i) **Name of PI:** Dr. Sandipan Ganguly

ii) **Name of the Co-**PI: Prof. Tomoyoshi Nozaki

iii) **Title of the Project:** ***Differential pathogenesis of Giardia : Role of Giardia Virus***

iv) **Objectives of the Project:**

- To identify the infection of different GLV among different isolates of *Giardia lamblia* in patients with differential infection, i.e. with and without symptomatic *Giardia lamblia* infection and with multiple infections along with *Giardia lamblia* as a co-infection and also with asymptomatic *Giardia lamblia* infection.
- To understand if there is any genetic variability in housekeeping as well as pathogenic genes (*e.g.* the excretory secretory protein genes of *Giardia lamblia*) before and after infection of GLV or the outcome of differential pathogenesis is only a result of differential transcriptomic expression by PCR, RT PCR and Microarray hybridization.

v) Plan of work for fiscal year 2014-2015

1. Further validation of the results obtained in transcriptomic analysis.
2. To find out the regulatory mechanism for these candidate genes in differential pathogenic regulation.
3. To further extend the same in other pathogenic enteric protozoa like *Entamoeba histolytica*.

We have used different in vitro procedures for mimicking human GUT, like high oxygen tension etc. to find out what are the differentially regulated factors in *Giardia* that helps the parasite to live inside the human GUT even at very high oxygen tolerance

level than they can withstand. We have used a genomic DNA microarray for hybridization procedure for fishing out these particulate candidate regulators.

Array analysis and RT PCR validation

The hybridized slides were scanned using ScanArray® software in the scanner and more than 200 clones were identified that show 2 folds or higher times up regulation or down regulation than the control set. The result was cross checked twice and also by dye swapping. The clone numbers that matched in all the cases have been chosen for further sequencing analysis. From the array analysis some interesting gene candidates were selected for Real time PCR validation (Fig.1A, 1B, 1C). Transcriptional regulation due to stress shows that it affects the parasite cell massively and can change several physiological activities. This has helped to generate basic knowledge about some pathways controlling the survival and evolution of this human enteric parasite.

