

RESEARCH

Open Access



# Horizontal transfer of $\beta$ -carbonic anhydrase genes from prokaryotes to protozoans, insects, and nematodes

Reza Zolfaghari Emameh<sup>1,2,3\*</sup>, Harlan R. Barker<sup>1</sup>, Martti E. E. Tolvanen<sup>4</sup>, Seppo Parkkila<sup>1,3</sup> and Vesa P. Hytönen<sup>2,3</sup>

## Abstract

**Background:** Horizontal gene transfer (HGT) is a movement of genetic information occurring outside of normal mating activities. It is especially common between prokaryotic endosymbionts and their protozoan, insect, and nematode hosts. Although beta carbonic anhydrase ( $\beta$ -CA) plays a crucial role in metabolic functions of many living organisms, the origin of  $\beta$ -CA genes in eukaryotic species remains unclear.

**Methods:** This study was conducted using phylogenetics, prediction of subcellular localization, and identification of  $\beta$ -CA, transposase, integrase, and resolvase genes on the MGEs of bacteria. We also structurally analyzed  $\beta$ -CAs from protozoans, insects, and nematodes and their putative prokaryotic common ancestors, by homology modelling.

**Results:** Our investigations of a number of target genomes revealed that genes coding for transposase, integrase, resolvase, and conjugation complex proteins have been integrated with  $\beta$ -CA gene sequences on mobile genetic elements (MGEs) which have facilitated the mobility of  $\beta$ -CA genes from bacteria to protozoan, insect, and nematode species. The prokaryotic origin of protozoan, insect, and nematode  $\beta$ -CA enzymes is supported by phylogenetic analyses, prediction of subcellular localization, and homology modelling.

**Conclusion:** MGEs form a complete set of enzymatic tools, which are relevant to HGT of  $\beta$ -CA gene sequences from prokaryotes to protozoans, insects, and nematodes.

**Keywords:** Horizontal gene transfer, Mobile genetic elements, Plasmid, Beta carbonic anhydrase, Transposase, Integrase, Resolvase, Endosymbionts, Parasite, Evolution

## Background

Horizontal, or lateral, gene transfer (HGT or LGT) refers to movement of genetic information across normal mating barriers, between more or less phylogenetically distinct organisms, and thus stands in distinction to the standard vertical transmission of genes from parent to offspring. HGT is proving to be a more influential evolutionary mechanism than 20th-century scientists ever thought [1]. Most early, and even current, evidence for HGT in eukaryotes comes from study of protists [2, 3].

Mobile genetic elements (MGEs) are segments of DNA, encoding enzymes and other proteins, which mediate the movement of DNA in HGT within genomes (intracellular

mobility) or between cells (intercellular mobility) [4]. Transposases and site-specific recombinases catalyse the intracellular movement of MGEs. Site-specific recombinases in bacteria fall into one of two very distinct families, the  $\lambda$  integrase-like enzymes and the resolvases/invertases [5]. Recombinase interacts with a specific site in the DNA, brings the sites together in a synapse, and religates exchanged DNA strand to the host genome. Homologous recombination systems of the host also enable them to function in chromosomal deletions and other rearrangements [6]. The majority of horizontally transferred genes are either eventually excluded or rapidly become nonfunctional in the recipient genome. However, there are some reports where horizontally transferred genes have shown high level of transcription [6, 7].

Many protists are phagotrophic and subsist by consuming bacteria. Subsequently, protozoan phagotrophs often

\* Correspondence: zolfaghari.emameh.reza.x@student.uta.fi

<sup>1</sup>School of Medicine, University of Tampere, Medisiinärinkatu 3, FI-33520 Tampere, Finland

<sup>2</sup>BioMediTech, University of Tampere, FI-33520 Tampere, Finland  
Full list of author information is available at the end of the article

live for long periods in environments where they are frequently exposed to bacterial DNA. One such example is the direct contact of bacteria and parasites in digestive system of ruminants [2].

In addition, previous literature has demonstrated numerous well-established endosymbiotic partnerships between a variety of eukaryotic hosts and prokaryotic or eukaryotic endosymbionts [8–19]. The close inter-organismal interaction between the host and endosymbiont also provides an opportunity for HGT. Two prominent endosymbiotic relationships in eukaryotic evolution resulted in adoption of mitochondria and plastids from  $\alpha$ -proteobacteria and cyanobacteria species, respectively. Among *Eubacteria*, HGT is involved in the evolution of antibiotic resistance, pathogenicity, and metabolic pathways [20]. Both endosymbiotic and pathogenic prokaryotes are usually considered as the HGT DNA donors to protozoans, insects, and nematodes [21] (Table 1).

Carbonic anhydrases (CAs) are ubiquitous metalloenzymes, which belong to six evolutionary divergent gene families, including  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\zeta$ , and  $\eta$  [22, 23]. The active site of most CAs contains a zinc ion ( $Zn^{2+}$ ) which plays a critical role in the catalytic activity of the enzyme. CAs are involved in many biological processes, such as respiration involving transport of  $CO_2$  and bicarbonate between metabolizing tissues, regulation of pH homeostasis, electrolyte transfer, bone resorption, calcification, tumor progression, gluconeogenesis, lipogenesis, and ureagenesis [24–27]. In the past decade, a large number of putative  $\beta$ -CAs have been discovered in protozoans, arthropods, and nematodes [28–32], as well as in bacteria, fungi, algae, and plants [33]. Despite the presence of  $\beta$ -CA sequences in genomes of many, if not most, living organisms, they are absent in vertebrate genomes [28, 29].

In this study, we investigated the possible origin of  $\beta$ -CA gene sequences in protozoans, insects, and nematodes by HGT from ancestral prokaryotes using phylogenetics, prediction of subcellular localization, and identification of  $\beta$ -CA, transposase, integrase, and resolvase genes on the MGEs of bacteria. We also structurally analyzed  $\beta$ -CAs from protozoans, insects, and nematodes and their putative prokaryotic common ancestors, by homology modelling. Our study suggests that HGT likely explains the presence of similar  $\beta$ -CA genes across multiple species living together in distinct environments.

## Methods

### Identification of $\beta$ -CA gene and protein sequences

We collected all  $\beta$ -CA protein expressing bacteria which are endosymbiotic or pathogenic to a protozoan, insect, or nematode species from Uniprot (<http://www.uniprot.org/>) and EMBL-EBI databases (<http://www.ebi.ac.uk/>) (Additional file 1). In addition, we included ten  $\beta$ -CA protein sequences from endosymbiotic bacteria of protozoans, insects, and nematodes to the identification process, including: *Afipia* spp. (K8NQ88), *Anaeromyxobacter* spp. (A7HD59), *Campylobacter* spp. (K0I0K3), *Salmonella* spp. (Q8ZRS0), *Gardnerella* spp. (E3D7T4), *Emticicia* spp. (I2EZ21), *Simkania* spp. (F8L9G5), *Nostoc* spp. (Q8YT17), *Exiguobacterium* spp. (K0ACL8), and *Fusobacterium* spp. (C6JPI1). Moreover, we performed protein homology BLAST search for  $\beta$ -CA protein sequences from protozoans, insects, and nematodes in the EMBL-EBI BLAST database (<http://www.ebi.ac.uk/Tools/sss/fast/>) to define bacterial  $\beta$ -CA protein homologs. A highly conserved region (102 amino acid residues, starting from three amino acid residues prior to the first highly conserved motif (CXDXR) was extracted from bacterial,

**Table 1** Examples of HGT of prokaryotic genes to protozoans, insects, and nematodes

| Prokaryotic gene donors                              | Protozoan, insect and nematode gene recipients  | Horizontally transferred genes   |
|--|---|--|
| <i>Wolbachia</i>                                     | <i>Aedes aegypti</i> (yellow fever mosquito), <i>Anopheles gambiae</i> (malaria mosquito), and <i>Drosophila melanogaster</i> | Many prokaryotic genes, such as gag-pol, D34 immunodominant antigen, actin and aminotransferase genes [65, 66] |
| <i>Escherichia coli</i>                              | <i>Caenorhabditis elegans</i>   | Antibiotic-resistance genes [67]   |
| Prokaryotes  | Anaerobic protozoans: <i>Trichomonas vaginalis</i> , <i>Entamoeba histolytica</i> , and <i>Naegleria gruberi</i>              | Alcohol dehydrogenase ( <i>adh</i> gene) and Pyruvate:ferredoxin oxidoreductase genes [1, 68]                  |
| Prokaryotes  | <i>Dictyostelium discoideum</i> (soil-living amoeba)  | 18 prokaryotic genes [48]  |
| Prokaryotes  | Trypanosomatids: <i>Leishmania</i> spp., <i>Angomonas deanei</i> and <i>Strigomonas culicis</i>                               | Bacterial amino acid pathways [58]   |
| $\alpha$ -proteobacteria                             | <i>Leishmania</i> spp.  | Mitochondria (initiation point of apoptosis) [69]  |
| $\beta$ -proteobacteria and $\gamma$ -proteobacteria | Trypanosomatids: <i>Leishmania</i> spp., <i>Angomonas deanei</i> and <i>Strigomonas culicis</i>                               | <i>Heme synthesis</i> gene [50]  |
| <i>Peptostreptococcus harei</i>                      | <i>Trichomonas vaginalis</i>  | Lateral gene transfer fragment (TvLF) [51]   |

protozoan, insect, and nematode  $\beta$ -CA protein sequences. These sequences were aligned using the Clustal Omega multiple sequence alignment (MSA) algorithm (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) [34], and the results were visualized in Jalview (<http://www.jalview.org/>) [35].

