Research Article

Bioinformatics structural and phylogenetic characterization of *Entamoeba histolytica* alcohol dehydrogenase 2 (EhADH2)

Katie M. Lowerre[1], Avelina Espinosa[1], Guillermo Paz-y-Miño-C[2], and Christopher Hemme[3]

[1]*Department of Biology, Roger Williams University, Bristol, Rhode Island, 02809;*

[2]*New England Center for the Public Understanding of Science, Roger Williams University, Bristol, Rhode Island, 02809;*

[3]*RI-INBRE Bioinformatics Core, Department of Biomedical and Pharmaceutical Sciences, University of Rhode Island, Kingston, Rhode Island, 02881*

Abstract. An amitochondriate parasite, *Entamoeba histolytica*, has a bifunctional ADHE enzyme (EhADH2) that contains separate acetaldehyde (ALDH) and alcohol (ADH) dehydrogenase activities. In a cluster of 25 bifunctional enzymes of single cell eukaryotes and bacteria, we present a phylogenetic analysis that suggests a lateral gene transfer event (prokaryotic ancestor to single-cell eukaryotic ancestor) and a complex structure that aligns with key homologs in the ADHE evolutionary history based on their similarity with bacterial alcohol dehydrogenases. We show that the ADHE in *Entamoeba* lineage diverged independently but shows significant similarities to the structure of ADHE in *Fusobacterium*, and a complex model that maps its ALDH and ADH domain well with bacteria such as *Geobaccillus thermoglucosidasius*. Our analyses likely support a lateral acquisition of an EhADH2-like ancestral gene from bacteria. Analysis using several evolutionary analyses software programs reveal that the enzyme structure is highly conserved, and maintains a similar function within a diverse set of pathogens, including *Escherichia coli* and *Clostridium spp*.

Introduction

Entamoeba histolytica is an obligate anaerobic parasite responsible for amebiasis. Amebiasis causes approximately 100,000 human deaths per year (Espinosa et al., 2001; 2004; Parsonage et al., 2014) and is estimated to be a major cause of death in the world, second among parasitic induced mortalities worldwide (St-Pierre et al., 2014). The disease continues to be one of the greatest major public health issues in developing countries mostly due to improper water sanitation, sewage disposition problems, and ingestion of infested food (Nowak et al., 2015; Saha et al., 2015). Ingestion of the parasite can result in passage of the cysts into the small intestine, where excystation processes occur and can result in the development of mature tropozoites that lyse cell tissues (Nowak et al., 2015). Additionally, *E. histolytica* can cause cognitive problems and brain abscess, affecting the central nervous system (Finsterer and Auer, 2013), and ingestion of *E. histolytica* usually results in loose or bloody stools inducing major diarrhea. Diarrhea

Correspondence to: klowerre1@gmail.com

Bios 90(1) 30–41, 2019

is a major contributor to childhood mortality and morbidity, causing an estimated 2.5 million deaths each year.

E. histolytica can survive within the host organism by growing within suitable pH levels of the human gastrointestinal tract. Within the tract, *E. histolytica* ferments and grows using a bifunctional alcohol dehydrogenase, EhADH2, to break down the host's glucose supply by the *E. histolytica's* own glycolytic pathway. EhADH2's ADHE enzyme is responsible for converting acetyl CoA to ethanol in a two-step reaction process and is essential for its growth (Espinosa et al., 2001; 2004; Yang et al., 1994). In single cell eukaryotes, oxidative phosphorylation enzymes that reduce NADH in the Krebs cycle are absent, and these key fermentation enzymes provide central roles in anaerobic unicellular eukaryotes that resemble proteins of anaerobic prokaryotes (Girbal et al., 1995; Reeves, 1984). Since some of these enzymes are essential to the organism and lack common homologs in more complex eukaryotes, they might be prime antimicrobial targets (Chen et al., 2003; Espinosa et al., 2001; 2004).

Fermentation enzymes such as *E. histolytica*'s EhADH2 belong to the ADHE iron-dependent family of alcohol dehydrogenase enzymes (Espinosa et al., 2001; Paz-y-Miño-C and Espinosa, 2010). These enzymes share conserved amino acid residues with other ADHE enzymes found in prokaryotes like *Firmucutes*, *Proteobacteria,* and cyanobacteria. EhADH2's 870-amino acid length includes a conserved linker sequence GGGS from residues 543 to 546, shared among protists and bacteria analyzed in the present study. Matching functional regions relate to moieties in the ADH domain. These include an EKLSP motif, a PHG motif, and a conserved GXGXXG motif at positions 431-436, a putative NAD binding domain (Genbank Accession number CAA54388; Espinosa et al., 2001; Paz-y-Miño-C and Espinosa, 2010). *E. histolytica* has shown metabolic adaptations shared with amitochondrial protist pathogens such as *Giardia lamblia* (Nixon et al., 2002) and *Cryptosporidium parvum* (Genbank Accession number XM_627078.1).