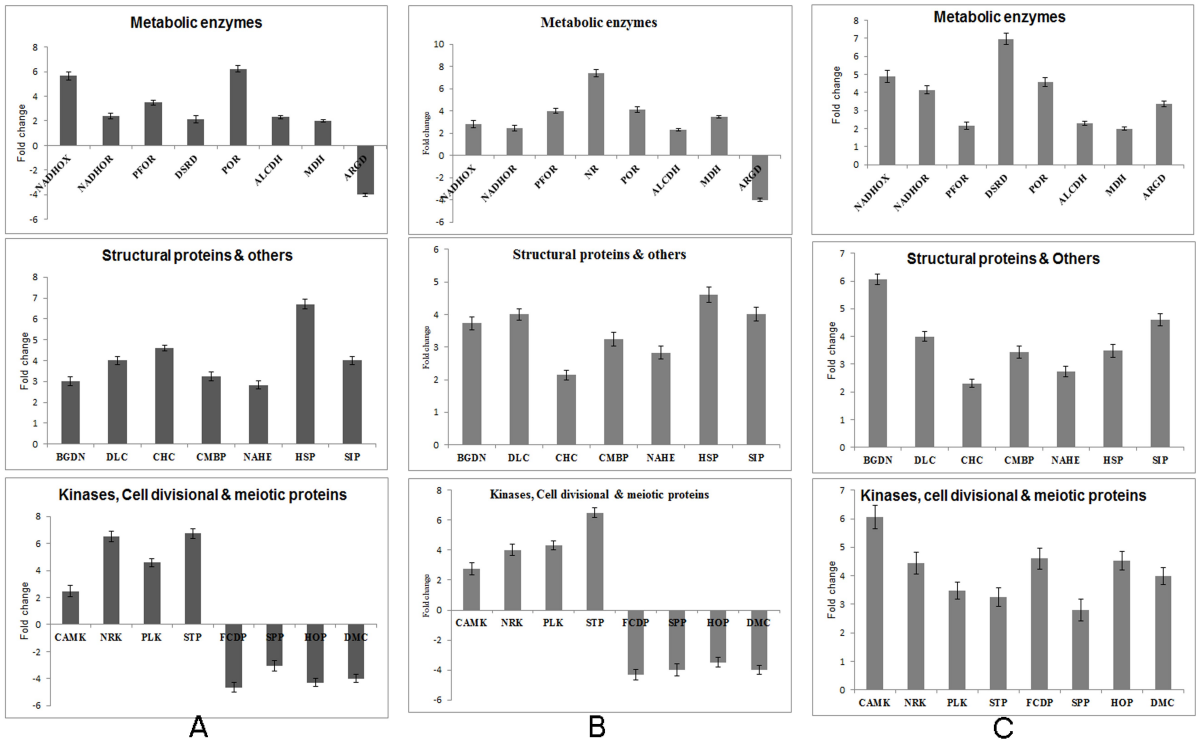


Fig. 1 Real time PCR graphs showing differential gene expressions under different modes of oxidative stresses. A. Gene expression (fold change) under H2O2 stress. **B.** Gene expression under metronidazole treatment. **C.** Gene expression during modified medium treatment. (Gene abbreviation used: **Metabolic enzymes:** NADHOX – NADH oxidase, NADHOR – NADH oxidoreductase, PFOR – Pyruvate-ferredoxin oxidoreductade, DSRD – Disulfide reductase, NR – Nitroreductase, POR – Peroxiredoxin, ALCDH – Alcohol dehydrogenase, MALDH – Malate dehydrogenase, ARGD – Arginine dihydrolase. **Structural proteins & others:** BGDN – β giardin, DLC – Dynein light chain, CHC – Clathrin light chain, CMBP – Cysteine rich membranr binding protein, NAHE – Na-H exchanger, HSP – Heat shock protein, SIP – Stress induced phosphoprotein. **Kinases, cell divisional & meiotic proteins:** CAMK – CAM kinase, PLK – Polo-like kinase, NRK – NIMA related kinase, STP – Serine threonine phosphatase, FCDP – Fts-J cell divisional protein, SPP - Spindle pole protein, HOP – Homologous pairing, DMC – Disrupted meiotic protein.

Mode of cell death

It was hypothesized that the origin of eukaryotic programmed cell death is a consequence of aerobic metabolism (Frade *et al,* 1997). In the previous paragraph, a hypothesis by Blackstone *et al,* 1999 was discussed partly. According to them, in metazoans, the mechanism of apoptosis involving Cytochrome c may be a vestige of the process where programmed cell death is triggered instead of sexual reproduction. In another study, we reported a protease independent programmed cell death in *Giardia* under oxidative stress (Ghosh *et al,* 2009). When oxidative stress exceeds the limit of the parasite's tolerance level, the cells commit suicide. In the dying cells, a protein named Programmed cell death protein has been found to be up-regulated. A time-kinetics has been done where the increased expression of this protein has been observed. This protein is a type of phosphatase. Another protein named protein required for cell viability is inversely regulated with the previous one i.e. when the cells undergo death phase, expression level of this protein gradually decreases. Cathepsin B precursor gene has also been observed to be regulated under cell death during oxidative stress. Cysteine Protease has been found to be down-regulated in the initial phase whereas, Signal Recognition Particle 64 (SRP 64) gets up-regulated in the primary stage but gets down-regulated during the commencement of cell death (Fig. 2). The total cell death regulation under oxidative stress has been found to be a unique feature of this parasite. In our previous study, we obtained some results on the changes in parasitic cell cycle, cell morphology and cell death under oxidative stress. Transcriptomic data from the Real-time PCR

study has further elaborated that concept indicating the name of some significant genes regulating cell death of this parasite during oxidative stress.

