

8-15-2021

## Illumination On The Structure And Characteristics Of Entamoeba Histolytica Genome.

Musafer H. Al-Ardi

Lecturer at The General Directorate for Education\Al-Qadisiyah -Ministry of education –Iraq.,  
mussafir78@yahoo.com

Follow this and additional works at: <https://qjps.researchcommons.org/home>

---

### Recommended Citation

Al-Ardi, Musafer H. (2021) "Illumination On The Structure And Characteristics Of Entamoeba Histolytica Genome.," *Al-Qadisiyah Journal of Pure Science*: Vol. 26: No. 4, Article 3.

DOI: 10.29350/qjps.2021.26.4.1311

Available at: <https://qjps.researchcommons.org/home/vol26/iss4/3>

This Article is brought to you for free and open access by Al-Qadisiyah Journal of Pure Science. It has been accepted for inclusion in Al-Qadisiyah Journal of Pure Science by an authorized editor of Al-Qadisiyah Journal of Pure Science. For more information, please contact [bassam.alfarhani@qu.edu.iq](mailto:bassam.alfarhani@qu.edu.iq).



## Illumination on the structure and characteristics of *Entamoeba histolytica* genome.

Authors Names	ABSTRACT
<p>Musafer H. Al-Ardi</p> <p><b>Article History</b>            Received on: 1/6/2021            Revised on: 15/6/2021            Accepted on: 17/6/2021</p> <p><b>Keywords:</b>  <i>Entamoeba histolytica</i> ;            Genome ; gene families ;            Transposable elements.</p> <p><b>DOI:</b> <a href="https://doi.org/10.29350/jops.2021.26.4.1311">https://doi.org/10.29350/jops.2021.26.4.1311</a></p>	<p><i>Entamoeba histolytica</i>, likes other organisms, is characterized by diversity and heterogeneity in its genetic content, which is one of the most paramount reasons for survival, and the increase in susceptibility to infection. Non-condensation of chromosomes during the process of cell division and the ambiguity of the chromosomal ploidy makes predicting the exact chromosomal numeral difficult. Genes distributed across 14 chromosomes as well as many extra-chromosome elements. Most Genes compose of one axon only, with Introns in 25% of Genes. This genome is characterized by the presence of Polymorphic internal repeat regions, and several gene families, one of these large families encoding Transmembrane kinas, Cysteine protease (CP), SREHP protein, and others.</p> <p>In conclusion genome of <i>Entamoeba</i> spp. similar all Eukaryotic Organisms when has introns. Transposable elements , un flexible ploidy and presence many Polymorphic internal repeat regions all these made it be strong Parasite.</p>

### Introduction

In 1893, Quincks and Ross discover the cyst phase of *Entamoeba*, whereas Schaudinn, in 1903, called it *Entamoeba histolytica* due to its ability to invasion tissues and distinguish it from *E. coli* (Pinilla et al., 2008).

Clifford Dobell first mentioned the existence of several amoeba species in 1919, each of which had a cystic phase with four nuclei. (Marianne, 2010). Emile Brumpt, in 1925, demonstrated the existence of two strains representing the same species (Dhaliwal and Juyal, 2013). Sargeant and Williams asserted that there could be two strains of this species, one capable of invading tissues and the other didn't (Marianne, 2010). In 1993, Diamond and Clark proved that the second strain is another species belonging to this genus, which was later called *E. dispar* (Zaki, 2002).

In 2005, the complete genetic sequence of *E. histolytica* and the strain HM-1: IMSS was revealed, which opened up new horizons of research towards a broader and more comprehensive study of this parasite (Battacharye and Battacharye, 2013).

*Entamoeba* spp. includes Several species that infect different species of hosts, *E. muris* that infect mice, *E. bovis* ( cows), *E. deblickei* ( goats and pigs), and other species (Hooshyar et al., 2015), in addition

to six different species found in the large intestine of humans (*E. histolytica*, *E. dispar*, *E. hartmanni*, *E. polecki*, *E. moshkovskii*, *E. coli*) (Tanyukse and Petri, 2003).