Analysis between diplomonad, fungal, and other alcohol dehydrogenases indicate a weaker cluster, suggesting that lateral gene transfer events occurred from prokaryotic to single-cell eukaryotic ancestors, and eukaryotic to eukaryotic ancestors (Andersson et al., 2006). These homologies infer that lateral acquisition played an important role in alcohol dehydrogenase evolution (Andersson et al., 2006; Paz-y-Miño-C and Espinosa, 2010). Prior analysis suggests a lateral transfer of bacterial genes into the *Entamoeba lineage*, with modifications and adaptations necessary in their metabolic repertoire (Nixon et al., 2002; Paz-y-Miño-C and Espinosa, 2010). Based on sequence similarities, previous maximum likelihood parsimony analysis suggests that the entire bifunctional enzyme related to other bacterial lineages and to the amitochondriate protozoan parasite, *Trichomonas vaginalis* (Rosenthal et al., 1997). The genes encoding for these proteins were likely laterally transferred from ancestral bacteria associated with *T. vaginalis* (Rosenthal et al., 1997). We have hypothesized using sequence alignments that each domain that had been acquired laterally fused to form a single gene expressing a bifunctional enzyme in the *Entamoeba* lineage. Selective pressures likely generated biochemically adapted enzymes in diverse environments and metabolic pathways (Paz-y-Miño-C and Espinosa, 2010). This hypothesis is supported by the analyses of EhADH2's two independent domains; there are fewer similarities within the ALDH domain; however, the enzyme's ADH domain may have similarity with the *Zymomonas mobilis* ADH (ZmADH4). It is known that the iron-containing ADH portion of *Z. mobilis* aligns with the ADH in *Saccharomyces cerevisiae* (Conway and Ingram, 1989; Williamson and Paquin, 1987). *Z. mobilis* and the ADH from *S. cerevisiae* belong to a unique class of alcohol dehydrogenases, which may include the ADH from *E. histolytica* (Williamson and Paquin, 1987). An ancestral prokaryotic origin through horizontal gene transfer also supports a potential drug target due to the lack of human homologs (Espinosa et al., 2001; 2004; 2009).

A crystal structure for the ADHE enzyme

family, including EhADH2, has not been elucidated. In this work, maximum likelihood methods were used to determine the phylogenetic similarities between EhADH2 and other respective bifunctional ADHE enzymes. In using primary sequences, phylogenetic analysis can indicate separate evolution patterns of ancestral eukaryotic bifunctional enzymes as well as understanding the evolution of the enzyme over larger periods of time. Modeling the enzyme can further reveal potential origins of each separate domain, given that the enzymes appeared to have fused before the acquisition by an early eukaryotic ancestor. In this study, phylogenetic analysis helped answer evolutionary history questions, homology modeling was used to predict structural similarity, and molecular evolutionary analysis suggested the rate of evolution between these enzymes.

Materials and Methods

Phylogeny

All primary protein sequences were found under the Uniprot database (Chen et al., 2017) and Genbank databases (Benson et al., 2013). Data were compiled into the Molecular Evolutionary Genetic Analysis (MEGA) tool to check for clear compilation of sequence data into respective phylogenies (Kumar et al., 2015). Phylogenies for these sequences were constructed in RAxML (Randomized Axerlated Maximum Likelihood), which formulated maximum likelihood values and SH-like values, and bootstrapping was an option utilized in the program (Stamatakis, 2014).

Sequence files were used as input in the RAxML program with a smaller subset of 25 bifunctional enzymes that generated an ML tree. This tree could be expressed in Newick format and then could be translated to MEGA for sequence analysis. Comparisons between the trees constructed in MEGA and those compiled by the RAxML program could be formulated. However, only RAxML was used for this data set. Tree algorithm files from RAxML were downloaded and exported to MEGA to preview a final output for each maximum likelihood phylogeny result.

Homology modelling and protein structure analysis

Homology modelling was accomplished using three programs plus an additional program to check the validity of the model. Chimera, Discovery Studio, and ZDOCK were used for homology modelling while SWISS-MODEL was used as a comparison for the protein (Biasini et al., 2014). UCSF Chimera is a free program that provides essential services and visualizations that include formats such as those found in protein database files (PDB; Berman et al., 2003; Pettersen et al., 2004). The program also includes extensions that include Multiscale, adding the ability to visualize larger scale molecular assemblies, including those that encapsulate high kilodalton structures (Pettersen et al., 2004). Multalign Viewer was used in homology modelling and can display multiple sequence alignments and associated structures, along with ViewDock, which can visualize docking ligand orientations (Pettersen et al., 2004). Chimera was used principally for structure visualization, as well as mapping other protein primary sequences to the overall protein structure of related homologs in its homology modeling inclusion. The interface to Chimera includes Modeller, which was utilized to model each domain of the protein.

Like Chimera, Discovery Studio and ZDOCK were used to visualize the protein. Discovery Studio was a program used on trial to create PDB files which could be inputted into the Chimera visualization system (Dassault Systèmes BIOVIA, 2016). ZDOCK, a site principally used for antibody docking, was utilized to form two similar clusters of the ALDH domain and the ADH domain into a compiled EhADH2 enzyme that could be traced in Chimera (Pierce et al., 2014). ZDOCK took the linker sequence of EhADH2 and transferred each domain into a separate PDB file which was visualized in Discovery Studio and transferred directly into Chimera for homology modelling. An exact model was not formulated; however, a structure almost aligned with the proper orientation was constructed.

SWISS-MODEL was an interactive homology modelling generator which took the prima-

ry sequence of the protein and mapped variable regions (Biasini et al., 2014). The protein was compared to the model created to the final construct created using Chimera, Discovery Studio and ZDOCK. The primary role of SWISS-MODEL was to generate two separate domains to assess composite QMEAN and Global Model Quality Estimation (GMQE) scores. QMEAN scores were local and global estimates of the quality of the model (Benkert et al., 2009). These include torsion angle potential, C-beta interactions, solvation potential, and pseudo energy potential (Benkert et al., 2009). QMEAN utilized a Z-score scheme, where lower deviations denoted relatedness to a high resolution like X-ray structure. GMQE scores were assessed on reliability and accuracy of the model via a target-templated based algorithm. These numbers ranged between 0 and 1, where higher numbers denoted higher accuracy.