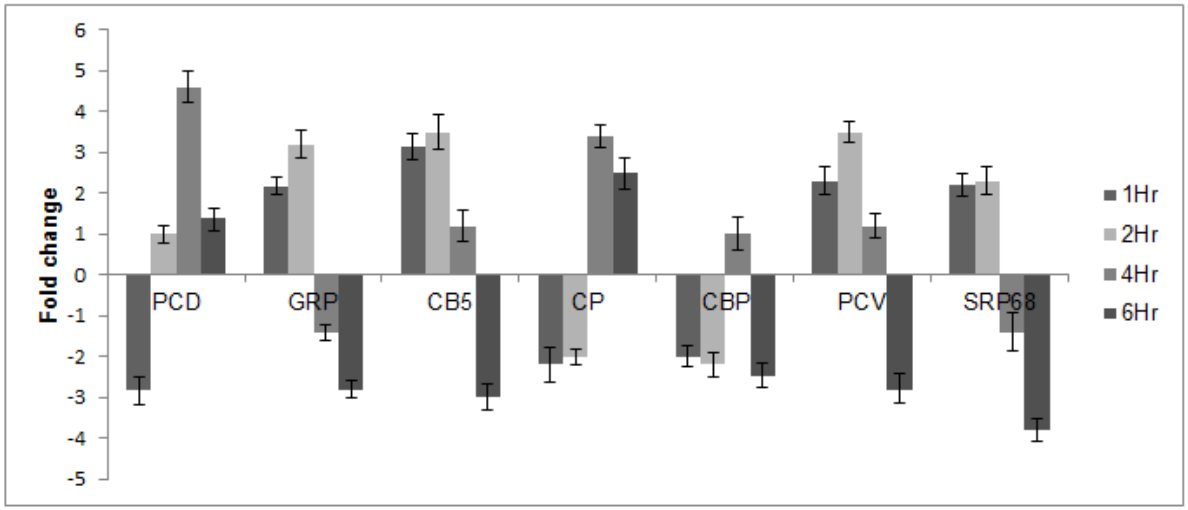


Fig: 2 Real time PCR graph of cell death regulating proteins

X axis: Name of the gene; Y axis: Fold change

PCD: Programmed cell death protein like protein, GRP: Glucose regulated protein 94KDa (Anti apoptotic death regulator), CB5: Cytochrome B5, CP: Cysteine protease, CBP: Cathepsin B precursor, PCV: Protein for cell viability, SRP68: Signal recognition particle 68

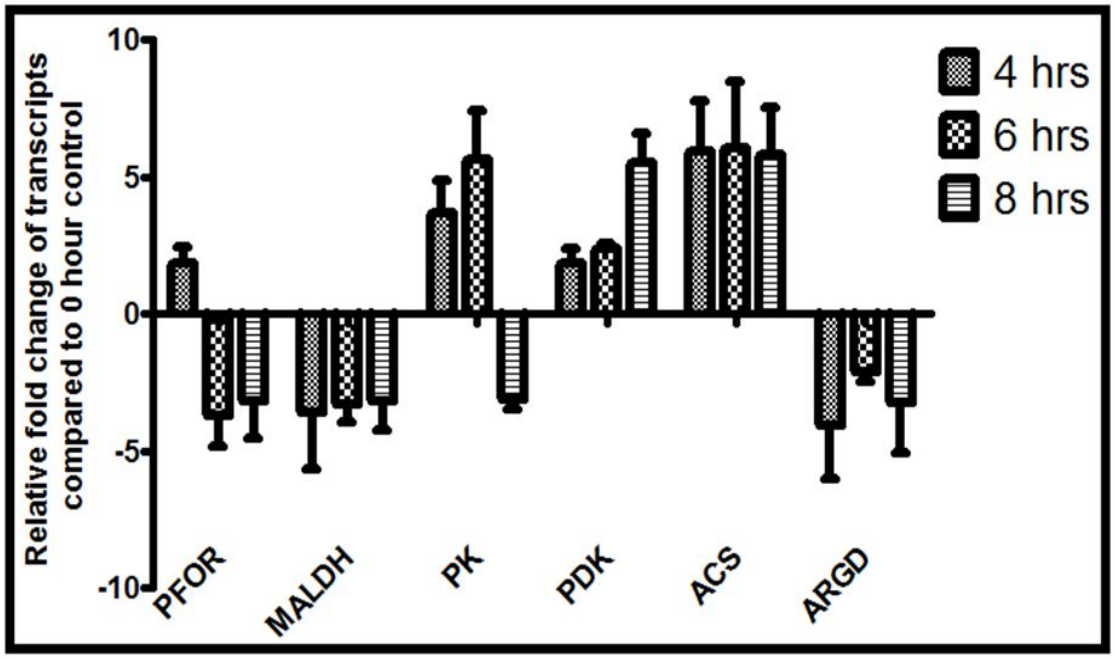
Modulation of different metabolic genes under differential pathogenic regulation