In addition to the inability to induce disease, the non-wide spread, and the small size of the non-pathogenic species, there are many important differences between the two sections (invasive and non-invasive) (Schuster and Visvesvara, 2004). The absence or difference of some genetic groups such as (SINE) in the non-pathogenic group which presence in *E. histolytica*, the presence and activity of some surface proteins such as cystine CP5, the activity of the amoebapore protein, all these variants are due to the genetic makeup differences among amoeba species genomes (Zaki, 2002).

## Whole Genome

*Entamoeba histolytica* genome is the first precursor genome which sequences is completed in 2005 (Das and Ganguly, 2014) using the Shotgun method which is the best method for sequencing. Its genome count estimates at 23.7 million base pairs compared with *Plasmodium falciparum* genome size (23 M bases), and the free-living amoeba *Dictyostelium discoideum* (34 M bases) (Clark et al., 2007).

The genome re-sequenced in 2010, with a genome size of 20.8 megabases (MB), distributes in 14-17 chromosomes. The difference occurs because of the removal of genes with repeating regions and the deletion of genes less than 300 base pairs which there is no clear evidence of their action (López-Camarillo et al., 2011).

The complete genome distributes in an estimated number of chromosomes (about 14 chromosomes) (Weedall and Hall, 2011) as well as, many circular extra-chromosome elements that include a gene region of rDNA that can be replicated (Episome). *E. histolytica* genome does not have Microsatellites, so, measuring Genetic diversity and estimation of gene community composition depends on other genetic markers (Black and Seed, 2004).

The genome of *Entamoeba* genus has 8300 genes, representing an average of 1260.9 base pairs (about 49.7% ) of genome size. The shortest gene is 147 base pairs and the longest is 15,210 base pairs. An Introns occupy 24.4% of the expected genome size (Clark et al., 2010), which means that a quarter of *E. histolytica* genes contain introns, as well as 6% of the genes with double introns, third of these genes (about 31.8% ), produce heterogenic proteins (Brendan et al., 2005).

Most genes contain a single exon, however, 25% of them may have splicing. In general, the size of the genes is small due to the absence of introns, as these genes encode proteins up to 389 amino acids in length (López-Camarillo et al. ., 2009).

---

Among the genes, there are many polymorphic tandem repeat regions, these regions encoding the amoeba-rich serine (SREHP) and chitinase, as well as, many sites for short repeats of tRNA (tRNA-STR loci) that are used as genetic markers and to distinguish between the Genotypes patterns, which are associated with various clinical markers, also indicate high levels of diversity in the *E. histolytica* population (Weedall and Hall, 2011).

The size and number of genes in *E. histolytica* are proportional to the metabolic adaptations required by the parasite, where the deletion or reduction of most cellular redox pathways in the mitochondria is observed with the presence of some enzymes similar to the oxidative enzymes in the eukaryotes (López-Camarillo et al., 2009).

It is expected that a large portion of the genes is accidentally transferred from bacteria to the *Amoeba* genome, and there are shreds of evidence that indicate the functions of these genes in the amoeba metabolism (Srivastava, 2005). Also, the genome encodes a large number of receptors such as kinases receptor and a variety of genes families that important in the parasite's virulence including Cysteine and Metallo-proteinases (López-Camarillo et al., 2009).

### **Ploidy and chromosomes**

The knowledge about the structure and composition of the chromosomes in the amoeba is little because the chromosomes do not condense during the mitosis of cell division, therefore, a kind of ambiguity surrounds the chromosome group (Chavez-Munguia et al., 2006). The huge variations between homologous chromosomes in different isolation makes predicting the accurate chromosome number is difficult, (Brendan et al., 2005) as it cannot be confirmed that it is haploid, diploid, or even quadruple with the presence of two or one groups of them in some studies (Mukherjee et al., 2008). This variation in the number and structure of chromosomes and the type of chromosomal ploidy even at the level of a single strain cell that growing in different conditions, whether in vivo or in vitro, maybe due to the ebb and flow of the subtelomeric repeats region as in the case of other Protista (Bagchi, 2001). Interestingly, this region contains tRNA microarrays in *E. histolytica* (Brendan et al., 2005). The heterogeneity of chromosomal ploidy may also be a delusion that the parasites are multi-chromosomal (Ghosh et al., 2000).