Evolutionary analysis

Evolutionary analysis was carried out by PAML (Phylogenetic Analysis by Maximum Likelihood; Yang, 2007). The program was used to search for omega (ω) values that described the ratio of nonsynonymous/synonymous substitution rates, as well as kappa (κ) values that describe transversion vs. transition ratios. The ω variation or ω ratio is a direct measure of natural selection acting upon the protein (Yang, 2007). An $\omega < 1$ suggests negative selection, an $\omega = 1$ indicates neutral selection, and an $\omega > 1$ indicates positive selection.

Different tests allowed for different fixations of the ω ratio (Yang, 2007). Branch models, site models, and branch site models were tested. Branch site models allow for a variation of ω using the variation by modifying the data file. The separation of the enzyme according to different branch lengths is outlined by specific branches to have different ωs, leading to what was used as a "two-ratios" model. These models required for branches in the tree to be labelled according to what ω was treated first based on branch length. These labels were specified with the "#" symbol. *E. histolytica* was labelled with a #1 label which suggests the internal branch with just *E. histolytica* has the ω1 ratio while the other branches have the default #0 label (Yang, 2007). The *Entamoeba* lineage was marked with a "$" symbol, separating the clade. The following shows the document tree from the file used to create the outputs:

```
(((Entamoebahistolytica:0.01285969#1,
Entamoebadispar:0.01949044):0.0583487
9,Entamoebamoshkovskii:0.08161586):0.
05049119,Entamoebainvadens:0.1748451
7$1,(Ecoli:0.36850382,(Clostridiumace
tobutylicum:0.18198586,Clostridium
carboxidivorans:0.16655217):0.05892
072):0.25744525);
```

Site models were used to test specific sites that allow ω to change along these sites. The variable NSsites was utilized to constructs variables for NSsites, which included NSsites = 0 2 7 and were different tests utilized to test for variable sites among the entire protein (M0 and M2 in Table 1), along with specific sites (M7 in Table 1). Positive selection tests could be checked under a chi square distribution that compares a variation (M2) with the null hypothesis (M0). Clade tests were done with branch-site models, which allowed for the ω ratio to vary among sites across the protein and across branches on the tree. This allowed for specific sites to be tested within the protein. This test was performed using model = 2, NSsites = 3, which displayed different models according to the branch. In the tests done in these models, the *Entamoeba spp*. lineage was separate from the other species in a codon model that included seven species across the 25 bifunctional enzymes, which included three other *Entamoeba* sequences, *Escherichia coli*, and two sequences that fill under the *Clostridium spp*. lineage (See separation above).

Results

Phylogeny

Phylogenetic analysis showed similar topologies between the bifunctional tree, the ALDH

Table 1. Results including kappa, omega (dN/dS) values, and parameters for each test. General parameters are included for the one ratio, neutral (M0) test, selection (M2), and M(7), with an exclusion of parameters for M(3). Parameters for the first two tests (M0 and M2) indicate the comparison between the null parameter (ω) and the four different parameters that determine positive selection. Parameters for the beta test in M7 were relatively comparable. The positive selected sites indicating "None" means that there are very few sites that are selected for. "Virtually" none indicates that hardly any are selected. The overall result from the branch tests (M0, M2), site tests (M7), and branch site test (M3) indicated no positive selection across the ADH domain, the ALDH domain, and the entire enzyme. Tests for individual sites and parameter information for positive selection are included at the bottom of the table.

Model	Kappa (ts/tv)	dN/dS	Parameters	Positively selected sites
M0	1.25	0.08	ω	None
M2	1.28	N/A	p0, p1 (p2 = 1 – p0 – p1) ω0 < 1, ω1 = 1, ω2 > 1	None
[a]M3	1.31	N/A	N/A	None
M7	1.28	N/A	[b]p= 0.85 q= 8.16	None
ALDH				
M0	1.23	0.06	ω	None
M2	1.21	N/A	p0, p1 (p2 = 1 – p0 – p1) ω0 < 1, ω1 = 1, ω2 > 1	None
[a]M3	1.22	N/A	N/A	None
M7	1.17	N/A	[b]p= 0.63 q=7.17	None
ADH				
M0	1.28	0.08	ω	None
M2	1.30	N/A	p0, p1 (p2 = 1 – p0 – p1) ω0 < 1, ω1 = 1, ω2 > 1	None
[a]M3	1.33	N/A	N/A	None
M7	1.31	N/A	[b]p= 0.88 q= 8.47	None
Iron binding site				
M0	0.40	0.01	ω	Virtually none
NAD binding site				
M0	0.56	0.03	ω	Virtually none

[a]Denotes model = 2, NSsites = 3, which is a branch-site model
[b]Denotes specific parameters for the selected beta test

tree, and the ADH tree (Fig. 1), with two major clusters identified. The first cluster included *Entamoeba spp*., *Fusobacterium spp*., *Geobacillus spp*., and *Staphylococcus spp*. group, along with the *Proteobacteria* and *Clostridium spp*. group (Fig. 1). Bootstrap values are relatively high on each tree, indicating high accuracy of each tree.