To better understand the effect of different oxidative stress in transcriptional regulation of gene expression in *Giardia lamblia*, we performed time course analysis of gene expression of pyruvate metabolism pathway upon H_2O_2 (0.10 μM) and cysteine-ascorbate deprived medium stress using a quantitative RT-PCR. We have chosen six genes, related to the pyruvate metabolism of *Giardia lamblia* modulated by at least 2 fold at one or more time points in response to H_2O_2 and cysteine-ascorbate deprivation. The metabolism of *Giardia* species is essentially fermentative, with glucose and amino acids being used as energy sources (Jaroll *et al*., 1995). Arginine is metabolized by Arginine dihydrolase pathway (Knodler *et al*., 1995) and aspartate can be metabolized to pyruvate via a variety of enzymes (Mendis *et al*., 1992). In our study, we have shown that arginine deiminase (ARGD)-encoding gene was up-regulated in *Giardia* trophozoites during cysteine-ascorbate deprived medium stress but remains down-regulated under hydrogen peroxide stress (Fig. 3B, 3A). Glucose is metabolized by a glycolytic pathway that is modified such that ATP is replaced by pyrophosphate at several key points (Mertens, 1993). This makes the pathway reversible and produces an increase in net ATP

synthesis. In *Giardia lamblia*, pyruvate can be produced by three different pathways. Phosphoenolpyruvate carboxyphosphotransferase together with malate dehydrogenase and malate dehydrogenase (decarboxylating), serves as a pathway to convert phosphoenol pyruvate (PEP) into pyruvate (Lindmark, 1980). Malate dehydrogenase (MDH) gene was up regulated at one or more time points upon cysteine-ascorbate deprivation (Fig. 3B). The gene showed a maximum induction at 6 h of cysteine-ascorbate deprivation. In contrast to the induction in response to hydrogen peroxide stress, the gene was down-regulated after (4-8 h) time points (Fig. 3A). However, pyruvate phosphate dikinase (PDK) (Bruderer *et al*., 1996; Park and Sinskey, 1997) and pyruvate kinase (PK) (Ellis *et al*., 1993) can also convert phosphoenolpyruvate into pyruvate in *Giardia*. Pyruvate phosphate dikinase, an enzyme that catalyzes the irreversible conversion of phosphoenolpyruvate (PEP) to pyruvate, the energy generating step of pyruvate biosynthetic pathway, was up-regulated till 8 h under both H_2O_2 stress and cysteine-ascorbate deprivation (Fig. 3A, 3B). Other genes that were slightly modulated by cysteine-ascorbate deprivation and H_2O_2 stress included pyruvate kinase, which was induced at early (4-6 h) time points under H_2O_2 stress (Fig. 3A) but it was getting down-regulated after (8 h) time points. In response to cysteine-ascorbate deprivation pyruvate kinase remained down-regulated after 6 h time points but up-regulated after 8 h time points (Fig. 3B). This has suggested that the three-enzyme pathway may have an alternative function, such as transferring equivalents from NADH to NADPH (Lindmark, 1980; Ellis *et al*., 1993). Pyruvate ferredoxin oxidoreductase (PFOR) (Townson *et al*., 1996), an enzyme system involved in the conversion of pyruvate to acetyl-coA and finally from acetyl-CoA to acetate is formed by the enzyme acetyl-CoA synthase (ACS). *Giardia* displays a significant sensitivity to O_2 (Lloyd *et al*. 2000) that are attributed to the expression of O_2-labile key

metabolic enzymes such as PFOR (Townson *et al*., 1996). In our study, PFOR-encoding gene was up-regulated during first couple of hours under cysteine-ascorbate deprived medium stress (Fig. 3B) but it was getting down-regulated after (6-8 h) time points upon H_2O_2 stress (Fig. 3A). The enzyme acetyl-CoA synthase transcript was up-regulated during both stresses. However, both hydrogen peroxide and cysteine-ascorbate deprivation has been shown to significantly modulate the metabolic flux across pyruvate metabolism in *Giardia lamblia*.

