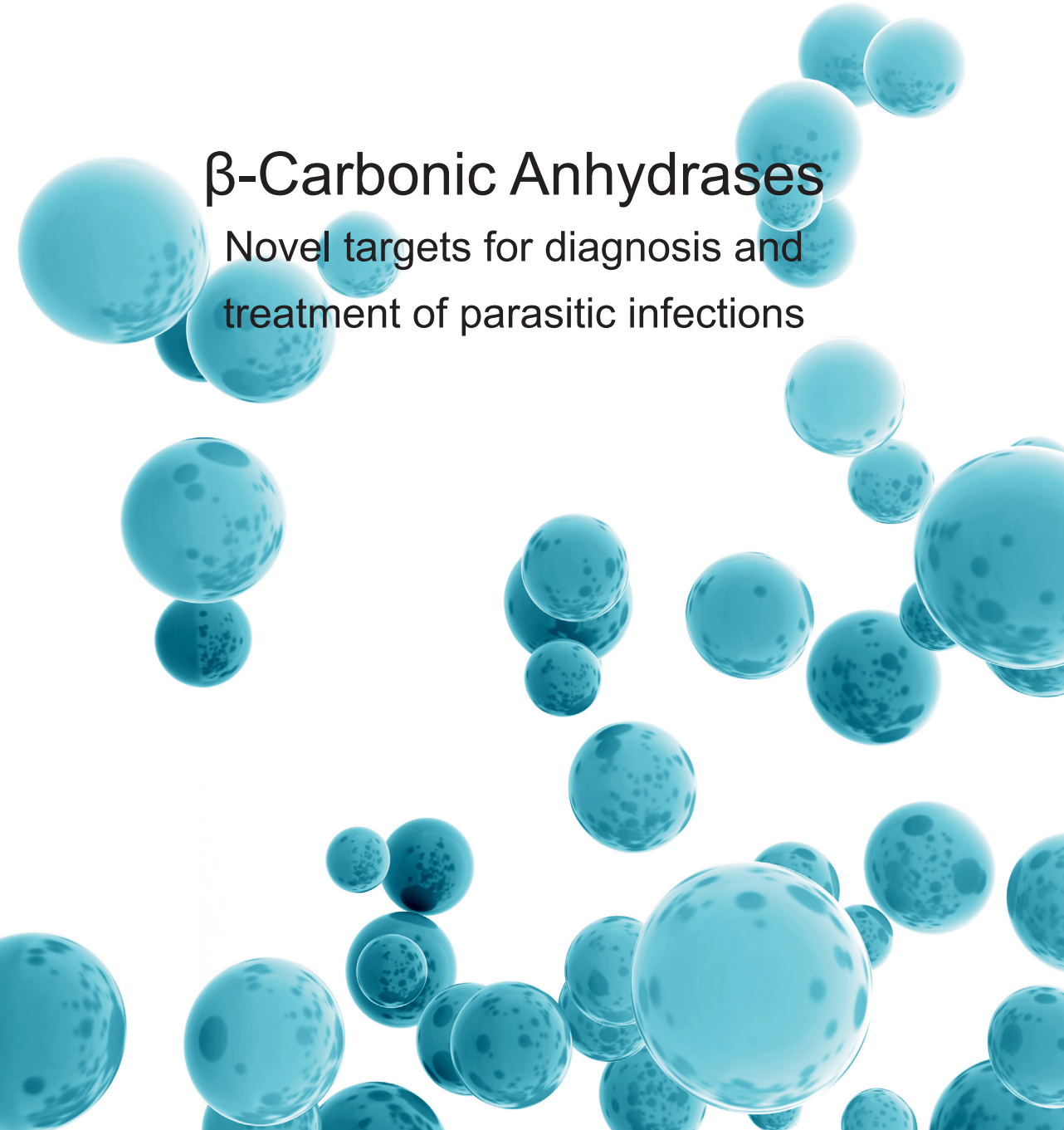
Acta Universitatis Tamperensis 2217

REZA ZOLFAGHARI EMAMEH

β-Carbonic Anhydrases

Novel targets for diagnosis and treatment of parasitic infections



REZA ZOLFAGHARI EMAMEH

β-Carbonic Anhydrases

Novel targets for diagnosis and treatment of parasitic infections

ACADEMIC DISSERTATION
To be presented, with the permission of
the Board of the BioMediTech of the University of Tampere,
for public discussion in the auditorium F114 of the Arvo building,
Lääkärinkatu 1, Tampere,
on 18 November 2016, at 13 o'clock.

UNIVERSITY OF TAMPERE

REZA ZOLFAGHARI EMAMEH

β-Carbonic Anhydrases

Novel targets for diagnosis and treatment of parasitic infections

Acta Universitatis Tamperensis 2217
Tampere University Press
Tampere 2016

ACADEMIC DISSERTATION
University of Tampere, BioMediTech
Finland

Supervised by
Professor Emeritus Markku Kulomaa
University of Tampere
Finland
Professor Seppo Parkkila
University of Tampere
Finland

Reviewed by
Docent Jukka Hytönen
University of Turku
Finland
Docent Lotta-Riina Sundberg
University of Jyväskylä
Finland

Cover design by
Mikko Reinikka

Acta Universitatis Tamperensis 2217
ISBN 978-952-03-0241-2 (print)
ISSN-L 1455-1616
ISSN 1455-1616

Acta Electronica Universitatis Tamperensis 1717
ISBN 978-952-03-0242-9 (pdf)
ISSN 1456-954X
http://tampub.uta.fi

Suomen Yliopistopaino Oy – Juvenes Print
Tampere 2016

Table of contents

List of original communications

This thesis is based on the following original communications, which are referred to in the text by their Roman numerals (I-VI).

I. **Zolfaghari Emameh R**, Barker H, Tolvanen ME, Ortutay C, Parkkila S. "Bioinformatic analysis of beta carbonic anhydrase sequences from protozoans and metazoans". Parasites & Vectors. 7:38. 2014.

II. **Zolfaghari Emameh R**, Barker H, Hytönen VP, Tolvanen ME, Parkkila S. " Beta carbonic anhydrases: novel targets for pesticides and anti-parasitic agents in agriculture and livestock husbandry". Parasites & Vectors. 7:403. 2014.

III. **Zolfaghari Emameh R**, Kuuslahti M, Vullo D, Barker HR, Supuran CT, Parkkila S. "Ascaris lumbricoides β carbonic anhydrase: a potential target enzyme for treatment of ascariasis". Parasites & Vectors. 8:479. 2015.

IV. **Zolfaghari Emameh R**, Kuuslahti M, Näreaho A, Sukura A, Parkkila S. "Innovative molecular diagnosis of Trichinella species based on β-carbonic anhydrase genomic sequence". Microbial Biotechnology. 9(2):172-9. 2016.

V. **Zolfaghari Emameh R**, Barker HR, Tolvanen ME, Parkkila S, Hytönen VP. "Horizontal transfer of β-carbonic anhydrase genes from prokaryotes to protozoans, insects, and nematodes". Parasites & Vectors. 9(1):152. 2016.

VI. **Zolfaghari Emameh R**, Barker HR, Parkkila S. "Mobilome genomics: new evidences in horizontal transfer of β-carbonic anhydrase genes from prokaryotes to parasitic and non-parasitic eukaryotes". [Submitted].

