DNA and amino acid sequences shows that cyanobacteria group is closely related with both bacteria and plants whereas fungi are closely related only with plants. Phylogeny using the amino acid sequence and hierarchical clustering on the basis of the amino acid frequencies of GK shows similar relationship among the taxa. Hierarchical clustering on the basis of GC% of GK encoding gene showing the unusual property like the RSCU value of the codons UUG and AGG are significantly low and CGA is significantly high in GC rich cluster. Correlation coefficient between GC% and the amino acids arginine, tryptophan and serine shows that the plants are different from the other selected species. ENc plot shows that except few GK genes from fungi and gammaproteobacteria all of them are under mutational bias. There is no as such codon usage similarity for the GK encoding gene from different organisms but they have similar degree of expression i.e, CAI (highest in plant) which is significantly low along with the amino acids lysine, phenylalanine, tyrosine, isoleusine and asparagine and serine in GC rich GK encoding gene.

Keywords: CAI: Codon Adaptation Index; ENc: Effective number of codon; RSCU: Relative Synonymous Codon Usage.

Introduction

There are variety of D-glycerate forming pathway which reflect the existence of three classes of glycerate kinases (GK)[1] which are phylogenetically distinct. In bacteria (e.g. *E coli*) with one reported exception[2], the Class I GKs are thought to produce glycerate 3-phosphate (3PGA) in bacterial glucarate and glycolate metabolism[3,4]. But recently seen that class I GKs in glucarate and glycolate metabolism[5] producing only glycerate 2-phosphate (2PGA) not 3PGA. Class II GKs form 2PGA in Archae by sugar degradation via non-phosphorylating branch of Entner–Doudoroff pathway[6] and in animals by serine degradation and fructose metabolism. Functionally defective human GK causes a hereditary disease D-glyceric aciduria[7]. Class III GKs is known as GLYK and only enzyme which form 3PGA[5] and it is one of the essential core enzymes which catalyses the terminal reaction of the photorespiration cycle in plants[8]. Photorespiration also occurs in cyanobacteria[9] like filamentous *Nostoc sp.* strain PCC 7120[5] which are the close relatives of endosymbiontic plastid ancestor[10] but surprisingly some unicellular cyanobacteria such as *Synechocystis sp.* strain PCC 6803 have class I GK's. The enzyme GLYK purified from *Saccharomyces cerivisiae*[11] and some plants[12] but recently identified in the genome of *Arabidopsis thaliana* where GLYK is positioned at the end of the chromosome 1 right arm comprising 11 exons and 1368 bp open reading frame corresponds to a protein of 456 amino acids. The protein also includes chloroplast transit peptide (cTP) of 118 amino acids and ATP/GTP binding site and also identified two highly specific domains GLYK -1domain found in GLYK and its homologous proteins of plant or non plant. But GLYK- 2 domain is specific for plant and yeast and absent in cyanobacteria[1]. The present study deals with the characterization of the GK based on amino acids frequency, encoding codons RSCU, genes GC%, evolutionary pattern of the genes in prokaryotes (proteobacteria and cyanobacteria) and eukaryotes (fungi and plants). An extensive analysis have been done to reveal the compositional variation and similarity (GC content, amino acid frequency and codon bias) and also to understand the mutational pressure on the GK genes.

Material and Methods

Collection of Data: Taxonomic and other related information of 145 organisms were collected (46 proteobacteria spp., 33 cyanobacteria spp., 52 fungi spp. and 11 plant spp.) and their nuclieotide sequence, amino acid sequence, 16S ribosomal RNA sequence and also GK gene and protein sequence collected from KEGG database (www.genome.jp/kegg/) and from NCBI.

Evolutionary Analysis: We have generated thousand time boot strapped phylogenetic tree[13,14] using ClustalW[15-17] (www.ebi.ac.uk/tools/msa/clustalw2), PHYLIP version 3.69 [18,19,20] and Tree view software[19] of nucleotide sequence, 16S rRNA sequence and amino acid sequence of GK.

Compositional Analysis: The parameters like GC content, amino acid frequencies and RSCU (Relative Synonymous

Codon Usage) were used for compositional analysis. We have generated GK encoding genes GC1%, GC2%, GC3%, A3%, U3%, Nc[22] and gravy score using codonW and inhouse PERL script. The GC1%, GC2%, GC3% of the organisms was collected from Codon Usage Database. We have also calculated amino acid frequency of GK protein, RSCU values of all codons encoding GK protein using in house or home based PERL programme. These provided useful information regarding existence of mutational pressures acting on the genes[23]. ENc, the expected effective number of codon were calculated from GC3s under H0 (Null hypothesis, i.e., no selection) according to the given equation, where S denotes GC3.

$$Enc=2+S+\{29/[S^2+(1-S)^2]\} \quad (1)$$

RSCU, the relative synonymous codon usage values close to one indicates lack of biasness where as much higher and lower values indicate preference and avoidance of those particular codons, respectively. Using codonW, the correspondence analysis[24] has been performed to investigate major trend in RSCU variation among genes and distribute the genes along continuous axes in accordance with these trends also we have calculated gravy score to know about the hydrophobicity / hydrophilicity of the protein.

Hierarchical Cluster Analysis: Hierarchical clustering based on gene GC%, amino acid frequencies and RSCU of GK were generated using programme DIANA of the package cluster of R-Statistical software[25,26].

Expressional Probability: The geometric mean of the weight associated to each codon over the length of the gene sequence (measured in codons) is known as Codon Adaptatation Index (CAI) i.e. the measure of gene's probable expression. We have calculated it by following Sharp and Li method[27] and using inhouse PERL script and MS Excel 2007.

Statistical Analysis: We have calculated correlation coefficient and RSQ values between GC content and amino acids frequencies of GK in different groups by using MS excel 2007. Another statistical significance (z) test was performed based on GC1%, GC2%, GC3%, GC%, G3%, C3%, A3%, U3%, CAI and all amino acids of GK protein sequence and GK encoding nucleotide sequence for different situations: i. between high and low GC poor clusters for the same genus, ii. within GC rich cluster for genus to genus, iii. within GC poor cluster for genus to genus (clusters were obtained through hierarchical clustering on the basis of GC% of the GK encoding gene) and iv. within the population (organism set considered for this study) for genus to genus.

$$Z_i = (AVERAGE_a - AVERAGE_b)/\sqrt{((STDEV_a^2+STDEV_b^2)/(N_a + N_b))} \quad (2)$$

Here, i denotes the parameter, a and b denotes the genus and N indicates the sample size.