### Phylogenetic analysis

A total of 220  $\beta$ -CA sequences were retrieved from various databases and sorted into sub-groups (clades) based on identification by the Conserved Domain Database server (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) [36]. Phylogenetic trees were constructed individually for each  $\beta$ -CA sub-group (clade A-D). The total numbers of sequences analyzed for each sub-group were 109(A), 53(B), 36(C), and 22(D). Four incomplete sequences were corrected, including three from *Naegleria gruberi*, which replace UniProt entries D2W4H2, D2W1R2, and D2W492, and one from *Leishmania braziliensis*, which replaces UniProt entry A4H4M7. In these corrections, the target species genome was analyzed by the Exonerate program [37], using complete  $\beta$ -CA sequences as queries, followed by a comparative analysis of a Clustal Omega alignment of the predictions [34]. For each of the clades A to D, the final set of protein sequences was aligned using Clustal Omega, and a corresponding alignment of coding sequences (CDS) was created by Pal2Nal [38]. Each set of sequences were analyzed using supercomputer resources

provided by the Finnish IT Center for Science. The first method applied was Bayesian inference within the MrBayes v3.2.3 program [39], using the General Time Reversible (GTR) nucleotide model until the standard deviation of split frequencies was  $<0.01$ . A second analysis by maximum likelihood was completed using PhyML with 1000 bootstrap replicates [40] (Table 2).

### Prediction of subcellular signals

Prediction of subcellular signals of defined protozoan, insect, and nematode  $\beta$ -CA protein sequences was performed using a subcellular signal prediction tool. Mitochondrial and secretory targeting peptides in  $\beta$ -CA protein sequences were predicted by TargetP 1.1 Server (<http://www.cbs.dtu.dk/services/TargetP/>) [41]. Even if these targeting systems are only found in eukaryotes, bacterial sequences were analyzed as well to see if they contain regions similar to eukaryotic targeting signals. Based on the phylogenetic tree results, we performed this analysis on only those bacterial  $\beta$ -CA protein sequences, which had a predicted common ancestor with protozoan, insect, or nematode  $\beta$ -CA protein sequences. Specifically, this included *Afipia felis* (K8NQ88), *Bradyrhizobium japonicum* (G7D846), *Cesiribacter andamanensis* (M7MX87), *Colwellia psychrerythraea* (Q47YG3), *Coralloccoccus coralloides* (H8MJ17), *Leptospira kirschneri* (M6X652), *Magnetospirillum magneticum* (Q2VZD0),

**Table 2** Predicted sources of the  $\beta$ -CA genes. The tentative prokaryotic endosymbionts and their hosts are listed

| $\beta$ -CA clades | Tentative prokaryotic endosymbiont (donor)                                 | Bacterial group     | Protozoan, insect, and nematode hosts (acceptor)   |
|--------------------|--|---------------------|--|
| A                  | <i>Cesiribacter andamanensis</i> (M7MX87)                                  | Bacteroidetes       | <i>Acanthamoeba castellanii</i> (L8GR38) (Fig. 2a)   |
| A                  | <i>Leptospira kirschneri</i> (M6X652)                                      | Spirochaetes        | <i>Naegleria gruberi</i> (Predicted 1, 2, 3)<br><i>Paramecium tetraurelia</i> (A0BD61, A0CEX6, A0C922, A0BDB1, A0E8I0) (Fig. 2a)   |
| A                  | <i>Colwellia psychrerythraea</i> (Q47YG3)                                  | Gammaproteobacteria | <i>Ichthyophthirius multifiliis</i> (G0QYZ1, G0QPN9)<br><i>Tetrahymena thermophila</i> (Q22U21, Q22U16, I7M0M0, I7M748, I7LWM1, I7MDL7, Q23AV1, I7MD92)<br><i>Dictyostelium</i> spp (Q555A3, Q55BU2, Q94473, F0Z7L1, F4PL43) (Fig. 2a) |
| A                  | <i>Magnetospirillum magneticum</i> (Q2VZD0)                                | Alphaproteobacteria | <i>Angomonas daenei</i> (S9WXX9)<br><i>Strigomonas culicis</i> (S9TM82)<br><i>Leishmania</i> spp (A4H4M7 as predicted, E9B8S3, A4HSV2, Q4QJ17, E9AKU0, S0CTX5) (Fig. 2a)   |
| B                  | <i>Myxococcales</i>  | Deltaproteobacteria | Insects and nematodes (F1LE18, G4V6B2, Q22460, Q5TU56, Q17N64, Q9VHJ5) (Fig. 2b)   |
| C                  | <i>Vesicomysocius okutanii</i> (A5CVM8)                                    | Gammaproteobacteria | <i>Entamoeba</i> spp (B0E7M0, 1C4LXK3, K2GQM0) (Fig. 2c)   |
| C                  | <i>Afipia felis</i> (K8NQ88)<br><i>Bradyrhizobium japonicum</i> (G7D846)   | Alphaproteobacteria | <i>Acanthamoeba castellanii</i> (L8GLS7) (Fig. 2c)   |
| D                  | <i>Selenomonas ruminantium</i> (I0GLW8)<br><i>Veillonella</i> spp (F9N508) | Firmicutes          | <i>Trichomonas vaginalis</i> (A2ENQ8, A2DLG4) (Fig. 2d)  |

*Selenomonas ruminantium* (I0GLW8), *Veillonella* spp. (F9N508), and *Vesicomysocius okutanii* (A5CVM8).

#### Identification of $\beta$ -CA, transposase, integrase, resolvase, and conjugation complex protein (CCP) genes on the prokaryotic MGEs

Identification of  $\beta$ -CA, transposase, integrase, resolvase, and CCP genes on the bacterial MGEs was carried out using the plasmid database from EMBL-EBI (<http://www.ebi.ac.uk/genomes/plasmid.html>), and the Jena Prokaryote Genome Viewer (JPGV) (<http://jpgv.fli-leibniz.de/cgi/index.pl>) [42]. JPGV contains a vast amount of information on most fully sequenced prokaryotic genomes and presents figures of linear and circular genome plots.

#### Identification of $\beta$ -CA gene sequences on protozoan, insect, and nematode genomic DNA

Analyses regarding determination of precise locations of protozoan, insect and nematode  $\beta$ -CA genes in genomic DNA were performed using National Center for Biotechnology Information (NCBI) database (<http://www.ncbi.nlm.nih.gov/>). Furthermore, we utilized the *Trichomonas vaginalis* genome project database (TrichDB version 1.3) (<http://trichdb.org/trichdb/>) [43] and EMBL-EBI database (<http://www.ebi.ac.uk/>), for detection of  $\beta$ -CA genes in *Trichomonas vaginalis* (a protozoan parasite and the causative agent of trichomoniasis) and *C. elegans* respectively. Analysis of mitochondrial coding genes in *Acanthamoeba castellanii* (the most common free-living amoeba in soil and water) was performed using the NCBI database (<http://www.ncbi.nlm.nih.gov/>).

#### Homology modelling

Homology models were prepared for  $\beta$ -CAs selected based on the phylogenetic analysis. The most similar eukaryotic and prokaryotic proteins within the phylogeny tree branch in question were selected using the percent identity matrix generated by Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) [34]. For each of the selected proteins, the most similar protein structure was obtained using BLAST search targeted for the PDB database (<http://www.rcsb.org/pdb/home/home.do>). For each protein pair (eukaryotic and prokaryotic) analyzed here, the BLAST search resulted in the same template protein as follows: Clade A: *Escherichia coli*  $\beta$ -CA PDB 1I6P; Clade B: *Pisum sativum*  $\beta$ -CA PDB 1EKJ; Clade C: *Mycobacterium tuberculosis*  $\beta$ -CA PDB 1YM3; and Clade D: *Methanobacterium thermoautotrophicum*  $\beta$ -CA PDB 1G5C.

Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) [34] was used to prepare a sequence alignment for the modelled protein and the template protein sequence. The homology models were prepared using Modeller program (version 9.14) [44]. The resulting models were

structurally aligned by the BODIL program [45]. A figure illustrating the homology models was prepared using the VMD program (version 1.9.1) [46] and edited with Adobe Photoshop (version 13.0.1).

The evaluation of the conserved residues in the homology models was performed by using multiple sequence alignments prepared by Clustal Omega algorithm (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) [34] and by inspecting the homology models using program VMD program (version 1.9.1) [46].

## Results

### Identification and phylogenetic analysis of $\beta$ -CA protein sequences from defined bacterial, protozoan, insect, and nematode species

Multiple sequence alignment (MSA) of  $\beta$ -CA protein sequences from protozoan, insect, nematode species with bacterial  $\beta$ -CA protein sequences, revealed that all the aligned sequences included both the first (CXDXR; C: Cysteine, D: Aspartic acid, R: Arginine, and X: any residue) and second (HXXC; H: Histidine, C: Cysteine, X: any residue) highly conserved motifs of the active site (Fig. 1).

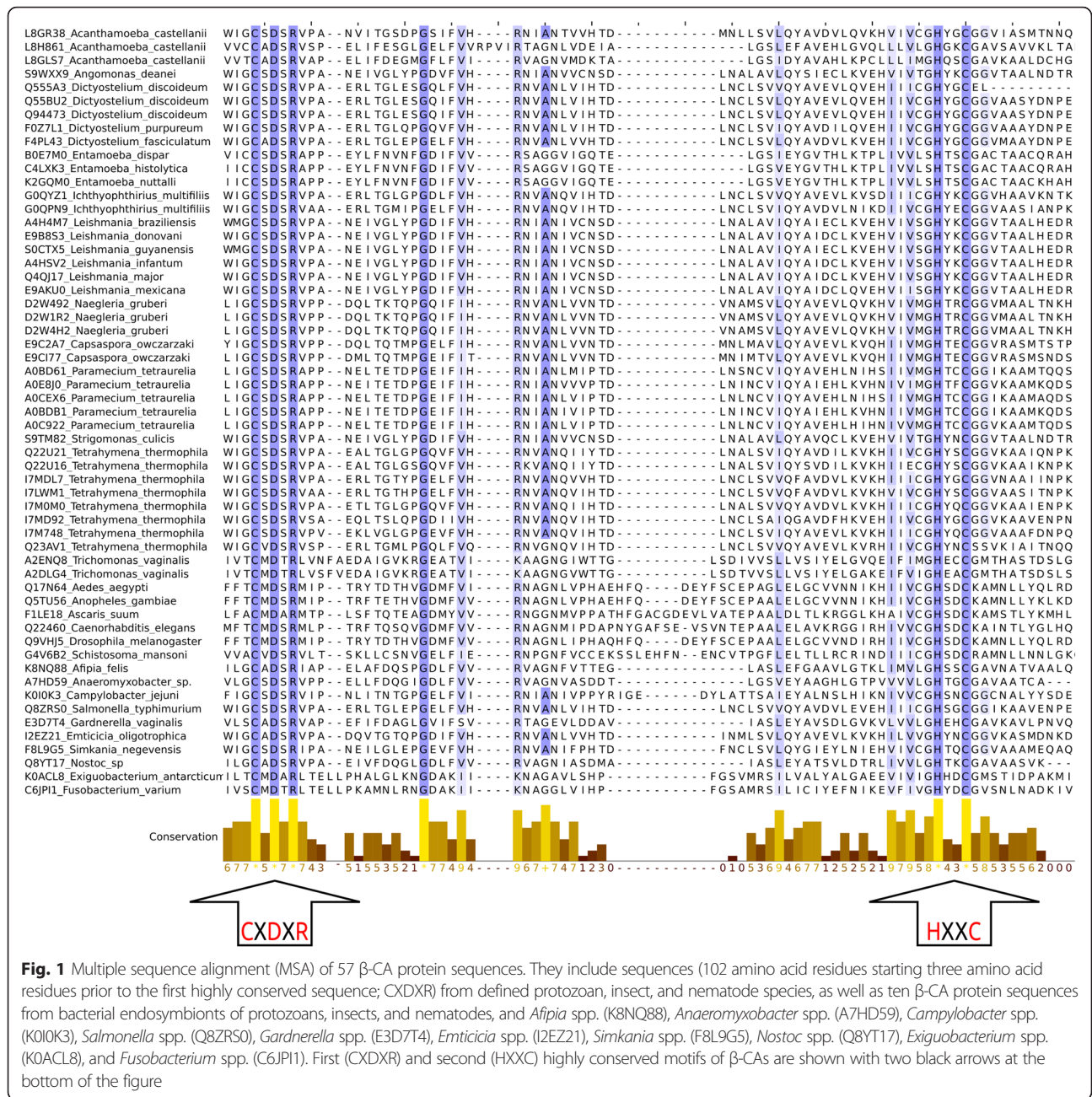
Phylogenetic analyses of clade A, B, C, and D of  $\beta$ -CA protein sequences revealed the common ancestor of protozoan, insect, and nematode  $\beta$ -CAs within bacterial  $\beta$ -CA protein sequences (Fig. 2a-d) (Table 3).

### Prediction of subcellular signals

Prediction of subcellular signals revealed that five protozoan (L8GR38, A4H4M7, S0CTX5, S9TM82, and I7MDL7) and three insect (Q5TU56, Q17N64, and Q9VHJ5)  $\beta$ -CA proteins probably contain mitochondrial targeting peptides. Even three bacterial  $\beta$ -CA proteins (K8NQ88, H8MJ17, and M6X652) contained N-terminal sequences sufficiently similar to mitochondrial targeting peptides so that mitochondrial prediction by TargetP 1.1 Server was positive. In addition, one protozoan  $\beta$ -CA protein (L8H861) sequence from *A. castellanii* is predicted to contain a signal peptide for the secretory pathway. The prediction tool provided no definitive localization for the other bacterial, protozoan, insect, and nematode  $\beta$ -CA proteins (Additional file 2).

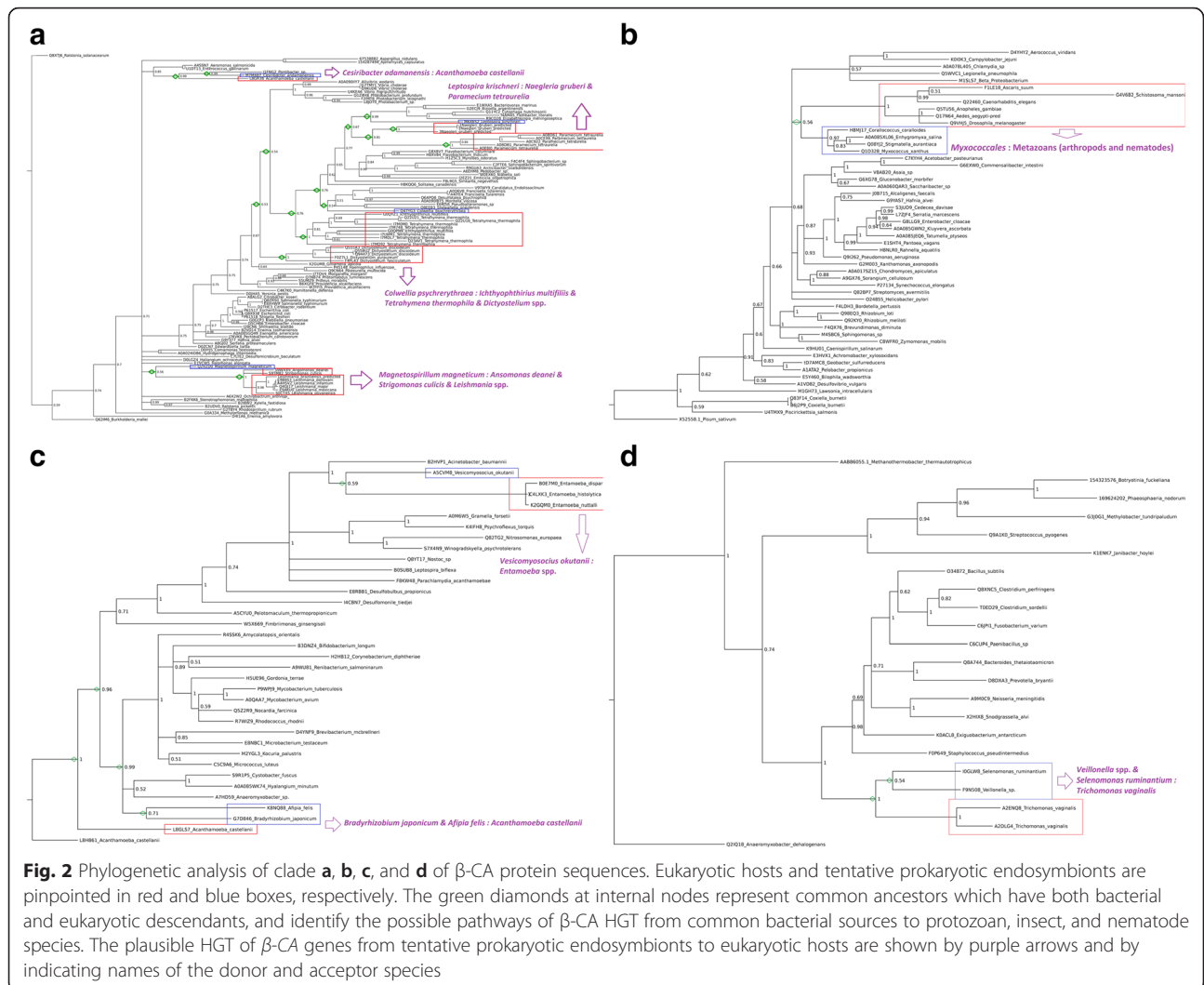
### Identification of $\beta$ -CA, transposase, integrase, resolvase, and conjugation complex protein (CCP) coding sequences on the bacterial MGEs