Abbreviations

AAZ	Acetazolamide
ACLAME	A classification of mobile genetic elements
AgaCA	*Anopheles gambiae* β-carbonic anhydrase
AlBCA	*Ascaris lumbricoides* β-CA
BLAST	Basic Local Alignment Search Tool
CATH	Protein structure classification database
CCP	Conjugation Complex Protein
cTP	Cytoplasmic Targeting Peptide
CynT	Cyanate-induced carbonic anhydrase
DmBCA	*D. melanogaster* β-CA
EE	Early Embryo
EBI	European Bioinformatics Institute
EMBL	European Molecular Biology Laboratory
EMBOSS	European Molecular Biology Open Software Suite
Gene3D	Structural analysis of whole genes and genomes
GI	Genomic Island
hCA	Human Carbonic Anhydrase
HGT	Horizontal Gene Transfer
InterPro	Protein sequence analysis and classification
IS	Insertion Sequence
JPGV	Jena Prokaryote Genome Viewer
k_{cat}	Turnover number
k_{cat}/K_m	Specificity constant

KEGG	Kyoto Encyclopedia of Genes and Genomes
K_m	Michaelis constant
L1, 2, 3, 4	Larval stage 1, 2, 3, 4
LdcCA	*Leishmania donovani chagasi* β-carbonic anhydrase
LE	Late Embryo
METACYC	Metabolic pathways from all domains of life
MGE	Mobile Genetic Element
MSA	Multiple Sequence Alignment
mTP	Mitochondrial Targeting Peptide
Ni-NTA	Nickel-Nitrilotriacetic Acid
PDB	Protein Data Bank
RuBisCO	Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase
SP	Signal Peptide
TrichDB	*Trichomonas vaginalis* Database
UniProt	Universal Protein Resource
VdW	Van der Waals
VMD	Visual Molecular Dynamics

Abstract

Carbonic anhydrases (CAs) are ubiquitous metalloenzymes, which catalyze the following reaction: $H_2O + CO_2 \leftrightarrow H^+ + HCO_3^-$. CAs are categorized in the six evolutionary divergent gene families α, β, γ, δ, ζ, and η. Metal ions such as zinc (mainly), cobalt, and cadmium (occasionally) play a critical role in the catalytic active site of CAs. While α-CAs have been reported in many prokaryotic and eukaryotic organisms, β-CAs have only been identified in prokaryotes, photosynthetic organisms, protozoans, and invertebrates (insects and nematodes). Hence, there is no report of β-CAs in any vertebrate. γ-CAs are present in some prokaryotes (archaea and bacteria) and eukaryotes (plants and yeasts). In addition, δ- and ζ-CAs were reported in some eukaryotic phytoplanktons, such as *Thalassiosira weissflogii*. Recently, η-CAs were identified and added to the other known CA families. η-CAs are present in *Plasmodium* species (the causative agents of malaria). β-CAs represent a suitable target for the treatment of infectious diseases, because several putative isoforms have been discovered in pathogenic prokaryotes, such as *Salmonella* spp. and eukaryotes, such as *Candida* spp., *Entamoeba histolytica*, and *Ascaris lumbricoides*. β-CAs in many parasitic organisms are attractive and suitable targets for identification, inhibition, diagnostic, and evolutionary studies.

In this study, we used several bioinformatic techniques for identification of new β-CAs from protozoans and metazoans (invertebrates, such as insects and nematodes). We also investigated β-CAs from parasites as potential targets for antibodies or CA inhibitors, and evaluated mobile genetic elements (MGEs) in horizontal transfer of *β-CA* gene sequences from ancestral prokaryotes to eukaryotes, including protozoans, insects, and nematodes. We produced recombinant *Ascaris lumbricoides* (causative agent of ascariasis) β-CA (AlBCA) in sf-9 insect cells and performed kinetic and inhibitory studies on AlBCA using acetazolamide (a well-known CA inhibitor). In addition, we applied a polymerase chain reaction (PCR) method based on β-CA genomic sequences for molecular diagnosis of *Trichinella* (causative agent of trichinellosis) contamination in meat samples.

As a result, 52 protozoan and metazoan β-CA protein sequences were reported for the first time, all of which contained the first (CXDXR) and second (HXXC) highly conserved sequence motifs characteristic of β-CAs in their catalytic active sites. Prediction of subcellular localization revealed that 37 β-CAs are probably mitochondrial proteins. Although prediction of antigenic sites in β-CAs showed that one of the highly conserved β-CA sequences (HXXC) is highly antigenic in parasites. Protein modelling results suggested that this part of the protein, at least in the *Ancylostoma caninum* β-CA, is buried and thus probably inaccessible for the immune

system. In our studies on horizontal gene transfer (HGT), we discovered that many prokaryotic mobilomes, including plasmids and genomic islands (GIs) contain enzymatic accessory tools, such as transposase, integrase, and resolvase, together with *β-CA* gene sequences. The results suggested that *β-CA* genes might have been horizontally transferred from archaeal GIs to other prokaryotic plasmids and then from prokaryotic plasmids to protozoans, insects, and nematodes.

We produced AlBCA as a recombinant protein, which was then subjected to kinetic and inhibitory studies. The kinetic analysis of AlBCA showed a reasonable enzymatic activity, which was comparable to the well-known human carbonic anhydrase I (hCA I) and also in the same range with the recently characterized enzymes present in *Anopheles gambiae* and *Leishmania donovani chagasi.* Acetazolamide, showed an inhibition constant of 84.1 nM against AlBCA.

In another set of experiments, we wanted to investigate whether *β-CA* genes can be used as a marker of parasitic infection. For this purpose, we used a mouse model of trichinellosis to demonstrate the presence of *Trichinella* species in meat samples. The results showed a good sensitivity for our PCR-based detection method. Further analysis showed no false-positive results on other samples laden with parasites including *Toxoplasma gondii*, *Parascaris equorum*, and *Toxocara cati.*

In summary, β-CAs are present in a broad spectrum of protozoans and metazoans, where they might play critical roles in biochemical and metabolic functions. Due to the crucial roles and their absence in vertebrates, β-CAs are considered as brilliant targets for molecular diagnostics and inhibition-based treatment of parasitic infections. These enzymes may also be vitally important for agricultural pests and vectors. Therefore, they can become important molecular targets in the fields of medicine, veterinary medicine, forestry, horticulture, and agriculture. Also, the HGT study of both prokaryotic and metazoan β-CAs, provided novel information about how *β-CA* gene sequences may have horizontally transferred via MGEs.

1. Introduction

Parasites affect about two billion people worldwide (Yones, Galal et al. 2015). The prevalence of some infections per year is beyond imagination, such as malaria (207 million) (Xia, Cai et al. 2015), leishmaniasis (10 million) (de Morais, Castro Lima et al. 2015), ascariasis (760 million) (Betson, Nejsum et al. 2014), and schistosomiasis (240 million) (Bockarie, Kelly-Hope et al. 2013). The devotion of US $10 billion by the G8 Hokkaido Toyako 2008 Summit to control or eliminate high-prevalence tropical neglected diseases within 56 countries (Hotez, Molyneux et al. 2008, Hotez 2010), indicates how much human and financial resources are invested for treatment of these parasitic diseases. Aside from cost, another bottleneck for controlling and treating of parasitic infection is access to safe and effective anti-parasitic agents.