There is no reliable information about the size and nature of the centromeres, nor about the peripheral region, due to it does not exist or branched in a way that cannot be distinguished. despite the presence of genes encoding for histone proteins H1, H2A, H3, H4, chromatin surveys with an electron microscope, it becomes evident the presence of nuclear particle-like structures consisting of a base protein bound with DNA, this protein differs from it in the other Eukaryotic cells (Black and Seed, 2004).

### **Extra-chromosomes structures genes**

Outside the chromatin, many rDNA molecules are circular in structures. These structures are important in the phenotype and this raises several questions, including whether the number of copies differs from it in the chromosome? Are these molecules isolated in the same way as the isolation of the chromosome in the case of cell division? (Weedall and Hall, 2011).

Mukherjee et al. (2008) find that these episomes comprise about 10-20% of the total cellular DNA (genome), approximately 15% of the readable genomic sequences among *E. histolytica* genome belong to these molecules that include 200 copies occupying about 25 kilobases. Lorenzi et al. (2010) suggests that the duplicated segments discovered along the genome represent some of these circular molecules.

In addition to these molecules, there are many less widespread molecules of different sizes (5.12 and 50 kb), that exact their function (Zakei, 2002), nor their genetic sequences known (Clark, et al., 2007).

Various circular rDNA molecules in *E. histolytica* have been described, while all previous studies have focused on molecules carrying rRNA genes, of which there are approximately 200 copies encoded for small and large Ribosomal units and 5.8S rRNAs, but not encoded for 5S rRNA nor protein. The size of these Molecules are 24.5 kilobases (Zakei, 2002).

When analyzing circular rDNA sequences in *E. histolytica* strains, two forms of arrangement are observed. In some strains, one rDNA unit is transcribed per cycle, while in other strains two units are arranged in opposite repeats. Also, these strains differ in the presence or absence of intergenic spacers (IGS). Where the upstream region is in two rDNA units strains are mismatched. These upstream sequences in the right direction of the single rDNA unit strains are present, while There are missing in the left direction (Bhattacharya and Bhattacharya, 2002).

## tRNA genes

The tRNA genes are organized as double and multiple arrays units that are separated from each other by rich repeats of thymine-adenine sequences. With evidence indicating their presence at the end of the chromosome and in the peripheral chromatin, giving them a function equivalent to the telomeres missing in *E. histolytica*. This suggests an effect of tRNA arrays in the structural organization of the nucleus (Irmer et al., 2010).

Most of the unique structural features that have been characterized in the *E. histolytica* genome are due to tRNA genes, as 10% of the readable sequences contain tRNA genes and these (with a few exceptions) are arranged in linear arrays (Clark et al., 2007).

The number of copies of the tRNA gene is about 4,500, which is ten times more than the human genome. They are arranged in repeating arrays that make up 10% of the genome. 25 distinct arrays containing repeating units that encode 1-5 types of receptors for tRNA, three arrays encode for 5S RNA, and one encodes for RNA which later snRNA (Tawari et al., 2008).

Clark et al. (2007) confirmed that it is not possible to accurately guess the size of the arrays due to the convergence of the arrays that have been identified. In general, there are 25 distinct arrays with a unit size ranging from 500-1750 base pairs. The regular arrays of tRNAs observed in some cases represent more than one repetitive unit read in one direction and other cases both readings are observed in both directions.

Intragenic regions among all the genetic arrays in *E. histolytica* contain complex structures of single tandem repeats (STRs). These repeats are variants in size (7-12 base pairs), although few reach more than 44 base pairs (Irmer et al., 2010). Some of these variances occur in the number of STR among the same array unit, but these variations are simple and do not appear when performing PCR for Inter-tRNA (Ali et al., 2008). This variation is meaningful. When comparing different strains, this can be used as a method for genotyping in this organism (Irmer et al., 2010).

## Gene families and their diversity

The efforts to know the complete genetic sequence of *E. histolytica* genome obtained amazing results, one of these is to distinguish some gene families, one of these large families encoding Transmembrane kinas (TMK) enzyme, which was previously found in higher Eukaryotic cells, the presence of this protein in the plasma membrane reflects Participation in extra-cellular signals, some members of this family affects phagocytosis (Clark et al., 2010). Series of Oligonucleotide encodes this enzyme that gives the highest expression rate in invading tissues and exposure to inappropriate conditions, that is, it is an indicator of response to the environment (Clark et al., 2010).