In order to study the genomic context of  $\beta$ -CA genes and to understand the molecular mechanisms involved in HGT, we explored the association of prokaryotic  $\beta$ -CA genes in MGEs. The ACLAME version 0.4 database (<http://aclame.ulb.ac.be/>) [47] enabled us to first identify a  $\beta$ -CA gene within the pSLT mobile genetic element of *Salmonella typhimurium* (str. LT2) (data not shown). Subsequent



analysis within other MGE browsers, including EMBL-EBI (<http://www.ebi.ac.uk/genomes/plasmid.html>) and Jena Prokaryote Genome Viewer (JPGV) (<http://jpgv.fli-leibniz.de/cgi/index.pl>) databases, led to discovery of 40  $\beta$ -CA genes located within MGEs in different prokaryotic species. Each bacterial MGE contained only one  $\beta$ -CA gene sequence and occasionally several transposase, integrase, resolvase, and CCP coding genes. MGEs were found to differ from each other by length, number of coding genes, and encoded proteins. Each  $\beta$ -CA, transposase, integrase, resolvase, and CCPs were identified by specific

coding IDs from ACLAME and GenBank and only one instance of each protein is listed (Additional file 3) for each bacterial species as a representative example. The study of ACLAME data shows that  $\beta$ -CA is found in evolutionary conserved modules of MGEs, even at the most stringent significance thresholds. The locations of  $\beta$ -CA, transposase, integrase, and resolvase gene sequences in plasmid pSLT from *S. typhimurium* (strain LT2) are shown in Fig. 3. The figure shows that pSLT expresses transposase, integrase, and resolvase as the main enzymatic tools, which facilitate the HGT of  $\beta$ -CA gene in this plasmid



and similar configuration was observed in the case of several other MGEs.

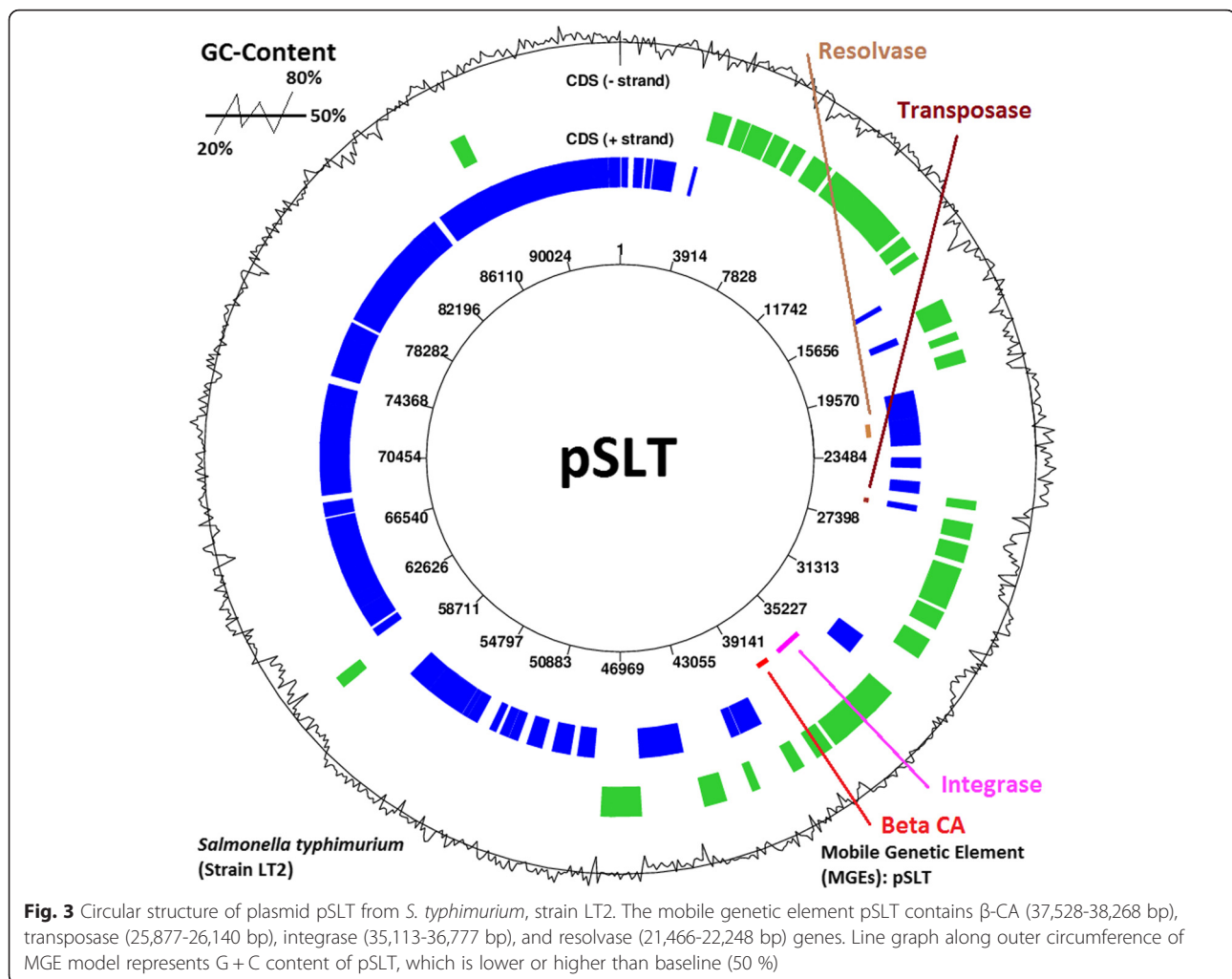
**Identification of  $\beta$ -CA gene sequences on protozoan, insect, and nematode genomic DNA**

Analysis of the precise location of  $\beta$ -CA gene sequences in protozoan, insect, and nematode genetic structures revealed that all were located in chromosomal DNA (Additional file 4). Exon counts, for the group of studied  $\beta$ -CA gene

sequences, vary in quantity from 1 to 11. The maximum exon counts were 8 for *A. castellanii* (Entry ID: L8GR38), and 11 for *P. pacificus* (Entry ID: H3EVA6) for protozoan and nematode species, respectively. Interestingly, some protozoan  $\beta$ -CA gene sequences included only one exon. The definitive locations of  $\beta$ -CA gene sequences are shown on linear genomic DNA from *T. vaginalis* (A2DLG4) (Additional file 5) and *C. elegans* (Q22460) (Additional file 6), whereas they are still unknown in many species.