A


Relative fold change of transcripts compared to 0 hour control
10
5
0
-5
-10
4 hrs
6 hrs
8 hrs
PFOR
MALDH
PK
PDK
ACS
ARGD

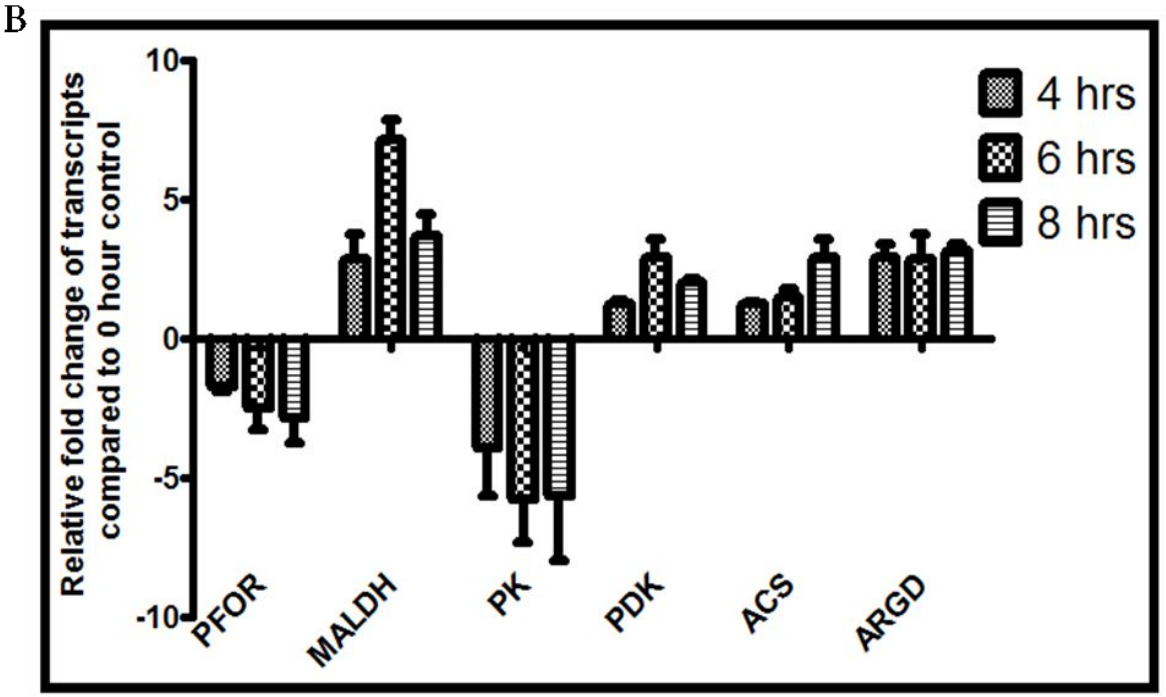


Fig: 3Effect of H_2O_2 and cysteine-ascorbate deprived medium on the expression of genes involved in pyruvate metabolism. Modulation of transcripts encoding enzymes involved in pyruvate metabolism. **A.** Gene expression (fold change) under H_2O_2 stress. **B.** Gene expression during modified medium treatment. Data are shown as fold change in relative expression compared with Actin on the basis of Comparative Ct ($2^{-\Delta\Delta Ct}$) method. Values are shown as mean ± SEM of three independent experiments, each performed in triplicate. (Gene abbreviation used: **Metabolic enzymes:** PFOR: Pyruvate-ferredoxin oxidoreductade, MALDH: Malate dehydrogenase, ARGD: Arginine deiminase, PK: Pyruvate kinase, PDK: Pyruvate dikinase, ACS: Acetyl coA synthase)