Due to the many current challenges, there is an urgent need to search and discover safe, effective, cheap, and single dose anti-parasitic agents with a broad spectrum of effectiveness on a variety of parasites. Achieving this goal requires identification of novel molecular targets in parasites. Optimally, targeting of these genes or proteins should have no, or only minimal, side effects on the normal physiology of the host. In addition, it would be beneficial if these parasite-specific genes or proteins would serve as targets for diagnostic purposes to detect parasitic infections.

Carbonic anhydrases (CAs) are ubiquitous metalloenzymes, which catalyze the conversion of water and carbon dioxide molecules to a proton and bicarbonate ion in a reversible chemical reaction (Koenig and Brown 1972, Zolfaghari Emameh, Syrjanen et al. 2015):

$$H_2O + CO_2 \leftrightarrow H^+ + HCO_3^-$$

CAs are categorized in six evolutionarily distinct gene families, including α, β, γ, δ, ζ, and η (Elleuche and Poggeler 2010, Del Prete, Vullo et al. 2014). Metal ions, mainly zinc (Zn^{2+}), play a critical role in the catalytic active site of CAs. Cobalt, iron (γ) and cadmium (ζ) cofactors are located with lower frequency in the catalytic active sites of CAs (Lane, Saito et al. 2005, De Simone, Di Fiore et al. 2015). α-CAs were reported in many prokaryotes and eukaryotes (Tripp, Smith et al. 2001, Supuran 2008, Elleuche and Poggeler 2009, Elleuche and Poggeler 2010, Huang, Hainzl et al. 2011, Del Prete, Vullo et al. 2014, Zolfaghari Emameh, Syrjanen et al. 2015). β-CAs are expressed in prokaryotes (bacteria and archaea) and some eukaryotes (yeast, fungi, algae, and plants). Some organisms, including prokaryotes (bacteria and archaea) and eukaryotes (plants) contain γ-CAs. δ- and ζ-CAs are expressed in some

eukaryotic marine organisms, such as phytoplankton. The newly reported η-CAs were recently identified in the eukaryotic malaria protozoa, *Plasmodium falciparum* (Del Prete, Vullo et al. 2014). In total, 13 enzymatically active α-CA enzymes have been identified in mammals, including cytosolic (CA I, CA II, CA III, CA VII, and CA XIII), membrane-bound (CA IV, CA IX, CA XII, CA XIV, and CA XV), mitochondrial (CA VA and CA VB); secreted (CA VI). Additionally there are three acatalytic CA-related proteins in mammals (CA VIII, CA X, and CA XI) (Esbaugh and Tufts 2006, Supuran 2008).

CAs play crucial roles in several metabolic pathways and biological functions including nitrogen metabolism, gluconeogenesis, ureagenesis, lipogenesis, calcification, bone resorption, pH homeostasis, and electrolyte transfer (Vullo, Franchi et al. 2004, Jaiswal, Prasanna et al. 2005, Vullo, Innocenti et al. 2005, Alterio, Vitale et al. 2006, Nishimori, Minakuchi et al. 2007, Supuran 2008).

Many putative β-CAs have been discovered since the 1990s, not only in photosynthetic organisms, but also in archaea, bacteria, and yeasts (Rowlett 2010). β-CA enzymes have been identified in a number of pathogenic bacteria, such as *Helicobacter pylori, Mycobacterium tuberculosis, Salmonella typhimurium* (*Salmonella enterica,* subspecies *enterica,* serovar *typhimurium*) (Smith and Ferry 1999, Rowlett 2010), *Haemophilus influenzae* (Rowlett, Tu et al. 2009, Supuran 2011), *Brucella suis* (Joseph, Turtaut et al. 2010, Supuran 2011, Tobal and Balieiro 2014), *Streptococcus pneumoniae* (Burghout, Vullo et al. 2011, Supuran 2011, Tobal and Balieiro 2014), *Salmonella enterica* (Supuran 2011, Vullo, Nishimori et al. 2011), and *Vibrio cholerae* (Abuaita and Withey 2009, Kovacikova, Lin et al. 2010, Supuran 2011). β-CAs have also been identified in fungi, such as *Candida albicans* (Klengel, Liang et al. 2005, Elleuche and Poggeler 2010, Tobal and Balieiro 2014), *Candida glabrata* (Innocenti, Leewattanapasuk et al. 2009, Elleuche and Poggeler 2010, Tobal and Balieiro 2014), *Cryptococcus neoformans* (Bahn, Cox et al. 2005, Elleuche and Poggeler 2010, Tobal and Balieiro 2014), *Sordaria macrospora* (Elleuche and Poggeler 2008, Elleuche and Poggeler 2009), and *Saccharomyces cerevisiae* (Cleves, Cooper et al. 1996, Gotz, Gnann et al. 1999, Amoroso, Morell-Avrahov et al. 2005). β-CAs have also been reported in cyanobacteria (*Synechocystis* sp. PCC6803) (So and Espie 1998), carboxysomes of chemoautotrophic bacteria (*Halothiobacillus neapolitanus*) (Sawaya, Cannon et al. 2006), green algae (*Chlamydomonas reinhardtii*) (Eriksson, Karlsson et al. 1996), and red algae (*Porphyridium purpureum*) (Mitsuhashi, Mizushima et al. 2000).

Two metazoan (invertebrate) β-CAs from two different model organisms *Caenorhabditis elegans* and *Drosophila melanogaster* were reported in 2010. During the study of *D. melanogaster* β-CA (DmBCA), 55 β-CA orthologs were detected in different invertebrate species, including *Aedes aegypti, Culex quinquefasciatus, Anopheles gambiae, Drosophila virilis, Tribolium castaneum, Nasonia vitripennis, Apis mellifera, Acyrthosiphon pisum, Daphnia pulex, Caenorhabditis elegans, Pristionchus pacificus, Trichoplax adhaerens, Caligus clemensi, Lepeophtheirus salmonis, Nematostella vectensis, Strongylocentrotus purpuratus,* and *Saccoglossus kowalevskii.* The absence of *β-CA* gene from the genomes of vertebrates has made it an attractive target for evolutionary and drug design studies (Fasseas, Tsikou et al. 2010, Syrjanen, Tolvanen et al. 2010).

The presence of β-CAs in a number of metazoan species encouraged us to study further β-CAs in parasites and evaluate these parasite-specific genes and proteins as potential targets for diagnostics and treatment of parasitic infections.

2. Review of the literature

2.1. Overview of carbonic anhydrases (CAs)

CAs are ubiquitous metalloenzymes, which are encoded by six evolutionary distinct gene families: α, β, γ, δ, ζ, and η (Supuran 2016). Active site evolution of each CA enzyme family has occurred independently (Banerjee and Deshpande 2016). Amino acid composition of the CA catalytic active site determines the functional properties of each enzyme including enzyme kinetics and binding of different metal ions, such as zinc, cobalt, iron, or cadmium. Despite the different amino acid structure and reactive metal ion of the six CA families, all of them catalyse the same reaction to produce bicarbonate and proton, and ultimately support various biochemical functions in all living organisms.

In the next sections, comprehensive details for each CA enzyme family are presented to define the vital role of CAs in various physiological functions.

2.1.1. α-CAs

Since 1972, dozens of research projects have been conducted on human and mouse α-CAs (Table 2). All vertebrate α-CAs except for CA XV are expressed in humans. The mRNA expression study of mouse CA XV revealed that this enzyme is expressed in the mouse kidney (Hilvo, Tolvanen et al. 2005). Previous studies have demonstrated that expression of α-CAs is vital for human health (Sly, Hewett-Emmett et al. 1983, Rebello, Ramesar et al. 2004, Ogilvie, Ohlemiller et al. 2007, Feldshtein, Elkrinawi et al. 2010, van Karnebeek, Sly et al. 2014). Human α-CAs are expressed in many organs (Table 1).