**Table 3** MrBayes/PhyML Settings and Results of Phylogenetic Analysis

| $\beta$ -CA Clade | Sequences | MrBayes Iterations | MrBayes Std Dev of Split Freq | MrBayes Trees sampled | PhyML Boot Straps |
|-------------------|-----------|--------------------|-------------------------------|-----------------------|-------------------|
| A                 | 109       | 365,000,000        | 0.0092                        | 273515                | 1000              |
| B                 | 53        | 35,000,000         | 0.0087                        | 25812                 | 1000              |
| C                 | 36        | 2,000,000          | 0.0077                        | 7501                  | 1000              |
| D                 | 22        | 350,000            | 0.0055                        | 1314                  | 1000              |



Analysis of the genes on circular mitochondrial DNA from *A. castellanii* revealed that none of the protozoan  $\beta$ -CAs were considered mitochondrial coding genes (data not shown).

#### Homology models

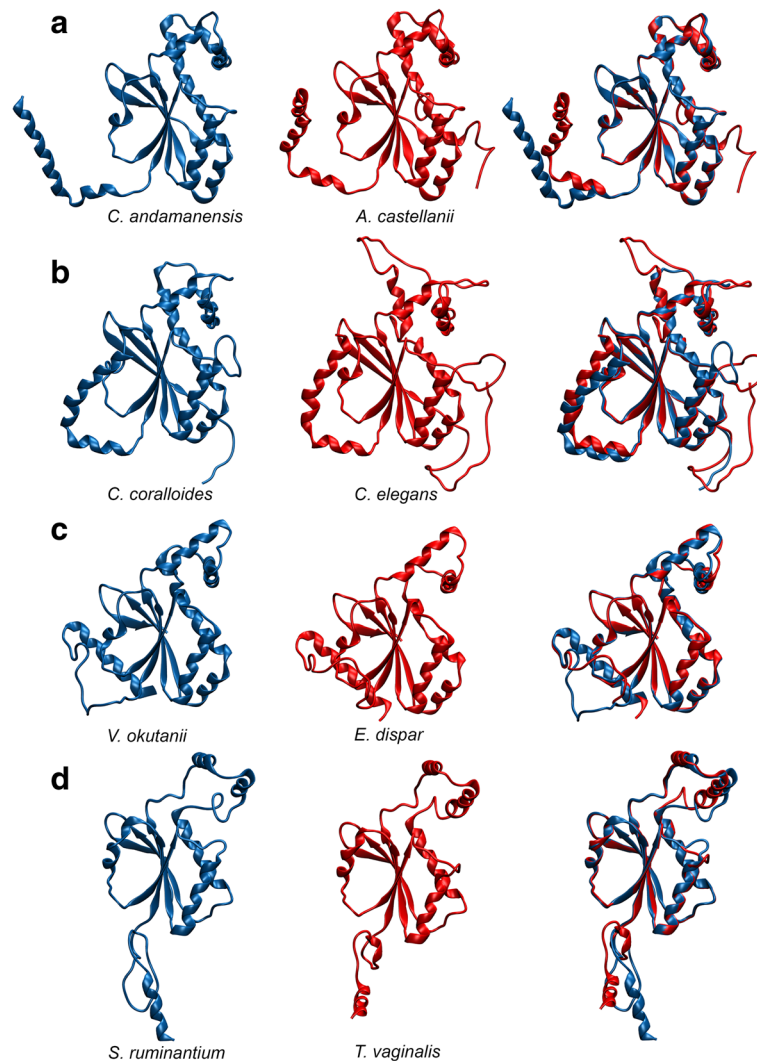
Homology modelling further supported the idea of high similarity within the inspected protein groups from prokaryotes and protozoans-metazoans (insects and nematodes). No large insertions or deletions were observed and the majority of structural variation is located in the termini of the polypeptide chains. The superimposed homology models created from a pair of proteins from each clade of the  $\beta$ -CAs are shown in Fig. 4.

#### Discussion

Throughout their evolution all eukaryotes have been in close contact with bacteria, and while eukaryotrophs are comparatively rare there are numerous identified bacterial endosymbionts which have adapted to intracellular endosymbiosis with protozoan host species [1, 48–52].

In general, HGT of prokaryotic genes to protozoan genomes is probably much more common than *vice versa* [2]. Interestingly, many bacteria are able to tolerate harsh conditions, such as presence of digestive enzymes in phagocytic vesicles, and survive inside protozoan species without any problems. The mechanisms of these efficient endosymbiotic and HGT phenomena are still unknown. There are multiple examples of highly efficient HGT, such as: from *E. coli* to protozoan ciliates, including *T. thermophila* and *T. pyriformis* [53]; from *Klebsiella* spp. to *Salmonella* spp. within the endosymbiotic environment of rumen protozoa of ruminants [54]; and from endosymbiont bacteria to *Leishmania* spp. during bacterial sepsis [55].

Multiple sequence alignment (MSA) of suspected protozoan, insect, and nematode  $\beta$ -CA protein sequences with previously defined bacterial  $\beta$ -CA proteins, revealed that all of the evaluated sequences contained the first (CXDXR) and second (HXXC) highly conserved motifs characteristic of  $\beta$ -CA proteins. Phylogenetic analysis revealed that protozoan, insect, and nematode  $\beta$ -CA protein sequences are



**Fig. 4** Homology models of representative pairs of  $\beta$ -CAs from clades **a**, **b**, **c**, and **d**. The blue protein models correspond to prokaryotic proteins and the red models to eukaryotic proteins. The superimposed models were shown in the third column at the right side

mostly categorized as clade A or B  $\beta$ -CA protein structures, respectively.

Based on our phylogenetic analysis, *A. castellanii* possesses two  $\beta$ -CA genes, one from clade A and one from clade C. Our results in Fig. 2c, suggest that the  $\beta$ -CA gene of *A. castellanii* (L8GLS7) was potentially horizontally transferred from a bacterial species, which probably was a common ancestor of *B. japonicum* (G7D846) and *A. felis* (K8NQ88). In addition, previous studies have shown that *B. japonicum* [56] and *A. felis* [57] are endosymbionts of *A. castellanii*.

Phylogenetic analysis of clade A  $\beta$ -CAs (Fig. 2a), showed that all  $\beta$ -CAs in *N. gruberi* and *P. tetraurelia* protozoa have a common source with the single  $\beta$ -CA from spirochaetes bacteria, *L. kirschneri* (M6X652). Potentially, after HGT of a  $\beta$ -CA gene from the common source to these two protozoan hosts, the gene duplicated and created

three different  $\beta$ -CA genes for *N. gruberi* (Predicted 1, 2, 3) and five for *P. tetraurelia* (A0BD61, A0CEX6, A0C922, A0BDB1, A0E810).

Among the various prokaryotic endosymbionts it is proposed that *I. multifiliis*, *T. thermophila*, and *Dictyostelium* spp. potentially have a distant common source with gammaproteobacteria *C. psychrerythraea* (Q47YG3), because there are multiple branch points between *C. psychrerythraea* and the other prokaryotic species. Gene duplication in these protozoans led to multiple copies of  $\beta$ -CA in *I. multifiliis* (G0QYZ1, G0QPN9), *T. thermophila* (Q22U21, Q22U16, I7M0M0, I7M748, I7LWM1, I7MDL7, Q23AV1, I7MD92), and *Dictyostelium* spp. (Q555A3, Q555BU2, Q94473, F0Z7L1, F4PL43) [28].

It has been shown earlier that essential amino acid and heme synthesis genes horizontally transferred from endosymbiont alpha, beta, and gammaproteobacteria to



*Trypanosomatidea* [9, 50, 58]. Our phylogenetic results (Fig. 2a) revealed that  $\beta$ -CA genes in *Trypanosomatidea*, including *Leishmania* spp. (A4H4M7, E9B8S3, A4HSV2, Q4QJ17, E9AKU0, and S0CTX5), *A. daenei* (S9WXX9), and *S. culicis* (S9TM82) have a common source with an alphaproteobacterium similar to *M. magneticum* (Q2VZD0).

The phylogenetic analysis (Fig. 2b) showed that insect and nematode  $\beta$ -CAs belong to clade B and suggests that they may have a common source with myxobacterial  $\beta$ -CAs. The various myxobacteria *Corallococcus*, *Enhygro-myxa*, *Stigmatella*, and *Myxococcus*, are part of the same subtree that contains insect and nematode  $\beta$ -CAs. However, a larger analysis with more insect, nematode, and plant  $\beta$ -CAs, which also belong to clade B, would be needed to fully resolve the relationships within this clade. Given the apparent distribution within insects and nematodes, in our limited analysis, this HGT would have occurred in the distant past. A single, very old transfer of  $\beta$ -CA gene to insects and nematodes would fit with the idea that heritable transfer to sexually reproducing organisms is significantly more difficult. Due to sequence divergence over 800 million years (estimated divergence time between nematodes and arthropods), our phylogenetic trees do not provide conclusive evidence for this, and it is thus possible to speculate that the  $\beta$ -CAs of clade B, which we see in insects and nematodes, have been retained from an ancestral eukaryote. However, it is tempting to assume that  $\beta$ -CAs of all four clades in protozoans, insects, and nematodes would have been derived by HGT from prokaryotes. In this context, we may also note that the HGT of  $\beta$ -CA gene sequences might have involved several mechanisms and genetic elements in addition to MGEs, such as genomic islands (GIs) and insertion sequence (IS) elements.

