Computational and Structural Biotechnology Journal 12 (2014) 26–33



# Mini Review Evolutionary genomics and population structure of *Entamoeba histolytica*

# Koushik Das, Sandipan Ganguly \*

Division of Parasitology, National Institute of Cholera and Enteric Diseases, P-33, CIT Road, Scheme XM, Beliaghata, Kolkata 700010, India

# ARTICLE INFO

# ABSTRACT

Article history: Received 11 July 2014 Received in revised form 2 October 2014 Accepted 3 October 2014 Available online 31 October 2014

Keywords: Genetic polymorphism Disease outcome Genetic recombination Genotyping Short tandem repeat loci Single nucleotide polymorphism

#### Contents

Amoebiasis caused by the gastrointestinal parasite <i>Entamoeba histolytica</i> has diverse disease outcomes. Study of
genome and evolution of this fascinating parasite will help us to understand the basis of its virulence and explain
why, when and how it causes diseases. In this review, we have summarized current knowledge regarding evolu-
tionary genomics of <i>E. histolytica</i> and discussed their association with parasite phenotypes and its differential
pathogenic behavior. How genetic diversity reveals parasite population structure has also been discussed.
Queries concerning their evolution and population structure which were required to be addressed have also
been highlighted. This significantly large amount of genomic data will improve our knowledge about this path-
ogenic species of Entamoeba

© 2014 Das and Ganguly. Published by Elsevier B.V. on behalf of the Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

1.	Introduction
2.	Whole-genome sequences of Entamoeba species
3.	Structure and organization of genome
4.	Genomic rearrangements and transposable elements
5.	Large gene families and their diversities
6.	Genetic diversity and population structure
7.	Conclusion
Ack	nowledgments
Refe	prences

# 1. Introduction

Amoebiasis 4caused by the gastrointestinal parasite *Entamoeba histolytica* is one of the major parasitic diseases after malaria and is responsible for approximately 100,000 human deaths per annum [1]. The parasite has an interchangeable two stage life cycle consisting of an infective cyst form and a motile pathogenic trophozoite form. Infection is endemic in many developing countries where poor sanitation and malnutrition are common. Infection can also be restricted to a certain population in some developed countries (among male homosexual population in Japan) [2,3]. The global prevalence of infection (estimated

in 1986) suggested that 10% of the world population was infected by this parasite [4]. E. histolytica infection develops variable disease outcomes. 90% of infected individuals remain asymptomatic, while only 10% develops symptoms of invasive amoebiasis [5,6]. However, the global prevalence was estimated prior to the differentiation of E. histolytica from its non-pathogenic sibling Entamoeba dispar in 1993 [7]. Regardless of this epidemiological modification, invasive amoebiasis is still relatively a rare outcome of E. histolytica infection. Specific determinants for the diverse outcomes of this infection still remain obscure. However, host genetics and parasite genotype could be two possible factors [8,9]. Exploring the hidden genetic trait of parasite, directly linked to its virulence or associated with disease outcome, motivates a substantial area of Entamoeba research. Intra and inter-specific genomic comparisons have been conducted to identify the parasites' genetic factor linked to its virulence or associated with differential disease causing abilities [10-13]. These studies also provide some interesting and

<sup>\*</sup> Corresponding author. Tel.: +91 33 2363 3855; fax: +91 33 2370 5066.

*E-mail addresses:* koushikdas55@gmail.com (K. Das), sandipanganguly@gmail.com (S. Ganguly).

http://dx.doi.org/10.1016/j.csbj.2014.10.001

<sup>2001-0370/© 2014</sup> Das and Ganguly. Published by Elsevier B.V. on behalf of the Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

valuable information concerning the evolution and population structure of this parasite. Recent information concerning evolutionary genomics of *E. histolytica* and their association with parasite phenotype and its virulence have been discussed. How parasite population structure is revealed by genetic diversity has also been discussed. Questions related to their evolution and population structure have also been emphasized in this review.

#### 2. Whole-genome sequences of Entamoeba species

Several species of Entamoeba infects a wide range of hosts [14]. The simplest morphological characteristic like the number of nuclei per cyst has been exploited to distinguish between species [15]. However, morphological variations do not always reflect species-level differences and significant genetic diversity exists among morphologically indistinguishable organisms [15]. Some species like the oral parasite Entamoeba gingivalis do not produce cysts [14]. Phylogenetic relationships among SSU rRNA gene sequences of Entamoeba species suggested that E. dispar, Entamoeba nuttalli and Entamoeba moshkovskii are closely related to E. histolytica, while Entamoeba invadens and Entamoeba coli are distantly related [15]. E. dispar, morphologically identical with E. histolytica is usually considered as an avirulent commensal of human gut [14]. However, a recent study suggested that a certain strain of E. dispar (ICB-ADO), isolated from a Brazilian patient can cause amoebic liver abscess (ALA) in hamsters [16]. E. moshkovskii is microscopically indistinguishable from E. histolytica and E. dispar in its cyst and trophozoite form. It was initially thought to be a free living protozoan species [17] but a recent study suggested that E. moshkovskii infects humans and causes diarrhea and colitis in infants [17]. E. dispar infection is, in general much more common than *E. histolytica* worldwide [18]. Since, worldwide prevalence of *E. histolytica* infection [4] was estimated prior to the genetic discrimination of E. histolytica from E. dispar, the prevalence value can be completely erroneous and E. dispar could be a potential contributor to the prevalence figures in endemic areas [19]. E. moshkovskii can be found more frequently in regions where amoebiasis shows high prevalence [19,20]. Entamoeba bangladeshi, recently discovered from Bangladesh was clearly grouped with the clade of Entamoeba infecting humans, including E. histolytica [21]. E. invadens is a reptilian parasite and is an important model for encystation process. E. invadens can be induced to encyst in axenic laboratory culture, while encystation has not yet been achieved in axenically grown E. histolytica trophozoites [14].

The genome sequence of E. histolytica strain HM1: IMSS was published and analyzed in 2005 [22-24]. The genome assembly contains 20, 800, 560 bp of DNA in 1496 scaffolds. The genome has a high AT content (approximately 75%). Approximately half of the assembled sequence is predicted to be coding, with 8333 annotated genes [14]. Genome assembly of E. dispar strain SAW760 is of a similar size to that of E. histolytica strain HM1:IMSS. It consists of 22,955,291 bp of DNA in 3312 scaffolds. AT content is also quite similar to that of E. histolytica strain HM1:IMSS (approximately 76.5%). 50% of the assembled sequence is predicted to be coding, with 8749 annotated genes [14]. Genome assembly of *E. invadens* strain IP1 appears to be larger than that of E. histolytica strain HM1:IMSS and E. dispar strain SAW760. It contains 40,888,805 bp of DNA in 1149 scaffolds. AT content is comparatively less (approximately 70%). Approximately 38% of the assembled sequence is predicted to be coding, with 11,549 annotated genes. As per AmoebaDB database version 4.1 [25, www.amoebadb. org], genome assembly of E. moshkovskii strain Laredo consists of 25, 250,000 bp of DNA in 1147 scaffolds. AT content is approximately 64%. A total of 12,518 annotated genes are present. According to AmoebaDB database version 4.1 [25, www.amoebadb.org], genome assembly of E. nuttalli strain P19 consists of 14, 399,953 bp of DNA in 5233 scaffolds. AT content is approximately 75%. A total of 6187 annotated genes are present.

### 3. Structure and organization of genome

Structure of *E. histolytica* genome has been extensively reviewed by Clark et al. [24]. Many interesting evolutionary features of *E. histolytica* genome have been highlighted. *E. histolytica* have gained a significant number of metabolic genes (at least 68) through horizontal gene transfer from bacteria [14,22,24]. Orthologues of these genes found in both *E. histolytica* and its evolutionary distant species *E. invadens* [15] indicate that gene transfer is ancient [14].

The haploid genome of *E. histolytica* strain, HK9 is  $3 \times 10^7$  bp in size, based on renaturation kinetics experiments [26]. Hybridization of gene marker to pulse field gels identified 14 linkage groups with 1-4 chromosomes per linkage group per nucleus [27]. Tetra-nucleated E. histolytica cyst must contain at least one to two genome copies (1n-2n) in each of the nuclei [28]. However, karyotype analysis of E. histolytica trophozoite revealed the presence of at least 4 functional copies of many structural genes and therefore probably a ploidy that is a multiple of four [28]. Ploidy can vary even within a cell lineage under different growth conditions [28]. However, this phenomenon was only studied in-vitro and whether this occurs in nature is not known. The rRNA gene occurs in circular DNA molecules that exist in multiple copies per nucleus [29]. These circular structures could be important for determining parasite phenotypes. The rDNA episome varies in size from 15 kb to 25 kb depending on E. histolytica strains. The rDNA episome in E. histolytica virulent strain HM1:IMSS has two rDNA units per circle, while E. histolytica avirulent strain Rahman has only a single rDNA unit in its episome [30]. Moreover, Jasson et al. reported that structural genes for hemolysins were present within the ribosomal RNA repeat on extra-chromosomal DNA element of *E. histolytica* [31].

Initial characterization of E. histolytica genome revealed some unusual features of its organization. E. histolytica genome is highly repetitive (about 40% of the sequences are assigned to repetitive elements). Among them, tRNA genes are exceptionally abundant; with an estimated 4500 copies (about 10 times of human genome) were present. Moreover, most of these tRNA genes are clustered and organized into 25 distinct arrays. The tRNA arrays are composed of tandemly repeated units encoding between 1 and 5 tRNA acceptor types [32]. The intergenic regions of these tRNA genes comprises of short tandemly repeated sequences (STRs) which resembles the micro/mini satellites of eukaryotic genomes. The only difference is that unlike randomly dispersed micro/mini satellites, STRs form a part of a larger unit which is itself tandemly arrayed [32]. tRNA genes are thought to be "hotspots" for recombination and mutation due to their unique structural organizations [32]. The arrangement of tRNA gene showed inter-specific variation. E. histolytica has 2 versions of tRNA array containing Asn<sup>GTT</sup> and Lys<sup>CTT</sup> genes [i.e. (N-K1) and (N-K2)], while E. dispar genome contains only 1 type of [N-K] array. E. moshkovskii array units are significantly smaller than their homolog in E. histolytica and E. dispar and their intergenic regions do not contain any STRs [32]. STR regions between these tRNA array units showed high degree of intra-specific variation in their repeat number, type and arrangement patterns [13]. These particular features make them very useful as population genetic markers for quantification of evolutionary divergence of this fascinating parasite. The only proposed function of this tRNA array unit is nuclear matrix binding [33]. Moreover, circumstantial evidence also suggests that they may be located either at subtelomeric or at chromosomal ends and could be functional replacements of traditional telomere repeats [32].

## 4. Genomic rearrangements and transposable elements

Unlike *Plasmodium* which has a stable genomic organization even among distantly related species, *Entamoeba* exhibit high degree of genomic plasticity and instability [14]. Genome rearrangement associated with tissue invasion and organ tropism has been reported as one possible explanation for the different tRNA STR genotypes identified in liver abscess and stool derived parasites from the same infected person [34]. Transposons and repetitive DNA molecule, which are present abundantly in Entamoeba genome, may be responsible for genome reorganization [14]. Transposable elements are organized in clusters, frequently found at syntenic break points providing insights into their contribution to chromosome instability and therefore, to genomic variation and speciation in these parasites [35]. Investigation of repetitive elements within genome from three Entamoeba species identified hundred copies of LINE (long interspersed elements) and SINE (short interspersed elements) elements and a large proportion of Entamoeba specific repeats (ERE1 and ERE2). ERE1 is spread across the three genomes and is associated with different repeats in a species-specific manner [35]. ERE2 sequence was present exclusively in E. histolytica [14]. LINEs and SINEs are class I transposons, propagated by reverse transcription [36]. EhLINEs (LINEs of E. histolytica) each has a single open reading frame with a putative nucleic acid binding motif (CCHC) and restriction enzyme-like endonuclease domain located downstream of the reverse transcriptase (RT) domain. Phylogenetic analysis of the RT domain placed the EhLINEs in the R4 clade of non-LTR elements, a mixed clade of non-LTR elements that includes members from nematodes, insects, and vertebrates [36]. EhLINEs share a common 3' end sequence with EhLSINEs (SINEs of E. histolytica) which indicates that they are involved in the retro transposition of EhLSINEs. EhLSINEs also have a conserved 5' end, involved in regulation of their transcription [36]. A genome-wide comparison based on location of LINEs and SINEs elements in E. histolytica and E. dispar genome suggested that SINE expansion has taken place after divergence of two species. However, the basic retrotransposition machinery is conserved in these two species [37]. Since, LINE and SINE can profoundly influence the expression of neighboring genes, their genomic location can affect the phenotypic consequences of parasites [37]. Moreover, a recent study by Yadav et al. [38] suggested that E. histolytica can form recombinant SINEs at high frequency during induced retro transposition in-vivo. DNA transposons (class II transposons) are rarely present in E. histolytica and E. dispar, but are much more prevalent in E. invadens and E. moshkovskii [14]. Representatives of three DNA transposase superfamilies (hobo/Activator/ Tam3, Mutator, and piggyBac) were identified in Entamoeba in addition to a variety of members of a fourth superfamily (Tc1/mariner), previously reported only from ciliates and Trichomonas vaginalis among protozoans [39]. Genomic rearrangement might be responsible for variation in number of transposable elements in different lineages [14].

# 5. Large gene families and their diversities

The genome of *E. histolytica* contains a number of large multi-gene families [14]. One such gene family encodes a group of AIG1 like proteins [23]. AIG1 protein family comprises of 29 members distributed in 3 clusters [23]. 18 of them are present near transposons, but whether their duplication and subsequent growth are encouraged by the proximity of transposons is required to be explored [23]. AIG1 proteins are associated with resistance to bacteria [40]. Another gene family encodes a group of leucine-rich-repeats (LRRs) containing proteins, homologous to bacterial fibronectin (BspA of Bacteroides forsythus) [41,23]. Lorenzi et al. identified 114 genes encoding for BspA-like proteins in the genome of E. histolytica strain HM1:IMSS. 41 of them are associated with transposable elements [23]. Proteins of the family contain conserved N-terminal domain. However, no classic membrane-targeting signal is present in the proteins [23]. Hence, it is tempting to speculate that conserved N-terminal domain of proteins might function as either an export signal or serve as a membrane-anchor domain or that export involves a non-classical transport mechanism, independent of the ER-Golgi pathway, similar to those that have been detected in yeast and mammalian cells [42]. At least one member of this family is expressed at the external surface of parasite [41]. Genome survey of E. invadens identified multiple copies of these leucine-rich-repeats (LRRs) containing genes and differential gene expression within gene families has also been reported [43]. However, it is quite unknown whether gene expression has been controlled in such a way that a single gene family is expressed at any one time, as observed in other parasites like *Trypanosoma* and *Plasmodium* [14].