Table 1. α-CAs from selected human organs.

Organs	Some representative references
Bone	(Vaananen 1984, Chang, Zheng et al. 2012)
Brain	(Kumpulainen and Nystrom 1981, Parkkila, Parkkila et al. 2001, Kida, Palminiello et al. 2006)
Ear	(Weber, Cunningham et al. 2001)
Eye	(Wolfensberger, Mahieu et al. 1994)
Gastrointestinal canal	(Parkkila, Parkkila et al. 1994)
Kidney	(Spicer, Sens et al. 1982, Parkkila, Parkkila et al. 2000, Purkerson and Schwartz 2007)
Liver	(Spicer, Sens et al. 1982, Harju, Bootorabi et al. 2013)
ovary and uterus	(Friedley and Rosen 1975, Catala 1997)
Pancreas	(Kumpulainen and Jalovaara 1981, Fujikawa-Adachi, Nishimori et al. 1999, Inada, Nienaber et al. 2008)
Salivary glands	(Noda, Sumitomo et al. 1986, Parkkila, Kaunisto et al. 1990)
Skeletal muscle	(Vaananen, Paloniemi et al. 1985, Sender, Decker et al. 1998, Geers and Gros 2000, Harju, Bootorabi et al. 2013)
Skin	(Mastrolorenzo, Zuccati et al. 2003)
Testis	(Mezquita, Mezquita et al. 1999)
Thyroid gland	(Hori, Yoshida et al. 1998)

α-CAs are not only present in mammals but also in many other prokaryotic and eukaryotic organisms, including bacteria, archaea, diatoms, fungi, algae, plants, protozoa, arthropods, and nematodes (Premkumar, Bageshwar et al. 2003, Fisher, Tariku et al. 2006, Elleuche and Poggeler 2009, Tachibana, Allen et al. 2011, Sherman, Rongali et al. 2012, Zabaleta, Martin et al. 2012, Kumar and Ferry 2014). Typically, histidine residues play a crucial role in the catalytically active sites of α-CAs (Tripp, Smith et al. 2001). The catalytic active site of some diatom α-CAs may not contain histidine residues (Tachibana, Allen et al. 2011). In this case, other substituting residues can coordinate the zinc ion within the active site cavity. Phylogenetic studies revealed that most of the *a-CA* genes in two diatom species, *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*, have been laterally transferred from green and red algal plastids through endosymbiosis, whereas the genes are distinct compared to those of brown algae (Moustafa, Beszteri et al. 2009). α-CAs from *P. tricornutum* are localized in the chloroplastic surrounding membranes (plastid), in order to conduct a critical role in the CO_2 concentrating mechanism, as well as periplastidial compartment (PPC) and endoplasmic reticulum (ER)

(Tachibana, Allen et al. 2011). A *T. pseudonana* α-CA (CA-1) is responsible for decreasing CO_2 concentration and works under CO_2-limited conditions.

Table 2. The list of human α-CAs and mouse CA XV.

α-CA	Host	Cellular localization	Year of discovery	References
CA I	Human	Cytoplasmic	1972	(Andersson, Nyman et al. 1972)
CA II	Human	Cytoplasmic	1972	(Liljas, Kannan et al. 1972)
CA III	Human	Cytoplasmic	1978	(Carter, Shiels et al. 1978)
CA IV	Human	Membrane-bound	1989	(Wistrand and Knuuttila 1989)
CA VA	Human	Mitochondrial	1993	(Nagao, Platero et al. 1993)
CA VB	Human	Mitochondrial	1999	(Fujikawa-Adachi, Nishimori et al. 1999)
CA VI	Human	Secretory	1987	(Murakami and Sly 1987)
CA VII	Human	Cytosolic	1991	(Montgomery, Venta et al. 1991)
CA IX	Human	Transmembrane	1996	(Opavsky, Pastorekova et al. 1996)
CA XII	Human	Transmembrane	1998	(Tureci, Sahin et al. 1998)
CA XIV	Human	Transmembrane	1999	(Fujikawa-Adachi, Nishimori et al. 1999)
CA XIII	Human	Cytosolic	2004	(Lehtonen, Shen et al. 2004)
CA XV	Mouse	Membrane-bound	2005	(Hilvo, Tolvanen et al. 2005)

2.1.2. β-CAs

In previous decades many CAs were identified for the first time and recorded in various databases. The first instance of CA activity in chloroplasts was discovered in *Trifolium pretense* (red clover) and *Onoclea sensibilis* (bead fern) by Neish in 1939 (Neish 1939), prokaryotic β-CA from *Neisseria sicca* (Veitch and Blankenship 1963), fungal β-CA from yeast *S. cerevisiae* (Cleves, Cooper et al. 1996), green algal β-CA from *Chlorophyta* spp., red algal β-CA from *Rhodophyta* spp. (Bowes 1969), first diatom β-CAs (PtCA1 and PtCA2) from *Phaeodactylum tricornutum* (Satoh, Hiraoka et al. 2001, Tanaka, Nakatsuma et al. 2005), first protozoan β-CA from *Leishmania donovani chagasi* (Syrjanen, Vermelho et al. 2013), first arthropod β-CA from *D. melanogaster* (Syrjanen, Tolvanen et al. 2010), and first nematode β-CA from *C. elegans* (Fasseas, Tsikou et al.

2010). Through the evolution of CA enzymes, the β-CA gene has never integrated into the genome of vertebrates (Syrjanen, Tolvanen et al. 2010).

Studies on the cDNA sequence of spinach (*Spinacea oleracea*) chloroplastic β-CA in 1990 found that this sequence is non-homologous to animal α-CAs (Burnell, Gibbs et al. 1990, Rowlett 2010). It is believed that the plant β-CAs are localized within thylakoid space of chloroplastic stroma. β-CA is an important accessory enzyme for many CO_2 or HCO_3^- utilizing enzymes e.g. RuBisCO (Ribulose-1,5-bisphosphate carboxylase/oxygenase) in chloroplasts, cyanase in *Escherichia coli* (Guilloton, Lamblin et al. 1993), urease in *Helicobacter pylori* (Nishimori, Onishi et al. 2008), and carboxylases in *Corynebacterium glutamicum* (Mitsuhashi, Ohnishi et al. 2004)). In cyanobacteria, β-CA is an essential component of the CO_2-concentrating carboxysome organelle (Fukuzawa, Suzuki et al. 1992, Rowlett 2010). β-CA activity is required for growth of *E. coli* bacteria in air (Merlin, Masters et al. 2003); it is also indispensable if the atmospheric partial pressure of CO_2 is high or during anaerobic growth in a closed vessel at low pH, where copious CO_2 is generated endogenously. β-CA is also needed for growth of *C. glutamicum* (Majeau and Coleman 1996, Mitsuhashi, Ohnishi et al. 2004) and some yeasts, such as *Saccharomyces cerevisiae* (Mitsuhashi, Mizushima et al. 2000). The genomes of higher plants, such as *Flaveria bidentis,* contain at least three β-CA genes: *CA1, CA2,* and *CA3* (Tetu, Tanz et al. 2007). C3 and C4 plants have different mechanisms for carbon fixation and photosynthesis and, thus, β-CAs might possess different roles, depending on the location of the enzyme and the type of plant (Ludwig 2011). For example, C3 and C4 perform photosynthesis in distinct *Flaveria* (yellowtops) species, such as C3 photosynthesis in *F. pringlei* and C4 photosynthesis in *F. trinervia.* Also, *F. pubescens* applies a combination of C3-C4 in the intermediate metabolism pathway of photosynthesis. Paying more attention to the role of C3 and C4 in *Flaveria* species revealed that some mutations led photosynthesis from C3 to C4. During this evolutionary process, non-photosynthetic phosphoenolpyruvate carboxylase from C3 was evolved to a C4 photosynthetic enzyme, which enhances the kinetic yield of the enzyme and diminished the inhibitory effects of malate and aspartate. In C3 photosynthesis is performed by the RuBisCO enzyme, which leads to formation of 3-phosphoglycerate, while the C4 photosynthesis mechanism prevents the loss of energy by using extra CO_2 concentrating mechanism. This additional mechanism enables plants to adapt to environmental factors, including high light and heat. C4 uses phosphoenolpyruvate carboxylase (PEPC) to form oxaloacetate from carboxylation of phosphoenolpyruvate (PEP), which is a bicarbonate-dependent mechanism. Then oxaloacetate passes two pathways: transformation to aspartate or reduction to malate. Finally, a high concentration of carbon dioxide is prepared at the active site of RuBisCO. Therefore, the rate and efficiency of photosynthesis are raised by application of other factors such as water, soil minerals, and nitrogen (Paulus, Schlieper et al. 2013).