Phylogenetic analysis of clade C (Fig. 2c) revealed that  $\beta$ -CA genes from *Entamoeba* spp. (B0E7M0, 1C4LXK3, K2GQM0) have a common source with the  $\beta$ -CA gene of gammaproteobacterium *V. okutanii* (A5CVM8). From this result, we propose that  $\beta$ -CA genes horizontally transferred from an ancestral enteric gammaproteobacteria to *Entamoeba* spp. through a symbiotic or pathogenic relationship in the gut of arthropods, nematodes, or animals.

Phylogenetic analysis of clade D (Fig. 2d) revealed that  $\beta$ -CA genes in *T. vaginalis* (A2ENQ8, A2DLG4) have a common source with  $\beta$ -CA genes from firmicutes bacteria *S. ruminantium* (I0GLW8) and *Veillonella* spp. (F9N508). Previous results have shown that *Clostridium sordellii* and *Veillonella* spp. from firmicutes phylum and *T. vaginalis* have a symbiotic living situation in sexual organs of animals [18, 19], providing the environment in which a transfer of firmicutes bacteria  $\beta$ -CA gene sequence into the *T. vaginalis* genome is possible.

Prediction of subcellular signals of  $\beta$ -CA protein sequences revealed that some bacterial species (*A. felis*, *L. kirschneri*, and *C. coralloides*), protozoan species (*A. castellanii*, *L. braziliensis*, *L. guyanensis*, *S. culicis*, and *T. thermophila*), and insect species (*A. gambiae*, *A. aegypti* and *D. melanogaster*) include mitochondrial signals or similar bacterial sequences in their  $\beta$ -CA protein sequences (Additional file 2). It is well established that prokaryotes and some anaerobic protozoa, such as *G. lamblia*, *E. histolytica*, *T. vaginalis*, *C. parvum*, *Blastocystis hominis*, *Encephalitozoon cuniculi*, *Sawyeria marylandensis*, *Neocallimastix patriciarum*, and *Mastigamoeba bala muthi* completely lack mitochondria. In anaerobic protozoan species, mitochondrion-related organelles (MROs, mitosomes, or hydrogenosomes) replaced mitochondria in oxygen-restricted environments. Many studies have hypothesized that a majority of the mitochondrial genes in anaerobic parasitic protozoa have been acquired from  $\alpha$ -proteobacterial genomes [59]. The *Monoamine oxidase* (a mitochondrial outer membrane enzyme for metabolism of neuromediators) gene is one such example, and its sequence has been investigated thoroughly from bacterial to vertebrate lineages [60]. Therefore, we hypothesize that sequences similar to mitochondrial localization signals emerged in  $\beta$ -CA proteins in prokaryotes, leading to their mitochondrial localization after HGT into protozoans and possibly insects. Supporting this idea, the  $\beta$ -CA of *D. melanogaster* has been experimentally shown to be localized in mitochondria [28, 29].

Identification of  $\beta$ -CA with transposase, integrase, resolvase, and CCP coding sequences in bacterial MGEs suggests that these genetic elements are a complete set of enzymatic tools, which are relevant to HGT. These accessory enzymes detect target sites on the genome of recipient protozoan species using complex mechanisms and create a conducive environment for integration of  $\beta$ -CA gene sequences. On the other hand, in some MGEs, including pSLT, pOU1113, pSCV50, and pKDSC50 from *S. typhimurium* (str. LT2), *S. enterica*, *S. enterica* (serovar *Choleraesuis*, str. SC-B67), and *S. enterica* (serovar *Choleraesuis*), respectively,  $\beta$ -CA is a virulence factor which is located at 5' end of the *resolvase* gene [61]. The MGEs from *E. histolytica* contain the coding sequence for B2 DNA polymerase [62]. Analysis of the full genomes of protozoans revealed that all  $\beta$ -CA gene sequences were located on a single chromosome, although the precise chromosomal location for some protozoan  $\beta$ -CA genes is still pending (Additional file 4).

In order to evaluate the structural features of the identified  $\beta$ -CA proteins, we first analyzed the functional roles of the conserved residues.  $\beta$ -CAs have only a limited number of conserved residues essential for the protein fold and function [63]. We demonstrated this by creating a MSA of the  $\beta$ -CAs included in the homology modelling analysis.

Indeed, this analysis indicated strict conservation of only the active site residues plus one glycine (Fig. 1). We then further analyzed the residues conserved in  $\beta$ -CAs where eukaryotic and prokaryotic versions grouped together in phylogenetic analysis, i.e. those that we suspected were the result of HGT. One would expect that high similarity between proteins in distantly related species would exist due to two reasons: (1) convergent evolution or (2) HGT. In the possible case of convergent evolution, there should be a selective pressure towards a particular structural or functional feature in certain locations of the protein sequence. We analyzed this by selecting residues, which were found to be conserved between each pairing of phylogenetically grouped eukaryotic and prokaryotic  $\beta$ -CAs, but not in the  $\beta$ -CAs used as a template in homology modelling. Because this excludes the well known functional active site residues, the remaining conserved residues (especially the side chains) should have a particularly important role in the protein structure to cause convergent evolution. Within the ten conserved residues from the protein core selected for analysis of each homology model, we typically observed only a few hydrophobic contacts and in particular polar interactions were almost completely missing, even when considering possible rotamers of the surrounding residues. The result of this analysis thus implies that there are no structurally important roles for the majority of the conserved residues common for the protein pairs observed in the phylogeny analysis. This suggests that the proteins share their identical residues due to their origins in a relatively recent identical genetic source (HGT), not because of selection pressure towards the particular residue observed in each position.

Our present findings may shed some light into the question of why  $\beta$ -CA gene sequences are completely absent in the genomes of vertebrates. In protozoan and invertebrate metazoans, including insect and nematode species,  $\beta$ -CA gene sequences have integrated in nuclear chromosomes through the aid of some enzymatic functions included in MGEs, such as transposase, integrase, and resolvase. These enzymes function as site-specific cutters and snip the DNA of the recipient eukaryotic host. There are some possible reasons for the lack of HGT of  $\beta$ -CA gene sequences in vertebrate genomes. First, there may not be a specific transposable element insertion site within vertebrate genomes for these enzymatic cutters. Second, vertebrates are complex multicellular organisms in which evolutionarily stable integration of  $\beta$ -CA gene sequences would need to have taken place in the germ cells that give rise to egg and sperm cells [64]. Finally, supposing successful integration of a  $\beta$ -CA gene sequence in the germ line, it may have then been removed by genetic assortment of the vertebrate hosts. Therefore, the lack of  $\beta$ -CA gene sequences from the vertebrate genomes is understandable, especially because there is no evolutionary pressure for

the adoption of another CA class due to the presence of several efficient  $\alpha$ -CAs in all vertebrates.

## Conclusions

Many prokaryotic MGEs contain necessary enzyme gene sequences, such as transposase, integrase, and resolvase, together with  $\beta$ -CA. These enzymes can facilitate HGT of  $\beta$ -CA genes from prokaryotes to other prokaryotes (Pro-Pro) and eukaryotes (Pro-Euk). The results from both mitochondrial targeting signal prediction and phylogenetic analysis supported our hypothesis of HGT of  $\beta$ -CA gene sequences from endosymbiotic bacteria to protozoan, insect, and nematode hosts by MGEs. The phylogenetic analysis suggests that different protozoan  $\beta$ -CA genes have various common ancestors among prokaryotes, divided between clades A, C and D of  $\beta$ -CAs. In contrast, the case of insect and nematode  $\beta$ -CA genes is more complex. We propose that they may have had a single common ancestor from a bacterial  $\beta$ -CA gene, however, their descent from an ancient eukaryote origin cannot be ruled out. In analysis of the conserved residues in the homology models of prokaryote/eukaryote pairs, we observed no particularly important structural reason for the high sequence homology. This finding speaks against convergent evolution as a reason for the high similarity between the proteins and supports the idea of HGT as a source of the  $\beta$ -CA gene in eukaryotic species.