We have a special interest on *Entamoeba histolytica* to understand if there is any genetic variability in housekeeping as well as pathogenic genes. Amoebiasis caused by protozoan parasite *Entamoeba histolytica* is one of the major enteric diseases in human. It, being one the foremost parasitic disease after malaria is responsible for approximately 100,000 human deaths per annum [WHO, 1997]. The outcomes of *E. histolytica* infection are highly variable. Majority of infected individuals remain asymptomatic. Only a fraction of the infected develops diarrhea, dysentery, and rare extra-intestinal complications like- amoebic liver abscess (ALA) [Stanley, 2003; Mehmet and Petri, 2003]. Specific determinants for these diverse disease outcomes still remain elusive; however, host genetics and parasite genotypes could be two possible factors [Ali *et.al*, 2007; Duggal *et.al*, 2004]. Genome information of infecting strains from endemic areas throughout the world is certainly crucial to determine the exact genetic traits of parasite affecting its virulence capacity as well as different disease causing abilities. Selection of suitable genetic markers is needed for an optimum genotyping system. Since *E. histolytica* does not appear to contain microsatellite, measurement of genetic diversity and estimation of population structure has relied upon other polymorphic repetitive markers like serine-rich *E. histolytica* protein (*SREHP*) gene, chitinase (*CHI*) gene, and tRNA-linked STR loci. [Weedall and Hall, 2011]. However, genotyping studies based upon variations in these repetitive DNA often indicate very high levels of genetic diversity in an *E. histolytica* population [Ayeh-Kumi *et.al*, 2001; Ghosh *et.al*, 2000; Simonishvili *et.al*, 2005; Nozaki *et.al*, 2006; Escueta-de Cadiz *et.al,* 2010; Ali *et.al,* 2012]. This also results in the identification of immense variety of novel genotypes from different parts of the world [Ali *et.al*, 2008a; Blessmann *et.al*, 2003; Watanabe *et.al*, 2011; Feng *et.al,* 2012], which are too complex to predict the actual genetic makeup of parasite responsible for

particular outcome. While these methods are useful for regional studies, they may be less suitable for analysis of global and diverse parasite population [Gilchirst *et.al*, 2012].

In contrast to highly repetitive tRNA arrays, non-repetitive loci are genetically stable and their locations have been mapped in *E. histolytica* genome [Weedall and Hall, 2011]. Analysis of these loci revealed very limited single nucleotide polymorphism (SNPs), providing some easily comparable and analyzable data to achieve a discrete conclusion [Beck *et.al*, 2002; Weedall *et.al*, 2012; Gilchirst *et.al*, 2012]. Moreover, SNP analyses could be a suitable strategy to identify potential recombination events within *E. histolytica* and other parasite population [Gilchirst *et.al*, 2012; Li *et.al*, 2013].In the present study, three coding [lysine and glutamic acid rich protein gene 1 (*kerp1*) (accession number EHI_098210), lysine glutamic acid rich protein gene 2 (*kerp2*) (accession number EHI5A_120210) and amoebapore C (*apc*) (accession number X76903)] as well as two non-coding [i.e. AY956435 intron and AY956439 intergenic region] non-repetitive genomic regions were analyzed among *E. histolytica* clinical isolates from different infection outcomes. KERP1 is a surface-associated protein of *E. histolytica*, and has been shown to be involved in the parasite's adherence to human enterocytes and an important virulence factor in liver abscess pathogenesis. Homolog of *kerp1* gene was found in *E. nuttali* P19 strain (ENU1_189420), but not in *E. dispar* [Santi-Rocca *et.al,* 2008; Perdomo *et.al,* 2013]. A total of 10 nucleotide substitutions have been identified within *kerp1* gene of *E. nuttalli* P19 strain with respect to that of *E. histolytica* HM1:IMSS strain (EHI_098210) (inter-species variation). However, no SNP has been identified within the *kerp1* region among various *E. histolytica*

strains (intra-species variation) based on AmoebaDB database [Aurrecoechea *et.al*, 2011]. No association of SNPs within *kerp1* gene with disease outcome has been reported so far. KERP2 has been identified in *E. histolytica* and is thought to interact with human enterocytes during tissue invasion [Seigneur *et.al*, 2005]. It is predicted to have α- helical structure and to be localized in the external surface of the parasite [Seigneur *et.al*, 2005]. However, its exact function in *E. histolytica* virulence is still unidentified [Seigneur *et.al*, 2005]. Homologs of *kerp2* gene (EHI5A_ 120210) were also found in *E. dispar* strain SAW760 (EDI_199100), *E. invadens* strain IP1 (EIN_239080), *E. moshkovskii* strain Laredo (EMO_055400) and *E. nuttali* strain P19 (ENU1_019550). A high degree of inter-species variation was evident among these homolog genes based on AmoebaDB database [Aurrecoechea *et.al*, 2011]. Moreover, study of intra-species variation among various *E. histolytica* strains has identified only a single SNP (168 T/*G*) within *kerp2* region of *E. histolytica* PVB and PVF strains (nucleotide of *kerp2* region represented in *italic* was identified in those particular strains). No association of SNP within *kerp2* gene with disease outcome has been reported so far. Amoebapore C (*apc*) is a virulence factor of *E. histolytica* [Bhattacharya *et.al*, 2005]. Homologs of this gene were also found in other *Entamoeba sp*. like *E. dispar* strain SAW760 (EDI_206610), *E. invadens* strain IP1 (EIN_133650) and *E. moshkovskii* Laredo (EMO_119370). High degree of inter-species variation was also evident among these homolog genes based on AmoebaDB database [Aurrecoechea *et.al*, 2011]. A single SNP (237 T/*C*) within the *apc* region has been previously reported among various *E. histolytica* strains (i.e. NIH:200, HK-9, Rahman, DS664, MS533046 and MS301047) (nucleotide of *apc* region represented in *italic* was identified in those particular strains) [Bhattacharya *et.al*, 2005]. Previous study has also identified SNPs within upstream region of *apc* gene, significantly associated with disease