In plants, there are some CA activities in the cytosol of mesophyll cells and the highest β-CA level is found in the chloroplastic stroma (Zabaleta, Martin et al. 2012). A study on two β-CAs (CAH4 and CAH5) from single cell green algae *Chlamydomonas*

reinhardtii demonstrated that the enzymes supply CO_2 to tricarboxylic acid or citric acid cycle (Eriksson, Villand et al. 1998, Giordano, Norici et al. 2003). Electron microscopic studies have localized two β-CAs of phytoplankton *P. tricornutum* (PtCA1 and PtCA2) to the pyranoid sub-cellular micro-compartments in the chloroplast or plastid, which are the specific zones for accumulation of RuBisCO enzyme (Matsuda, Nakajima et al. 2011, Tachibana, Allen et al. 2011). Environmental CO_2 must be rapidly hydrated by β-CA and converted into HCO_3^- for the phosphoenolpyruvate carboxylase enzyme. β-CAs play a major role in photosynthesis in the chloroplast by catalyzing HCO_3^- dehydration to supply CO_2 for RuBisCO (Majeau and Coleman 1996). Interestingly, both RuBisCO and β-CA expression levels increase together when *Pisum sativum* is transferred from an environment with high levels of CO_2 to one with low levels. Crystal structures of β-CAs have revealed that a Zn^{2+} is ligated by two conserved cysteines and one conserved histidine (Tripp, Smith et al. 2001). The molecular masses of β-CA range from 45 to 200 kDa, depending on the state of oligomerization (Mitsuhashi, Mizushima et al. 2000). The first β-CA from arthropods was reported from *D. melanogaster*. The recombinant β-CA from *D. melanogaster* (DmBCA) was produced in Sf-9 insect cells, and its kinetic and inhibition profiles were determined. CO_2 was hydrated with a k_{cat} of 9.5×10^5 s^{-1} and a k_{cat}/K_m of 1.1×10^8 $M^{-1}s^{-1}$. Also, DmBCA was inhibited by acetazolamide, (inhibition constant of 49 nM). Green fluorescent labeled *D. melanogaster* β-CA (DmBCA-GFP) method and subcellular localization prediction of DmBCA showed it as an active CA in the mitochondria (Syrjanen, Tolvanen et al. 2010, Syrjanen, Parkkila et al. 2014).

2.1.3. γ-CAs

γ-CAs may contain Fe^{2+} (iron) or Co^{2+} (cobalt) ions, instead of Zn^{2+} (zinc), in their catalytic active site (Ferry 2010). Organisms containing γ-CAs include many plants, algae, bacteria, and archaea. The first plant γ-CA was studied in *Arabadopsis thaliana* (mouse-ear cress) in 2004 (Parisi, Perales et al. 2004).

Crystal structure analysis of two archaeon *Methanosarcina thermophila* and *Pyrococcus horikoshii* γ-CAs described a homotrimeric structure with parallel left-handed β-sheet structures (Ferry 2010). Arg59 is essential for the stability and activity of the trimeric γ-CAs. Three conserved histidines, from the monomer constituents of trimeric γ-CAs, are attached to a metal ion of the catalytic active site by a network of hydrogen bonds. The active site of γ-CA includes three histidines, including His81 and His112 from one enzyme monomer and His117 from another monomer, which together coordinate to the metal ion (Tripp, Smith et al. 2001). γ-CA from some diatoms have no histidine residues in their catalytic active sites (Parisi, Perales et al. 2004). Instead of histidines, these enzymes have other alternative residues to coordinate with the Zn^{2+} cofactor ion (Iverson, Alber et al. 2000). *A. thaliana* γ-CA contains highly

conserved residues His81, His117, His122, Arg59, Asp61, Gln75, and Asp76, which are in close contact with the enzyme's catalytic active site. A hydrophobic pocket at the surface of γ-CAs may function as a CO_2 trapper for the enzyme. Previous experimental studies defined that γ-CA from *A. thaliana* is encoded by a nuclear gene and imported into mitochondria. Prediction of sub-cellular localization revealed that two γ-CAs from marine diatom *P. tricornutum* are also localized in the mitochondria (Tachibana, Allen et al. 2011).

2.1.4. δ-CAs

The first δ-CA (TWCA1) was isolated from the marine diatom *Thalassiosira weissflogii* (Tripp, Smith et al. 2001). Although the function of this enzyme is probably identical to the other CA families, the analysis of coding genes of different families indicated that *δ-CA* gene is different from α-, β-, and γ-CAs. Therefore, it has been categorized as the new CA class containing a cadmium cofactor.

The distribution pattern of Cd^{2+} ion in marine environments has revealed that this metal ion is taken up by phytoplanktons from surface ocean water and used as metal cofactor for δ-CAs of diatoms (Xu, Feng et al. 2008). In phytoplankton, δ-CA catalyses the hydration of CO_2 for photosynthesis. It has been demonstrated that TWCA1 contains a signal peptide responsible for targeting the enzyme into the chloroplast as the final localization site.

2.1.5. ζ-CAs

ζ-CA is another distinct CA class, containing Cd^{2+} (cadmium) ion in the enzyme's catalytic active site (Lane and Morel 2000, Xu, Feng et al. 2008). The first isoenzyme of this class was isolated from marine diatom, *T. weissflogii,* in 2000 for the first time. The study showed that ζ-CA is able to exchange metal ions automatically and use Zn^{2+} or Cd^{2+} depending on the availability of the ions. The study also showed that the non-zinc ζ-CA is less efficient than the zinc-containing enzyme.

Diatom *T. pseudonana* cytosolic ζ-CA is localized in the periplasmic space (Tachibana, Allen et al. 2011). Under CO_2-limited conditions, cytosolic diatom ζ-CA (CA-7) is suggested to aquire CO_2 from seawater to compensate the concentration of HCO_3^- for pH regulation.