Another important gene family is a gene family encoding Cysteine protease (CP) with a total of 86 genes, as 50 genes encode to papain family, 22 genes encode to Metallo-protease (MP), 10 genes encode to Serine protease (SP), and 4 genes encode to Aspartic protease (AP), although this group was not expressed in vivo (Tillack et al. ., 2007).

SREHP protein, which is known as K2 membrane protein consists of the leading sequences at the amniotic end and an anchor basis of the hydrophobic sequences at the carboxylic end. The expression method of this group of genes vary according to the infection stage and the host type, so it is used to distinguish the isolated and growing strains (Zhiming and Samuelson, 1998).

Gal / GalNAc lectin protein (260 KD) is a heterodimeric glycoprotein with a bi-sulfur bridge linking its two subunits. It is located on the Trophozoites plasma membrane. Two gene families are coded, one for the heavy unit (KD 170), and the other for the light (KD31-35). This gene group performs many functions, including attachment to the host cell surface and pathogenesis of the parasite (Caron and Seve, 2004).

---

Lorenzi (2010) asserted that one of the large gene families encodes for the AIG1-like GTPases group. This gene family consists of 29 members distributed in three groups, 18 genes of which possess repetitive elements with no known function, but heterogeneity in the expression of these genes may be associated with virulence. Their products act as a heat shock protein.

Other gene families include a family that encodes for the LRR protein (Das and Ganguly, 2014), and a family that encodes for Rab GTPases (Weedall and Neil, 2011).

### **Transposable elements (TES)**

Transposable (jumping) elements are known as pieces of DNA that can be introduced into new locations of any chromosome, as well as they can produce duplicate copies of these elements. It is found in prokaryotic and eukaryotic cells, Barbara McClintock discovered it in eukaryotes in the corn plant in 1940. (Pray, 2008)

These elements play an important role in the biology of living organisms, as they can cause the mutation in the gene when inserts on it and can influence the genetic regulation if inserts in a place near the promoter in addition to being a basic material for the rearrangement of genes, which makes it an important indicator in the evolution of the genome (Griffith et al., 2000).

*E. histolytica*, like other organisms, have two types of these elements (TES), long and short (Dewannieux and Heidmann, 2005), three independent gene families are long tandem repeat (LTR) called *E. histolytica* long interspersed nuclear elements (EhLINE), the others, three gene families are short tandem repeat (STR), non-independent, and called *E. histolytica* short intersperses nuclear elements (EhSINE). In *E. dispar* that have the same number of these large families just. (Mandal et al., 2004).

LINEs have 4.8 kgb, while SINEs have 0.5 -0.7 kgb in size, which make up 11.2% of the genome in *E. histolytica* (Mandal et al., 2004). These elements are often found in intergenic regions, with thymine-rich sequences. Fifty base pairs are found upstream of insertion site for a limited number of EhSINE1, those occupy different genome sites in *E. histolytica* and *E. dispar*, and vice versa, which may indicate the relationship of these elements to the virulence of the parasite and its ability to infect (Kumari et al., 2011).

These transposable elements affect the gene expression of nearby or neighboring genes through various mechanisms, including alternative promotor processing, splicing, and others. SINEs are distinguished by being stable across generations, and attempts to insert them into the gene are rare, so genetic analyzes using SINEs are fairly accurate from RFLPs and Microsatellite sites (Kumari et al., 2013).

### **Conclusion:**

Each organism has its own distinctive characteristics that reflect the image of its existence. Genes and the genome are the most expressive images of the activity and functions of the organism. *E. histolytica* shows a variety in the structure and characteristics of the genome, in terms of the number of chromosomes and the chromosome set, the number of genes that packed into these chromosomes, the structural details of these The genes, and the method by it these genes are used to keep the parasite survive and enable it to invade different hosts.

### **References**

---

[1] IK. Ali, S. Solaymani-Mohammadi, J. Akhter, S. Roy and C. Gorrini, "Tissue Invasion by *Entamoeba histolytica*: Evidence of Genetic Selection and/or DNA Reorganization Events in Organ Tropism". PLoS. Neg. Trop. Dis., vol. 2(4), (2008),p.219.