Results and Discussion

Phlogenetic Analysis: Phylogenetic tree (figure - 1A) on the basis of 16s ribosomal RNA sequences showning that the species of the plant group are related closely. Then the plant group clustered with cyanobacteria group more closely than gammaproteobacteria and alphaproteobacteria groups.

Phylogenetic tree on the basis of GK amino acid sequence shown in figure - 1B also have two main clusters, in the first one chlorophyta group and streptophyta group are clustered closely and these two groups at first clustered with fungi group and then with cyanobacteria and some organisms of alphaproteobacteria group. Second cluster formed between gammaproteobacteria and other organisms of alphaproteobacteria group denoting their close relation with each other.

Phylogenetic tree on the basis of GK encoding DNA sequence shown in figure - 1C have two major clusters, in first cluster some organisms belong to fungi group clustered closely with some organisms of cyanobacteria group then this two clustered with some organisms of gammaproteobacteria group and with some other organisms of cyanobacteria, fungi and streptophyta group. In second cluster some organisms of cyanobacteria group clustered closely with some organisms of gammaproteobacteria group and this two groups clustered at first clustered with alphaproteobacteria group and then with the other organisms gammaproteobacteria and cyanobacteria groups and then this two clusters joined with the major number of organisms of fungi group.

Compositional Variability: Hierarchical clustering on the basis of GC%, amino acid frequencies and RSCU values of codons in GK encoding DNA sequence gives the similar variation. The relation obtained through hierarchical clustering on the basis of the amino acid frequencies shows that fungi, chlorophyta and streptophyta are closely clusterd with some organisms of cyanobacterial group. The other organisms of cyanobacterial group are separately clustered with alphaproteobacteria and gammaproteobacteria groups.

Statistical significance test (at $P<0.05$) shows (table - 2) that within the GC (60% ± 7%) rich cluster (Cluster IandII in figure - 2A) the parameters like: CAI, K, F, Y, M, I, N and S of the GK encoding gene are significantly low where as the parameters like: R, W, P, G, A, V, Q are significantly high for all genus (with some non significance like CAI, W and Q in FA and V in PG) than the GC (40% ± 7%) poor cluster. In the GC rich cluster, the parameters C, H, T are significantly high in CY and low in PG and within FA, C is high and E is low than the GC poor cluster but D has no significant variation.

Table–1
List of organisms selected to study GK

HC*	Organisms Full Name	PC#	HC	Organisms Full Name	PC	HC	Organisms Full Name	PC
1	Chlamydomonas reinhardtii	cre_CL	137	Teredinibacter turnerae	ttu_PG	25	Prochlorococcus marinus MIT 9211	pmj_CY
95	Caulobacter sp. K31	cak_PA	33	Synechococcus sp. PCC7002	syp_CY	17	Nostoc punctiforme PCC 73102	npu_CY
130	Pseudoxanthomonas spadix	psd_PG	127	Nitrosococcus watsonii	nwa_PG	47	Botryotinia fuckeliana	bfu_FA
106	Sphingobium japonicum	sjp_PA	41	Aspergillus fumigatus	afm_FA	108	Alteromonas sp. SN2	alt_PG
43	Ashbya gossypii (Eremothecium gossypii)	ago_FA	111	Alteromonas macleodii Deep ecotype	amc_PG	112	Alteromonas macleodii English Channel 673	amg_PG
105	Sphingobium chlorophenolicum	sch_PA	113	Alteromonas macleodii Black Sea 11	amk_PG	109	Alteromonas macleodii ATCC 27126	amac_PG
134	Stenotrophomonas maltophilia K279a	sml_PG	126	Nitrosococcus oceani	noc_PG	122	Marinomonas mediterranea	mme_PG
135	Stenotrophomonas maltophilia R551-3	smt_PG	38	Phytophthora infestans	pif_EK	91	Zygosaccharomyces rouxii	zro_FA
114	Stenotrophomonas maltophilia JV3	buj_PG	119	Marinobacter adhaerens	mad_PG	117	Glaciecola sp. 4H-3-7+YE-5	gag_PG
96	Caulobacter crescentus CB15	ccr_PA	55	Cyanidioschyzon merolae	cme_FA	53	Coccidioides immitis	cim_FA
97	Caulobacter crescentus NA1000	ccs_PA	64	Lachancea thermotolerans	lth_FA	51	Cryptococcus gattii	cgi_FA
104	Phenylobacterium zucineum	pzu_PA	57	Cryptococcus neoformans JEC21	cne_FA	128	Pseudoalteromonas atlantica	pat_PG
98	Caulobacter segnis	cse_PA	44	Aspergillus niger	ang_FA	142	Zea mays (maize)	zma_ST
107	Sphingomonas wittichii	swi_PA	138	Oryza sativa japonica (Japanese rice)	dosa_ST	66	Moniliophthora perniciosa	mpr_FA
144	Arabidopsis thaliana (thale cress)	ath_ST	56	Cryptococcus neoformans B-3501A	cnb_FA	110	Alteromonas macleodii Balearic Sea AD45	amb_PG
49	Coprinopsis cinerea	cci_FA	82	Sordaria macrospora	smp_FA	42	Aspergillus flavus	afv_FA
46	Aspergillus oryzae	aor_FA	87	Ustilago maydis	uma_FA	116	Frateuria aurantia	fau_PG
67	Myceliophthora thermophila	mtm_FA	2	Ostreococcus lucimarinus	olu_CL	121	Methylophaga sp. JAM7	mec_PG
101	Novosphingobium sp. PP1Y	npp_PA	141	Sorghum bicolor (sorghum)	sbi_ST	143	Vitis vinifera (wine grape)	vvi_ST
14	Gloeobacter violaceus PCC7421	gvi_CY	145	Brachypodium distachyon	bdi_EK	124	Marinomonas posidonica	mpc_PG
100	Hyphomonas neptunium	hne_PA	71	Nectria haematococca	nhe_FA	58	Coccidioides posadasii	cpw_FA
136	Stenotrophomonas maltophilia D457	smz_PG	6	Anabaena sp. PCC7120	ana_CY	21	Prochlorococcus marinus MIT 9303	pmf_CY
34	Synechococcus sp. RCC307	syr_CY	74	Penicillium chrysogenum	pcs_FA	84	Sclerotinia sclerotiorum	ssl_FA
54	Clavispora lusitaniae	clu_FA	132	Pseudoxanthomonas suwonensis	psu_PG	77	Phaeosphaeria nodorum	pno_FA
12	Cyanothece sp. PCC 8801	cyp_CY	60	Debaryomyces hansenii	dha_FA	27	Prochlorococcus marinus MIT 9313	pmt_CY

