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

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

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

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

ABSTRACT

Amoebiasis caused by the gastrointestinal parasite *Entamoeba histolytica* 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 evolutionary genomics of *E. histolytica* 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 pathogenic species of *Entamoeba*.

Contents

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

* 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

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 3×10^7 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^{GTT} and Lys^{CTT} 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]. *Eh*LINEs (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 *Eh*LINEs in the R4 clade of non-LTR elements, a mixed clade of non-LTR elements that includes members from nematodes, insects, and vertebrates [36]. *Eh*LINEs share a common 3′ end sequence with *Eh*LSINEs (SINEs of *E. histolytica*) which indicates that they are involved in the retro transposition of *Eh*LSINEs. *Eh*LSINEs 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 *Eh*Rab1, *Eh*Rab2, *Eh*Rab5, *Eh*Rab7, *Eh*Rab8, *Eh*Rab11, and *Eh*Rab21 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 (*Eh*RabB) during phagocytosis of target cells, suggesting the potential role of *Eh*RabB protein in phagocytosis process [48]. *Eh*RabB 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 inter-isolate 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^{TGA}*-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

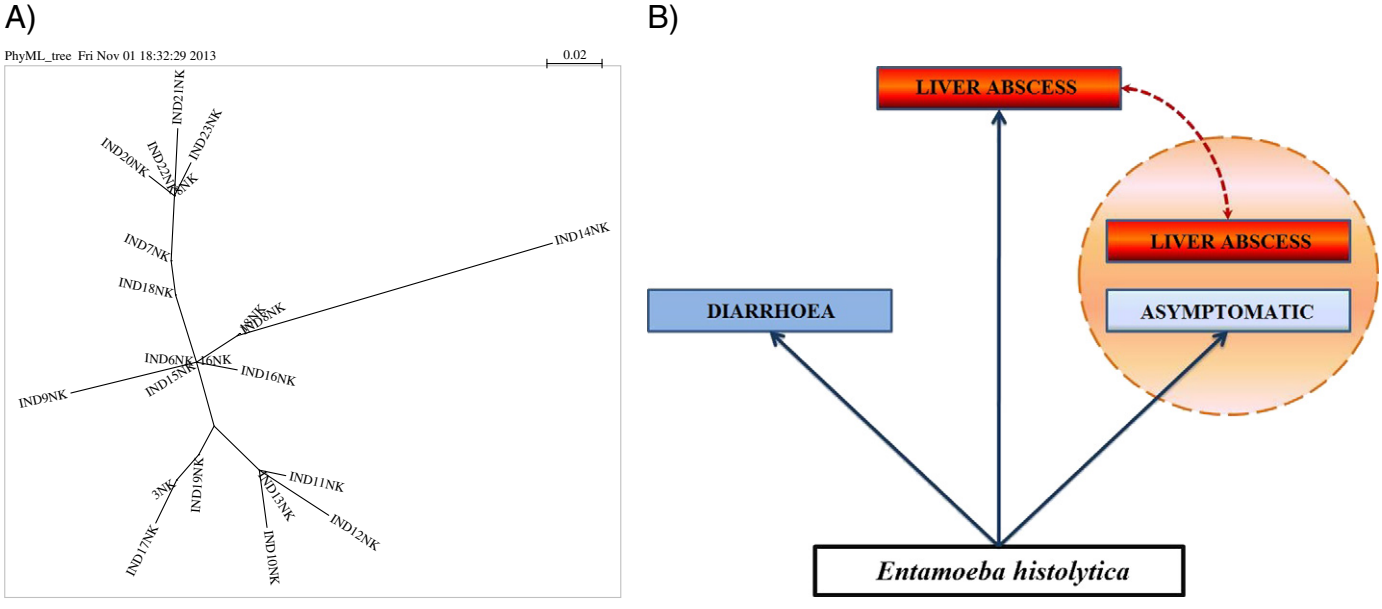

Fig. 1. *E. histolytica* isolates remaining asymptomatic are genetically closer to those causing liver abscess: depicted by (A) phylogenetic tree, (B) graphical representation. Phylogeny was based on tRNA linked *N-K2* (STR) locus. The sequences of all (a total of 22) representative STR patterns from *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 exclusively in diarrheal outcome. 'LA' group contains STR patterns found only in liver abscess outcome. 'AS + LA' group contains STR patterns exclusive for asymptomatic (AS) and liver abscess (LA) outcome.