- [2] A. Bagchi “Studies on structure and organisation of chromosomes in *Entamoeba histolytica*” .Ph.D. thesis, School of Environmental Sciences , Jawaharlal Nehru University .India,(2001), p 144.
- [3] A. Bhattacharya and P. J. Bhattacharya “Close sequence identity between Ribosomal DNA episomes of the nonpathogenic *Entamoeba dispar* and pathogenic *Entamoeba histolytica*”. J. Bio.sci., vol 2 (3), ( 2002), 619–627.
- [4] S. Bhattacharya and A. Bhattacharya “Amoebiasis and *Entamoeba* species: unexplored liaisons”. Trop. Ga., vol. 34(2), (2013),pp.55–57.
- [5] S. J. Black, and J. R. Seed ”The pathogenic enteric protozoa: *Giardia*, *Entamoeba*, *Cryptosporidium* and *Cyclospora*” . kluwer academic publishers J.. New York: (2004),P.184.
- [6] L. Brendan , I. Anderson, R. Davies, U.C. Alsmark, J. Samuelson, P. Amedeo, P. Roncaglia, M. Berriman, RP. Hirt, B.J. Mann, T. Nozaki, B. Suh, M. Pop, M. Duchene and J. Ackers “The genome of the protist parasite *Entamoeba histolytica*”. Natu., vol.43(3) (2005),pp.865-868.
- [7] M. Caron and A. Seve “Lectins and pathology” , Harwood academic publishers , British. (2004), p.289.
- [8] B. Chavez -Munguia , V. Tsutsumi and A. Martinez-Palomo “Research brief *Entamoeba histolytica*: Ultrastructure of the chromosome and the mitotic spindle” . Expe. Parasitol., vol. 114(2006), pp. 235–239.
- [9] C.G. Clark, p.j. Johnson and D. Adam “Anaerobic parasitic protozoa genomics and molecular biology”, Caister academic press, UK., (2010),p.562.
- [10] C.G. Clark, UC. Alsmark, M. Tazreiter, Y. Saito-Nakano, V. Ali and S. Marion “Structure and Content of the *Entamoeba histolytica* Genome” Advances in Parasitology, Vol. 65,(2007), p.451.
- [11] K. Das and S. Ganguly “Evolutionary genomics and population structure of *Entamoeba histolytica* ‘, Compu. and Stru. Biotechn. J., 12 (2014), pp. 26–33.
- [12] M. Dewannieux and T. Heidmann “LINEs, SINEs and processed pseudogenes: parasitic strategies for genome modeling”, Cytog. Gen. Res., vol.110(1-4),( 2005), pp. 35-48.
- [13] B. B. Dhaliwal and P. D. Juyal “Parasitic Zoonoses”, Springer, India,(2013), p.157.
- [14] S. Ghosh, M. Frisardi, L. Ramirez-Avila, S. Descoteaux, K. Sturm-Ramirez, A. Newton-Sanchez, JI. Santos-Preciado, C. Ganguly, A. Lohia, S. Reed, and J. Samuelson, “Molecular epidemiology of *Entamoeba* spp. evidence of a bottleneck (demographic sweep) and transcontinental spread of diploid” parasites . J. Clin. Microbiol., vol. 38( 10),(2000). pp. 3815–3821.
- [15] A. Griffiths, J. Miller and D. Suzuki “An Introduction to Genetic Analysis” 7th edition. New York., USA.,(2000), p.157.
- [16] H. Hooshyar, P. Rostam khani and M. Rezaeian “An annotated checklist of the Human and Animal *Entamoeba* ) Amoebida: Endamoebidae) species- a review article. Iran J. parasitol., vol. 10( 2),(2015),pp.146-156.