HC*	Organisms Full Name	PC#	HC	Organisms Full Name	PC	HC	Organisms Full Name	PC
36	Synechococcus sp. WH7803	syx_ CY	89	Vanderwaltozyma polyspora	vpo_ FA	73	Paracoccidioides brasiliensis	pbl_ FA
35	Synechococcus sp. WH8102	syw_ CY	139	Medicago truncatula (barrel medic)	mtr_ ST	75	Puccinia graminis	pgr_ FA
81	Schizophyllum commune	scm_ FA	7	Anabaena variabilis ATCC 29413	ava_ CY	39	Arthroderma benhamiae	abe_ FA
4	Volvox carteri f. nagariensis	vcn_ CL	15	Microcystis aeruginosa NIES-843	mar_ CY	63	Laccaria bicolor	lbc_ FA
92	Asticcacaulis excentricus	aex_ PA	62	Kluyveromyces lactis	kla_ FA	86	Trichophyton verrucosum	tve_ FA
3	Ostreococcus tauri	ota_ CL	140	Glycine max (soybean)	gmx_ ST	61	Fusarium graminearum	fgr_ FA
120	Methylococcus capsulatus	mca_ PG	16	Anabaena azollae 0708	naz_ CY	79	Pyrenophora teres	pte_ FA
29	Synechococcus sp. CC9605	syd_ CY	99	Hirschia baltica	hba_ PA	85	Tuber melanosporum	tml_ FA
65	Magnaporthe oryzae	mgr_ FA	123	Marinomonas sp. MWYL1	mmw _PG	13	Cyanothece sp. ATCC 51142	cyt_ CY
28	Synechococcus elongatus PCC6301	syc_ CY	118	Glaciecola nitratireducens	gni_P G	37	Trichodesmium erythraeum IMS101	ter_ CY
31	Synechococcus elongatus PCC7942	syf_ CY	131	Pseudoalteromonas sp. SM9913	psm_ PG	20	Prochlorococcus marinus NATL1A	pme_ CY
90	Yarrowia lipolytica	yli_ FA	10	Cyanothece sp. PCC 7822	cyj_ CY	26	Prochlorococcus marinus NATL2A	pmn_ CY
40	Aspergillus clavatus	act_ FA	70	Neosartorya fischeri	nfi_ FA	18	Prochlorococcus marinus AS9601	pmb_ CY
80	Saccharomyces cerevisiae (budding yeast)	sce_ FA	93	Alpha proteobacterium HIMB59	apc_ PA	94	Alpha proteobacterium HIMB5	apm_ PA
68	Neurospora crassa	ncr_ FA	83	Schizosaccharomyces pombe (fission yeast)	spo_ FA	102	Candidatus Pelagibacter sp. IMCC9063	pel_ PA
72	Podospora anserina	pan_ FA	78	Pichia pastoris	ppa_ FA	19	Prochlorococcus marinus MIT 9515	pmc_ CY
5	Acaryochloris marina	amr_ CY	76	Meyerozyma guilliermondii	pgu_ FA	22	Prochlorococcus marinus MIT 9301	pmg_ CY
88	Uncinocarpus reesii	ure_ FA	129	Pseudoalteromonas haloplanktis TAC125	pha_ FA	23	Prochlorococcus marinus MIT 9215	pmh_ CY
125	Nitrosococcus halophilus	nhl_ PG	8	Cyanothece sp. PCC 7424	cyc_ CY	24	Prochlorococcus marinus MIT9312	pmi_ CY
133	Saccharophagus degradans	sde_ PG	52	Candida glabrata	cgr_ FA	103	Candidatus Pelagibacter ubique	pub_ PA
11	Cyanothece sp. PCC 7425	cyn_ CY	69	Naumovozyma castellii	ncs_ FA	48	Candida albicans	cal_ FA
45	Aspergillus nidulans	ani_ FA	9	Cyanothece sp. PCC 8802	cyh_ CY	50	Candida dubliniensis	cdu_ FA
30	Synechococcus sp. CC9902	sye_ CY	115	Colwellia psychrerythraea 34H	cps_ FA	59	Candida tropicalis	ctp_ FA
32	Synechococcus sp. CC9311	syg_ CY	*H.C.: Code number used in Hierarchical Clustering; #P.C.: Code used in Phylogeny					

P.C.: triple letter code from KEGG _ genus (chlorophyta (CL-plant), streptophyta (ST-plant), cyanobacteria (CY), gammaproteobacteria (PG), alphaproteobacteria (PA), fungi (FA).

Table-2

Score of different genomic and proteomic parameters of the GK encoding gene from CY, FA, PG, PL (plant). Z – score which are greater/less than ±1.96 (at P<0.05) are significant and marked as bold. Positive significant values indicating that the parameters are high/greater in GC rich cluster than GC poor cluster for the respective (same) genus in column no. II to IV and the parameters of the first genus are high/greater than the later genus within the corresponding cluster/population in the rest columns V to XV