2.1.6. η-CAs

For many years, it was believed that parasites contain only α-CAs. Recently, a phylogenetic and sequence analysis of malaria protozoa, *P. falciparum*, revealed a new distinct coding gene sequence for CA (Del Prete, Vullo et al. 2014). The encoded protein was nominated as η-CA.

P. falciparum η-CA contains histidine residues including His94, His96, and His118, which match with the catalytic active site residues of human α-CAs. The zinc ion coordination pattern of η-CA is, however, slightly different from human α-CAs. These recent studies have demonstrated that even one of the "classical" histidines, His94, has been replaced in murine malaria protozoa by Asn (aspargine), which is not able to coordinate to Zn^{2+}. In human α-CAs, His64 accepts a proton from an H_2O molecule in the so called "proton shuttle", which is located along the pathway to the active site cavity of the enzyme. This critical residue is not available in η-CAs as the major difference with α-CAs. Also, two other critical amino acids including Glu106 (glutamic acid) and Thr199 (threonine) are available in α-CAs, but absent in η-CAs. Sequence analysis identified that the full sequence length of η-CAs (>400 amino acids) is longer than α-CAs (250-280 amino acids). In spite of these differences, η-CAs seem to be active CAs and are now considered a structurally distinct family of CAs (Del Prete, Vullo et al. 2014, De Simone, Di Fiore et al. 2015).

2.2. CAs in protozoan and metazoan species

Several previous studies have focused on different families of CAs in free-living and parasitic protozoans and metazoans, which were shown in Table 3.

Table 3. The list of CA-containing free-living and parasitic protozoans and metazoans.

CA containing protozoans and metazoans	CA class	Function	References
Acanthamoeba castellanii (free-living protozoa)	γ	-	(Gawryluk and Gray 2010)
Anopheles gambiae (malaria mosquito)	β	-	(Syrjanen, Kuuslahti et al. 2015)
Apis mellifera (honeybee)	-	Regulation of acid-base balance in brain and ionic homeostasis and metabolism in retinal glial cells	(Walz 1988)
Aplysia californica (California sea hare)	-	pH regulation in statocysts of statoconia	(Pedrozo, Schwartz et al. 1995)
Atta capiguara (leaf-cutter ant)	-	-	(Cantagalli, Mangolin et al. 2010)
Branchiostoma lanceolatum (lancelet)	α	pH regulation in the gut cells of youngs and skin cells of adults	(Pederzoli, Mandrioli et al. 2014)
Caenorhabditis elegans (a free-living nematode)	β	Development and stress conditions	(Fasseas, Tsikou et al. 2010)
Cherax quadricarinatus (Australian red claw crayfish)	α and β	pH regulation in the gill	(Ali, Pavasovic et al. 2015)
Daphnia pulex (water flea)	α and β	-	(Colbourne, Pfrender et al. 2011, Culver and Morton 2015)
Dictyostelium discoideum (soil-living amoeba)	-	Preparation of protons for ATPase activity in the contractile vacuoles	(Marchesini, Ruiz et al. 2002)
Drosophila melanogaster (fruit fly)	β	In gonads and fertility of female fruit fly	(Syrjanen, Tolvanen et al. 2010, Syrjanen, Valanne et al. 2015)
Hirudo medicinalis (leech)	-	Preparation of bicarbonate across the glial cells	(Riehl and Schlue 1990)
Leishmania donovani chagasi (leishmaniasis protozoa)	β	-	(Syrjanen, Vermelho et al. 2013)
Nephila clavipes (banana spider)	-	pH gradient in silk gland	(Andersson, Chen et al. 2014)
Ostertagia ostertagi (brown stomach worm)	α	Development in early stage of exsheathment	(DeRosa, Chirgwin et al. 2008)
Periplaneta americana (American cockroach)	-	Preparation of protons for ATPase activity in the salivary gland cells	(Just and Walz 1994)
Plasmodium falciparum (malaria protozoa)	α and η	-	(Krungkrai, Scozzafava et al. 2005, Del Prete, Vullo et al. 2014)
Plutella Xylostella (diamondback moth)	α	-	(Cheng, Du et al. 2015)
Saccoglossus kowalevskii (acorn worm)	-	Biomineralization by using calcium carbonate	(Cameron and Bishop 2012)
Schistosoma spp. (blood-flukes)	α	pH regulation	(Castro-Borges, Dowle et al. 2011)
Spodoptera littoralis (African Cotton Leafworm)	-	Preparation of protons for ATPase activity in the male reproductive tract	(Kotwica, Ciuk et al. 2006)
Strongylocentrotus purpuratus (purple sea urchin)	-	Biomineralization by using calcium carbonate	(Mann, Wilt et al. 2010)
Tribolium castaneum (red flour beetle)	α	Detoxification of high concentration of carbon dioxide and preparation of highly toxic concentration of carbonyl sulphate	(Haritos and Dojchinov 2005)
Trypanosoma cruzi (chagas disease protozoa)	α	-	(Pan, Vermelho et al. 2013)

-: not defined.

These reports mostly cover information about protozoan and metazoan α-CA family, while they also describe a few β-CAs in *D. melanogaster*, *C. elegans*, *A. gambiae*, and *L. donovani chagasi*. In the following paragraphs, some detailed information is presented on protozoan and metazoan β-CAs, including the enzymes of some insects and nematodes.

2.2.1. *D. melanogaster*

The first β-CA from insects was identified and characterized from *D. melanogaster* in 2010 (Syrjanen, Tolvanen et al. 2010). Phylogenetic studies showed sequence similarity to plant-type β-CAs, and the enzyme was named as DmBCA. To further characterize this enzyme, it was produced as a recombinant protein by transfection of sf-9 (*Spodoptera frugiperda*) insect cells. Signal peptide prediction of the DmBCA suggested that it may be targeted to mitochondria. Analysis by the DmBCA-GFP method supported the mitochondrial subcellular location prediction. Publicly available DNA microarray datasets revealed the highest β-CA mRNA levels in spermatheca (female), fat body, and heart of adults. A time course study showed upregulated levels in early larval stages and late in metamorphosis. The specific expression pattern of DmBCA suggested a tight regulation of expression and a specific functional role during development. Recent studies on *D. melanogaster β-CA* gene silencing by RNAi (Ribonucleic acid interference) technique showed a delayed migration of ovarian border cells and infertility of females (Syrjanen, Valanne et al. 2015). This result indicated that β-CA is a prerequisite for normal ovarian function and fertilization in female *D. melanogaster*. Therefore, inhibition of this function could be an attractive target for restricting the spread of different parasites and vectors.

An inhibitory study with various sulfonamide derivatives including benzolamide (BZA), brinzolamide (BRZ), celecoxib (CLX), diuretic hydrochlorothiazide (HCT), dorzolamide (DZA), indisulam (IND), sulpiride (SLP), and valdecoxib (VLX) showed weak inhibitory action on DmBCA (Hilvo, Salzano et al. 2009, Syrjanen, Parkkila et al. 2014). Some other inhibitors, such as acetazolamide (AAZ) and dichlorophenamide (DCP), showed moderate inhibitory action on DmBCA. The most effective DmBCA inhibitors recorded to date include methazolamide (MZA), ethoxzolamide (EZA), and sulthiazide (SLT).