Entamoeba also encodes a large number of Rab GTPase (like another protozoan parasite T. vaginalis), involved in vesicular trafficking in the cell [44,45]. A total of 102 Rab GTPase distributed in 16 subfamilies have been annotated in genome of E. histolytica [44,45]. Majority of them showed moderate similarity to Rab from other organisms, while only 22 amoebic Rab proteins including EhRab1, EhRab2, EhRab5, EhRab7, EhRab8, EhRab11, and EhRab21 showed significant similarity to Rab from other organisms [44]. E. invadens has over 100 Rab genes similar to E. histolytica [45]. A comparison of Rab GTPase from E. histolytica and E. invadens revealed that most Rab subfamilies are conserved among these two Entamoeba species [45]. This indicates that Rab GTPase-controlled vesicular trafficking machinery is well conserved among them and expansion of the gene family largely occurred before the divergence of these two species [45]. Rab GTPases have been involved in the regulation of cysteine protease secretion and transport [46,47]. E. histolytica differentially expressed their RabB protein (EhRabB) during phagocytosis of target cells, suggesting the potential role of EhRabB protein in phagocytosis process [48]. EhRabB protein has been mutated experimentally at 118 amino acid position and thus the resulted protein (RabBN118I) was unable to bind guanine nucleotide and became constitutively inactive [49]. Over-expression of such mutated RabB protein within E. histolytica trophozoites resulted in a significant reduction of parasite phagocytosis, cytopathic activity and ability to produce liver abscess in hamster [49]. Hence, Rab-regulated vesicular trafficking is important for parasite biology and pathogenesis. Gene families encoding heavy (hgl) and light (lgl) chain subunits of virulence determinant Gal/GalNAc lectin present in multiple Entamoeba species, but genes for intermediate chain subunit (igl) are only detected in E. histolytica and E. dispar [24]. Bioinformatics comparison among members of this gene family from E. histolytica and E. dispar identified the evidence of gene conversion within the lineages, which may play an important role in molecular evolution of these parasites [50]. Cysteine protease-5, the key virulence factor of E. histolytica is present as a pseudogene in E. dispar [14]. Over-expression of specific cysteine protease genes (ehcp-b8, ehcp-b9 and ehcp-c13) within parasite cells also confers pathogenicity to non-pathogenic *E. histolytica* clone A1 [51]. Southern blot analysis indicates that the ariel surface proteins of E. histolytica are either not present or highly divergent in E. dispar [14].

# 6. Genetic diversity and population structure

Since E. histolytica genome does not appear to contain any microsatellite like elements, measurement of genetic diversity and estimation of population structures greatly rely on other genetic markers like Serine Rich E. histolytica protein (SREHP) gene and chitinase [14]. SREHP is an immune dominant surface antigen, involved in phagocytosis of apoptotic host cells to prevent inflammatory responses by host [52] whereas chitinase is only expressed during encystations of amoeba [53]. Both genes contain tandem repeats which showed high degree of interisolate diversity based on their repeat types and arrangement patterns [2,3,54]. However, SREHP gene showed comparatively high degree of polymorphism than chitinase [3]. Since SREHP is highly immunogenic, such high genetic diversity within SREHP gene may suggest that it has a biological role like immune evasion [55]. However, PCR amplification of SREHP gene often produces multiple and mixed PCR bands from a single strain due to allelic variation [18]. Direct sequencing of such mixed PCR products (without cloning of PCR product into a vector prior to sequencing) gives rise to a chromatogram showing multiple peaks at a single nucleotide position. Multiple variations of a single sequence can be obtained from the analysis of such a sequence and this can be misinterpreted as genetic diversity. tRNA linked STR loci of E. histolytica has proved to be a useful population genetic marker and has been used to identify the parasite genotypes associated with different disease outcomes [56]. Studies of genetic diversity based on 6 tRNA linked STR loci (i.e. *D-A*, *S<sup>TGA</sup>-D*, *N-K2*, *R-R*, *A-L* and *S-Q*) have identified few parasite genotypes associated with disease outcomes [8,13,57,12, 58]. For example- 5RR of *R-R* locus was associated with asymptomatic outcome, while 10RR was associated with symptomatic outcome [59]. J1DA and VEN2DA of *D-A* locus were associated with asymptomatic and symptomatic outcomes respectively [60]. Even though tRNA linked STR loci showed few associations with disease outcomes, they are actually surrogate marker and their variations are not at all directly linked to parasite virulence [59]. Moreover, these loci are frequently mutated to form new genotypes and hence any significant association of parasite genotype with disease outcome would be lost over time [18].

However, patterns of polymorphism within these repetitive DNA sometimes reflect the population structure of parasite [14]. For example, in Japan, diversity among parasite population infecting homosexual men was high, while diversity was much more limited among parasite infecting residents of institution [2]. Similarly, low diversity among parasite population infecting residents of institution was seen in the Philippines, where clear population structure was observed within and between locations [54]. In South Africa, genotypes clustered within households but showed extensive diversity among different households [61]. Recently, Zermeno et al. have proposed the worldwide genealogy and population structure of *E. histolytica* based on two tRNA linked STR loci (i.e. *D-A* and *N-K2*) [60]. Majority of these genotypes were found to be exclusive for a particular country. Only few were shared by isolates from different countries. For example- 18NK, 17NK, 10NK, and 11NK of N-K2 locus and 5DA and 6DA of D-A locus were the only genotypes distributed in many regions. Among them, 18NK and 6DA, corresponding to the genotype of E. histolytica strain HM1:IMSS were the most abundant and widely distributed in many countries like Mexico, Bangladesh, Japan, China and the USA. However, genealogies based on these two individual loci (i.e. D-A and N-K2) suggested that there were no parasite lineages related with a particular geographic region. Moreover, concatenated analysis of two tRNA linked STR loci (i.e. D-A and N-K2) revealed the possibility of genetic recombination among the population studied [60]. Genetic organization of E. histolytica population from stool and liver abscess samples of same patients were also studied [34]. The study revealed that *E. histolytica* population from stool and liver abscess samples were genetically distinct [34]. However, few opposite but interesting scenarios have also been reported. E. histolytica population isolated from amoebic liver abscess (ALA) patients was genetically identical with those isolated from asymptomatic patients [57]. This finding was further supported by recent STR loci based genotyping study of E. histolytica from India. E. histolytica isolates remaining asymptomatic are genetically closer to those causing liver abscess rather than the diarrheal isolates (Fig. 1) [12]. Repetitive DNA markers appear to be stable enough to link closely related parasites recently transmitted among members of a household, an institution or recent sexual partners [14]. However, extensive population diversity in limited geographic regions and frequent occurrences of novel genotypes limit the efficiency of repetitive loci to probe large scale, long term population structure of E. histolytica [14]. SNP (single nucleotide polymorphism) markers may be preferable in these situations.

SNPs within non-repetitive loci arising under neutral, positive and negative pressure are genetically stable and inherited by their descendents [11]. SNP analysis could be a successful strategy to identify the potential virulence marker of parasite linked to infection outcome [11,62]. Comparison between genome sequences of various E. histolytica strains deposited onto AmoebaDB database version 4.1 [25, www.amoebadb.org] have identified a total of 2613 genes, which contain intra-species SNPs within them. Most of the proteins encoded by these genes are hypothetical in nature, while the functions of some genes are known. Few of such genes with known and hypothetical functions are listed in Table 1. A large number of SNPs have been identified in serine threonine isoleucine rich protein (EHI\_073630), gene for Gal/Gal NAc lectin lgl2 (EHI\_065330), heat shock protein70 (EHI\_159140), tyrosine kinase (EHI\_124500), gene for AIG1 family protein (EHI\_144270), gene for Rab family GTPase (EHI\_059670), etc. Gal/Gal NAc lectin is a surface antigen of Entamoeba and involved in parasite adhesion with intestinal epithelium [50]. AIG1 proteins are associated with resistance to bacteria [40]. Rab GTPases are involved in vesicular trafficking machinery of parasite [44,45]. However, further investigation is required to



**Fig. 1.** *E. histolytica* isolates remaining asymptomatic are genetically closer to those causing liver abscess: depicted by (A) phylogenetic tree, (B) graphical representation. Phylogeny was based on tRNA linked *N-K2* (STR) locus. The sequences of all (a total of 22) representative STR patterns from *N-K2* locus, obtained from the genetic analysis of 51 study isolates were aligned using ClustalW multiple alignment program of MEGA Version 4 software. Phylogenetic tree was constructed from the alignment through "Generalized Time Reversal (GTR) + gamma" substitution model of SeaView Graphical Interface Version 4 software using a maximum likelihood matrix algorithm. One distinct "D" group, one distinct "LA" group and one mixed "AS + LA" group can be assigned, 'D' group contains STR patterns found only in liver abscess outcome. 'AS + LA' group contains STR patterns found only in liver abscess outcome. 'AS + LA' group contains STR patterns found only in liver abscess outcome. 'AS + LA' group contains STR patterns found only in liver abscess outcome. 'AS + LA' group contains STR patterns found only in liver abscess (LA) outcome.

# Table 1

Genes of E. histolytica, contain intra-species single nucleotide polymorphisms (SNPs).