outcomes [Bhattacharya *et.al*, 2005]. Inter-isolate polymorphisms within intron and intergenic regions were also previously reported. A total of 2 SNPs (369T/G, 664T/C) has been previously identified within intron region (AY956435 intron). Among them, 369T/*G* was found among *E. histolytica* DS1-1036 and DS315 strains, while 664 T/*C* was found among *E. histolytica* DS1727, DS315 and MS533046 strains (nucleotides of intron region represented in *italics* were identified in those particular strains) [Bhattacharya *et.al*, 2005]. A total of 5 SNPs (236T/G, 240A/G, 364A/G, 550T/G, 561 T/G) were identified within the intergenic region (AY956439 intergenic region). Among them, 236T/*G*, 240A/*G*, 550T/*G* and 561T/*G* were found among *E. histolytica* MS27291, MS301047 and MS153394 strains. 364 A/*G* was present among *E. histolytica* NIH: 200, HK-9 and Rahman strains (nucleotides of intergenic region represented in *italics* were identified in those particular strains) [Bhattacharya *et.al*, 2005]. However, three SNPs within intergenic regions (236T/*G*, 240A/*G* and 561T/*G*) were found to be associated exclusively with asymptomatic outcome [Bhattacharya *et.al*, 2005]. The goals of our study were to identify potential genetic markers having SNPs, significantly associated with specific clinical outcome and to determine whether our study population had experienced any recombination events in their history or not. Intragenic and interlocus linkage disequilibrium (LD) of target loci from study isolates were assessed to identify potential recombination events within them. Comparative genetics between *E. histolytica* populations from different disease outcomes were also evaluated through interlocus LD, population differentiation (F_{ST}) and phylogenetic analyses.

Genetic diversity:

The haplotype diversity (Hd, represents the probability that two randomly sampled alleles are different) of individual polymorphic loci ranges from 0.117 to 0.621. Number of haplotypes ranges between 2 to 8 (sample number 49). Highest Hd value (0.621) and highest number of haplotypes (8) were observed in *kerp2* locus among the five examined loci (Table 4).

Concatenated nucleotide sequences of all five target loci have a length of 2463 bp and contained 26 polymorphic sites. A total of 11 multilocus genotypes (MLGs) have been identified. It has an Hd value of 0.703 (Table 5). Since majority of identified polymorphic sites were detected in *kerp2* locus (18 among 26) (Table 4), analysis of genetic diversity was also performed using concatenated nucleotide sequences of all target loci excluding *kerp2*. Comparatively, few (8) polymorphic sites were observed and only 5 MLGs were identified. A lower level of Hd (0.420) was detected (Table 5). Hence, the obvious observation from the above analysis is that *kerp2* locus has high degree of genetic diversity compare to other target loci. Similar type of observation was previously reported in glycoprotein 60 (*gp60*) locus of *Cryptosporidium hominis* [Li *et.al*, 2013].

Overall linkage disequilibrium (LD) and recombination analyses of target loci:

Intragenic LD and number of potential recombination events were evaluated for each target locus. The analysis was feasible only for 2 loci (*kerp2* and AY956435 intron) that had at ≥ 3 polymorphic sites. Incomplete intragenic LD value (|D'| Y= 0.9854 + 0.0426X, where Y is the LD value and X is the nucleotide distance in kilobases) was identified at *kerp2* locus. Comparatively a complete LD value (|D'| Y= 1.0000 – 0.0000X) was found at the AY956435 intron region. Intragenic recombination test has identified 1 potential

intragenic recombination event (Rm) only at *kerp2* locus. No recombination event was implicated at AY956435 intron region (Table 4).