Parameters/ Sample/ Genus Combinations	Between high and low cluster			GC HIGH CLUSTER			GC LOW CLUSTER			Population				
	CY	FA	PG	CY FA	CY PG	FA PG	CY FA	CY PG	FA PG	CY FA	CY PG	FA PG	PL CY	PL FA
I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV
GC1%	**13.53**	**11.38**	**10.65**	1.16	-1.82	**-3.12**	**-3.07**	**-4.29**	-0.67	**-2.83**	**-3.22**	-1.21	**3.88**	**2.47**
GC2%	**14.12**	**12.99**	**9.16**	0.92	**-2.47**	**-3.75**	**-2.61**	**-6.3**	**-3.78**	**-2.6**	**-4.36**	**-2.74**	**4.47**	**3.43**
GC3%	**13.1**	**11.6**	**9.5**	0.21	**-2.46**	**-3.09**	**-3.69**	**-5.23**	-1.58	**-3.62**	**-4.02**	-1.5	**4.56**	**2.69**
GC	**14.25**	**12.67**	**10.45**	0.98	**-2.15**	**-3.39**	**-3.31**	**-5.48**	-1.81	**-3.01**	**-3.74**	-1.69	**4.38**	**3.02**
G3%	**10.6**	**6.93**	**6.84**	0.92	-0.3	-1.16	**-3.34**	**-2.71**	0.49	**-3.14**	**-2.4**	0.26	**3.8**	**2.17**
C3%	**12.81**	**11.13**	**8.12**	-0.96	**-3.71**	**-3.91**	**-3.03**	**-5.73**	**-2.82**	**-3.54**	**-5.02**	**-2.69**	**4.66**	**2.46**
A3%	**-11.78**	**-10.52**	**-7.58**	**-3.63**	-0.17	**2.76**	-0.54	**2.24**	**3.29**	-0.08	1.57	1.96	**-4.11**	**-5.36**
U3%	**-9.59**	**-7.95**	**-8.54**	**3.91**	**4.22**	**2.86**	**5.95**	**5.32**	-1.1	**6.69**	**5.77**	0.74	**-4.14**	0.19
CAI	**-2.15**	-0.68	**-6.61**	0.25	**2.78**	**2.67**	1.62	**-4.17**	**-5.07**	1.56	0.88	-0.31	**2.83**	**3.78**
asp(D)	-1.02	-1.79	-0.78	0.22	0.39	0.31	-0.32	0.47	0.87	-0.43	1.42	1.95	-0.11	-0.35
glu(E)	-1.54	**2.11**	-0.74	**-5.02**	-1.28	**3.2**	**-2.29**	-0.94	1.42	**-5.48**	-1.4	**3.83**	0.9	**-4.05**
his(H)	1.39	1.59	**-2.72**	**-3.91**	-1.51	1.61	**-3.71**	**-6.35**	**-3.04**	**-6.22**	**-5.41**	-0.55	**2.08**	**-2.28**
lys(K)	**-12.08**	**-9.14**	**-10.81**	**-9.23**	-0.17	**7.42**	0.21	**2.06**	**3.03**	-1.29	1.75	**3.91**	1.44	1.2
arg(R)	**8.98**	**8.78**	**9.59**	1.51	**-2.18**	**-4.25**	-1.07	-0.64	0.38	-1.2	**-2.06**	-1.45	0.69	-0.22
phe(F)	**-5.58**	**-4.37**	**-7.63**	**-4.83**	-1.17	**3.79**	-0.07	-0.87	-1.23	-1.38	-0.67	0.62	0.17	-0.88
tyr(Y)	**-5.97**	**-4.33**	**-7.77**	**-10.93**	**-2.72**	**10.5**	**-6.39**	**-3.42**	**3.98**	**-9.98**	**-2.99**	**6.57**	**8.09**	0.65
trp(W)	**2.58**	0.23	**3.01**	**13.94**	**7.72**	**-5.93**	**10.22**	**7.24**	**-4.24**	**16.87**	**10.75**	**-7.11**	**-9.88**	**3.94**
pro(P)	**6.56**	**6.47**	**5.49**	**4.81**	1.52	-1.83	**2.69**	**2.11**	-0.59	**3.21**	1.68	-1.02	**-2.01**	-0.04
cys(C)	**5.95**	-1.93	**-3.95**	**7.95**	**5.52**	-1.36	-1.35	**-3.84**	**-3.95**	**2.59**	-0.42	**-4.16**	-0.47	1.79
met(M)	**-3.22**	**-5.36**	-1.22	1.76	-1.09	**-3.1**	0.12	-0.38	-0.5	1.72	-0.44	-1.66	**2.37**	**3.76**
gly(G)	**3.94**	**3.45**	**3.79**	-0.84	**3.13**	**5.19**	**-2.35**	**3.14**	**6.33**	**-3.52**	**3.98**	**8.77**	0.69	-1.55
ala(A)	**10.54**	**8.81**	**6.89**	1.89	**-4**	**-5.9**	0.06	**-5.19**	**-5.88**	0.05	**-5.78**	**-6.63**	**2.95**	**3.56**
val(V)	**3.53**	**5.29**	-0.73	**-7.09**	**-3.21**	**3.8**	**-5.3**	**-7.33**	**-2.06**	**-8.93**	**-8.5**	1.16	**6.92**	0.91
ile(I)	**-11.32**	**-7.77**	**-5.61**	1.05	**2.97**	**3.03**	**4.05**	**7.57**	**3.01**	**4.28**	**6.53**	**3.19**	**-5.11**	**-2.1**
leu(L)	-0.13	-0.69	**7.75**	**12.67**	0.99	**-9.98**	**10.97**	**9.01**	0.51	**17.9**	**6.46**	**-4.83**	**-13.76**	**-2.68**
gln(Q)	**3.76**	0.87	**2.24**	**4.84**	0.45	**-4.21**	0.29	-1.34	**-2.61**	**2.9**	-1.22	**-5.26**	**-4.03**	**-3.7**
asn(N)	**-6.28**	**-8.12**	**-5.82**	**-6.73**	-1.91	**4.44**	-0.16	**2.58**	**5.33**	-0.65	**2.07**	**4.35**	-0.35	-1.36
ser(S)	**-2.29**	**-4.46**	**-8.55**	1.01	**3.12**	**3.31**	0.91	0.15	-1.24	1.77	**2.69**	1.63	0.06	1.83
thr(T)	**2.22**	0.37	**-3.34**	**-3.72**	-1.09	**1.99**	**-5.35**	**-6.15**	**-2.61**	**-6.98**	**-6.05**	-0.88	**4.25**	**-2.21**

Except G3% and D in GC high cluster and D, R, F, M and S in GC low cluster and Dand F in selected organism set for the GK encoding gene the rest calculated parameters were significantly varied from genus to genus. Z-test on the basis of the parameters also showing that GK's from plant is closer to fungi than cyanobacteria and CAI is highest in plant.