ID     SNPs     S	AmoebaDB	Protein product for this gene	Total SNPs	Non-synonymous	Synonymous	Non-sense	Non-coding	Non-synonymous	SNPs per
EHL 07-80Cale of the three is backwards of the transformed of tran	ID			SNPs	SNPs	SNPs	SNPs	SNP/synonymous SNP ratio	kb (CDS)
EHL0530GL/GAI MAC Heatin MgG7G7G7G7G4OOO.8338.14EHL0500GL/GAI MAC Heatin MgGGG.543.89GG.543.89EHL05100MAC polymarcas, putative13OGG1.313.15H114100MAC polymarcas, putative12GGO1.233.75H1142200Prospholing Intraporting p-typeOGGO0.573.75H114200Machan Bile prosein, putativeOGG3.75	EHI_073630	Serine threonine isoleucine rich protein, putative	70	46	24	0	0	1.92	4.6
BH 1.9104   Heat shock protein 70, putative   14   12   2   0   0   6   6.694     BH 1.04980   Galcal Nak lettin 101   13   9   4   0   0   2.25   1.88     BH 1.04980   Galcal Nak lettin 101   13   9   4   0   0   2.25   1.88     BH 1.04980   Galcal Nak lettin 101   12   6   0   0   1.2   1.513     BH 1.04940   Atmin 105 protein partice   9   0   0   0   2.25   3.66     Control Intering Protein partice   9   6   3   0   0   0   5.97     BH 1.02200   Protein knase domain containing protein   8   4   4   0   0   0   5.97     BH 1.03400   Incine rind repart protein BpA family   8   4   4   0   0   0   6.86     BH 1.03400   Prosphafdy Insord 13 kinase, putative   8   3   5   0   0   0   1.8     BH 1.03400   Prosphafdy Insord 13 kinase, putative   8   3   5   0   0	EHI_065330	Gal/Gal NAc lectin lgl2	27	13	14	0	0	0.93	8.14
BIL00980   GLGAI MA (exclus hgl1   13   9   4   0   0   2.25   3.88     BIL12400   WA polymers as putative   12   6   6   0   0   1.313     BIL124020   Work polymers as putative   12   6   6   0   0   1.2   1.2734     BIL123220   Prospholing Intransporting p-type   9   6   3   0   0   2   2.55     HIL03500   Calcutore inhishabe Horien 35 k0a   8   4   4   0   0   1   5.59     BHL13340   Lexcine rich repeat potein BsA I maily   8   3   5   0   0   0   5.59     BHL13480   Lexcine rich repeat potein BsA I maily   8   3   5   0   0   0   1.8     BHL042401   Lexcine rich repeat potein BsA I maily   8   4   4   0   0   1   3.06     BHL042401   Lexcine rich repeat potein BsA I maily   8   4   4   0   0   0   1.8   4.10     BHL042401   Lexcine rich repeat potein AsA I maily   1.6<	EHI_159140	Heat shock protein70, putative	14	12	2	0	0	6	6.94
BIL 2480   Tyrosine knase, putative   13   9   4   0   0   225   1.88     BIL 14202   ACI family protein   11   6   5   0   0   1.2   1.313     BIL 14202   ACI family protein   16   6   5   0   0   1.2   1.333     BIL 14202   Protein knase domain containing protein   9   6   3   0   0   2   2.555     BIL 05205   Protein knase domain containing protein   9   6   3   0   0   0   .599     BIL 01240   Chrose inhibibable lectin 35 KDa   8   4   4   0   0   0   .599     BIL 01240   Chrose inhibibable lectin 35 KDa   8   4   4   0   0   0   .599     BIL 01240   Chrose inhibibable lectin 35 KDa   8   4   4   0   0   0   .599     BIL 01240   Chrose inhibibable lectin 35 KDa   8   3   5   0   0   0   .509     BIL 01240   Chrose inhibibable lectin 35 KDa   8   3	EHI_006980	Gal/Gal NAc lectin lgl1	13	9	4	0	0	2.25	3.89
Ell L6419   DNA polymerske, pictative   12   6   6   0   0   1   31.31     EHL 16424   Action line protein, partative   9   0   8   0   0   0.57     EHL 15224   Action line protein, partative   9   6   3   0   0   0.5   2.05     EHL 02520   Potein kinase domain containing protein   9   6   3   0   0   0   5.59     EHL 02580   Galactose inhibitable (Erin 35 RDA astice)   8   4   4   0   0   1.306     BHL 02430   Cocyony Hydrolise affinitity   8   3   5   0   0.6   3.96     EHL 02430   Cocyony Hydrolise affinity   8   3   5   0   0   1.8     BHL 02580   Calactose specific adhesin 10 Rate, putative   8   3   5   0   0   0   1.8     BHL 02580   Calactose specific adhesin 10 Rate, putative   8   4   4   0   0   1.8   2.06     BHL 02580   Calactose specific adhesin 10 Rate, putative   8   4   4	EHI_124500	Tyrosine kinase, putative	13	9	4	0	0	2.25	1.68
H1, H444   Aik 1 laminy protein   11   b   5   0   0   1.2   1.2/1.5     BHL 1322   Prospholipit transporting prope   9   0   0   0   5.74     BHL 1322   Prospholipit transporting prope   9   6   0   0   2.55     BHL 1323   Calcrose inhibitable lettin 35 Bab   8   4   0   0   0   5.74     BHL 12021   Diogation factor alpha 1   8   0   8   0   0   0   5.99     BHL 13243   Licence in chrospat protein Bap family   8   3   5   0   0   0.66   1.86     BHL 01243   Galacces specifit Adhiesin 10 kDa   8   6   2   0   0   0.5   2.66     BHL 01380   Phosphatidy linesitol 3 kinase, putative   8   3   5   0   0   0.5   2.66     BHL 01380   Phosphatidy linesitol 3 kinase, putative   6   2   4   0   0   2   1.51     BHL 01380   Phosphatidy linesitol 3 kinase, putative   5   1   4   0   0	EHI_164190	DNA polymerase, putative	12	6	6	0	0	1	3.13
EHL 1944     Anclini me protein, plataive     9     0     9     0     0     0     0     0.5     3.03       EHL 3232     Propholigit answording protein     9     6     3     0     0     2     2.55       EHL 3334     Face indication     35 kDa     8     4     0     0     0     5.57       EHL 3434     Stomit prevator     8     0     0     0     5.59       EHL 3434     Exercise fact regator factar answording 31 protein     8     0     0     0     1.59       EHL 01343     Exercise fact regator factar answording 31 protein     8     3     5     0     0     0     1.88       EHL 03434     Exercise fact regator factar answording	EHI_144270	AlG1 family protein	11	6	5	0	0	1.2	12.75
DIL 1322     Prins/pring trainsporting proper     9     5     6     0     0     0.5     3.55       HU 02385     Calcrosse inhibitable ledin 35 kDa     8     4     4     0     0     1     8.57       BHL 02385     Calcrosse inhibitable ledin 35 kDa     8     0     8     0     0     0     5.99       BHL 02342     Clocosy Hubitable ledin 35 kDa     8     0     0     0     0.66     3.96       BHL 02342     Clocosy Hubitable ledin 35 kDa     8     6     2     0     0     0.66     1.86       BHL 02340     Clocosy Hubitable ledin 10 kDa     8     6     2     0     0     0.5     2.66       BHL 03580     Enophatidy linositol 3 kinase, parative     8     3     5     0     0     0     5     2.06       BHL 03580     Enophatidy linositol 3 kinase, parative     6     2     4     0     0     0     2.05     2.06       BHL 03580     Enosito for kinas kontance calcrosesci adhy kontase kontance calcrosesci adhy kontase kontance cal	EHI_164440	Actinin like protein, putative	9	0	g	0	0	0	5./4
EHL0306Ordein Kinase domain containing protein9630022.55EHL03506Calcetor highbal lectin 35 Kha8000.663.96EHL13404Corogaton factor alpha 1835000.663.96EHL03435Corogaton factor alpha 1835000.663.96EHL03435Corogaton factor alpha 1835000.663.96EHL03435Corogaton factor alpha 1835000.661.86EHL03435Corogatin factor alpha 1835000.661.86EHL03456Corogatin factor alpha 18350002.052.05HU1356Indiguin carboxy terminal hydrolase, pataive8330002.061.86EHL0356Corotatining protein6420021.11HU1357Corotatining protein pataive642002.022.02HU14587Corotatining protein pataive642002.421.11HU1358Magnetin Explorating protein pataive514000.252.52HU1359Magnetin Explorating protein pataive514000.252.52HU1359Magnetin Explorating protein pataive514 <t< td=""><td>EHI_135220</td><td>ATPase, putative</td><td>9</td><td>3</td><td>0</td><td>0</td><td>0</td><td>0.5</td><td>3.05</td></t<>	EHI_135220	ATPase, putative	9	3	0	0	0	0.5	3.05
EHL 0580Galactos inhibitable lectin 35 (ba)899918.57EHL 01210Elongation factor alpha 18080005.99EHL 02321Calctors excific danabasis8440013.06EHL 02323Calctors excific danabasis8440013.06EHL 02324Calctors excific danabasis8350001.8Calctors excific danabasis70001.8Calctor excific danabasis70002.66EHL 01500Noisoli al Stass, purtaive6240002.66Calcal MAC (etcin heavy suburit6420021.11EHL 0500Insidi al Staps protein6420021.11EHL 0511US sakNP specific 200 kD protein, purtaive6420022.57EHL 0512Inco aliturit favoring protein purtaive614000.252.57EHL 0520Inco aliturit favoring protein purtaive5140002.421.11Interbase darbase actific transporting5140002.552.57EHL 0520Incordasin favoring protein17116001.533.51Interbase darbase protein protein <td>EHI_023050</td> <td>Protein kinase domain containing protein</td> <td>9</td> <td>6</td> <td>3</td> <td>0</td> <td>0</td> <td>2</td> <td>2.55</td>	EHI_023050	Protein kinase domain containing protein	9	6	3	0	0	2	2.55
whatw	EHI_035690	Galactose inhibitable lectin 35 kDa	8	4	4	0	0	1	8.57
EHLEHLS080005.99EHL22430Glycos/I Mydrolase family 31 protein8440013.06EHL.02430Glycos/I Mydrolase family 31 protein8620013.06EHLGlactores specific adhesin 170 kOa8350001.8EHLDisophaticly I Insistol 3 Kinase, putative8350001.8EHLDisophaticly I Insistol 3 Kinase, putative624000.52.66Disotiol polyhophate 5 phosphatase,6240021.55EHLGLGAI NA'C Etchin heavy subunit6420021.11EHLDisotiol polyhophate 5 phosphatase, putative514000.52.92EHLDisotiol polyhophate 5 phosphatase, putative514000.52.92EHLDisotial polyhophate 5 phosphatase, putative514000.52.92EHLDisotial polyhophate 5 phosphatage, putative514000.52.92EHLDisotial polyhophate 5 phosphatage, putative514000.52.92EHLDisotial polyhophate 5 phosphatage, putative514000.52.92EHLDisotial polyhophate 6 putative </td <td></td> <td>subunit precursor</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>		subunit precursor							
EHL 33430Luckne inth repeat protein RspA family835000.63.96EHL 02430Glaktose specific adhesin 170 kDa8620032.05subuinit putative835000.61.86EHL 013900Ubiquinit arbacylu terminal hydrolase7000.52.66Chamain Catalining protein624000.52.66EHL 10500Ubiquinit arbacylu terminal hydrolase700021.55EHL 105207Gal/Gal NKA lectin heavy subunit6420021.55EHL 015207SakNA specific 200 kDa protein putative642002.52.06BrudarieSakNA specific 200 kDa protein putative514000.252.92EHL 01520Insus amenchane calcium transporting514000.252.92EHL 01520Magna menchane calcium transporting514000.252.92EHL 01580Magna menchane calcium transporting514000.252.92EHL 01580Magna menchane calcium transporting2136001.443.4EHL 01580Magna Gribas activating protein putative514002.45.1EHL 01580MagnA Gribas activating protein putative	EHI_011210	Elongation factor alpha 1	8	0	8	0	0	0	5.99
EHL 02340Glycosyl hydrolase family 31 protein8440013.06EHL 04370Glactores specific adhesin 170 k0a862000.61.86EHL 01900Displaticly limital hydrolase707000.61.86EHL 05070Rab family OTPase624000.52.66Distribution on ania containing protein6240021.55EHL 05070Rab family OTPase6240021.11EHL 05070Rab family OTPase6420021.11EHL 05070In suffur flavorottin pseudopene5410041.36EHL 05170In suffur flavorottin pseudopene514000.252.92EHL 05100In suffur flavorottin pseudopene514000.252.92EHL 05100Inog chain fatty actic Ani kase514000.252.92EHL 05180Inog chain fatty actic Ani kase514000.252.92EHL 05180Inog chain fatty actic Ani kase514000.252.92EHL 05180Inog chain fatty actic Ani kase514000.258.93IL 10580Myonbreical protein1711600<	EHI_139430	Leucine rich repeat protein BspA family	8	3	5	0	0	0.6	3.96
BHL 042370     Calactose specific adhesin 70 kDa     8     6     2     0     0     3     205       BHL 013900     Mosphatidy linosital 3 kinase, putative     8     3     5     0     0     0.6     1.86       BHL 119300     Molynthic adhesing protein     7     0     0     0.5     2.66       BHL 015070     Bal family CTPase     6     2     0     0     2     1.51       BHL 045170     CalaCal MAc lectin heavy subunit     6     4     2     0     0     2     1.11       BHL 04510     CalsCal MAc lectin heavy subunit     6     4     2     0     0     2.52     4.42       BHL 04510     CalsCal MAc lectin heavy subunit     6     4     2     0     0     2.52     2.92       BHL 045170     Damas membrane calcium transporting     5     1     4     0     0     2.52     2.92       BHL 045170     Damas membrane calcium transporting     5     1     4     0     0     2.4     2.51	EHI_023430	Glycosyl hydrolase family 31 protein	8	4	4	0	0	1	3.06
abunit, putative     is a banit, putative     is a banit, putative     is a banit, putative     is a banit, putative       EHI, 19380     Ubiquiti narboxyl terminal hydrolase     7     0     7     0     0     1.8       EHI, 05860     Rab family GTPase     6     2     4     0     0     5     2.06       Putative	EHI_042370	Galactose specific adhesin 170 kDa	8	6	2	0	0	3	2.05