An overall interlocus LD and number of potential recombination events were also assessed from the concatenated multilocus sequence data (including *kerp2* locus) of our study isolates. Incomplete interlocus LD value (|D'| Y= 0.9909 + 0.0032X) with 2 potential interlocus recombination events were identified. Among 300 pair wise comparisons, 61 were significant by Fisher exact test and 38 were significant after Bonferroni correction (Table 5). In contrast, interlocus LD analysis from the concatenated multilocus sequence data (excluding *kerp2* locus) of our study isolates has identified a complete interlocus LD value (|D'| Y= 1.0000 + 0.0000X) and no recombination events. Only 10 were significant by Fisher exact test (after Bonferroni correction) among 21 pair wise comparisons (Table 5). Hence, obvious observation from both intragenic and interlocus LD analyses was that genetic recombination was mostly occurring at *kerp2* locus.

Table 1: Intragenic linkage disequilibrium (LD) and recombination analyses of each target locus.

Genetic locus	Haplotype diversity (Hd)	Number of Haplotypes	Number of polymorphic sites analyzed	Number of pairwise comparisons	Number of significant pairwise comparisons [a]	Intragenic LD (|D'|) [b]	Rm [c]
kerp1 (EHI_098210)	0.219	2	1	0	0	NA [d]	NA [d]
kerp2 (EHI5A_120210)	0.621	8	18	153	30 (9)	Y= 0.9854 + 0.0426X	1
apc (X76903)	0.292	3	2	1	0	NA [d]	NA [d]
Intron (AY956435)	0.187	2	3	3	3 (3)	Y= 1.0000 – 0.0000X	0
Intergenic region (AY956439)	0.117	2	2	0	0	NA [d]	NA [d]

[a] Number of significant pairwise comparisons by Fisher's exact test (after Bonferroni correction), [b] Intragenic linkage disequilibrium (LD), where Y is the LD value and X is the nucleotide distance in kilobases, [c] Minimum number of intragenic recombination events, [d] Analysis is not applicable for the locus, which has <3 polymorphic sites.

Table 2: Estimation of the genetic diversity from concatenated multilocus sequences of *E. histolytica* isolates.

Concatenated multilocus sequences	Haplotype diversity (Hd)	Number of multilocus genotypes	Number of polymorphic sites analyzed	Number of pairwise comparisons	Number of significant pairwise comparisons [a]	Interlocus LD (|D'|) [b]	Rm [c]
Including *kerp2* (EHI5A_120210) locus [d]	0.703	11	26	300	61 (38)	Y= 0.9909 + 0.0032X	2
Excluding *kerp2* (EHI5A_120210) locus [e]	0.420	5	8	21	10 (10)	Y= 1.0000 + 0.0000X	0

[a] Number of significant pairwise comparisons by Fisher's exact test (after Bonferroni correction), [b] Interlocus linkage disequilibrium (LD), where Y is the LD value and X is the nucleotide distance in kilobases, [c] Minimum number of interlocus recombination events, [d] Based on concatenated multilocus gene sequences of all five target loci (2463 bp), [e] Based on concatenated multilocus gene sequences of all loci excluding kerp2 (1975 bp).

vi. Future goals and Plan for fiscal year 2014-2015

1. To find out the regulatory mechanism under differential pathogenic regulation.
2. To understand the multiple infections along with *Giardia* as a co-infection and also with asymptomatic *Giardia* infection.
3. To further extend the same in other pathogenic enteric protozoa like *Entamoeba histolytica*.

Publications from the project (2014-15):

1. Dibyendu Raj, Prasanta Saini, Tomoyshi Nozaki and Sandipan Ganguly. Involvement of pyruvate on oxidative stress management in the microaerophilic protozoan parasite *Giardia lamblia*. Int J Adv Res, 3. 1148-1166; 2015.

2. Raj D[1], Ghosh E[1], Mukherjee A. K., Nozaki T. and Ganguly S. Differential Gene Expression in *Giardia lamblia* under Oxidative Stress: Significance in Eukaryotic Evolution. *Gene* 535.131–139; 2014.