Table - 3 depicting that RSCU of the codons within GK having G or C in third position are high in GC rich clusters (IandII in figure - 2B and I in figure - 2C) and low in AT rich clusters (IIIandIV in figure - 2B and IIandIII in figure - 2C) with some exceptions like: the RSCU value of the codons UUG(L) and AGG(R) are significantly low and CGA(R) is significantly high in GC rich cluster.

Correlation coefficients of Gene GC% with amino acids frequencies: Figure - 3 showing that within GK the amino acids like aspartate(D) and glutamate(E) are significantly negatively correlated with GC% in chlorophyta and significantly positively correlated in alphaproteobacteria. Arginine(R) and tryptophan(W) are negatively correlated with GC% in chlorophyta, and streptophyta, and positively correlated to GC% in cyanobacteria, fungi, alphaproteobacteria and gammaproteobacteria whereas serine(S) is negatively correlated with GC% in cyanobacteria, fungi, alphaproteobacteria and also in gammaproteobacteria and positively correlated in chlorophyta and streptophyta. Tyrosine(Y) is only positively correlated with GC% in streptophyta among the six groups whereas isoleusine(I) is only positively correlated with GC% in chlorophyta. Cystein(C), valine(V), glutamine(Q) and threonine(T) are the amino acids present in GK have no significant relation with GC% within these six groups. Within GK the other amino acids like proline(P) and leucine(L) significantly positively correlated with GC%. Glycine(G), alanine(A) significantly positively correlated with GC% except streptophyta. Lysine(K), phenylalanine(F) significantly negatively correlated with GC%. Methionine(M) and asparagine(N) significantly negatively correlated with GC% except streptophyta.

Expressional Variability: Correspondence Analysis: Figure - 4A showing that there is no as such codon usage similarity for the GK encoding gene from different organisms / genus. Fungi is separated from cyanobacteria and proteobacteria gamma along the first major axis (axis1), thereby depicting the difference in codon usage pattern.

ENc plot analysis: The ENc plot analysis (figure - 4B: ENc/Nc ploted against $(G+C)_3$) used to search patterns of synonymous codon usage which shows that except some organisms from fungi group like Arthroderma benhamiae, Cryptococcus gattii, Cryptococcus neoformans JEC21, Paracoccidioides brasiliensis, Phaeosphaeria nodorum, Pyrenophora teres, Schizosaccharomyces pombe (fission yeast) and Trichophyton verrucosum and few from gammaproteobacteria group like Alteromonas sp. SN2 and Colwellia psychrerythraea 34H all of them are under mutational bias.

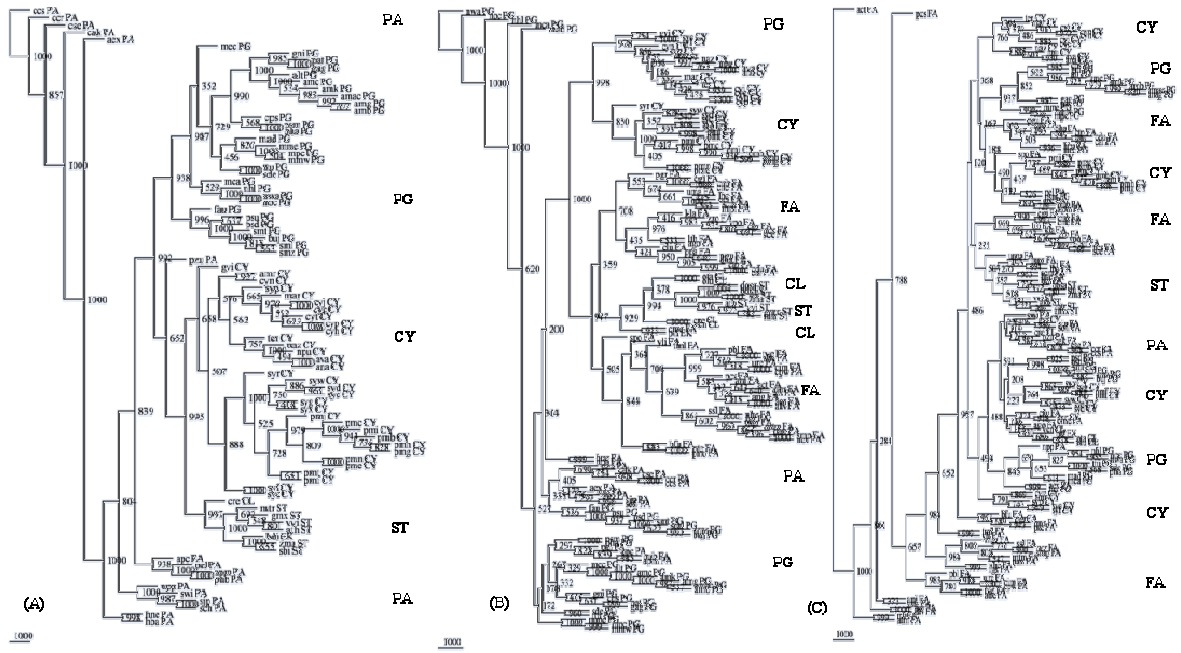

Figure- 1A, 1B, 1C
Phylogenetic trees based on 1A: 16s ribosomal RNA sequences; 1B: GK amino acid sequences; 1C: GK encoding DNA sequences