## Additional files

**Additional file 1:**  $\beta$ -CA expressing prokaryotes and their endosymbiotic protozoan, insect, and nematode hosts. (PDF 301 kb)

**Additional file 2:** Prediction of subcellular localization of in vitro-approved prokaryotic endosymbionts and protozoan  $\beta$ -CA protein sequences. (PDF 394 kb)

**Additional file 3:** Bacterial MGEs containing  $\beta$ -CA, transposase, integrase, resolvase, and CCP coding sequences. (PDF 338 kb)

**Additional file 4:** Genomic location of  $\beta$ -CA gene sequences from protozoan, insect, and nematode species. (PDF 363 kb)

**Additional file 5:** Location of  $\beta$ -CA gene sequence (TVAG\_268150) in *T. vaginalis*. This gene (Entry ID: A2DLG4) has been located on the linear main genomic DNA sequence from 151,119 to 151,673 nt. Analysis revealed that it consists of only one exon (Additional file 4). (TIF 45 kb)

**Additional file 6:** Location of  $\beta$ -CA gene sequence (bca-1) in *C. elegans*. This gene (Entry ID: Q22460) has been located on linear main genomic DNA sequence from 23,095 to 25,694 nt. Analysis revealed that it consists of seven exons (Additional file 4). (TIF 132 kb)

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

All authors participated in the design of the study. RZE carried out the bioinformatics searches on bacterial, protozoan, insect, and nematode species, as well as identification of  $\beta$ -CA, transposase, integrase, resolvase, and conjugation complex protein genes from bacterial mobile genetic elements and genomic location of protozoan  $\beta$ -CAs. RZE and HRB participated in the multiple sequence alignment. HRB made protein sequence corrections and predictions. RZE and HRB performed the phylogenetic analysis. RZE performed the prediction of subcellular

localization signals of  $\beta$ -CAs. RZE and VPH participated in the homology modelling. RZE, HRB and VPH drafted the first version of the manuscript. All authors participated in writing further versions and read and approved the final manuscript.

#### Acknowledgments

To perform these studies RZE received a scholarship support from the Ministry of Science, Research and Technology, and National Institute of Genetic Engineering and Biotechnology of Islamic Republic of Iran. Also, this work was supported by the Academy of Finland, Finnish Cultural Foundation (Pirkanmaa Regional Fund for RZE and Maili Autio Fund for HRB), Sigrid Juselius Foundation, Jane and Aatos Erkko Foundation, Tampere Tuberculosis Foundation, and Competitive Research Funding of the Tampere University Hospital for SP.

#### Author details

<sup>1</sup>School of Medicine, University of Tampere, Medisiininkatu 3, FI-33520 Tampere, Finland. <sup>2</sup>BioMediTech, University of Tampere, FI-33520 Tampere, Finland. <sup>3</sup>Fimlab Laboratories Ltd and Tampere University Hospital, FI-33520 Tampere, Finland. <sup>4</sup>Department of Information Technology, University of Turku, FI-20520 Turku, Finland.

Received: 3 February 2016 Accepted: 1 March 2016

Published online: 16 March 2016

#### References

- Andersson JO. Lateral gene transfer in eukaryotes. *Cell Mol Life Sci*. 2005; 62(11):1182–97.
- Keeling PJ, Palmer JD. Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet*. 2008;9(8):605–18.
- Top EM, Springael D. The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. *Curr Opin Biotechnol*. 2003; 14(3):262–9.
- Frost LS, Leplae R, Summers AO, Toussaint A. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol*. 2005;3(9):722–32.
- Thorpe HM, Smith MC. In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc Natl Acad Sci U S A*. 1998;95(10):5505–10.
- Toussaint A, Merlin C. Mobile elements as a combination of functional modules. *Plasmid*. 2002;47(1):26–35.
- Hou Q, He J, Yu J, Ye Y, Zhou D, Sun Y, et al. A case of horizontal gene transfer from C6/36 cell line. *Mob Genet Elements*. 2014;4(1):e28914.
- Wernegreen JJ. Endosymbiosis: lessons in conflict resolution. *PLoS Biol*. 2004;2(3):E68.
- Catta-Preta CM, Brum FL, da Silva CC, Zuma AA, Elias MC, de Souza W, et al. Endosymbiosis in trypanosomatid protozoa: the bacterium division is controlled during the host cell cycle. *Front Microbiol*. 2015;6:520.
- Greub G, Raoult D. Microorganisms resistant to free-living amoebae. *Clin Microbiol Rev*. 2004;17(2):413–33.
- Skrivan C, Fajardo M, Hagele S, Horn M, Wagner M, Michel R, et al. Various bacterial pathogens and symbionts infect the amoeba *Dictyostelium discoideum*. *Int J Med Microbiol*. 2002;291(8):615–24.
- Bertelli C, Greub G. Lateral gene exchanges shape the genomes of amoeba-resisting microorganisms. *Front Cell Infect Microbiol*. 2012;2:110.
- Sun HY, Noe J, Barber J, Coyne RS, Cassidy-Hanley D, Clark TG, et al. Endosymbiotic bacteria in the parasitic ciliate *Ichthyophthirius multifiliis*. *Appl Environ Microbiol*. 2009;75(23):7445–52.
- Beier CL, Horn M, Michel R, Schweikert M, Gortz HD, Wagner M. The genus *Caedibacter* comprises endosymbionts of *Paramecium* spp. related to the Rickettsiales (Alphaproteobacteria) and to *Francisella tularensis* (Gammaproteobacteria). *Appl Environ Microbiol*. 2002;68(12):6043–50.
- Fujishima M, Kodama Y. Endosymbionts in *Paramecium*. *Eur J Protistol*. 2012;48(2):124–37.
- Nakajima T, Sano A, Matsuoka H. Auto-/heterotrophic endosymbiosis evolves in a mature stage of ecosystem development in a microcosm composed of an alga, a bacterium and a ciliate. *Biosystems*. 2009;96(2):127–35.
- Siegmund L, Burmester A, Fischer MS, Wostemeyer J. A model for endosymbiosis: interaction between *Tetrahymena pyriformis* and *Escherichia coli*. *Eur J Protistol*. 2013;49(4):552–63.
- Fichorova RN, Buck OR, Yamamoto HS, Fashemi T, Dawood HY, Fashemi B, et al. The villain team-up or how *Trichomonas vaginalis* and bacterial vaginosis alter innate immunity in concert. *Sex Transm Infect*. 2013;89(6):460–6.
- Smutna T, Goncalves VL, Saraiva LM, Tachezy J, Teixeira M, Hrdy I. Flavodiiron protein from *Trichomonas vaginalis* hydrogenosomes: the terminal oxygen reductase. *Eukaryot Cell*. 2009;8(1):47–55.
- Boucher Y, Douady CJ, Papke RT, Walsh DA, Boudreau ME, Nesbo CL, et al. Lateral gene transfer and the origins of prokaryotic groups. *Annu Rev Genet*. 2003;37:283–328.
- Kjeldsen KU, Obst M, Nakano H, Funch P, Schramm A. Two types of endosymbiotic bacteria in the enigmatic marine worm *Xenoturbella bocki*. *Appl Environ Microbiol*. 2010;76(8):2657–62.
- Elleuche S, Poggeler S. Carbonic anhydrases in fungi. *Microbiology*. 2010; 156(Pt 1):23–9.
- Del Prete S, Vullo D, Fisher GM, Andrews KT, Poulsen SA, Capasso C, et al. Discovery of a new family of carbonic anhydrases in the malaria pathogen *falciparum*- the eta-carbonic anhydrases. *Bioorg Med Chem Lett*. 2014; 24(18):4389–96.
- Alterio V, Vitale RM, Monti SM, Pedone C, Scozzafava A, Cecchi A, et al. Carbonic anhydrase inhibitors: X-ray and molecular modeling study for the interaction of a fluorescent antitumor sulfonamide with isozyme II and IX. *J Am Chem Soc*. 2006;128(25):8329–35.
- Nishimori I, Minakuchi T, Onishi S, Vullo D, Scozzafava A, Supuran CT. Carbonic anhydrase inhibitors. DNA cloning, characterization, and inhibition studies of the human secretory isozyme VI, a new target for sulfonamide and sulfamate inhibitors. *J Med Chem*. 2007;50(2):381–8.
- Vullo D, Franchi M, Gallori E, Antel J, Scozzafava A, Supuran CT. Carbonic anhydrase inhibitors. Inhibition of mitochondrial isozyme V with aromatic and heterocyclic sulfonamides. *J Med Chem*. 2004;47(5):1272–9.
- Vullo D, Innocenti A, Nishimori I, Pastorek J, Scozzafava A, Pastorekova S, et al. Carbonic anhydrase inhibitors. Inhibition of the transmembrane isozyme XII with sulfonamides—a new target for the design of antitumor and anti-glioma drugs? *Bioorg Med Chem Lett*. 2005;15(4):963–9.
- Zolfaghari Emameh R, Barker H, Tolvanen ME, Ortutay C, Parkkila S. Bioinformatic analysis of beta carbonic anhydrase sequences from protozoans and metazoans. *Parasit Vectors*. 2014;7:38.
- Syrjanen L, Tolvanen M, Hilvo M, Olatubosun A, Innocenti A, Scozzafava A, et al. Characterization of the first beta-class carbonic anhydrase from an arthropod (*Drosophila melanogaster*) and phylogenetic analysis of beta-class carbonic anhydrases in invertebrates. *BMC Biochem*. 2010;11:28.
- Zolfaghari Emameh R, Syrjanen L, Barker H, Supuran CT, Parkkila S. *Drosophila melanogaster*: a model organism for controlling Dipteran vectors and pests. *J Enzyme Inhib Med Chem*. 2015;30(3):505–13.
- Zolfaghari Emameh R, Barker H, Hytonen VP, Tolvanen ME, Parkkila S. Beta carbonic anhydrases: novel targets for pesticides and anti-parasitic agents in agriculture and livestock husbandry. *Parasit Vectors*. 2014;7:403.
- Zolfaghari Emameh R, Kuuslahti M, Vullo D, Barker HR, Supuran CT, Parkkila S. *Ascaris lumbricoides* beta carbonic anhydrase: a potential target enzyme for treatment of ascariasis. *Parasit Vectors*. 2015;8:479.
- Smith KS, Jakubczik C, Whittam TS, Ferry JG. Carbonic anhydrase is an ancient enzyme widespread in prokaryotes. *Proc Natl Acad Sci U S A*. 1999; 96(26):15184–9.
- Sievers F, Higgins DG. Clustal omega. *Curr Protoc Bioinformatics*. 2014;48:3.13.1–16.
- Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25(9):1189–91.
- Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res*. 2015; 43(Database issue):D222–6.
- Slater GS, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*. 2005;6:31.
- Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. 2006;34(Web Server issue):W609–12.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61(3):539–42.
- Guindon S, Delsuc F, Dufayard JF, Gascuel O. Estimating maximum likelihood phylogenies with PhyML. *Methods Mol Biol*. 2009;537:113–37.
- Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc*. 2007;2(4):953–71.
- Romualdi A, Felder M, Rose D, Gausmann U, Schilhabel M, Glockner G, et al. GenColors: annotation and comparative genomics of prokaryotes made easy. *Methods Mol Biol*. 2007;395:75–96.