Table 1
Genes of *E. histolytica*, contain intra-species single nucleotide polymorphisms (SNPs).

AmoebaDB ID	Protein product for this gene	Total SNPs	Non-synonymous SNPs	Synonymous SNPs	Non-sense SNPs	Non-coding SNPs	Non-synonymous SNP/synonymous SNP ratio	SNPs per kb (CDS)
EHI_073630	Serine threonine isoleucine rich protein, putative	70	46	24	0	0	1.92	4.6
EHI_065330	Gal/Gal NAc lectin lgl2	27	13	14	0	0	0.93	8.14
EHI_159140	Heat shock protein70, putative	14	12	2	0	0	6	6.94
EHI_006980	Gal/Gal NAc lectin lgl1	13	9	4	0	0	2.25	3.89
EHI_124500	Tyrosine kinase, putative	13	9	4	0	0	2.25	1.68
EHI_164190	DNA polymerase, putative	12	6	6	0	0	1	3.13
EHI_144270	AIG1 family protein	11	6	5	0	0	1.2	12.75
EHI_164440	Actinin like protein, putative	9	0	9	0	0	0	5.74
EHI_135220	Phospholipid transporting p-type ATPase, putative	9	3	6	0	0	0.5	3.05
EHI_023050	Protein kinase domain containing protein	9	6	3	0	0	2	2.55
EHI_035690	Galactose inhibitable lectin 35 kDa subunit precursor	8	4	4	0	0	1	8.57
EHI_011210	Elongation factor alpha 1	8	0	8	0	0	0	5.99
EHI_139430	Leucine rich repeat protein BspA family	8	3	5	0	0	0.6	3.96
EHI_023430	Glycosyl hydrolase family 31 protein	8	4	4	0	0	1	3.06
EHI_042370	Galactose specific adhesin 170 kDa subunit, putative	8	6	2	0	0	3	2.05
EHI_013980	Phosphatidyl linositol 3 kinase, putative	8	3	5	0	0	0.6	1.86
EHI_119600	Ubiquitin carboxyl terminal hydrolase domain containing protein	7	0	7	0	0	0	1.8
EHI_059670	Rab family GTPase	6	2	4	0	0	0.5	2.66
EHI_160860	Inositol polyphosphate 5 phosphatase, putative	6	5	1	0	0	5	2.06
EHI_012270	Gal/Gal NAc lectin heavy subunit	6	4	2	0	0	2	1.55
EHI_045170	U5 SnRNP specific 200 kDa protein, putative	6	4	2	0	0	2	1.11
EHI_164520	Iron sulfur flavoprotein pseudogene	5	4	1	0	0	4	11.36
EHI_001160	Plasma membrane calcium transporting ATPase, putative	5	1	4	0	0	0.25	4.42
EHI_000430	Rap/Ran GTPase activating protein, putative	5	1	4	0	0	0.25	2.92
EHI_188590	Long chain fatty acid CoA ligase, putative	5	1	4	0	0	0.25	2.57
EHI_190880	Thioredoxin domain containing protein 2, putative	5	4	1	0	0	4	2.42
EHI_061870	Hypothetical protein	22	13	9	0	0	1.44	3.4
EHI_033550	Hypothetical protein	17	11	6	0	0	1.83	6.93
EHI_023320	Hypothetical protein	17	12	5	0	0	2.4	5.51
EHI_072500	Hypothetical protein	15	3	12	0	0	0.25	8.93
EHI_013060	Hypothetical protein	15	10	5	0	0	2	3.11
EHI_077290	Hypothetical protein	14	8	6	0	0	1.33	5.03
EHI_018390	Hypothetical protein	13	8	5	0	0	1.6	4.73
EHI_121060	Hypothetical protein	12	10	2	0	0	5	5.75
EHI_059870	Hypothetical protein	12	8	4	0	0	2	4.54
EHI_172000	Hypothetical protein	12	3	9	0	0	0.33	3.57
EHI_050660	Hypothetical protein	12	5	7	0	0	0.71	2.28
EHI_174540	Hypothetical protein	11	7	1	0	3	7	2.65
EHI_196760	Hypothetical protein	10	9	1	0	0	9	8.71
EHI_174560	Hypothetical protein	10	8	2	0	0	4	7.38
EHI_111770	Hypothetical protein	10	9	1	0	0	9	5.45
EHI_006990	Hypothetical protein	9	6	3	0	0	2	7.27
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	0.17	13.75
EHI_106320	Hypothetical protein	7	4	3	0	0	1.33	11.69
EHI_107040	Hypothetical protein	7	0	6	0	1	0	7.99
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.