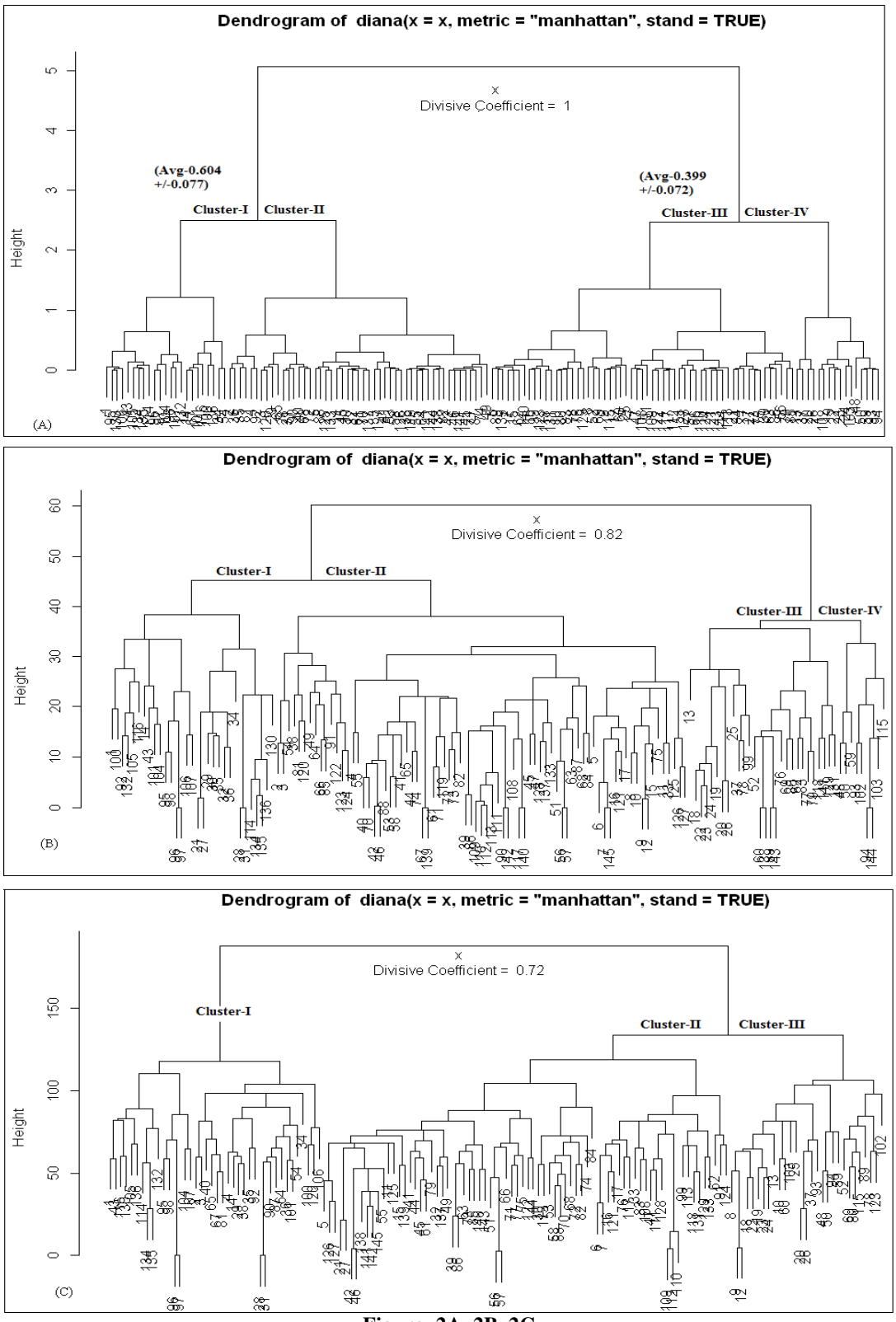

Figure- 2A, 2B, 2C

Dendograms based on 2A: total guanine and cytosine percentage in GK gene (GC%); 2B: amino acid frequencies of GK protein (Gene amino acid); 2C: RSCU values of the codons of GK encoding gene

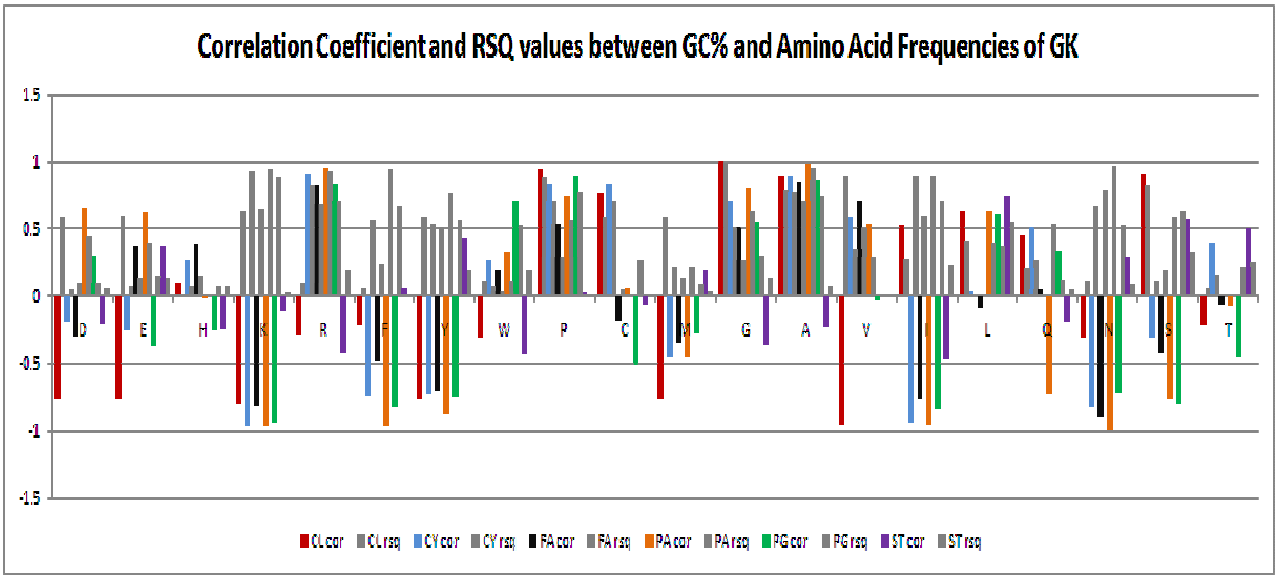

Figure-3
Correlation and RSQ of amino acids frequencies with total guanine and cytosine percentage in GK encoding gene (GC%) in different genus