EHL 013900   Prinosphatagy Infositol & Maske, putative   8   3   5   0   0   0.6   1.86     EHL 119300   Ubiquith carboxyli terminal hydrolase   7   0   0   0.5   2.66     EHL 05807   Nab family CTPase   6   2   4   0   0   2.26     EHL 01207   Cal/Cal MAc lectin heavy subunit   6   4   2   0   0   2   1.11     EHL 01210   Cal/Cal MAc lectin heavy subunit   6   4   2   0   0   2.2   1.11     EHL 014517   Usins membrane calcium transporting, putative   5   4   1   0   0   0.25   2.92     EHL 08150   Hagnifan CTPase activiting protein, putative   5   1   4   0   0   0.25   2.92     EHL 08180   Lange fan Entry actic Anligase, putative   5   1   4   0   0   0.25   2.92     EHL 08180   Hypothetical protein anton anima containing protein 2, putative   5   1   4   0   0   0.25   2.92     EHL 08180   Hypothetical protein	FUL 012000	subunit, putative	0	2	-	0	0	0.0	1.00
EHL 1980   Undquint catroxy Herminal nytrotase   7   0   7   0   0   0   1.8     EHL 0980   Rab family CTPase   6   2   4   0   0   0.5   2.66     EHL 0980   Inosito polyphosphate 5 phosphates,   6   5   1   0   0   2   1.55     EHL 01220   Cal/Cal NAE lettin heavy subunit   6   4   2   0   0   2   1.11     FHL 04520   US SnRNP specific 200 kDa protein, putative   6   4   2   0   0   2   1.13     FHL 04520   Uro sulfur flavoprotein pusatore approtein putative   5   1   4   0   0   0.25   2.92     EHL 08300   Thoredoxin domain contalning protein 2, putative   5   1   4   0   0   4.242     EHL 08300   Thoredoxin domain contalning protein 2, putative   5   1   4   0   0   1.83   6.93     EHL 03300   Hypothetical protein   17   12   5   0   0   2.4   5.13     EHL 02300   Hypothetical protein	EHI_013980	Phosphatidyi linositoi 3 kinase, putative	8	3	5	0	0	0.6	1.86
EHL 059670   Rah family CTPase   6   2   4   0   0   0.5   2.66     EHL 16860   nositol polyphosphate 5 phosphatase,   6   5   1   0   2   1.55     EHL 22707   Cal/Cal Mck lectin heavy subunit   6   4   2   0   0   2   1.11     EHL 045170   Vois SinkPT Specific 200 k0a protein, putative   6   4   2   0   0   2.2   1.11     EHL 045170   Vois SinkPT Specific 200 k0a protein, putative   5   4   1   0   0   0.25   2.92     EHL 00150   Insark membrane calcium transporting   5   1   4   0   0   0.25   2.92     EHL 005370   Ingredani farty acid CoA ligase, putative   5   1   4   0   0   0.43   3.693     EHL 033050   Hypothetical protein   17   11   6   0   1.83   6.93     EHL 033050   Hypothetical protein   15   3   12   0   0   2.4   5.175     EHL 032050   Hypothetical protein   13   8	EHI_119600	domain containing protein	/	0	/	0	0	0	1.8
EHL00000   inositol polyphosphate 5 phosphates,   6   2   1   0   0   5   2.06     PHL01220   Cal/Cal NAL Cletin heavy subunit   6   4   2   0   0   2   1.15     EHL02270   Cal/Cal NAL Cletin heavy subunit   6   4   2   0   0   2   1.11     EHL01270   Cal/Cal NAL Cletin heavy subunit   6   4   2   0   0   2   1.11     EHL004010   Plasma membrane calcium transporting   5   1   4   0   0   0.25   2.92     EHL00430   Rap/Ran CTPase activating protein, putative   5   1   4   0   0   0.25   2.92     EHL08305   Thioredoxin domain containing protein 2, putative   5   1   4   0   0   0.25   2.92     EHL03350   Hypothetical protein   17   11   6   0   0   1.83   6.93     EHL03350   Hypothetical protein   17   12   5   0   0   2.2   3.11     EHL03306   Hypothetical protein   15	FHI 059670	Rah family CTPase	6	2	4	0	0	0.5	2.66
Introduction     putative     putative     putative     putative     putative     putative       EHL_01270     Gal/Gal MAc lectin heavy subunit     6     4     2     0     0     2     1.55       EHL_05170     Cal/Gal MAc lectin heavy subunit     6     4     2     0     0     2     1.11       EHL_05170     Cal/Gal MAC lectin heavy subunit     6     4     2     0     0     2.5     4.42       EHL_00160     Resma membrance calcium transporting     5     1     4     0     0     0.25     2.92       EHL18980     Long chain farty acid CoA ligase, putative     5     1     4     0     0     0.43     3.43       EHL031370     Hypothetical protein     17     11     6     0     0     1.44     3.4       EHL03230     Hypothetical protein     15     3     12     0     0     2.5     8.93       EHL03301     Hypothetical protein     13     8     5     0     0     1.3     5.03	EHI_055070	Inositol polyphosphate 5 phosphatase	6	5	1	0	0	5	2.00
EHL_012270     Cal/Cal M/x lectin heavy subunit     6     4     2     0     0     2     1.55       EHL_045170     US snRMP specific 200 kDa protein, putative     6     4     1     0     0     2     1.11       EHL_01620     Iron sultir flavoprotein pseudogene     5     4     1     0     0     2.5     .42       EHL_000430     Rap/Ran CTPase activating protein.putative     5     1     4     0     0     0.25     .252       EHL_108480     Thioredoxin domain containing protein 2.putative     5     4     1     0     0     4     .242       EHL_00380     Hypothetical protein     17     11     6     0     1.83     6.633       EHL_03320     Hypothetical protein     15     3     12     0     0     0.25     8.33       EHL_03306     Hypothetical protein     15     10     5     0     0     1.6     4.73       EHL_03306     Hypothetical protein     12     10     2     0     5     <	2111_100000	nutative	0	5		0	0	5	2.00
EHL B4170US \$nRNP specific 200 [Da protein, putative B41,011606410021.11EHL,01160Plasma membrane calcium transporting ATPase, putative51400.254.42EHL00180RayRan (TPase activating protein, putative B478889514000.252.52EHL18809Long chain fatty acid CoA ligase, putative B47888514000.252.52EHL18080Hypothetical protein Momain containing protein 2, putative 	EHI 012270	Gal/Gal NAc lectin heavy subunit	6	4	2	0	0	2	1.55
EHL 164520   Iron suffur havoprotein pseudogene   5   4   1   0   0   4   11.36     EHL 00100   Plasma membrane calcium transporting   5   1   4   0   0   0.25   2.92     EHL 000430   Rap/Ran CTPase activating protein, putative   5   1   4   0   0   0.25   2.57     EHL 198050   Unog chain fait ya cid Co ligase, putative   5   4   1   0   0   4   2.42     EHL 081370   Hypothetical protein   17   11   6   0   1.44   3.4     EHL 03350   Hypothetical protein   17   11   6   0   0   2.4   5.51     EHL 03350   Hypothetical protein   15   3   12   0   0   2.25   8.33     EHL 07250   Hypothetical protein   15   10   5   0   0   1.33   5.03     EHL 07250   Hypothetical protein   12   18   4   0   0   2.4   4.4     EHL 07260   Hypothetical protein   12   3   9	EHI_045170	U5 SnRNP specific 200 kDa protein, putative	6	4	2	0	0	2	1.11
EHL_00110Plasma membrane calcum transporting51400.254.42EHL_00130Rap/Ran CPrase activating protein, putative51400.252.92EHL_18850Long chain fatty cid CoA ligase, putative514000.252.92EHL_18080Thioredoxin domain containing protein 2, putative5410042.42EHL.01301Hypothetical protein22139001.836.93EHL023320Hypothetical protein17116002.45.51EHL02304Hypothetical protein15312002.45.51EHL03030Hypothetical protein1486001.64.73EHL03104Hypothetical protein12102000.333.57EHL03050Hypothetical protein1284000.333.57EHL12104Hypothetical protein1277000.333.57EHL121050Hypothetical protein10910098.71EHL12104Hypothetical protein10910095.45EHL121050Hypothetical protein10910095.45EHL121050Hypothetical protein97200.813.36 <td< td=""><td>EHI_164520</td><td>Iron sulfur flavoprotein pseudogene</td><td>5</td><td>4</td><td>1</td><td>0</td><td>0</td><td>4</td><td>11.36</td></td<>	EHI_164520	Iron sulfur flavoprotein pseudogene	5	4	1	0	0	4	11.36
ATTase, putative     No     No     No     So	EHI_001160	Plasma membrane calcium transporting	5	1	4	0	0	0.25	4.42
EHL EHL 1890Ray/Ray GTPase activating protein, putative514000.252.57EHL 189080Injoredoxin domain containing protein 2, putative5410042.42EHL 1081081Hypothetical protein171166001.836.93EHL033520Hypothetical protein17125002.45.51EHL073200Hypothetical protein15105002.43.11EHL07200Hypothetical protein15105002.33.11EHL073050Hypothetical protein1385001.64.73EHL013060Hypothetical protein12102002.24.54EHL121060Hypothetical protein1239000.333.57EHL05060Hypothetical protein1239000.333.57EHL174500Hypothetical protein10910098.71EHL174500Hypothetical protein10910098.71EHL174500Hypothetical protein963003.533.64EHL174500Hypothetical protein972000.353.64EHL174500Hypothetical protein972000.35 <t< td=""><td></td><td>ATPase, putative</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>		ATPase, putative							
EHL 188590Long chain fatty acid CoA ligase, putative514000.252.57EHL 190800Thioredoxin domain containing protein 2, putative5110042.42EHL 061870Hypothetical protein17116001.836.93EHL 032350Hypothetical protein17116000.258.93EHL 032300Hypothetical protein15312000.258.93EHL 072500Hypothetical protein151050023.11EHL07280Hypothetical protein1385001.64.73EHL07280Hypothetical protein121020024.54EHL078200Hypothetical protein12840024.54EHL058670Hypothetical protein1239000.333.57EHL058670Hypothetical protein1239000.333.57EHL058670Hypothetical protein10910098.71EHL05060Hypothetical protein10910027.25EHL174500Hypothetical protein10910027.45EHL174500Hypothetical protein972000.84.34EH	EHI_000430	Rap/Ran GTPase activating protein, putative	5	1	4	0	0	0.25	2.92
EHL 190880   Thioredoxin domain containing protein 2, putative   5   4   1   0   0   4.44   3.4     EHL 061870   Hypothetical protein   17   11   6   0   0   1.43   6.93     EHL 023320   Hypothetical protein   17   12   5   0   0   2.4   5.51     EHL 0273200   Hypothetical protein   15   10   5   0   0   2.3   3.11     EHL 073200   Hypothetical protein   13   8   6   0   0   1.6   4.73     EHL 073200   Hypothetical protein   12   10   2   0   0   2.3   5.03     EHL 013300   Hypothetical protein   12   10   2   0   0   2.4   5.75     EHL 059870   Hypothetical protein   12   3   9   0   0   0.33   3.57     EHL 059870   Hypothetical protein   10   9   1   0   0   9   8.71     EHL 1745400   Hypothetical protein   10   9   1   0	EHI_188590	Long chain fatty acid CoA ligase, putative	5	1	4	0	0	0.25	2.57
EHL 061870Hypothetical protein22139001.443.4EHL 032320Hypothetical protein1711601.836.93EHL 023320Hypothetical protein15312000.258.93EHL 0172500Hypothetical protein15312000.258.93EHL 017290Hypothetical protein1486001.335.03EHL 018390Hypothetical protein121020055.75EHL 0198970Hypothetical protein12840024.54EHL 112000Hypothetical protein1239000.333.57EHL 1126060Hypothetical protein1257000.712.28EHL 114560Hypothetical protein10910098.71EHL 117700Hypothetical protein10910027.27EHL 1196760Hypothetical protein945000.833.64EHL 117700Hypothetical protein972001.676.79EHL 117800Hypothetical protein972003.622.3EHL 117800Hypothetical protein972001.676	EHI_190880	Thioredoxin domain containing protein 2, putative	5	4	1	0	0	4	2.42
EHL 033550   Hypothetical protein   17   11   6   0   0   1.83   6.93     EHL 033550   Hypothetical protein   15   3   12   0   0   2.24   5.51     EHL 073200   Hypothetical protein   15   10   5   0   0   2.2   3.11     EHL 073200   Hypothetical protein   15   10   5   0   0   1.33   5.03     EHL 073200   Hypothetical protein   12   10   2   0   0   5   5.75     EHL 05870   Hypothetical protein   12   10   2   0   0   2.4   5.45     EHL 05870   Hypothetical protein   12   3   9   0   0   0.33   3.57     EHL 058660   Hypothetical protein   11   7   1   0   3   7   2.26     EHL 174540   Hypothetical protein   10   9   1   0   0   9   5.51     EHL 174540   Hypothetical protein   9   7   2   0   0   3.64	EHI_061870	Hypothetical protein	22	13	9	0	0	1.44	3.4
EHL 023320   Hypothetical protein   17   12   5   0   0   2.4   5.51     EHL 027320   Hypothetical protein   15   3   12   0   0   0.25   8.93     EHL 071200   Hypothetical protein   15   10   5   0   0   1.33   50.3     EHL 013000   Hypothetical protein   13   8   5   0   0   1.6   4.73     EHL 121000   Hypothetical protein   12   10   2   0   0   5   5.75     EHL 050660   Hypothetical protein   12   3   9   0   0   0.33   3.57     EHL 050660   Hypothetical protein   12   5   7   0   0   0.71   2.28     EHL 174500   Hypothetical protein   10   9   1   0   0   9   5.71     EHL 050609   Hypothetical protein   10   8   2   0   0   4   7.38     EHL 050509   Hypothetical protein   9   7   2   0   0   3.57	EHI_033550	Hypothetical protein	17	11	6	0	0	1.83	6.93
EHL_072500   Hypothetical protein   15   10   5   0   0   0.25   8.93     EHL_01306   Hypothetical protein   15   10   5   0   0   1.33   5.03     EHL_01306   Hypothetical protein   13   8   5   0   0   1.6   4.73     EHL_12106   Hypothetical protein   12   10   2   0   0   5   5.75     EHL05066   Hypothetical protein   12   3   9   0   0   0.33   3.57     EHL050666   Hypothetical protein   12   5   7   0   0   0.71   2.28     EHL174506   Hypothetical protein   11   7   1   0   3   7   2.65     EHL174506   Hypothetical protein   10   9   1   0   9   5.45     EHL17170   Hypothetical protein   9   6   3   0   0   2.2   7.27     EHL07505   Hypothetical protein   9   7   2   0   0   3.55   3.64	EHI_023320	Hypothetical protein	17	12	5	0	0	2.4	5.51