43. Aurrecochea C, Brestelli J, Brunk BP, Carlton JM, Dommer J, Fischer S, et al. GiardiaDB and TrichDB: integrated genomic resources for the eukaryotic protist pathogens *Giardia lamblia* and *Trichomonas vaginalis*. *Nucleic Acids Res.* 2009;37(Database issue):D526–30.
44. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, et al. Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics.* 2006;Chapter 5:Unit 5.6.
45. Lehtonen JV, Still DJ, Rantanen VV, Ekholm J, Bjorklund D, Iftikhar Z, et al. BODIL: a molecular modeling environment for structure-function analysis and drug design. *J Comput Aided Mol Des.* 2004;18(6):401–19.
46. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph.* 1996;14(1):33–8. 27–8.
47. Leplae R, Lima-Mendez G, Toussaint A. ACLAME: a CLAssification of Mobile genetic Elements, update 2010. *Nucleic Acids Res.* 2010;38(Database issue): D57–61.
48. Eichinger L, Pachebat JA, Glockner G, Rajandream MA, Suggang R, Berriman M, et al. The genome of the social amoeba *Dictyostelium discoideum*. *Nature.* 2005;435(7038):43–57.
49. Huang J, Mullanpudi N, Lancto CA, Scott M, Abrahamsen MS, Kissinger JC. Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in *Cryptosporidium parvum*. *Genome Biol.* 2004;5(11):R88.
50. Alves JM, Voegtly L, Matveyev AV, Lara AM, da Silva FM, Serrano MG, et al. Identification and phylogenetic analysis of heme synthesis genes in trypanosomatids and their bacterial endosymbionts. *PLoS One.* 2011;6(8): e23518.
51. Strese A, Backlund A, Alsmark C. A recently transferred cluster of bacterial genes in *Trichomonas vaginalis*-lateral gene transfer and the fate of acquired genes. *BMC Evol Biol.* 2014;14:119.
52. Motta MC, Martins AC, de Souza SS, Catta-Preta CM, Silva R, Klein CC, et al. Predicting the proteins of *Angomonas deanei*, *Strigomonas culicis* and their respective endosymbionts reveals new aspects of the trypanosomatidae family. *PLoS One.* 2013;8(4):e60209.
53. Matsuo J, Oguri S, Nakamura S, Hanawa T, Fukumoto T, Hayashi Y, et al. Ciliates rapidly enhance the frequency of conjugation between *Escherichia coli* strains through bacterial accumulation in vesicles. *Res Microbiol.* 2010; 161(8):711–9.
54. McCuddin ZP, Carlson SA, Rasmussen MA, Franklin SK. *Klebsiella* to *Salmonella* gene transfer within rumen protozoa: implications for antibiotic resistance and rumen defaunation. *Vet Microbiol.* 2006;114(3–4):275–84.
55. Endris M, Takele Y, Woldeyohannes D, Tiruneh M, Mohammed R, Moges F, et al. Bacterial sepsis in patients with visceral leishmaniasis in Northwest Ethiopia. *Biomed Res Int.* 2014;2014:361058.
56. La Scola B, Mezi L, Auffray JP, Berland Y, Raoult D. Patients in the intensive care unit are exposed to amoeba-associated pathogens. *Infect Control Hosp Epidemiol.* 2002;23(8):462–5.
57. La Scola B, Raoult D. *Afpia felis* in hospital water supply in association with free-living amoebae. *Lancet.* 1999;353(9161):1330.
58. Alves JM, Klein CC, da Silva FM, Costa-Martins AG, Serrano MG, Buck GA, et al. Endosymbiosis in trypanosomatids: the genomic cooperation between bacterium and host in the synthesis of essential amino acids is heavily influenced by multiple horizontal gene transfers. *BMC Evol Biol.* 2013;13:190.
59. Makiuchi T, Nozaki T. Highly divergent mitochondrion-related organelles in anaerobic parasitic protozoa. *Biochimie.* 2014;100:3–17.
60. Lander ES, Linton LM, Birren B, Nussbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001;409(6822): 860–921.
61. Stavrinides J, Kirzinger MW, Beasley FC, Guttman DS. E622, a miniature, virulence-associated mobile element. *J Bacteriol.* 2012;194(2):509–17.
62. Pastor-Palacios G, Lopez-Ramirez V, Cardona-Felix CS, Bribea LG. A transposon-derived DNA polymerase from *Entamoeba histolytica* displays intrinsic strand displacement, processivity and lesion bypass. *PLoS One.* 2012;7(11):e49964.
63. Sawaya MR, Cannon GC, Heinhorst S, Tanaka S, Williams EB, Yeates TO, et al. The structure of beta-carbonic anhydrase from the carboxysomal shell reveals a distinct subclass with one active site for the price of two. *J Biol Chem.* 2006;281(11):7546–55.
64. Andersson JO, Doolittle WF, Nesbo CL. Genomics. Are there bugs in our genome? *Science.* 2001;292(5523):1848–50.
65. Baldini F, Segata N, Pompon J, Marcenac P, Robert Shaw W, Dabire RK, et al. Evidence of natural *Wolbachia* infections in field populations of *Anopheles gambiae*. *Nat Commun.* 2014;5:3985.
66. Dunning Hotopp JC, Clark ME, Oliveira DC, Foster JM, Fischer P, Munoz Torres MC, et al. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science.* 2007;317(5845):1753–6.
67. Portal-Celhay C, Nehrke K, Blaser MJ. Effect of *Caenorhabditis elegans* age and genotype on horizontal gene transfer in intestinal bacteria. *FASEB J.* 2013;27(2):760–8.
68. Loftus B, Anderson I, Davies R, Alsmark UC, Samuelson J, Amedeo P, et al. The genome of the protist parasite *Entamoeba histolytica*. *Nature.* 2005; 433(7028):865–8.
69. Taylor-Brown E, Hurd H. The first suicides: a legacy inherited by parasitic protozoans from prokaryote ancestors. *Parasit Vectors.* 2013;6:108.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