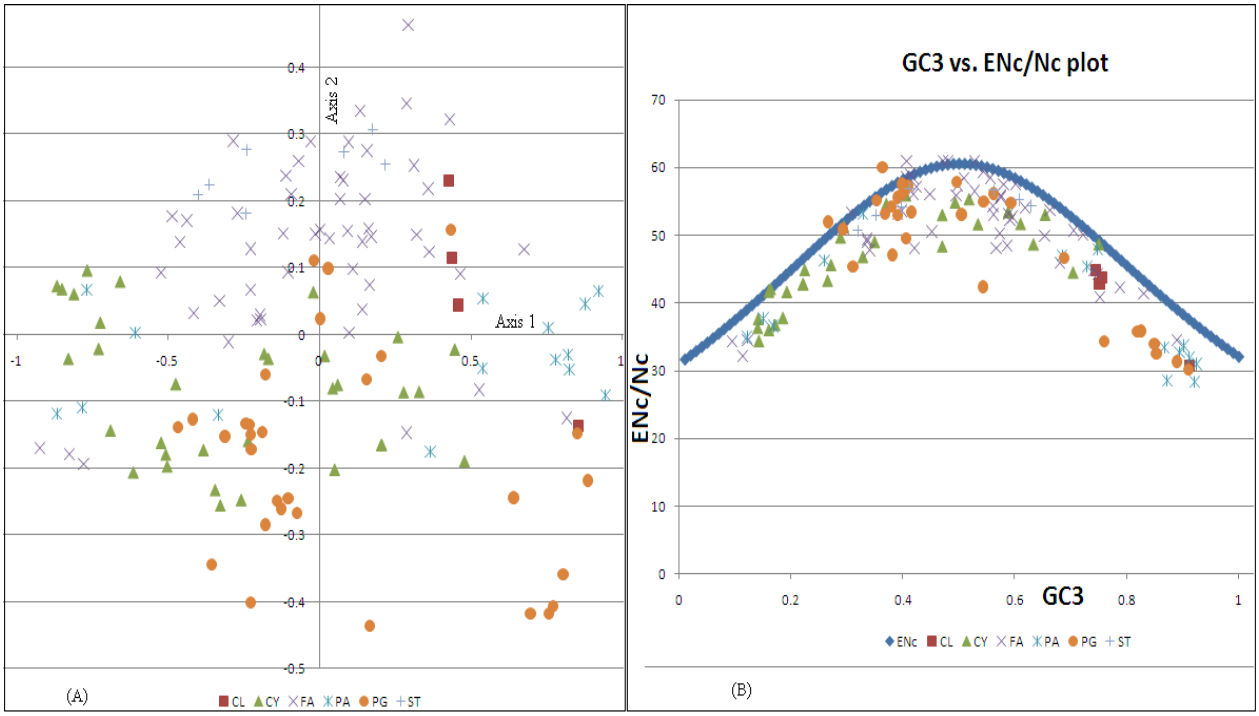

Figure-4A, 4B
4A: Correspondence analysis of the GK gene from the six groups based on its RSCU values; 4B ENc/Nc vs. GC_3 plot of the GK gene from the six groups

CAI and Gravy Score: For these selected genus the CAI is 0.5 and above (as shown in figure - 5A) which predicting the moderate degree of expression of GK encoding gene. In chlorophyta, streptophyta, cyanobacteria, alphaproteobacteria and gammaproteobacteria the gravy score of GK is around -0.3 and for fungi it is -0.45 (as shown in figure - 5B) which indicating that GK from fungi is more hydrophilic (due to negative gravy score) in nature than the GK's from rest of the selected groups.

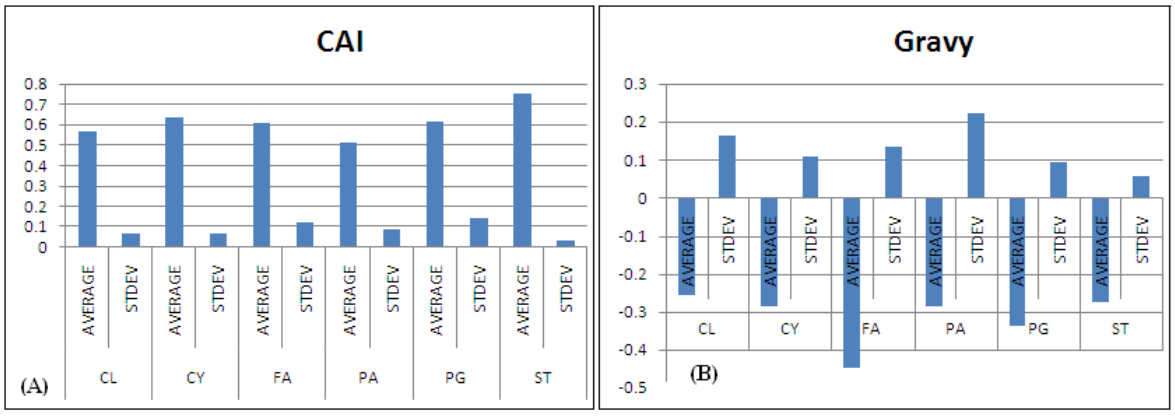

Figure-5A, 5B
Average and standard deviation of 5A: CAI and 5B: Gravy score of the GK encoding gene of the six groups

Conclusion

The genomic and proteomic comparative study of GK from different groups of species highlighted few outcomes. First of all, the results show that GK and its encoding DNA sequences from plant species are similar to fungi and then to cyanobacteria in comparison to others. In addition to that the hierarchical clustering on the basis of different genomic and proteomic parameter of GK giving similar variation. The hierarchical clustering on the basis of the amino acid frequencies of GK shows similar relationship among the groups as obtained through phylogenetic study and some unusual property like the RSCU value of the codons UUG(L) and AGG(R) are significantly low and CGA(R) is significantly high in GC rich cluster. The correlation coefficient between GC% and the amino acids arginine, tryptophan and serine shows that the GK's from plants are different from the other selected species. Enc plot shows that except few GK genes from fungi and gammaproteobacteria all of them are under mutational biasness. There is no as such codon usage similarity for the GK encoding gene from different organisms but they have similar degree of expression (highest in plant) which is low along with some amino acids K, F, Y, I, N and S in GC rich GK encoding gene.

Acknowledgement

We are thankful to Dr. Sudip Kundu, Department of Biophysics, Molecular Biology and Bioinformatics, University of Calcutta for his valuable suggestion about the calculation of the parameters used within this manuscript.