Despite great similarities, each tree was not exactly the same. Although a distinct lineage for *Entamoeba spp*. was identified in the bifunctional tree (Fig. 1A), the overall assemblage was slightly different within each tree, including placing *E. histolytica* along with *Entamoeba moshkovskii* in the ALDH and ADH segments of the enzyme (Fig. 1B and C). There were also individual differences between the accuracy of clustering between *Staphyloccocus spp*. and *Geobacillus spp*. given by different bootstrap values in these trees (Fig. 1). The highest cluster was shown by a bootstrap value of 76 compared to that of 72 in the bifunctional tree and 56 in the ALDH portion (Fig. 1).

Homology modeling

In individual folding, it was shown that EhADH2 has a similar folding and structure to the acyl CoA domain from an alcohol dehydrogenase tetramer in *Vibrio parahaemolyticus* with around 40% similarity (PDB: 3MY7.1_A). Both enzymes share residues involved in the oxidoreductase pathway, requiring NADH as a cofactor. The two enzymes

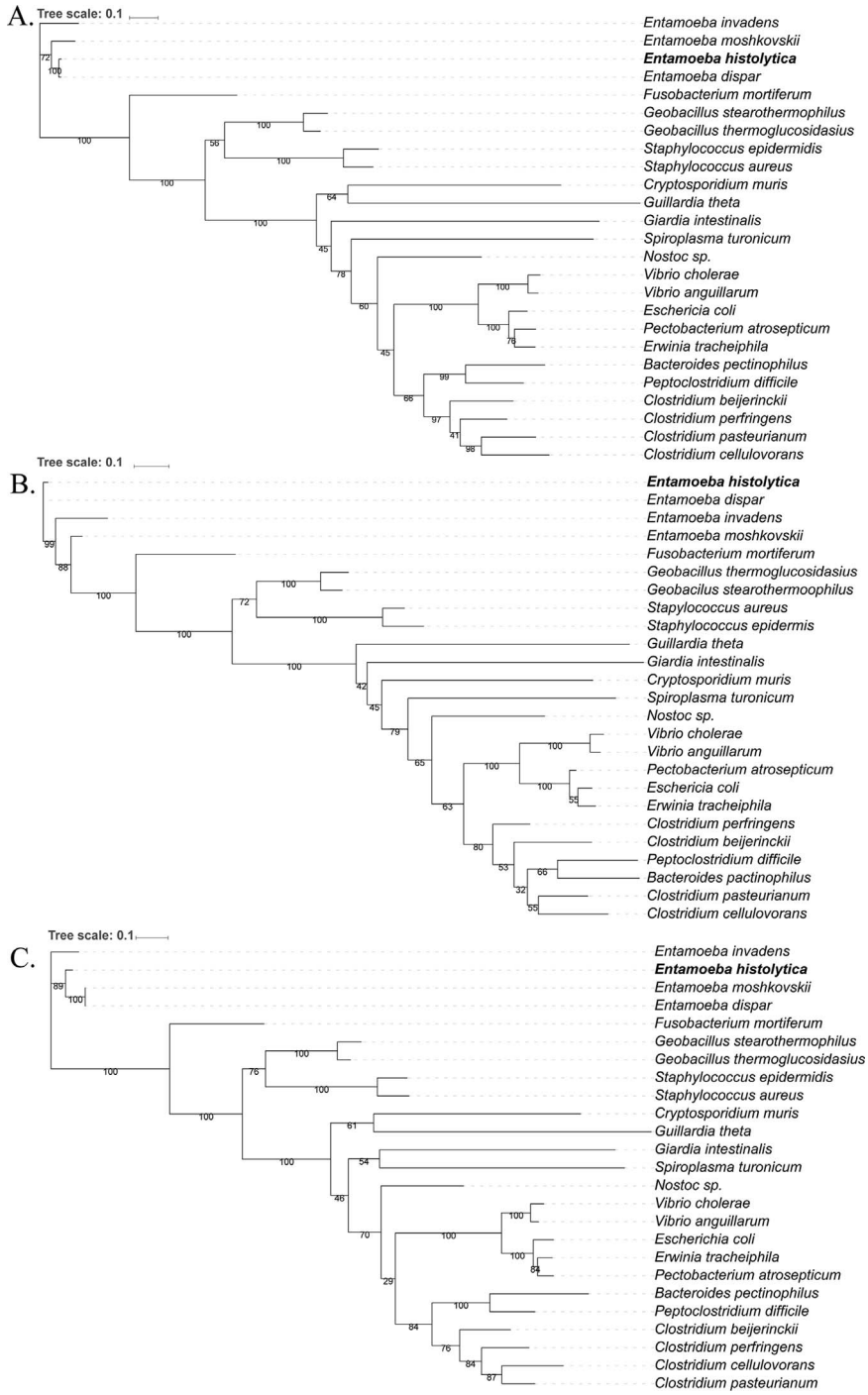

Figure 1. Maximum likelihood trees of a fixed subset of ADHE enzymes using RAx ML (Stamatakis, 2014). The bootstrap value of the branch separating each of the two major clusterings within each tree is 100, which represents both a high accuracy and a clear distinction between groups. A. ALDH domain of the ADHE enzyme modelled for each of the 25 species. B. ADH domain of the ADHE enzyme modelled for each of the 25 species. C. The bifunctional enzyme modelled for each species.

share the catalytic CASEQ motif present in EhADH2's ALDH domain.