- [17] H. Irmer , I. Hennings, I. Bruchhaus and E. Tannich” tRNA Gene sequences are required for transcriptional silencing in *Entamoeba histolytica*. Euk. Cell., vol. 9( 2 ),(2010), pp. 306–314.
- [18] V. Kumari, L. R. Iyer, R. Roy, V. Bhargava, S. Panda, J. Paul and J. Verweij “Genomic distribution of SINEs in *Entamoeba histolytica* strains: implication for genotyping “, BMC. Genom., vol.14(1),(2013), p.432.
- [19] V. Kumari, R. Sharma, V.P. Yadav, K. Gupta, A. Bhattacharya and S. Bhattacharya “Differential distribution of a SINE element in the *Entamoeba histolytica* and *Entamoeba dispar* genomes: Role of the LINE-encoded endonuclease”, BMC. Genom., vol. 12(1),(2011), p.267.
- [20] C. López-Camarillo, N. Guillen, C. Weber, E. Orozco and L. Marchat “A genomics approaches in the understanding of *Entamoeba histolytica* virulence and gene expression regulation”, Afr. J. Bio., vol. 8 (8),(2009), pp.1363-1369.
- [21]C. López-Camarillo, C. Zamorano, H. E la Cruz, I. Rosas and L. Marchat “Genomics and proteomics approaches to understand virulence of *Entamoeba histolytica*”, A Méndez-Vilas (Ed). ( 2011), p.511.
- [22] H.A. Lorenzi, D. Puiu, J.R. Miller, L.M. Brinkac, P. Amedeo, N. Hall and E.V. Caler” New assembly reannotation and analysis of the *Entamoeba histolytica* genome reveal new genomic features and protein content information”, PLoS. Negl. Trop. Dis., vol. 4(6),(2010),p.716.
- [23] P. K. Mandal, A. Bagchi and S. Bhattacharya”An *Entamoeba histolytica* LINE/SINE pair inserts at common target sites cleaved by the restriction Enzyme-like LINE-encoded endonuclease”, Ame. Soc. Microbiol., vol. 3(1),(2004), pp.170–179.
- [24] L. Marianne”Molecular diagnosis and characterization of two Intestinal Protozoa: *Entamoeba histolytica* & *Giardia intestinalis*”, Karolinska Institutet, Stockholm, (2010),pp.76.
- [25] C. Mukherjee, G. C. Clark and A. Lohia “*Entamoeba* shows reversible variation in ploidy under different growth conditions and between life cycle phases”, PLoS. Negl. Trop. Dis., vol. 2(8),(2008), p. 28.
- [26] E. Pinilla, M. Consuelo and D. Fernando(2008)” History of the *Entamoeba histolytica* protozoan”, Rev. Méd. Chile., vol. 136(1), (2008), pp. 118-124.
- [27] L. Pray”Transposons, or jumping genes: Not junk DNA?” Nat. Edu., vol. 1(1),(2008), p.32.
- [28] F.L. Schuster and G.S. Visvesvara “ Amebae and ciliated protozoa as causal agents of waterborne Zoonotic disease”, Vet. Parasitol., vol. 126 (1-2),(2004),pp.91–120.
- [29] S. Srivastava “Identification of strains of *Entamoeba histolytica* and *Entamoeba dispar* from natural isolates”, Ph.D. thesis, school of life science, Jawaharlal Nehru University. India, (2005), p.107.
- [30] M. Tanyukse and W.A. Petri ” Laboratory diagnosis of Amebiasis” . Clin. Microbial. Rev.,vol. 16( 4),(2003),pp. 713–729.



- [31] B. Tawari, M. Ali, C. Scott, A. Michael, M.B. Quail, N. Hall and G.C. Clark "Patterns of Evolution in the unique tRNA Gene arrays of the genus *Entamoeba*", Mol. Biol. Evol., vol. 25(1), (2008),pp.187–198.
- [32] M. Tillack, L.Biller, H. Irmer, M. Freitas, M.A. Gomes, E. Tannich and I. Bruchhaus "The *Entamoeba histolytica* genome: primary structure and expression of proteolytic enzymes", BMC. Genom., vol.8(1), (2007),p.170 .
- [33] G. D. Weedall and N. Hall "Evolutionary genomics of *Entamoeba*", Res. in Microbiol., vol. 162(2-6), (2011),pp. 637-645.
- [34] M. Zaki "Characterisation of polymorphic DNA and its application to typing of *Entamoeba histolytica* and *Entamoeba dispar*", Ph.D. thesis , University of London,(2002), p.278.
- [35] M . Zhiming and J. Samuelson "A new gene family (ariel) encodes Asparagine-rich *Entamoeba histolytica* antigens, which resemble the Amebic vaccine candidate Serine-rich *E. histolytica* Protein", Inf. and Immu., vol 66( 1) (1998),pp.353–355.