References

1. Boldt R., Edner C., Kolukisaoglu U., Hagemann M., Weckwerth W., Wienkoop S., Morgenthal K. and Bauwe, H., D-Glycerate 3-kinase, the last unknown enzyme in the photorespiratory cycle in Arabidopsis, belongs to a novel kinase family. *Plant Cell*, **17**, 2413–2420 **(2005)**
2. Hubbard B.K., Koch M., Palmer D.R., Babbitt P.C. and Gerlt J.A., Evolution of enzymatic activities in the enolase superfamily: characterization of the (D)-glucarate/galactarate catabolic pathway in Escherichia coli. *Biochemistry*, **37**, 14369–14375 **(1998)**
3. Aghaie A., Lechaplais C., Sirven P., Tricot S., Besnard-Gonnet M., Muselet D., de Berardinis V., Kreimeyer A., Gyapay G., Salanoubat M. and Perret A., New insights into the alternative D-glucarate degradation pathway, *J. Biol. Chem*., **283**, 15638–15646 **(2008)**
4. Cusa E., Obradors N., Baldoma L., Badia J. and Aguilar J., Genetic analysis of a chromosomal region containing genes required for assimilation of allantoin nitrogen and linked glyoxylate metabolism in Escherichia coli, *J. Bacteriol*, **181**, 7479–7484 **(1999)**
5. Bartsch O., Hagemann M., Bauwe H., Only plant –type (GLYK) glycerate kinases produce D-glycerate 3-phosphate. *FEBS Lett*., **582**, 3025–3028 **(2008)**
6. Reher M., Bott M., and Scho¨nheit P., Characterization of glycerate kinase (2-phosphoglycerate forming), a key enzyme of the nonphosphorylative Entner–Doudoroff pathway, from the thermoacidophilic euryarchaeon Picrophilus torridus. *FEMS Microbiol. Lett.,* **259**, 113–119 **(2006)**
7. Van Schaftingen E., D-glycerate kinase deficiency as a cause of D-glyceric aciduria. *FEBS Lett,*. **243**, 127–131 **(1989)**
8. Husic D.W., Husic H.D. and Tolbert N.E., The oxidative photosynthetic carbon cycle or C_2 cycle, *Crit. Rev. Plant Sci*., **5**, 45–100 **(1987)**
9. Eisenhut M., Ruth W., Haimovich M., Bauwea H., Kaplan A. and Hagemann M., The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbiontically to plants, *PNAS,* **105**, 17199-17204 **(2008)**

10. Deusch O., Landan G., Roettger M., Gruenheit N., Kowallik K.V., Allen J.F., Martin W. and Dagan T., Genes of cyanobacterial origin in plant nuclear genomes point to a heterocyst-forming plastid ancestor. *Mol. Biol. Evol*., **25**, 748–761 **(2008)**

11. Black S. and Wright N.G., Enzymatic formation of glyceryl and phosphoglyceryl methylthiol esters, *J. Biol. Chem*., **221**, 171–180 **(1956)**

12. Kleczkowski L.A., Randall D.D. and Zahler W.L., The substrate specificity, kinetics, and mechanism of glycerate-3- kinase from spinach leaves, *Arch. Biochem. Biophys,* **236**, 185–194 **(1985)**

13. Bhattacharya A., Power J. and Davey M., Genetic Manipulation of Gibberellin (GA) Oxidase Genes in Nicotiana sylvestris using constitutive promoter to modify Plant Architecture, *Res.J.Recent Sci*., **1(5)**, 1-7 **(2012)**

14. Maithri S.K., Ramesh K.V. and Mutangana D., Theoretical structure prediction of TcaA from Photorhabdus luminescens and aminopeptidase N receptor from Helicoverpa armigera, *Res. J. Recent Sci.,* **2(2)**, 40-49 **(2013)**

15. Bhatt T.K., Phylogenetic studies on tRNA dependent amidotransferase from *Plasmodium falciparum*, *ISCA J.Biological Sci.,* **1(3)**, 20-24 **(2012)**

16. Dwivedi V. D., Sharma T., Mishra Sarad K. and Pandey A.K., Insight to sequence information of lactoglutathione lyase enzyme from different source organism, *I. Res. J. Biological Sci,* **1(6)**, 38-42 **(2012)**

17. Kumar A. and Dwivedi V.D., Evolutionary analysis and motif discovery in rhodopsin from vertebrates, *ISCA J.Biological Sci*, **2(7)**, 6-11**(2013)**

18. Felsenstein J., PHYLIP: Phylogeny interference package (version 3.69) *Department of Genome Sciences and Department of Biology. University of Washington.* Washington, USA), 164-166 **(1989)**

19. Swofford D. L., Olsen G. J., Waddell P. J. and Hillis D. M., Phylogenetic inference. In D M Hillis, C Moritz and B K Mable (Eds.), *Molecular systematics*, Sunderland, USA: Sinauer Associates, Inc., Publishers. 2nd edn, 407-514 **(1996)**

20. Mondal S. K., Shit S. and Kundu S., A comparative computational study of the '*rbc*L' gene in plants and in the three prokaryotic families-Archaea, cyanobacteria and proteobacteria, *IJBT,* **12**, 58-66 **(2013)**

21. Saldanha A. J., Java Treeview-extensible visualization of microarray data. *BIOINFORMATICS APPLICATIONS NOTE*, **20(17)**, 3246–3248 **(2004)** doi:10.1093/bioinformatics/bth349.

22. Fu C., Xiong J. and Miao W., Genome wide identification and characterization of cytochrome P450 monooxygenase genes in the ciliat *Tetrahymena thermophila*, BMC genomics. **10**, 208 **(2009)** doi: 10.1186/1471-2164-10-208

23. Meng Z., Wei L. and Xia L., Analysis of synonymous codon usage in chloroplast genome of *Populus alba, J For Res.,***19**, 293-297 **(2008)**

24. Sharp P. M., Tuohy T. M. F. and Mosurski, K. R., Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes, *Nucleic Acids Research,* **14**, 5125-5143 **(1986)**

25. Kaufman L. and Rousseeuw P. J., Finding groups in data: An introduction to cluster analysis, *John Wiley and Sons, Inc*.New Jersey, USA) **(1990)**

26. Sharma A. and Sharma P., Genetic and Phytochemical analysis of Cluster bean (Cyamopsis tetragonaloba (L.) Taub) by RAPD and HPLC, *Res.J.Recent Sci.,* **2(2)**, 1-9 **(2013)**

27. Sharp P. M. and Li W. H., The codon adaptation index a measure of directional synonymous codon usage bias, and its potential applications, *Nucleic Acids Research,* **15**, 1281-1295 **(1987)**