2.2.2. *C. elegans*

The first β-CA from nematodes was identified and characterized in *C. elegans* in 2010. The *C. elegans* β-CA was retrieved by searching through bioinformatics databases. Studies revealed that *C. elegans* contains two different *β-CA* genes, *y116a8c.28* and *bca-*

1, which were located on chromosomes IV and X, respectively. The three dimensional comparison of these enzymes with *P. sativum* and *Methanobacterium thermoautotrophicum* β-CAs showed a greater identity to bacterial β-CA than plant-type. Under normal conditions, three larval stages of *C. elegans*, L1 to L3, showed the highest expression levels for both *β-CA* genes. Although *bca-1* gene showed overexpression during lack of CO_2, *y116a8c.28* gene followed a normal expression pattern. Also, both genes actually behave in a quite similar fashion under two diffrent experimental conditions, low expression at two hours and a little higher at eight hours, which suggested that the expression controlling factors for both *β-CA* genes are different. In addition, knockdown of either gene by RNAi technique produced no visible phenotype, suggesting that other CA families such as *C. elegans* α-CAs compensated the lack of β-CAs (Fasseas, Tsikou et al. 2010).

2.2.3. *A. gambiae* (malaria mosquito)

The first characterization and inhibition study on β-CA from *A. gambiae* (AgaCA) was performed in 2015. The enzyme was found to be fairly similar to the orthologous enzymes of *C. elegans* and *D. melanogaster*. The recombinant *A. gambiae* β-CA was produced in sf-9 insect cells. Inhibitory studies showed that AAZ inhibits AgaCA with higher efficiency than DmBCA or hCA I. Inhibitory studies on AgaCA further revealed that sulfamate topiramate (TPM) had no inhibitory efficacy. Among various sulfonamide derivatives, AAZ, BRZ, BZA, HCT, and saccharin (SAC) showed the most effective inhibitory action on AgaCA (Syrjanen, Kuuslahti et al. 2015).

2.2.4. *L. donovani chagasi* (leishmaniasis protozoa)

The *L. donovani chagasi β-CA* (LdcCA) gene was recently cloned. The catalytic activity of the purified LdcCA was determined by CO_2 hydration reaction, and inhibitory effects of some 5-mercapto-1,3,4-thiadiazoles and sulfonamides were assayed against the recombinant enzyme. The results showed that inhibitory efficacy of some heterocyclic thiols was higher than sulfonamides. Importantly, some of the investigated thiols efficiently inhibited the *in vivo* growth of *L. chagasi* and *L. amazonensis* promastigotes, by impairing the flagellar pocket and movement of the parasites and causing their death. Therefore, β-CA inhibitors were proposed as potential drug candidates to fight against leishmaniasis infection (Syrjanen, Vermelho et al. 2013).

2.3. Diagnosis of meat-borne parasites

Some protozoans and helminths are considered as the causative agents of zoonotic parasitic infections in humans around the world, especially in poor and third world countries. Control for the presence of meat-borne parasites of raw and processed meat products is a high priority of quality assessments at both slaughterhouses and food companies. In recent decades, with the development of molecular biology techniques and medical engineering, new molecular diagnostic tools have been developed to detect meat-borne parasites. These replace conventional time-consuming methods, including direct microscopic detection (Jenkins, Castrodale et al. 2013). The major meat-borne parasites, and the corresponding transmitted infections, are mentioned in the Table 4.

Table 4. The list of major meat-borne parasites.

Meat-borne parasites	Phylum	Class	Infection	References
Alaria alata (digestive tract trematode)	Platyhelminthes	Trematoda	Alariosis	(Tabaran, Sandor et al. 2013, Takeuchi-Storm, Al-Sabi et al. 2015)
Anisakis simplex (seafood roundworm)	Nematoda	Secernentea	Anisakiasis	(Armentia, Martin-Gil et al. 2006)
Ascaris lumbricoides (small roundworm)	Nematoda	Secernentea	Ascariasis	(Schneider, Obwaller et al. 2015)
Clonorchis sinensis (human liver fluke)	Platyhelminthes	Trematoda	Clonorchiasis	(Hong and Fang 2012)
Cryptosporidium spp. (respiratory and intestinal protozoa)	Apicomplexa	Conoidasida	Cryptosporidiosis	(Johnston, Ballard et al. 2003)
Dicrocoelium dendriticum (lancet liver fluke)	Platyhelminthes	Trematoda	Dicrocoeliosis	(Otranto and Traversa 2002)
Diphyllobothrium spp. (fish tapeworm)	Platyhelminthes	Cestoda	Diphyllobothriosis	(Waki, Oi et al. 1986)
Echinococcus spp. (hydatid worm or dog tapeworm)	Platyhelminthes	Cestoda	Echinococcosis	(Eckert and Deplazes 2004)
Entamoeba histolytica (anaerobic parasitic protozoa)	Sarcomastigophora	Lobosea	Amebiasis	(Tanyuksel and Petri 2003)
Fasciola hepatica (common liver fluke or sheep liver fluke)	Platyhelminthes	Trematoda	Fascioliasis	(Rapsch, Schweizer et al. 2006)
Giardia spp. (anaerobic flagellated parasitic protozoa)	Sarcomastigophora	Zoomastigophora	Giardiasis	(Johnston, Ballard et al. 2003)
Gnathostoma spp. (parasitic roundworm)	Nematoda	Secernentea	Gnathostomiasis	(Graeff-Teixeira, da Silva et al. 2009)
Hymenolepis nana (dwarf tapeworm)	Platyhelminthes	Cestoda	Hymenolepiasis	(Steinmann, Cringoli et al. 2012)
Opisthorchis viverrini (Southeast Asian liver fluke)	Platyhelminthes	Trematoda	Opisthorchiasis	(Worasith, Kamamia et al. 2015)
Paragonimus spp. (lung fluke)	Platyhelminthes	Trematoda	Paragonimiasis	(Slesak, Inthalad et al. 2011)
Pentastomida (*Armillifer armillatus, A. grandis, A. moniliformis, Linguatula serrata,* and *Porocephalus crotali*) (tongue worms)	Arthropoda	Maxillopoda	Pentastomidiasis	(Tappe and Buttner 2009)
Sarcocystis spp. (parasitic protozoa)	Apicomplexa	Conoidasida	Sarcocystosis	(Bunyaratvej, Unpunyo et al. 2007)
Spirometra spp. (pseudophyllid tapeworm)	Platyhelminthes	Cestoda	Sparganosis	(Yamasaki, Nakamura et al. 2014)
Taenia spp. (beef and pork tapeworm)	Platyhelminthes	Cestoda	Taeniasis	(Mayta, Gilman et al. 2008)
Toxocara spp. (dog and feline roundworm)	Nematoda	Secernentea	Toxocariasis	(Khademvatan, Abdizadeh et al. 2014)
Toxoplasma gondii (obligate intracellular protozoa)	Apicomplexa	Conoidasida	Toxoplasmosis	(Udonsom, Buddhirongawatr et al. 2010)
Trichinella spp. (pork roundworm)	Nematoda	Adenophorea	Trichinellosis	(Nockler, Pozio et al. 2000)

2.4. Horizontal gene transfer (HGT)

Horizontal gene transfer (HGT) refers to the movement of genetic sequences across normal mating barriers between phylogenetically distinct organisms, and as the result stands in distinction to the vertical transmission of genes from a parent to offspring. HGT is considered a very important evolutionary mechanism (Andersson 2005). Among eubacteria, the HGT process is involved in the evolution of pathogenicity, metabolic pathways, and antibiotic resistance (Boucher, Douady et al. 2003). HGT is conducted via three different pathways in prokaryotes: conjugation, transformation, and transduction, which are mediated by sex pili, natural competent systems, and bacteriophages respectively (Takeuchi, Kaneko et al. 2014).