For the ADH domain, the EhADH had a similar cofactor binding configuration compared ω to ADH4 in *Z. mobilis* (Fig. 2). The binding pockets, which include the linker sequence, contribute to three alpha helices and are relatively conserved in both sequences. In ZmADH4, the NAD^+ binds to these alpha helices similar to that of EhADH2. A cysteine residue at a binding site of ZmADH4 is absent in EhADH2. Comparisons for an alignment with EhADH2 and *Z. mobilis* ADH4 (zmADH4), which corresponds closely to EhADH2's ADH domain, indicated that almost every residue that is associated with NAD cofactor binding in zmADH4 is in EhADH2 (Fig. 2).

This confirms that binding sites and some structural folding were similar to that of ZmADH4 (Moon et al., 2011), along with the conserved EKLSP motif and a PHG motif for iron binding and the conserved GXGXXG motif for NAD binding. EhADH2 had greater similarity, over 50%, to that of the *G. thermoglucosidasius* ADH replicate (Genbank accession number 550544862; PDB: 3ZDR.1), which indicated 59% similarity overall compared to 37.5% in *Z. mobilis*'s ADH4.

Based on each domain homology model, a final construct (Complex 1; Fig. 3) was created to model EhADH2 according to overall conservation and out of ten ZDOCK outputs (Fig. 3). This model shows that EhADH2 is most likely a dimer, principally based on its properties and inclusion of single monomers in its ALDH domain and its ADH domain. A map of a single monomer unit from the acyl CoA domain from *V. parahaemolyticus* along with the formation of the ADH portion from *G. thermoglucosidasius* with various conserved regions can form the dimer that constructs the EhADH2 enzyme (Fig. 3)

Comparisons of the model (Fig. 3) were utilized with SWISS-MODEL. While displaying a similar format, the model showed relatedness to separate enzymes in *G. thermoglucosidasius*, which included the bifunctional enzyme found in this organism. The first domain portion in the acetylating acetaldehyde dehydrogenase homo-tetramer (PDB: 5J78_A) in *G. thermoglucosidasius* showed 46% similarity with the ALDH domain in EhADH2 and composite QMEAN deviation that was less than 1, which was at -0.77 with a GMQE score of 0.65. The similarity was around 61% for the alcohol domain in the bifunctional enzyme of the bacteria compared to EhADH2. The QMEAN deviation for the ADH domain indicated was also higher at -1.36, while the GMQE score was 0.79. In performing an overall enzyme check using SWISS-MODEL, some similarity for the acyl CoA domain from *V. parahaemolyticus* was indicated containing around 47% similarity with a QMEAN deviation of -1.14 and a GMQE score of 0.41.

Evolutionary analysis

A total of three tests were used (M0, M2, and M7) along with a clade specific test (model= 2, M3) were used in testing the variation of ω among these enzymes (Table 1). Each test accounted for the entire bifunctional enzyme, the ALDH domain, the ADH domain, and each highly conserved area (iron binding domain and NADH binding site). The transversion vs. transition value (kappa value) were relatively comparable and were especially low for the iron binding and NADH binding sites. The non-synonymous/synonymous substitution rates (dN/dS) or ω values for the bifunctional, ALDH, and ADH domains were all similar and below 1. This indicates negative selection. The iron binding site and the NADH binding sites were even lower, and closer to 0. Therefore, each primary test (M0, M2, and M7) showed little to no variation of ω based on the ω ratios that are below a statistical value of 1.

Discussion

Most crystal structure studies focus on unifunctional enzymes, avoiding the difficulties intrinsic to the analyses of bifunctional enzymes. In some of these studies, evolution of the ADHE lineage has been included, and it has

Figure 2. Alignment of the ADH domain in *E. histolytica*'s EhADH2 with *Z. mobilis*'s ADH4 from Multiple sequence alignment (MUSCLE; Edgar, 2004). Most residues are conserved, relatively similar, or somewhat similar by property, as indicated with a star (*), a (:), or a (.) symbol respectively. Residues that are associated with iron binding and associated with NAD cofactor binding are in larger size and italics respectively. Amino acids underlined represent three essential alpha helices for the NAD binding pocket.

shown speculation for lateral transfer, but there has been no direct evidence supporting a common ancestor or a fusion event. Similar to previous data, the phylogeny substantiates a common ancestor, but most likely favors a predated fusion event based on the data given (Andersson et al., 2006; Paz-y-Miño-C and Espinosa, 2010). Since the enzyme's *Entamoeba* lineage is closer to bacterium rather than the parasite *Giardia intestinalis*, this suggests that lateral acquisition played a part in the development of the enzyme. The *Entamoeba* lineage is also a separate lineage from other parasites, suggesting further that a fusion event must have predated the enzyme before its "full evolution." Further study would indicate whether a common ancestor predates the enzyme, since the information on this is somewhat unclear based on the results presented in Figure 1.