E. histolytica HM1:IMSS		*E. dispar* SAW760		*E. invadens* IP1		*E. moshkovskii* Laredo	
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
EHI_073630	Serine threonine isoleucine rich protein, putative	EDI_083900	Hypothetical protein	EIN_092260	Hypothetical protein	EMO_033950	Serine threonine isoleucine rich protein, putative
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
EHI_159140	Heat shock protein 70, putative	EDI_012650	Heat shock protein 70 kDa, putative	–[a]	–[a]	EMO_060560	Heat shock protein 70, putative
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
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
EHI_164190	DNA polymerase, putative	EDI_056410	Hypothetical protein, conserved	EIN_032840	Hypothetical protein, conserved	EMO_057600	DNA polymerase, putative
EHI_144270	AIG1 family protein	EDI_001050	Hypothetical protein, conserved	–[a]	–[a]	–[a]	–[a]
EHI_164440	Actinin like protein, putative	EDI_207850	Grainin, putative	EIN_037840	Grainin, putative	EMO_010570	Actinin like protein, putative
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
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
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
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
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
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
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_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
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
EHI_059670	Rab family GTPase	EDI_156940	Trichohyalin, putative	EIN_157460	Trichohyalin, putative	EMO_059660	Rab family GTPase
EHI_160860	Inositol polyphosphate 5 phosphatase, putative	EDI_159070	Type II inositol-1,4,5-trisphosphate 5-phosphatase precursor, putative	EIN_020640	Type II inositol-1,4,5-trisphosphate 5-phosphatase precursor, putative	EMO_012640	Inositol polyphosphate-5-phosphatase, putative
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_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_164520	Iron sulfur flavoprotein pseudogene	EDI_064980	Hypothetical protein, conserved	EIN_091700	Hypothetical protein, conserved	EMO_098730	Iron–sulfur flavoprotein, putative
EHI_001160	Plasma membrane calcium transporting ATPase, putative	EDI_013570	Plasma membrane calcium-transporting ATPase, putative	EIN_222480	Plasma membrane calcium-transporting ATPase, putative	EMO_006020	Plasma membrane calcium-transporting ATPase, putative
EHI_000430	Rap/Ran GTPase activating protein, putative	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 ligase, putative	EDI_093250	Long-chain-fatty-acid—CoA ligase, putative	EIN_016090	Long-chain-fatty-acid—CoA ligase, putative	EMO_002990	Long-chain-fatty-acid—CoA ligase, putative
EHI_190880	Thioredoxin domain containing protein 2, putative	EDI_197960	Hypothetical protein, conserved	EIN_163620	Hypothetical protein	EMO_099010	Thioredoxin domain-containing protein 2, putative

[a] 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

[1] 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-Domı'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.