EHL_07300   Hypothetical protein   13   10   5   0   0   2   5.11     EHL_07290   Hypothetical protein   13   8   5   0   0   1.63   4.73     EHL_018390   Hypothetical protein   12   10   2   0   0   5   5.75     EHL_050870   Hypothetical protein   12   8   4   0   0   2   4.54     EHL_17200   Hypothetical protein   12   3   9   0   0   0.33   3.57     EHL_05060   Hypothetical protein   12   5   7   0   0   0.71   2.28     EHL_174540   Hypothetical protein   10   9   1   0   0   9   8.71     EHL_174560   Hypothetical protein   10   9   1   0   0   9   5.45     EHL005900   Hypothetical protein   9   6   3   0   0   2   7.27     EHL005900   Hypothetical protein   9   7   2   0   0.8   4.34 <t< td=""><td>EHI_0/2500</td><td>Hypothetical protein</td><td>15</td><td>3</td><td>12</td><td>0</td><td>0</td><td>0.25</td><td>8.93</td></t<>	EHI_0/2500	Hypothetical protein	15	3	12	0	0	0.25	8.93
EHL_07759   Hypothetical protein   14   5   0   0   1.53   5.05     EHL_071590   Hypothetical protein   12   10   2   0   0   5   5.75     EHL_01830   Hypothetical protein   12   10   2   0   0   2   4.54     EHL_172000   Hypothetical protein   12   3   9   0   0   0.33   3.57     EHL050600   Hypothetical protein   12   5   7   0   0   0.711   2.28     EHL17400   Hypothetical protein   11   7   1   0   3   7   2.65     EHL17450   Hypothetical protein   10   9   1   0   9   8.71     EHL017500   Hypothetical protein   10   8   2   0   0   4   7.38     EHL020510   Hypothetical protein   9   4   5   0   0   8.5   3.64     EHL020510   Hypothetical protein   9   7   2   0   0   3.55   3.64     EHL0	EHI_013000	Hypothetical protein	13	0	5	0	0	2 1 2 2	5.11
EHL_210055   Hypothetical protein   12   0   2   0   0   5   5.75     EHL_21006   Hypothetical protein   12   8   4   0   0   2   4.54     EHL_21006   Hypothetical protein   12   3   9   0   0   0.33   3.57     EHL050660   Hypothetical protein   12   5   7   0   0   0.71   2.28     EHL174500   Hypothetical protein   11   7   1   0   3   7   2.65     EHL196760   Hypothetical protein   10   9   1   0   0   9   8.71     EHL174500   Hypothetical protein   10   8   2   0   0   4   7.38     EHL117707   Hypothetical protein   9   6   3   0   0   2   7.27     EHL050609   Hypothetical protein   9   7   2   0   0   8.54   3.64     EHL010400   Hypothetical protein   9   5   4   0   0   1.67   6.79	EHI_077230	Hypothetical protein	13	8	5	0	0	1.55	4.73
EHL_050870   Hypothetical protein   12   8   4   0   0   2   4.54     EHL_050660   Hypothetical protein   12   3   9   0   0   0.33   3.57     EHL_050660   Hypothetical protein   12   5   7   0   0   0.71   2.28     EHL_174540   Hypothetical protein   11   7   1   0   3   7   2.65     EHL_196760   Hypothetical protein   10   9   1   0   0   9   8.71     EHL_19760   Hypothetical protein   10   9   1   0   0   9   5.45     EHL_06990   Hypothetical protein   10   9   1   0   0   2   7.27     EHL006990   Hypothetical protein   9   7   2   0   0   .8   4.34     EHL007750   Hypothetical protein   9   7   2   0   0   .8   .2   .3   .6   .3   .6   .2   .3   .6   .2   .3   .6   .6	FHI 121060	Hypothetical protein	12	10	2	0	0	5	5.75
HL_172000   Hypothetical protein   12   3   9   0   0   0.33   3.57     EHL_050660   Hypothetical protein   12   5   7   0   0   0.71   2.28     EHL_174540   Hypothetical protein   11   7   1   0   3   7   2.65     EHL_196760   Hypothetical protein   10   9   1   0   9   8.71     EHL_196760   Hypothetical protein   10   8   2   0   0   4   7.38     EHL_111770   Hypothetical protein   10   9   1   0   0   9   5.45     EHL_006990   Hypothetical protein   9   6   3   0   0   2   7.27     EHL_005930   Hypothetical protein   9   7   2   0   0   3.53   3.64     EHL_07530   Hypothetical protein   9   7   2   0   0   3.62   2.5   2.3     EH_01400   Hypothetical protein   8   2   6   0   0   0.33   1.52 </td <td>EHI_059870</td> <td>Hypothetical protein</td> <td>12</td> <td>8</td> <td>4</td> <td>0</td> <td>0</td> <td>2</td> <td>4 54</td>	EHI_059870	Hypothetical protein	12	8	4	0	0	2	4 54
EH_050660Hypothetical protein1257000.712.28EHI_174540Hypothetical protein11710372.65EHI_196760Hypothetical protein10910098.71EHI_174560Hypothetical protein10820047.38EHI_17170Hypothetical protein10910095.45EHI_00590Hypothetical protein9630027.27EHI_025310Hypothetical protein972000.884.34EHL_077750Hypothetical protein972003.553.64EHI_1110Hypothetical protein954001.252.3EHI_114110Hypothetical protein954000.331.52EHI_114110Hypothetical protein826000.331.52EHI_11410Hypothetical protein734000.7521.28EHI_11990Hypothetical protein716000.1713.75EHI_11990Hypothetical protein743001.3311.69EHI_119620Hypothetical protein743001.3311.69EHI_1106320Hypothetical protein7	EHI 172000	Hypothetical protein	12	3	9	0	0	0.33	3.57
EHLHypothetical protein11710372.65EHL196760Hypothetical protein10910098.71EHL174560Hypothetical protein10820047.38EHL174500Hypothetical protein10910095.45EHL10909Hypothetical protein9630027.27EHL025310Hypothetical protein972003.53.64EHL1090Hypothetical protein972001.676.79EHL1111111Hypothetical protein954000.331.52EHL1111111Hypothetical protein953000.7521.28EHL114110Hypothetical protein734000.7521.28EHL114110Hypothetical protein716000.1713.75EHL114110Hypothetical protein716001.3311.69EHL114100Hypothetical protein716001.3311.69EHL114100Hypothetical protein716001.3311.69EHL114300Hypothetical protein761<	EHI_050660	Hypothetical protein	12	5	7	0	0	0.71	2.28
EHL196760Hypothetical protein10910098.71EHLHypothetical protein10820047.38EHLHypothetical protein10910095.45EHL109630027.27EHLHypothetical protein9630027.27EHLHypothetical protein972000.8.34.34EHLPypothetical protein972003.53.64EHLHypothetical protein9030603.62EHLHypothetical protein954001.252.3EHLHypothetical protein853001.676.79EHLHypothetical protein734000.171.52EHLHypothetical protein734000.171.375EHLHypothetical protein716001.3311.69EHLHypothetical protein7060107.99EHLHypothetical protein76100.46.56EHLHypothetical protein72500.46.56	EHI_174540	Hypothetical protein	11	7	1	0	3	7	2.65
EHL_174560Hypothetical protein10820047.38EHL_111770Hypothetical protein10910095.45EHL_006990Hypothetical protein9630027.27EHL_025310Hypothetical protein945000.8.53.64EHL_025370Hypothetical protein972003.53.64EHL_013400Hypothetical protein9030603.62EHL_114110Hypothetical protein954001.252.3EHL_016900Hypothetical protein853001.676.79EHL_004180Hypothetical protein734000.7521.28EHL_119790Hypothetical protein734000.1713.75EHL_114540Hypothetical protein743001.3311.69EHL_10720Hypothetical protein7060107.99EHL_14340Hypothetical protein76100.46.56EHL_017780Hypothetical protein72500.46.56	EHI_196760	Hypothetical protein	10	9	1	0	0	9	8.71
EHL_111770   Hypothetical protein   10   9   1   0   0   9   5.45     EHL_006990   Hypothetical protein   9   6   3   0   0   2   7.27     EHL_025310   Hypothetical protein   9   4   5   0   0   0.8   4.34     EHL_077750   Hypothetical protein   9   7   2   0   0   3.64     EHL_10400   Hypothetical protein   9   7   2   0   0   3.62     EHL_114110   Hypothetical protein   9   5   4   0   0   1.67   6.79     EHL_016900   Hypothetical protein   8   5   3   0   0   1.67   6.79     EHL_11970   Hypothetical protein   7   3   4   0   0   0.17   13.75     EHL_145400   Hypothetical protein   7   1   6   0   0   1.69   1.69     EHL_145400   Hypothetical protein   7   4   3   0   0   1.69   1.69     EH	EHI_174560	Hypothetical protein	10	8	2	0	0	4	7.38
EHL_006990   Hypothetical protein   9   6   3   0   0   2   7.27     EHL_025310   Hypothetical protein   9   4   5   0   0   0.8   4.34     EHL_025310   Hypothetical protein   9   7   2   0   0   0.8   4.34     EHL_077750   Hypothetical protein   9   7   2   0   0   3.64     EHL_01400   Hypothetical protein   9   7   2   0   0   3.62     EHL_114110   Hypothetical protein   9   5   4   0   0   1.25   2.3     EHL_016900   Hypothetical protein   8   5   3   0   0   1.67   6.79     EHL_014180   Hypothetical protein   7   3   4   0   0   0.75   21.28     EHL_145460   Hypothetical protein   7   1   6   0   0   1.69   1.69     EHL_116320   Hypothetical protein   7   0   6   0   1.33   1.69     EHL_107040	EHI_111770	Hypothetical protein	10	9	1	0	0	9	5.45
EHI_025310   Hypothetical protein   9   4   5   0   0   0.8   4.34     EHI_077750   Hypothetical protein   9   7   2   0   0   3.5   3.64     EHI_103400   Hypothetical protein   9   0   3   0   6   0   3.62     EHI_114110   Hypothetical protein   9   5   4   0   0   1.25   2.3     EHI_016900   Hypothetical protein   8   5   3   0   0   1.67   6.79     EHI_004180   Hypothetical protein   8   2   6   0   0   0.33   1.52     EHI_119790   Hypothetical protein   7   3   4   0   0   0.75   21.28     EHI_145460   Hypothetical protein   7   1   6   0   0   1.375     EHI_106320   Hypothetical protein   7   4   3   0   0   1.69     EHI_106320   Hypothetical protein   7   0   6   0   1   0   7.99     EH	EHI_006990	Hypothetical protein	9	6	3	0	0	2	7.27
EHI_07/750   Hypothetical protein   9   7   2   0   0   3.5   3.64     EHI_103400   Hypothetical protein   9   0   3   0   6   0   3.62     EHI_103400   Hypothetical protein   9   0   3   0   6   0   3.62     EHI_016900   Hypothetical protein   9   5   4   0   0   1.25   2.3     EHI_016900   Hypothetical protein   8   5   3   0   0   1.67   6.79     EHI_10790   Hypothetical protein   7   3   4   0   0   0.33   1.52     EHI_119790   Hypothetical protein   7   1   6   0   0   0.17   13.75     EHI_145400   Hypothetical protein   7   4   3   0   0   1.69   7.99     EHI_106320   Hypothetical protein   7   0   6   0   1   0   7.99     EHI_1043490   Hypothetical protein   7   6   1   0   0.4   6.56	EHI_025310	Hypothetical protein	9	4	5	0	0	0.8	4.34
EHI_103400Hypothetical protein9030603.62EHI_114110Hypothetical protein954001.252.3EHI_016900Hypothetical protein853001.676.79EHI_004180Hypothetical protein826000.331.52EHI_119790Hypothetical protein734000.7521.28EHI_145460Hypothetical protein716000.1713.75EHI_106320Hypothetical protein743001.3311.69EHI_107040Hypothetical protein761067.39EHI_14390Hypothetical protein725000.46.56	EHI_077750	Hypothetical protein	9	/	2	0	0	3.5	3.64
EHI_114110Hypothetical protein54001.252.5EHI_016900Hypothetical protein853001.676.79EHI_004180Hypothetical protein826000.331.52EHI_119790Hypothetical protein734000.7521.28EHI_145460Hypothetical protein716000.1713.75EHI_106320Hypothetical protein743001.3311.69EHI_107040Hypothetical protein761067.39EHI_144390Hypothetical protein72500.46.56	EHI_103400	Hypothetical protein	9	0	3	0	6	1.25	3.62
EHI_000180Hypothetical protein656661.676.79EHI_004180Hypothetical protein826000.331.52EHI_119790Hypothetical protein734000.7521.28EHI_145460Hypothetical protein716000.1713.75EHI_106320Hypothetical protein743001.3311.69EHI_107040Hypothetical protein7060107.99EHI_14390Hypothetical protein725000.46.56	ETI_114110	Hypothetical protein	8	5	4	0	0	1.20	2.5 6.70
EHL_107100Hypothetical protein736000.531.52EHL_119790Hypothetical protein734000.7321.28EHL_145460Hypothetical protein716000.1713.75EHL_106320Hypothetical protein743001.3311.69EHL_107040Hypothetical protein7060107.99EHL_14390Hypothetical protein72500.46.56	EHI_010900	Hypothetical protein	0 8	2	5	0	0	0.33	1.52
EHL_145460Hypothetical protein71600.1713.75EHL_10520Hypothetical protein743001.3311.69EHL_107040Hypothetical protein7060107.99EHL_14390Hypothetical protein7610067.39EHL_017780Hypothetical protein725000.46.56	EHI 119790	Hypothetical protein	7	2	4	0	0	0.75	21.22
EHI_106320Hypothetical protein743001.3311.69EHI_107040Hypothetical protein7060107.99EHI_14390Hypothetical protein7610067.39EHI_017780Hypothetical protein72500.46.56	EHI 145460	Hypothetical protein	, 7	1	6	0	0	0.17	13.75
EHI_107040   Hypothetical protein   7   0   6   0   1   0   7.99     EHI_14390   Hypothetical protein   7   6   1   0   0   6   7.39     EHI_017780   Hypothetical protein   7   2   5   0   0.4   6.56	EHI_106320	Hypothetical protein	7	4	3	0	0	1.33	11.69
EHI_144390     Hypothetical protein     7     6     1     0     0     6     7.39       EHI_017780     Hypothetical protein     7     2     5     0     0.4     6.56	EHI_107040	Hypothetical protein	7	0	6	0	1	0	7.99
EHI_017780     Hypothetical protein     7     2     5     0     0.4     6.56	EHI_144390	Hypothetical protein	7	6	1	0	0	6	7.39
	EHI_017780	Hypothetical protein	7	2	5	0	0	0.4	6.56