Phylogenetic analysis (Fig. 1) has shown distinct differences between the *Entamoeba spp*. lineage and traits associated with the *Giardia* lineage; this analysis suggests that gene acquisition occurred independently by both *Entamoeba* and *Giardia* ancestors. The phylogenetic distance between *Vibrio spp*. and *Geobacillus spp*. in respect to EhADH2 based on homology modeling found relative similarities, implying that ancestral bacterial genes gave origin to the bifunctional enzyme in the three lineages and residues in the three enzymes are conserved as seen in the phylogenetic tree. This further confirms previous results of EhADH2 having 48% similarity with the more distantly related ADHE homolog from *E. coli*, in addition to lesser similarities between the ALDH domain compared to the *V. parahaemolyticus* enzyme and higher similarity between the ADH domain and *G. thermoglucosidasius*. Furthermore, since there was a consistent result between EhADH2 being phylogenetically closer to the enzyme found in *Fusobacterium mortiferum,* an ancestral bacterium to *F. mortiferum* may have transferred its gene to an *Entamoeba* spp

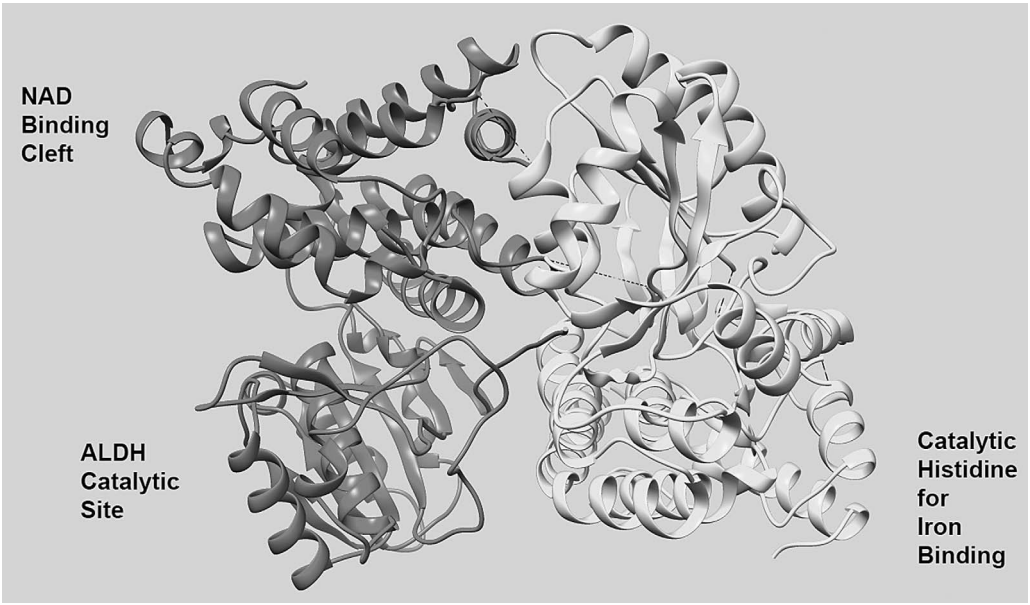

Figure 3. ZDOCK model construct of Complex 1 for EhADH2 with labelled cofactors, substrate, linker and domains sites (model from Chimera). The domain in red is the ALDH domain, while the green domain is the ADH domain. Modelling this final construct did not include the proper orientation but was the closest out of the ten constructs created.

ancestor and probably possessed similar metabolic pathways. This finding hints to a greater evolutionary connection between ancient bacterial and *Entamoeba spp* genes based on how each enzyme behaves in each lineage. The lack of clustering between various fungal and bacterial alcohol dehydrogenases also strongly supports distinctive evolutionary eukaryotic gene transfer events. The divergence in clustering among ADHE enzymes corroborates the presence of alternative metabolisms with overlapping ADH enzymes (Andersson et al., 2006).

Despite relatively low similarities, the EhADH2 enzyme modeled well with the conserved residues and acyl CoA domain of the *V. parahaemolyticus*, which represents a similar conserved pathway in both enzymes for the acetaldehyde dehydrogenase activity (ALDH) domain. This is also supported by high overall conservation and residues that are observed between the acetaldehyde dehydrogenase in *G. thermoglucosidasius* and the ALDH domain in EhADH2 using a simple MUSCLE alignment (Edgar, 2004; data not shown). This portion must convert acetyl CoA into acetaldehyde using NADH as a cofactor. Our results suggest that this pathway is relatively conserved, along with the mechanism of action; however, the enzymes across the phylogeny in Figure 1 may have additional enzymes that contribute to its metabolism. Although antimicrobial compounds targeting these enzymes may not be effective in preventing growth and survival due to the several ADH alternatives in some of the ADHE hosts, previous studies showed that blocking the expression of EhADH2 in *E. histolytica* affects trophozoite growth and survival (Espinosa et al., 2001). It is at least likely that ADHE could be important and/or essential for the persistence of some organisms within the *Entamoeba* spp. phylogenetic clade (Espinosa et al. 2001; 2004; 2009; Paz-y-Miño-C and Espinosa, 2010).

The modeling of EhADH2's ADH domain in Chimera closely corresponds with the results indicated with *V. parahaemolyticus*. Homology modeling using Chimera indicated close correlation the ADH domain closest to *G. thermoglucosidasius*, which again signifies similarities between *E. histolytica*'s ADHE enzyme and other bacterial ADHE enzymes. This also well defines the conservation of the EKLSP and the

GXGXXG motifs which signify that the NADH binding and iron binding respectively are similar, given the additional similarities between the bifunctional enzyme in *E. histolytica* compared to the one in *E. coli* in this particular ADHE class of enzymes.

However, while the results from SWISS-MODEL showed some overall similarity between the Acetyl CoA domain in *V. parahaemolyticus* and EhADH2, the SWISS-MODEL implies closer similarity with separate enzymes such as the acetylating acetaldehyde dehydrogenase in *G. thermoglucosidasius*. Despite having about the same similarity scores, there was a lower deviation (-0.77 compared to -1.14) and a higher GMQE (0.65 compared to 0.41) for the acetaldehyde dehydrogenase in *G. thermoglucosidasius* compared to the Acetyl CoA domain in *V. parahaemolyticus*. Phylogenetic analysis (Fig 1. A.) indicated a closer overall cluster that included *Geobacillus spp*. and *Entamoeba spp*. where *Vibrio spp*. was in another distinct lineage. The acetaldehyde dehydrogenase in *G. thermoglucosidasius* included the CASEQ motif, however, the invariable cysteine has been identified as unknown, but both enzymes include the EKLSP motif. Deviations for the alcohol domain were higher, signifying more minor differences in loop regions compared to that of the acetaldehyde domain. The high GMQE scores (over 0.50), especially for the alcohol domain (GMQE score: 0.79), for the separate enzymes in *G. thermoglucosidasius* indicate that both separate terminus models display a solid match for each domain in EhADH2.

Since the evolutionary analysis showed little variation of ω across the enzyme sequence, as well as for each separate domain (Table 1), we presume that any sequence and/or structural changes of the EhADH2 after the prokaryote-to-eukaryote lateral gene transfer were small. This also means that the EhADH2 enzyme did not significantly change after the lateral acquisition of the enzyme. Because the omega values for specific binding were much lower than 1, this confirms high homology between the conserved sites in EhADH2 to other homologs such as those found in *V. parahaemolyticus* and *G. thermoglucosidasius*, especially with the NADH and iron binding sites. Therefore, mechanism of action is conserved in this class of ADHE enzymes. The slight changes in the ω value may be explained by an enzyme function under new environments (=hosts) that would require slight adaptations. We suggest that the slight changes observed could have occurred at the metabolic proficiency level. Kinetic analyses of ADHE enzymes confirm this inference (Espinosa, unpublished data).

If the enzyme and its major components remain essential to an organism, theoretically, these enzymes can be inhibited by similar compounds. Specific carbinols, pyrazoline inhibitors, and iron chelators designed to block activities of ADHE have been shown to be effective on trophozoite cells *in vitro* (Espinosa et al., 2004; 2009; 2012). The current drug used to eradicate these pathogens, metronidazole, can be absorbed through multiple routes and is efficient in the treatment of several diseases, as it is used in patients with amebiasis and infections with *Clostridium spp*. The common side effects for the drug include severe headaches, vomiting, and fatigue. Pyrazoline inhibitors used to manage amebiasis could potentially be used to inhibit other anaerobic pathogens, such as those containing ADHE enzyme homologs (Fig. 1) and have little to no side effects. Further research would relate to the enzymes full relation to other enzymes in its class.

Acknowledgments: Espinosa and C. Hemme are supported by NIH-NIGMS grant 2P20GM103430. K. Lowerre is supported by Roger Williams University Senior Thesis and Honors programs. This research is based in part upon work conducted using the Rhode Island IDeA Network for Excellence in Biomedical Research Bioinformatics Core which is funded by the National Institutes of Health under grant 2P20GM103430.

References

Andersson, J.O., Hirt, R.P., Foster, P.G., and Roger, A.J. (2006). Evolution of four gene families with patchy

phylogenetic distributions: influx of genes into protist genomes. BMC evolutionary biology 6(27), 1–18.

Benkert, P., Kunzli, M., and Schwede, T. (2009). QMEAN server for protein model quality estimation. Nucleic Acids Res. 37, W510–514.

Benson, D.A., Cavanaugh, M., Clark, K. Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. Genbank. (2013). Nucleic Acids Res, 41(Database issue):D36–42.

Berman, H.M., Henrick, K., and Nakamura, H. (2003) Announcing the worldwide Protein Data Bank. Nature Structural Biology 10(12), 980.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., and Schwede, T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information Nucleic Acids Res. 2014 (1 July 2014) 42 (W1), W252–W258.

Chen, D., Hackbarth, C., Ni, Z.J., Wu, C., Wang, W., Jain, R., He, Y., Bracken, K., Weidmann, B., Patel, D.V., Trias, J., White, R.J., and Yuan, Z. (2003). Peptide deformylase inhibitors as antibacterial agents: Identification of VRC3375, a proline-3-alkylsuccinyl hydroxamate derivative, by using an integrated combinatorial and medicinal chemistry approach. Antimicrobial agents and chemotherapy. 48, 250–261.

Chen, C. Huang, H., and Wu, C.H. Protein bioinformatics databases and resources. (2017). Methods Mol. Biol. 1558, 3–39.

Conway, T., and Ingram, L.O. (1989). Similarity of *Escherichia coli* propanediol oxidoreductase (fucO product) and an unusual alcohol dehydrogenase from *Zymomonas mobilis* and *Saccharomyces cerevisiae*. J. Bacteriol. 171(7), 3754–3759.

Dassault Systèmes BIOVIA, Discovery Studio Visualizer, San Diego: Dassault Systèmes, 2016.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5), 1792–1797.

Espinosa A., Socha A.M., Ryke E., and Rowley D.C. 2012. Antiamoebic properties of the actinomycete metabolites echinomycin and tirandamycin. Parasitol. Res. 111(6), 2473–2477.

Espinosa, A., Perdrizet, G., Paz-y-Miño-C, G., Lanfranchi, R., and Phay, M. (2009). Effects of iron depletion on *Entamoeba histolytica* alcohol dehydrogenase 2 (EhADH2 and trophozoite growth: implications for antiamoebic therapy. J. Antimicrob. Chemother. 63(4), 675–678.