determine the precise function of hypothetical proteins listed in Table 1. Homologs for some of these genes are also found in AmoebaDB database. Few genes of *E. histolytica* and their homologs are listed in Table 2. High degree of inter-species genetic variability is also observed among genes of *E. histolytica* and their homologs. A total of 326 SNPs have been identified within *E. dispar hsp70* gene (EDI\_012650) in comparison to that of *E. histolytica* (EHI\_159140). Similarly, 520 inter-species SNPs have been detected in the homologous gene of *lgl2* (EDI\_244250) present in *E. dispar* SAW760 strain.

Homologous gene for AIG1 family protein (EDI\_001050) contains 144 inter-species SNPs. A total of 103 SNPs were identified in the homologous gene for actinin like protein (EDI\_207850) present in *E. dispar* SAW760 strain. Homologous gene for elongation factor alpha 1 (EDI\_134610) also contains a total of 90 SNPs. A total of 254 SNPs have also been identified in the homologous gene for inositol polyphosphate-5-phosphatase (EDI\_159070). Another important virulence factor of *E. histolytica* is lysine and glutamic acid rich protein (KERP1). KERP1 is a surface-associated protein of

#### Table 2

Genes of E. histolytica and their homologs present in AmoebaDB database.

Amene basel     Protein product for this protein product for this protein products for this protein products for this protein products for this protein products for this protein protein protein protein protein protein     Protein product for this protein protein     Protein     Protein	E. histolytica HM1:IMSS		E. dispar SAW760		E. invadens IP1		E. moshkovskii Laredo		
BHL03500     Sende theorem end proteines     PSU-08300     Ignotesical end proteines     Psu-08300     Resultance     Psu-08300     Resultance     Psu-08300     Resultance     Psu-08300     Concord and proteines     Concord and proteines <thconcord and="" proteines<="" th="">     Concord and pr</thconcord>	AmoebaDB ID	Protein product for this gene	AmoebaDB ID	Protein product for this gene	AmoebaDB ID	Protein product for this gene	AmoebaDB ID	Protein product for this gene	
FH1.0533GulCal Nak letrin Ig12FUAFurit report containing protein, putative protein, putative protein, putative protein, putativeFUA.05358Furit report containing protein, putative protein, putative protein, putative protein, putativeFUA.05358Furit report containing protein, putative protein, putative protein, putative protein, putativeFUA.05358Furit report containing protein, putative protein, putativeFUA.05358Furit report containing protein, putative protein, putativeFUA.05358Furit report containing protein, putativeFUA.05358	EHI_073630	Serine threonine isoleucine rich protein, putative	EDI_083900	Hypothetical protein	EIN_092260	Hypothetical protein	EMO_033950	Serine threonine isoleucine rich protein, putative	
EHIL 1944     Hear shock protein 70.     EMIC 00590     Earl Shock protein 70.     Putative       EHIL 006990     GalGal Nac Icerin Ig11     EDI 20120     First Shock protein 70.     First Shock protein 70.       EHIL 19420     Tyrssine kinase, putative     EDI 20150     Erine Attrinomine protein, putative     ENI 00210     First Shock protein 70.       EHIL 19440     And polymerase, putative     EDI 20150     Erine Attrinomine protein     ENI 00210     First Shock protein 70.       EHIL 19442     And polymerase, putative     EDI 20150     Erine Attrinomine protein     ENI 00210     First Shock protein 70.       EHIL 19442     And polymerase, putative     EDI 201500     First Shock protein 70.     ENI 00210     First Shock protein 70.       EHIL 19442     And polymerase, putative     EDI 201500     First Shock protein 70.     ENI 00200     First Shock protein 70.       EHIL 19442     And polymerase, putative     EDI 201200     First Shock protein 70.     ENI 00200     First Shock protein 70.       EHIL 19442     And polymerase, putative     EDI 201200     First Shock protein 70.     ENI 00200     First Shock protein 70.       EHIL 19442     And polymerase, putati	EHI_065330	Gal/Gal NAc lectin lgl2	EDI_244250	Furin repeat containing protein, putative	EIN_065850	Furin repeat containing protein, putative	EMO_010790	Gal/Gal Nac lectin lgl2	
H11.00889Gal.Gal Nac lectin Ig11ED.244250Priorite practor protein, putativeEIN.05820Furni repeat containing protein, putativeEM0.010790Gal/Gal Nac lectin subunit Ig12EH1.12450Tyrsine kinase, putativeED.00150Serine/threonine protein, insare ITI, putativeEN.03204Hypothetical protein, conservedEM0.05700DNA polymerase, putativeEH1.14472AIG I family proteinED.01050Hypothetical protein, conserved-*-*-*-*-*EH1.14542Pospholipid ransporting portein, putativeED.010700Galinin, putativeEN.013700Pospholipid ransporting portein, putativeEM0.010700Antini lite protein, portativeEH1.013500Galinin putativeED.010707Serine-threonine protein kinase, putativeSerine-threonine protein putativeED.010707Serine-threonine protein putativeEM0.010700Galinin lite protein, containing proteinEH1.01120ED.010710Calactose -inhibitable lectin sito a subunit precursor putativeED.010710Serine-threonine protein putative-*-*EM0.01700Galicose- mutativeEH1.01120ED.010710Calactose inhibitable lectin tinase, putativeED.010710Calactose inhibitable lectin precursor, putative-*-*-*EM0.01700Galicose- mutativeEH1.01202ED.010701ED.114700ED.014700Calactose inhibitable lectin precursor, putativeEM0.01700Galicose- 	EHI_159140	Heat shock protein 70, putative	EDI_012650	Heat shock protein 70 kDa, putative	_a	_a	EMO_060560	Heat shock protein 70, putative	
LHI.12450Tyrosine kinase, putativeEN.00210Proteine srine-threonineEM.0.09220Tyrosine kinase, putativeLII.16193DA polymerase, putativeHypothetical protein, conserveda.*	EHI_006980	Gal/Gal Nac lectin lgl1	EDI_244250	Furin repeat containing protein, putative	EIN_065850	Furin repeat containing protein, putative	EMO_010790	Gal/Gal Nac lectin subunit lgl2	
EHL 16419   DNA polymerase, putative   EUD.05541   Hypothetical protein, conserved   END.02570   DNA polymerase, putative   END.057601   Phypothetical protein, conserved     EHL 16442   AC1 family protein   EDU.01050   Hypothetical protein, putative   -* <td>EHI_124500</td> <td>Tyrosine kinase, putative</td> <td>EDI_004150</td> <td>Serine/threonine protein kinase HT1, putative</td> <td>EIN_000210</td> <td>Protein serine/threonine kinase, putative</td> <td>EMO_009220</td> <td>Tyrosine kinase, putative</td>	EHI_124500	Tyrosine kinase, putative	EDI_004150	Serine/threonine protein kinase HT1, putative	EIN_000210	Protein serine/threonine kinase, putative	EMO_009220	Tyrosine kinase, putative	
EHLHerLHerLHerL-'EHL1000000000000000000000000000000000000	EHI_164190	DNA polymerase, putative	EDI_056410	Hypothetical protein, conserved	EIN_032840	Hypothetical protein, conserved	EMO_057600	DNA polymerase, putative	
EHI. 14440Actian like protein, putativeEJL.207850Grainin, putativeEJN.037840Grainin, putativeEJN.037840Grainin, putativeEJN.037840Grainin, putativeFM.03704Phospholipid ransporting putativePhospholipid ransporting ransportingPhospholipid ransporting ransportingPhospholipid ransporting ransportingPhospholipid ransporting ransportingPhospholipid ransporting 	EHI_144270	AIG1 family protein	EDI_001050	Hypothetical protein, conserved	_a	_a	_ <sup>a</sup>	_ <sup>a</sup>	
EHI.1352/20 Phospholipid transporting EIN.03730 Phospholipid transporting Phospholipid transporting   PH.020350 Protein Kinase domain ATFase, putative EIN.016310 Serinethreonine protein Kinase, putative EIN.016310 Serinethreonine protein EIN.015210 Phospholipid transporting   PH.020350 Calactose inhibitable EID.012170 Serinethreonine protein Kinase, putative EIN.016210 Serinethreonine protein EIN.016210 EIN.017620 EIN.0176200 EIN.017620 EIN.017620 EIN.017620 EIN.017620 EIN.017620 EIN.017620 EIN.017620 EIN.017620 EIN.017620 EIN.0176400 EIN.0	EHI_164440	Actinin like protein, putative	EDI_207850	Grainin, putative	EIN_037840	Grainin, putative	EMO_010570	Actinin like protein, putative	
EHL_03260Protein kinase domain kinase, putativeENL_012370Serinethreonine protein kinase, putativeEM_0.01230Protein kinase domain containing protein sinase, putativeEM_0.01230Protein kinase domain containing protein sinase, putativeEM_0.01230Protein kinase domain containing protein sinase, putativeEM_0.01230Protein kinase domain containing protein 	EHI_135220	Phospholipid transporting p-type ATPase, putative	EDI_018000	Phospholipid transporting ATPase, putative	EIN_038730	Phospholipid transporting ATPase, putative	EMO_035200	Phospholipid transporting p-type ATPase, putative	
EHL035690   Galactose inhibitable   EDL023210   Galactose-inhibitable   EMO_050130   Galactose-inhibitable   Sk Das subunit precursor   EMO_0123750   Elongation factor 1-alpha   Elongation factor 1-alpha   EMO_0123750 <td>EHI_023050</td> <td>Protein kinase domain containing protein</td> <td>EDI_012370</td> <td>Serine-threonine protein kinase, putative</td> <td>EIN_016310</td> <td>Serine-threonine protein kinase, putative</td> <td>EMO_012200</td> <td>Protein kinase domain containing protein</td>	EHI_023050	Protein kinase domain containing protein	EDI_012370	Serine-threonine protein kinase, putative	EIN_016310	Serine-threonine protein kinase, putative	EMO_012200	Protein kinase domain containing protein	
EHLEINEI	EHI_035690	Galactose inhibitable lectin 35 kDa subunit precursor	EDI_023210	Galactose-inhibitable lectin 35 kDa subunit precursor, putative	_a	_a	EMO_050130	Galactose-inhibitable lectin 35 kDa subunit precursor	
HL139439Levcine rich repeat protein BspA family BJP conservedEND_64421END_054241Mouterial protein conservedEND_054241Mouterial protein conservedEND_07680Leucine rich repeat protein BspA family BJP consort, putative BJP consort, putative BLP consort, putativeEND_054241Mouterial alpha-glucosidase AB precursor, putative BLN_058210END_0112400Glycosyl hydrolase, family BJP consortEHL02330Glactose specific adhesin kinase, putativeED_147070Noka surface lexin precursor, putative precursor, putative mam, putative putativeINO 08210TO KDS surface lexin proteinMO.0670700EMO.067700Cal/GalNAc lexin heavy subunit adma, putative gamma, putative putativeMouterial alph-glucosidase admase catalytic subunit gamma, putative gamma, putativeEMO.058710Mouterial alph-glucosidase admase catalytic subunit gamma, putative gamma, putative gamma, putativeEMO.058701Mouterial alph-glucosidase admase catalytic subunit gamma, putative gamma, putative gamma, putativeEMO.058701Mouterial alph-glucosidase admase catalytic subunit gamma, putative gamma, putative gamma, putative putativeEMO.058701Mouterial alph-glucosidase admase catalytic subunit gamma, putative 	EHI_011210	Elongation factor alpha 1	EDI_134610	Elongation factor 1-alpha	EIN_146970	Elongation factor 1-alpha, putative	EMO_123750	Elongation factor 1-alpha 1	
EHL_023230Glycosyl hydrolase family 31 proteinEDL_13780Neutral alpha-glucosidase AB precursor, putative AB precursor, putative 	EHI_139430	Leucine rich repeat protein BspA family	EDI_284090	Hypothetical protein, conserved	EIN_054420	Hypothetical protein, conserved	EMO_007680	Leucine rich repeat protein BspA family	
EHLCalactose specific adhesin 170 kDa subini, putative precursor, putative precursor, putative precursor, putative precursor, putative gamma, putativeEMO_066770Cal/CalNAc lectin heavy subunitEHL_119600Ubiquitin carboxyl terminal hydrolase domain containing proteinEDL_02310Ubiquitin specific protease, putativeEIN_20010Hyothetical proteinEMO_066770Cal/CalNAc lectin heavy subunitEHL_05670Ubiquitin carboxyl 	EHI_023430	Glycosyl hydrolase family 31 protein	EDI_137800	Neutral alpha-glucosidase AB precursor, putative	EIN_108320	Neutral alpha-glucosidase AB precursor, putative	EMO_112400	Glycosyl hydrolase, family 31 protein	
EHLPhosphatidyl linositol 3 kinase, putativeEDL147070 3-kinase catalytic subunit gamma, putative gamma, putativePhosphatidylinositol 3-kinase catalytic subunit gamma, putative gamma, putativeEMO_071620Phosphatidylinositol 3-kinase, catalytic subunit gamma, putativeEHL_0196700Vibiquitin carboxyl terminal hydrolase domain containing proteinEDL_150907Trichohyalin, putative 	EHI_042370	Galactose specific adhesin 170 kDa subunit, putative	EDI_213670	170 kDa surface lectin precursor, putative	EIN_068210	170 kDa surface lectin precursor, putative	EMO_066770	Gal/GalNAc lectin heavy subunit	
EHI_119600Ubiquitin carboxyl terminal hydrolase domain containing proteinEDI_023410Ubiquitin specific protease, putativeEIN_200010Hypothetical proteinEMO_025900Ubiquitin carboxyl- terminal hydrolase domain containing proteinEHI_059670Rab family GTPase b fosphatase, putativeEDI_156940Trichohyalin, putativeEIN_157460Trichohyalin, putative trisphosphate 5- phosphatase precursor, putativeEMO_059660Rab family GTPaseRab family GTPaseEHI_012270Gal/Gal Nac lectin heavy subunitEDI_213670Trichohyalin, putative trisphosphate 5- 	EHI_013980	Phosphatidyl linositol 3 kinase, putative	EDI_147070	Phosphatidylinositol 3-kinase catalytic subunit gamma, putative	EIN_020710	Phosphatidylinositol 3-kinase catalytic subunit gamma, putative	EMO_071620	Phosphatidylinositol 3-kinase, putative	
EHLENLStabEDL156940Trichohyalin, putativeEIN_157460Trichohyalin, putativeEMO_059660Rab family GTPaseEHLInositol polyphosphateEDLType II inositol-1,4,5-ENO_012640Inositol polyphosphate5-phosphatase precursor,phosphatase precursor,phosphatase precursor,phosphatase precursor,phosphatase precursor,putativeFIN_0068210Gal/GalNac lectin heavySubunit	EHI_119600	Ubiquitin carboxyl terminal hydrolase domain containing protein	EDI_023410	Ubiquitin specific protease, putative	EIN_200010	Hypothetical protein	EMO_025900	Ubiquitin carboxyl- terminal hydrolase domain containing protein	
5 phosphatase, putative   trisphosphate 5-   phosphatase precursor, phosphatase precursor, putative   putative     EHI_012270   Gal/Gal Nac lectin heavy subunit   ED_213670   170 kDa surface lectin   EIN_068210   170 kDa surface lectin   EMO_066770   Gal/GalNac lectin heavy subunit     EHI_045170   U5 SnRNP specific   ED_076220   U5 small nuclear   EIN_09340   U5 small nuclear   EMO_014940   U5 snRNP-specific 200 kDa protein, putative     EHI_045170   Iron sulfur flavoprotein   ED_064980   Hypothetical protein, conserved   ribonucleoprotein 200 kDa helicase, putative   EIN_091700   Hypothetical protein, conserved   FMO_0029730   Iron-sulfur flavoprotein, putative     EHI_001160   Plasma membrane   EIN_017670   Plasma membrane   EIN_22480   Plasma membrane calcium-transporting   EIM_03200   Rap CRap GTPase, putative   ATPase, putative   Rap CRan GTPase   Rap CRan GTPase   Rap CRan GTPase, putative   ATPase, putative   ATPase, putative   Raprotein, putative   Rap CRan GTPase- activa	EHI_059670 EHI_160860	Rab family GTPase Inositol polyphosphate	EDI_156940 EDI_159070	Trichohyalin, putative Type II inositol-1,4,5-	EIN_157460 EIN_020640	Trichohyalin, putative Type II inositol-1,4,5-	EMO_059660 EMO_012640	Rab family GTPase Inositol polyphosphate-5-	
EHI_012270Gal/Gal Nac lectin heavy subunitEDI_213670170 kDa surface lectin precursor, putativeEIN_068210170 kDa surface lectin precursor, putativeEMO_066770Gal/GalNAc lectin heavy subunitEHI_045170U5 SnRNP specific 200 kDa protein, putativeED_076220U5 small nuclear ribonucleoprotein 200 kDa helicase, putativeEIN_093940U5 small nuclear ribonucleoprotein 200 kDa helicase, putativeEMO_014940U5 snRNP-specific 200 kDa protein, putativeEHI_164520Iron sulfur flavoprotein pseudogeneED_064980Hypothetical protein, conservedEIN_091700Hypothetical protein, conservedEMO_098730Iron-sulfur flavoprotein, putativeEHI_000160Plasma membrane calcium transporting ATPase, putativeED_013570Plasma membrane calcium-transporting ATPase, putativeEIN_023200Rap GTPase-activating protein, putativeEIN_03200Rap GTPase-activating protein, putativeATPase, putative ATPase, putativeEHI_000430Rap/Ran GTPase activating protein, putativeED_09250Rap GTPase-activating protein, putativeEIN_016090Rap GTPase-activating protein, putativeEMO_002900Long-chain-fatty-acid—CoA ligase, putativeEMO_002900Long-chain-fatty-acid—CoA ligase, putativeEHI_19080Thioredoxin domain containing protein 2,ED_197960Hypothetical protein, conservedEIN_163620FMO_002900Long-chain-fatty-acid—CoA ligase, putativeEHI_19080Thioredoxin domain containing protein 2,ED_197960Hypothetical protein, conservedEIN_163620		5 phosphatase, putative		trisphosphate 5- phosphatase precursor, putative		trisphosphate 5- phosphatase precursor, putative		phosphatase, putative	
EHI_045170U5 SnRNP specific 200 kDa protein, putativeEDI_076220U5 small nuclear ribonucleoprotein 200 kDa helicase, putativeEIN_093940U5 small nuclear ribonucleoprotein 200 kDa helicase, putativeEMO_014940U5 snRNP-specific 200 kDa protein, putativeEHI_06420Iron sulfur flavoprotein pseudogeneEDI_064980Hypothetical protein, conservedEIN_091700Hypothetical protein, conservedEMO_006200EMO_006202Iron-sulfur flavoprotein, putativeEHI_001160Plasma membrane calcium transporting ATPase, putativeEDI_013570Plasma membrane 	EHI_012270	Gal/Gal Nac lectin heavy subunit	EDI_213670	170 kDa surface lectin precursor, putative	EIN_068210	170 kDa surface lectin precursor, putative	EMO_066770	Gal/GalNAc lectin heavy subunit	
EHI_164520Iron sulfur flavoprotein pseudogeneEDI_064980Hypothetical protein, conservedEIN_091700Hypothetical protein, conservedEIM_0908730Iron-sulfur flavoprotein, putativeEHI_001160Plasma membrane calcium transporting ATPase, putativeEDI_013570Plasma membrane calcium-transporting ativeEIN_222480Plasma membrane calcium- transporting ATPase, putativeEIN_222480Plasma membrane calcium- transporting ATPase, putativeEIN_0006020Plasma membrane 	EHI_045170	U5 SnRNP specific 200 kDa protein, putative	EDI_076220	U5 small nuclear ribonucleoprotein 200 kDa helicase, putative	EIN_093940	U5 small nuclear ribonucleoprotein 200 kDa helicase, putative	EMO_014940	U5 snRNP-specific 200 kDa protein, putative	
EHI_001160Plasma membrane calcium transporting ATPase, putativeEDI_013570Plasma membrane calcium-transporting ATPase, putativeEIN_222480Plasma membrane calcium- transporting ATPase, putativeEIN_0006020Plasma membrane calcium-transporting ATPase, putativeEHI_000430Rap/Ran GTPase activating protein, putativeEDI_026850Rap GTPase-activating protein, putativeEIN_013200Rap GTPase-activating protein, putativeEIN_03200Rap GTPase-activating protein, putativeEIN_016090Long-chain-fatty-acid-CoAEMO_002290Rap/Ran GTPase-activating protein, putativeEHI_188590Long chain fatty acid CoA ligase, putativeEDI_093250Long-chain-fatty-acid- CoA ligase, putativeEIN_016090Long-chain-fatty-acid-CoA ligase, putativeEMO_002990Long-chain-fatty-acid-CoA ligase, putativeEHI_190880Thioredoxin domain 	EHI_164520	Iron sulfur flavoprotein pseudogene	EDI_064980	Hypothetical protein, conserved	EIN_091700	Hypothetical protein, conserved	EMO_098730	Iron–sulfur flavoprotein, putative	
EHI_000430   Rap/Ran GTPase   EDI_026850   Rap GTPase-activating protein, putative   EIN_033200   Rap GTPase-activating protein, putative   EMO_022230   Rap/Ran GTPase-activating protein, putative     eHI_188590   Long chain fatty acid CoA   EDI_093250   Long-chain-fatty-acid—   EIN_016090   Long-chain-fatty-acid—CoA   EMO_002290   Long-chain-fatty-acid—CoA     Igase, putative   CoA ligase, putative   EIN_163620   Hypothetical protein   EIN_016090   EMO_002900   Long-chain-fatty-acid—CoA     EHI_190880   Thioredoxin domain containing protein 2,   Conserved   EIN_163620   Hypothetical protein   EIN_009010   Thioredoxin domain-containing protein 2,	EHI_001160	Plasma membrane calcium transporting	EDI_013570	Plasma membrane calcium-transporting	EIN_222480	Plasma membrane calcium- transporting ATPase,	EMO_006020	Plasma membrane calcium-transporting	
EHI_18850   Long chain fatty acid CoA   EDI_093250   Long-chain-fatty-acid—   EIN_016090   Long-chain-fatty-acid—CoA   EMO_002990   Long-chain-fatty-acid—CoA     Igase, putative   CoA ligase, putative   ligase, putative   ligase, putative   ligase, putative     EHI_190880   Thioredoxin domain   EDI_197960   Hypothetical protein, conserved   EIN_163620   Hypothetical protein   EMO_099010   Thioredoxin domain-containing protein 2,	EHI_000430	Rap/Ran GTPase activating protein,	EDI_026850	Rap GTPase-activating protein, putative	EIN_033200	Rap GTPase-activating protein, putative	EMO_022230	Rap/Ran GTPase-activating protein, putative	
Ingase, putativeCOA ligase, putativeligase, putativeligase, putativeEHI_190880Thioredoxin domainEDI_197960Hypothetical protein,EIN_163620Hypothetical proteinEMO_099010Thioredoxin domain- containing protein 2,ConservedConservedConservedConservedConservedConserved	EHI_188590	Durative Long chain fatty acid CoA	EDI_093250	Long-chain-fatty-acid—	EIN_016090	Long-chain-fatty-acid—CoA	EMO_002990	Long-chain-fatty-acid—CoA	
putative putative	EHI_190880	ngase, putative Thioredoxin domain containing protein 2, putative	EDI_197960	COA ligase, putative Hypothetical protein, conserved	EIN_163620	ngase, putative Hypothetical protein	EMO_099010	ngase, putative Thioredoxin domain- containing protein 2, putative	

<sup>a</sup> Homolog of the corresponding gene is not found in the particular Entamoeba species [as per AmoebaDB database (www.AmoebaDB.org)].

*E. histolytica* and has been shown to be involved in the parasite adherence to human enterocytes. It is also an important virulence factor in liver abscess pathogenesis [63,64]. *kerp1* gene has been found in both *E. histolytica* (AmoebaDB i.d. EHI\_098210) and *E. nuttalli* (AmoebaDB i.d. ENU1\_189420) but not in *E. dispar* [64,65]. Analysis of AmoebaDB database version 4.1 [25, www.amoebadb.org] revealed that inter-species genetic variability within *kerp1* gene was present among *E. histolytica* 

and *E. nuttalli*. A total of 10 SNPs were identified within *E. nuttalli kerp1* sequence (ENU1\_189420) in comparison to that of *E. histolytica* (EHI\_098210). However, no intra-species genetic variability has been observed within the gene (EHI\_098210). Genome sequencing of *E. histolytica* clinical isolates has also identified SNPs within *cyclicin-2* gene, significantly associated with asymptomatic and liver abscess outcomes [62]. This indicates that *cyclicin-2* could be an important

virulence determinant of E. histolytica. Studies of comparative genomic hybridization of E. histolytica and E. dispar strains suggested relatively low genomic diversity among *E. histolytica* [10]. A recent study by Weedall et al. has also identified a low level of single nucleotide diversity within E. histolytica populations [66]. Sequence analysis of defined regions also suggests similar observations [11,67]. Such low level of genetic diversity suggests a relatively recent common ancestor for E. histolytica [14]. However, this observation was quite incongruous with a recent report by Gilchirst et al. [62]. Multilocus sequence typing of E. histolytica clinical isolates identified extensive population diversity, suggesting that the genotypes of individual parasites do not contain consistent phylogenetic signals. They have blamed genetic recombination events for such a result, since it can break down the linkage between target loci and assist to form loci with different genealogies [62]. Hence, an important question regarding the population structure of Entamoeba is whether the parasite populations are predominantly clonal or sexual.

Sexual reproduction can help parasite to improve the fitness of their progeny [68]. Parasitic protists are continuously exposed to exogenous environmental factors and host immune pressure, which can alter the chemical structure and stability of their genome [68]. Parasites should repair structural alteration in their genome, since it can lead to mutations, deletions, insertions, translocation and loss of essential genetic information [68]. Parasites remove their DNA damage by recombinational DNA repair mechanism and this allows greater survival of offspring with undamaged DNA [68]. It is also an important mechanism to generate genetic diversity used by parasites to evade host immune response [68]. This particular feature of parasite is quite important, since sexual reproduction can exchange genes, responsible for drug resistance and parasite virulence. This could generate selectively advantageous genotypes that can spread very rapidly through host population [14]. Sexual reproduction can also help in the removal of deleterious genes. Current deleterious mutations brought together by sexual reproduction create unfit individuals that are eliminated from the population [68]. The genome of E. histolytica contains meiotic genes like SPO11, DMC1, and MND1 and many homologous recombination (HR) specific genes like MLH1, MSH2, RAD21 and RAD51 [22,69,68]. Moreover, ploidy changes and unscheduled gene amplification, which indicate the possibility of recombination have also been reported in Entamoeba [68]. E. histolytica contain a large number of retrotransposons in its genome, which also indicates their ability to reproduce by sexual means [68]. Organisms which reproduce solely by asexual means would eventually lose these retrotransposons from their genome [68]. However, Singh et al. recently provide the first direct demonstration of HR in Entamoeba using a construct with inverted repeats, which upon recombination results in sequence inversion. Increased rate of genetic recombination has been reported in Entamoeba under stress conditions and during encystation process [68]. Stage inter-conversion between cyst and trophozoite is crucial for disease transmission and pathogenesis in E. histolytica [68]. In addition to this, few indirect evidences of genetic recombination have also been identified in Entamoeba through population genetic studies. Complete genome sequencing of 10 axenic E. histolytica cell lines has identified pattern of polymorphism, indicates that recombination has occurred in the history of the population studied [66]. Concatenated genealogy based on repetitive loci (i.e. D-A and N-K2) also revealed the possibility of genetic recombination among E. histolytica population [60]. Bioinformatics comparison of Gal/GalNAc lectin among E. histolytica and its non-pathogenic sibling E. dispar also identified the evidence of gene conversion within the lineages [50].

Transposable elements constitute a significant portion of *E. histolytica* genome and they can affect the expression of adjacent genes [37]. Phenotypic characteristic of this parasite is also influenced by their genomic location [37]. Variability in genomic distribution of SINE1 and SINE2 among *E. histolytica* clinical isolates has been recently studied by Kumari et al. [70]. Several loci with extensive polymorphism of SINE occupancy among *E. histolytica* strains have been identified [70].

#### 7. Conclusion

Queries related to evolution and population structure of *E. histolytica* still remains to be investigated. One of the concerning issue is whether E. histolytica population is sexual or clonal. Circumstantial evidence suggested that Entamoeba might engage in genetic recombination at some stage in their life-cycle. However, further detailed investigations with Entamoeba and other early branching protists are required to understand the origin of their sexual reproduction and to determine the variety of mechanisms by which these organisms exchange their DNA. Another major question that arises is whether E. histolytica population from ALA patients is genetically closer to that of asymptomatic individuals. If they are close (few studies suggested this), then individuals with persistent asymptomatic E. histolytica infection may be under high risk of developing ALA in the future. Prompt preventive measures should be undertaken for such individuals. Advanced whole genome sequencing of *E. histolytica* clinical isolates can be helpful to address this question.

### Acknowledgments

This study was supported by grant from the Okayama University Program of Founding Research Centre for Emerging and Re-emerging Infectious Disease (OUP 2- 5), Ministry of Education, Culture, Sports, Science and Technology of Japan. The authors would like to thank Mrs. Debarati Ganguly for her immense help regarding proof-reading of this manuscript.