Espinosa A., Clark, D.P, and Stanley, S.L. Jr. (2004). *Entamoeba histolytica* alcohol dehydrogenase 2 (EhADH2) as a target for anti-amoebic agents. J. Antimicrob. Chemother. 54(1), 56–59.

Espinosa A., Yan, L., Zhang, Z., Foster, L., Clark, D.P., Li, E., and Stanley, S.L. Jr. (2001). The bifunctional *Entamoeba histolytica* alcohol dehydrogenase 2 (EhADH2) protein is necessary for amebic growth and survival and requires an intact C-terminal domain for both alcohol dehydrogenase and acetaldehyde dehydrogenase activity. The Journal of Biological Chemistry 276(23), 20136–20143.

Finsterer, J. and Auer, H. (2013). Parasitoses of the human central nervous system. Journal of Helminthology 87(3), 257–270.

Girbal, L., Croux, C., Vasconcelos, I., and Soucaille, P. (1995). Regulation of metabolic shifts in *Clostridium acetobutylicum* ATCC 824. FEMS Microbiol. Rev. 17, 287–297.

Kumar, S., Stecher, G., and Tamura, K. (2015) MEGA7: Molecular evolutionary genetics analysis version 7.0. Molecular Biology and Evolution 33(7), 1870–1874.

Moon, J., Lee, J.H., Suk-Youl, P., Jung-Mi, S., Mi-Young, P., Hye-Mi, P., Jiali, S., Jeong-Hoh, P., Bo, Y.K., and Jeong-Sun, K. (2011). Structures of iron-dependent alcohol dehydrogenase 2 from *Zymomonas mobilis* ZM4 with and without NAD+ cofactor. J. Mol. Biol. 407, 413–424.

Nixon, J.E., Wang, A., Field, J., Morrison, H. G., McArthur, A.G., Sogin, M.L., Loftus, B.J., and Samuelson, J. (2002). Evidence for lateral transfer of genes encoding ferredoxins, nitroreductases, NADH oxidase, and alcohol dehydrogenase 3 from anaerobic prokaryotes to *Giardia lamblia* and *Entamoeba histolytica*. Eukaryotic cell 1(2), 181–190.

Nowak, P., Mastalska, K., and Loster, J. (2015). *Entamoeba histolytica* - Pathogenic protozoan of the large intestine in humans. J. Clin. Microbiol. Biochem. Technol. 1(1), 10–17.

Parsonage, D., Sheng, F., Hirata, K., Debnath, A., McKerrow, J.H., Reed, S.L., Abagyan, R., Poole, L.B., and Podust, L.M. (2014). X-ray structures of thioredoxin and thioredoxin reductase from *Entamoeba histolytica* and prevailing hypothesis of the mechanism of auranofin action. Journal of Structural Bio. 194(2), 180–190.

Paz-y-Miño-C G. and Espinosa A. (2010). Integrating horizontal gene transfer and common descent to depict evolution and contrast it with "common design." Journal of Eukaryotic Microbiology 57(1), 11–18.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S, Greenblatt, D.M., Meng, E.C, and Ferrin, T.E. (2004). UCSF Chimera: a visualization system for exploratory research and analysis. Computational chemistry 25 (13), 1605–1612. Available at: http://www.rbvi.ucsf.edu/chimera/

Pierce, B.G., Wiehe, K., Hwang, H., Kim, B.H., Vreven, T., and Weng, Z. (2014) ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. Bioinformatics 30(12), 1771–1773.

Reeves, R.E. (1984). Metabolism of *Entamoeba histolytica*. Adv. Parasitol. 23, 105–142.

Rosenthal, B., Mai, Z., Caplivski, D., Ghosh, S., Del la Vega, H., Graf, T., and Samuelson, J. (1997). Evidence for the bacterial origin of genes encoding fermentation enzymes of the amitochondriate protozoan parasite *Entamoeba histolytica*. Journal of Bacteriology 179(11), 3736–3745.

Saha, A., Gauruv, A. K., Bhattacharya, S., and Bhattacharya, A. (2015). Molecular basis of pathogenesis in amebiasis. Curr. Clinc. Micro. Rep. 2(4), 143–154.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Available at: http://bioinformatics.oxfordjournals.org/content/early/2014/01/21/bioinformatics.btu033.abstract? keytype=ref&ijkey=VTEqgUJYCDcf0kP

St-Pierre J., Moreau, F., and Chadee, K. (2014). *Entamoeba histolytica* evades innate immunity by triggering the degradation of macrophage cytoskeletal-associated proteins. The FASEB J. 28(1), Supplement 152.4.

Williamson, V.M., and Paquin, C.E. (1987). Homology of

Saccharomyces cerevisiae ADH4 to an iron-activated alcohol dehydrogenase from *Zymomonas mobilis*. Molecular Genetics and Genomics. 209(2), 374–381.

Yang, W., Li, E., Kairong, T., and Stanley, S.L., Jr. (1994). *Entamoeba histolytica* has an alcohol dehydrogenase homologous to the multifunctional *adhE* gene product of *Escherichia coli*. Molecular and Biochemical Parasitology 64(2), 253–260.

Yang, Z. (2007). PAML 4: a program package for phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24, 1586–1591. Available at: http://abacus.gene.ucl.ac.uk/software/paml.html).

Received 12 September 2017; revised 3 July 2018; accepted 1 November 2018.