#### References

- W. H. O./PAHO/UNESCO. A consultation with experts on amebiasis. Epidemiol Bull 1997;18:13–4.
- [2] Haghighi A, Kobayashi S, Takeuchi T, Masuda G, Nozaki T. Remarkable genetic polymorphism among *Entamoeba histolytica* isolates from a limited geographic area. J Clin Microbiol 2002;40:4081–90.
- [3] Haghighi A, Kobayashi S, Takeuchi T, Thammapalerd N, Nozaki T. Geographic diversity among genotypes of *Entamoeba histolytica* field isolates. J Clin Microbiol 2003; 41:3748–56.
- [4] Walsh JA. Problems in recognition and diagnosis of amebiasis: estimation of the global magnitude of morbidity and mortality. Rev Infect Dis 1986;8:228–38.
- [5] Stanley Jr SL. Amoebiasis. Lancet 2003;361:1025–34.
- [6] Mehmet T, Petri Jr WA. Laboratory diagnosis of amebiasis. Clin Microbiol Rev 2003; 16(4):713–29.
- [7] Diamond LS, Clark CG. A redescription of *Entamoeba histolytica* Schaudinn, 1903 (Emended Walker, 1911) separating it from *Entamoeba dispar* Brumpt, 1925. J Eukaryot Microbiol 1993;40:340–4.
- [8] Ali IKM, Mondal U, Roy S, Haque R, Petri Jr WA, et al. Evidence for a link between parasite genotype and outcome of infection with *Entamoeba histolytica*. J Clin Microbiol 2007;45:285–9.
- [9] Duggal PR, Haque R, Roy S, Mondal D, Sack RB, et al. Influence of human leukocyte antigen class II alleles on susceptibility to *Entamoeba histolytica* infection in Bangladeshi children. J Infect Dis 2004;189:520–6.
- [10] Shah PH, MacFarlane RC, Bhattacharya D, Matese JC, Demeter J, et al. Comparative genomic hybridizations of *Entamoeba* strains reveal unique genetic fingerprints that correlate with virulence. Eukaryot Cell 2005;4:504–15.
- [11] Bhattacharya D, Haque R, Singh U. Coding and noncoding genomic regions of *Entamoeba histolytica* have significantly different rates of sequence polymorphisms: implications for epidemiological studies. J Clin Microbiol 2005;43:4815–9.
- [12] Das K, Mukherjee AK, Chowdhury P, Sehgal R, Bhattacharya MK, et al. Multilocus sequence typing system (MLST) reveals a significant association of *Entamoeba histolytica* genetic patterns with disease outcome. Parasitol Int 2014;63:308–14.
- [13] Escueta-de Cadiz A, Kobayashi S, Takeuchi T, Tachibana H, Nozaki T. Identification of an avirulent *Entamoeba histolytica* strain with unique tRNA-linked short tandem repeat markers. Parasitol Int 2010;59:75–81.
- [14] Weedall GD, Hall N. Evolutionary genomics of *Entamoeba*. Resmic, 162; 2011 637-45.
- [15] Clark CG, Kaffashian F, Tawari B, Windsor JJ, Twigg-Flesner A, et al. New insights into the phylogeny of *Entamoeba* species provided by analysis of four new small-subunit rRNA genes. IJSEM 2006;56:2235–9.
- [16] Dolabella SS, Serrano-Luna J, Navarro-García F, Cerritos R, Ximénez C, et al. Amoebic liver abscess production by *Entamoeba dispar*. Ann Hepatol 2012;11(1):107–17.
- [17] Shimokawa C, Kabir M, Taniuchi M, Mondal D, Kobayashi S, et al. Entamoeba moshkovskii is associated with diarrhea in infants and causes diarrhea and colitis in mice. J Infect Dis 2012;206(5):744–51.
- [18] Ali IKM, Clark CG, Petri Jr WA. Molecular epidemiology of amoebiasis. Infect Genet Evol 2008;8(5):698–707.

- [19] Ximénez C, Morán P, Rojas L, Valadez A, Gómez A. Reassessment of the epidemiology of amebiasis: state of the art. Infect Genet Evol 2009;9(6):1023–32.
- [20] Heredia RD, Fonseca JA, López MC. Entamoeba moshkovskii: perspectives of a new agent to be considered in the diagnosis of amebiasis. 2012;123(3):139–45.
- [21] Royer TL, Gilchrist C, Kabir M, Arju T, Ralston KS, et al. Entamoeba bangladeshi nov. sp., Bangladesh. Emerg Infect Dis 2012;18(9):1543–5.
- [22] Loftus B, Anderson I, Davies R, Alsmark UC, Samuelson J, et al. The genome of the protist parasite *Entamoeba histolytica*. Nature 2005;433:865–8.
- [23] Lorenzi HA, Puiu D, Miller JR, Brinkac LM, Amedeo P, et al. New assembly, reannotation and analysis of the *Entamoeba histolytica* genome reveal new genomic features and protein content information. PLoS Negl Trop Dis 2010;4:716.
- [24] Clark CG, Alsmark UC, Tazreiter M, Saito-Nakano Y, Ali V, et al. Structure and content of the *Entamoeba histolytica* genome. Adv Parasitol 2007;65:51–190.
- [25] Aurrecoechea C, Barreto A, Brestelli J, Brunk BP, Caler EV, et al. AmoebaDB and MicrosporidiaDB: functional genomic resources for Amoebozoa and Microsporidia species. Nucleic Acids Res 2011;39(Database issue):D612–9.
- [26] Gelderman AH, Keister DB, Bartgis IL, Diamond LS. Characterization of the deoxyribonucleic acid of representative strains of *Entamoeba histolytica*. J Parasitol 1971; 57:906–11.
- [27] Willhoeft U, Tannich E. The electrophoretic karyotype of *Entamoeba histolytica*. Mol Biochem Parasitol 1999;99:41–53.
- [28] Mukherjee C, Clark CG, Lohia A. Entamoeba shows reversible variation in ploidy under different growth conditions and between life cycle phases. PLoS Negl Trop Dis 2008;2:281.
- [29] Bhattacharya S, Som I, Bhattacharya A. The ribosomal DNA plasmids of entamoeba. Parasitol Today 1998;14(5):181–5.
- [30] Ghosh S, Zaki M, Clark CG, Bhattacharya S. Recombinational loss of a ribosomal DNA unit from the circular episome of *Entamoeba histolytica* HM-1:IMSS. Mol Biochem Parasitol 2001;116:105–8.
- [31] Jansson A, Gillin F, Kagardt U, Hagblom P. Coding of hemolysins within the ribosomal RNA repeat on a plasmid in *Entamoeba histolytica*. Science 1994;263:1440–3.
- [32] Tawari B, Ali IKM, Scott C, Quail MA, Berriman M, et al. Patterns of evolution in the unique tRNA gene arrays of the genus *Entamoeba*. Mol Biol Evol 2008;25:187–98.
- [33] Clark CG, Ali IK, Zaki M, Loftus BJ, Hall N. Unique organisation of tRNA genes in Entamoeba histolytica. Mol Biochem Parasitol 2006;146:24–9.
- [34] Ali IKM, Solaymani-Mohammadi S, Akhter J, Roy S, Gorrini C, et al. Tissue invasion by Entamoeba histolytica: evidence of genetic selection and/or DNA reorganization events in organ tropism. PLoS Negl Trop 2008;2:e219.
- [35] Lorenzi H, Thiagarajan M, Haas B, Wortman J, Hall N, et al. Genome wide survey, discovery and evolution of repetitive elements in three *Entamoeba* species. BMC Genomics 2008;9:595.
- [36] Dellen KV, Field J, Wang Z, Loftus B, Samuelson J. LINEs and SINE-like elements of the protist Entamoeba histolytica. Gene 2002;297:229–39.
- [37] Kumari V, Sharma R, Yadav VP, Gupta AK, Bhattacharya A, et al. Differential distribution of a SINE element in the *Entamoeba histolytica* and *Entamoeba dispar* genomes: role of the LINE-encoded endonuclease. BMC Genomics 2011;12:267.
- [38] Yadav VP, Mandal PK, Bhattacharya A, Bhattacharya S. Recombinant SINEs are formed at high frequency during induced retrotransposition in vivo. Nat Commun 2012;3:854.
- [39] Pritham EJ, Feschotte C, Wessler SR. Unexpected diversity and differential success of DNA transposons in four species of entamoeba protozoans. Mol Biol Evol 2005;22: 1751–63.
- [40] Gilchrist CA, Houpt E, Trapaidze N, Fei Z, Crasta O, et al. Impact of intestinal colonization and invasion on the *Entamoeba histolytica* transcriptome. Mol Biochem Parasitol 2006;147:163–76.
- [41] Davis PH, Zhang Z, Chen M, Zhang X, Chakraborty S, et al. Identification of a family of BspA like surface proteins of *Entamoeba histolytica* with novel leucine rich repeats. Mol Biochem Parasitol 2006;145:111–6.
- [42] Prudovsky I, Tarantini F, Landriscina M, Neivandt D, Soldi R, et al. Secretion without Golgi. J Cell Biochem 2008;103:1327–43.
- [43] Wang Z, Samuelson J, Clark CG, Eichinger D, Paul J, et al. Gene discovery in the Entamoeba invadens genome. Mol Biochem Parasitol 2003;129:23–31.
- [44] Saito-Nakano Y, Loftus BJ, Hall N, Nozaki T. The diversity of Rab GTPases in Entamoeba histolytica. Exp Parasitol 2005;110:244–52.
- [45] Nakada-Tsukui K, Saito-Nakano Y, Husain A, Nozaki T. Conservation and function of Rab small GTPases in *Entamoeba*: annotation of *E. invadens* Rab and its use for the understanding of *Entamoeba biology*. Exp Parasitol 2010;126:337–47.
- [46] Nakada-Tsukui K, Saito-Nakano Y, Ali V, Nozaki T. A retromerlike complex is a novel Rab7 effector that is involved in the transport of the virulence factor cysteine protease in the enteric protozoan parasite *Entamoeba histolytica*. Mol Biol Cell 2005; 16(11):5294–303.

- [47] Mitra BN, Saito-Nakano Y, Nakada-Tsukui K, Sato D, Nozaki T. Rab11B small GTPase regulates secretion of cysteine proteases in the enteric protozoan parasite *Entamoeba histolytica*. Cell Microbiol 2007;9(9):2112–25.
- [48] Hernandes-Alejandro M, Calixto-Gálvez M, López-Reyes I, Salas-Casas A, Cázares-Ápatiga J, et al. The small GTPase EhRabB of *Entamoeba histolytica* is differentially expressed during phagocytosis. Parasitol Res 2013;112(4):1631–40.
- [49] Juárez-Hernández LJ, García-Pérez RM, Salas-Casas A, García-Rivera G, Orozco E, et al. *Entamoeba histolytica*: the over expression of a mutated EhRabB protein produces a decrease of in vitro and in vivo virulence. Exp Parasitol 2013;133(3): 339–45.
- [50] Weedall GD, Sherrington J, Paterson S, Hall N. Evidence of gene conversion in genes encoding the Gal/GalNac lectin complex of *Entamoeba*. PLoS Negl Trop Dis 2011; 5(6):1209.
- [51] Matthiesen J, Bär AK, Bartels AK, Marien D, Ofori S, et al. Overexpression of specific cysteine peptidases confers pathogenicity to a nonpathogenic *Entamoeba histolytica* clone. Mol Biol 2013;4(2).
- [52] Teixeira JE, Huston CD. Participation of the serine-rich Entamoeba histolytica protein in amebic phagocytosis of apoptotic host cells. Infect Immun 2008;76(3):959–66.
- [53] de la Vega H, Specht CA, Semino CE, Robbins PW, Eichinger D, et al. Cloning and expression of chitinases of Entamoebae. Mol Biochem Parasitol 1997;85:139–47.
- [54] Rivera WL, Santos SR, Kanbara H. Prevalence and genetic diversity of *Entamoeba histolytica* in an institution for the mentally retarded in the Philippines. Parasitol Res 2006;98:106–10.
- [55] Zhang T, Stanley Jr SL. DNA vaccination with the serine-rich Entamoeba histolytica protein (SREHP) prevents amebic liver abscess in rodent models of disease. Vaccine 1999;18:868–74.
- [56] Ali IKM, Zaki M, Clark CG. Use of PCR amplification of tRNA gene-linked short tandem repeats for genotyping *Entamoeba histolytica*. J Clin Microbiol 2005;43:5842–7.
- [57] Feng M, Cai J, Yang B, Fu Y, Min X, et al. Unique short tandem repeat nucleotide sequences in *Entamoeba histolytica* isolates from China. Parasitol Res 2012;111: 1137–42.
- [58] Jaiswal V, Ghoshal U, Mittal B, Dhole TN, Ghoshal UC. Association between allelic variation due to short tandem repeats in tRNA gene of *Entamoeba histolytica* and clinical phenotypes of amoebiasis. Act Trop 2014;133:1–7.
- [59] Ali IKM, Haque R, Alam F, Kabir M, Siddique A, et al. Evidence for a link between locus R–R sequence type and outcome of infection with *Entamoeba histolytica*. Clin Microbiol Infect 2012;18:235–7.
- [60] Zermeno V, Ximenez C, Moran P, Valadez A, Valenzuela O, Rascon E. Worldwide genealogy of *Entamoeba histolytica*: an overview to understand haplotype distribution and infection outcome. Infect Genet Evol 2013;17:243–52.
- [61] Zaki M, Reddy SG, Jackson TF, Ravdin JI, Clark CG. Genotyping of *Entamoeba* species in South Africa: diversity, stability, and transmission patterns within families. J Infect Dis 2003;187:1860–9.
- [62] Gilchrist CA, Ali IKM, Kabir M, Alam F, Scherbakova S, et al. A Multilocus Sequence Typing System (MLST) reveals a high level of diversity and a genetic component to *Entamoeba histolytica* virulence. BMC Microbiol 2012;12:151.
- [63] Seigneur M, Mounier J, Prevost MC, Guillen N. A lysine- and glutamic acid-rich protein, KERP1, from *Entamoeba histolytica* binds to human enterocytes. Cell Microbiol 2005;7:569–79.
- [64] Santi-Rocca J, Weber C, Guigon G, Sismeiro O, Coppée JY, et al. The lysine- and glutamic acid-rich protein KERP1 plays a role in *Entamoeba histolytica* liver abscess pathogenesis. Cell Microbiol 2008;10:202–17.
- [65] Perdomo D, Baron B, Rojo-Domi'nguez A, Raynal B, England P, et al. The α-helical regions of KERP1 are important in *Entamoeba histolytica* adherence to human cells. Nat Sci Reports 2013;3:1171.
- [66] Weedall GD, Clark CG, Koldkjær P, Kay S, Bruchhaus I, et al. Genomic diversity of the human intestinal parasite *Entamoeba histolytica*. Genome Biol 2012;13:38.
- [67] Ghosh S, Frisardi M, Ramirez-Avila L, Descoteaux S, Sturm-Ramirez K, et al. Molecular epidemiology of *Entamoeba* spp.: evidence of a bottleneck (Demographic sweep) and transcontinental spread of diploid parasites. J Clin Microbiol 2000;38:3815–21.
- [68] Singh N, Bhattacharya A, Bhattacharya S. Homologous recombination occurs in *Entamoeba* and is enhanced during growth stress and stage conversion. PLoS One 2013;8(9):e74465.
- [69] Stanley Jr SL. The Entamoeba histolytica genome: something old, something new, something borrowed and sex too? Trends Parasitol 2005;21:451–3.
- [70] Kumari V, Iyer LR, Roy R, Bhargava V, Panda S, et al. Genomic distribution of SINEs in Entamoeba histolytica strains: implication for genotyping. BMC Genomics 2013;14: 432.