In conjugation, bacteria containing F plasmid (F^+) horizontally transfers a strand of DNA via a sex pilus to bacteria lacking F plasmid (F^-) (Ghigo 2001). Subsequently, F^- bacteria are changed to F^+ bacteria. Induction of biofilm formation by conjugative factors is an example, which was studied in *E.coli* K12 and TG1. Natural transformation was characterized for the first time in *H. pylori* (Nedenskov-Sorensen, Bukholm et al. 1990), which was followed by a transformation study of *comB* genes (Hofreuter, Odenbreit et al. 1998). Later, this DNA transfer method was reported in many gram-negative and -positive bacteria, such as transformation of cell-cell signalling and nutritional regulatory factors in *S. pneumoniae, Bacillus subtilis, Neisseria gonorrhoeae,* and *Haemophilus influenzae* (Solomon and Grossman 1996). In transduction, different genetic factors are transferred from phages to prokaryotes. These genes encode, for example, extracellular enzymes diphtheria toxin from *Corynbacterium diphtheriae* and neurotoxin from *Clostridium botulinum* (Brussow, Canchaya et al. 2004).

Mobile genetic elements (MGEs) are segments of DNA, encoding proteins such as enzymes, which conduct the movement of DNA fragments in HGT between genomes (Frost, Leplae et al. 2005). Mobilomes include plasmids, genomic islands (GIs), and transposable elements, and are the backbone of the horizontal transfer of a specific gene, or genetic clusters, often using site-specific recombinase or integrase and transposase (Osborn and Boltner 2002). GIs are large genomic sequences, which have been obtained through HGT or are capable of being transferred to other genomic structures (Langille, Hsiao et al. 2010). The presence of individual integrase and transposase genes, as well as IS elements and direct repeats (DRs) adjacent to a *tRNA* gene, are the main charachteristics of GIs. Based on the mode of transposition, transposable elements are divided into two major categories: retrotransposons (class I elements) and DNA transposons (class II elements). Retrotransposons apply a copy-past method to transpose a gene through an RNA intermediate, while DNA transposons use a cut-paste method to transpose a gene directly as DNA (Munoz-Lopez and Garcia-Perez 2010). On the other hand, insertion elements (IS elements) are an important part of prokaryotic and eukaryotic genomes, which originated from retroviruses and bacteriophages and encode

transposase and integrase (Mahillon and Chandler 1998). IS elements contain left inverted repeat (IRL), right inverted repeat (IRR), and single or double open reading frames (ORFs) for coding transposase and integrase. IS elements contain a promoter for transposase or integrase ORFs, which is located in an IRL. In fact, gene copying or cutting from a target genetic structure and pasting in another genetic location are the main functions of IS element-encoded enzymes.

Transposases and site-specific recombinases mediate the intracellular movement of MGEs (Thorpe and Smith 1998). Site-specific recombinases in bacteria fall into one of two very distinct families, the λ integrase-like enzymes and the resolvases/invertases. Homologous recombinations within the host genome enable them to function in some rearrangements, such as chromosomal deletions (Toussaint and Merlin 2002). Phage integrases perform site-specific recombination by two different enzyme families: tyrosine and serine recombinases (Groth and Calos 2004). Tyrosine recombinase, such as λ integrase, applies tyrosine at the enzyme's catalytic active site to cleave short host DNA strands, as well as other complementary proteins, from the phage or host as helpers; serine recombinase uses serine at the enzyme's catalytic active site and do not require host cofactors to cleave long host DNA strands, especially the mammals. Transposases are categorized as large serine recombinases and classified based on their protein sequence similarities (Smith and Thorpe 2002).

One of the most famous forms of MGEs is the transposable element or transposon (Tn). The Tn3 family is the most abundant transposable element in prokaryotes and consists of: (I) IS elements, (II) variety of a single Tn (Tn3, Tn501, Tn4430, and Tn5652 subfamilies), and (III) a site-specific recombinase (resolvase), which catalyse transposition based on resolution of an intermediate form (cointegrate) (Szuplewska, Czarnecki et al. 2014). This means that the Tn carries an IS element to cover the need for accessory enzymes, such as transposase and integrase, to conduct HGT.

Site-specific recombination includes the following simplified steps: excision of a gene from a donor genetic structure, monomerization of the recipient genetic structure, alteration of gene expression in the recipient genetic structure, and fusion of horizontally transferred gene into a functional location of the recipient genetic structure (Figure 1) (Reed and Moser 1984).

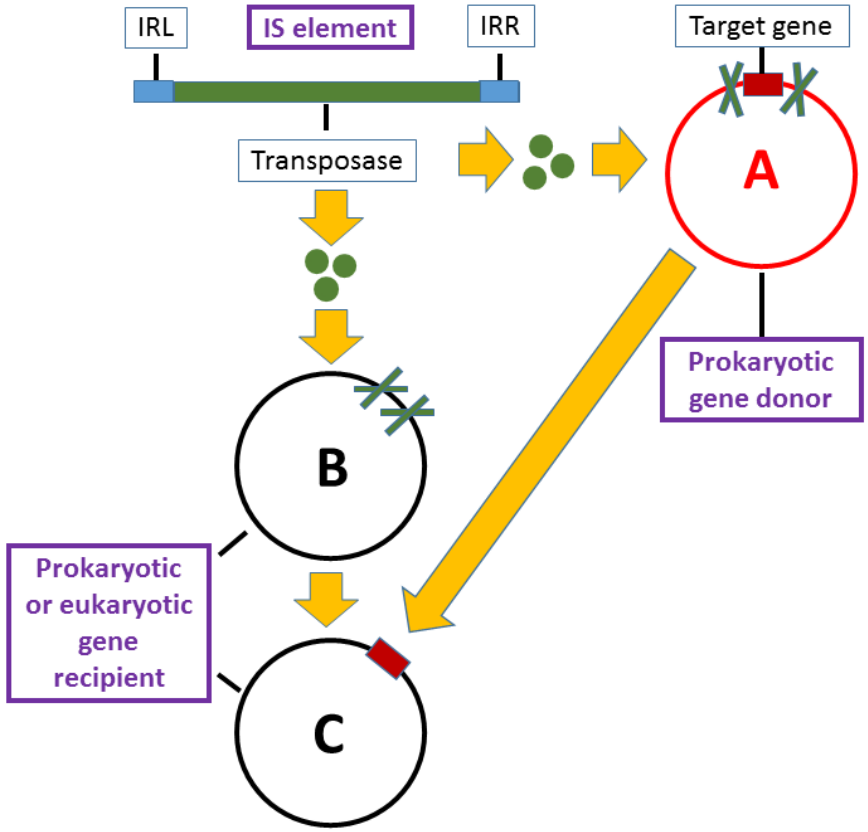


Figure 1. Schematic molecular mechanism of HGT of a target gene. An IS element includes: IRL and IRR, a promoter for transposase or integrase in IRL, and transposase or integrase ORF. (A) The encoded transposase will cut or copy a target gene from prokaryotic genome; (B) The encoded transposase cut a specific sequence on another prokaryotic or eukaryotic genome; and (C) The target gene is horizontally transferred from prokaryotic genome to another prokaryotic or eukaryotic genome (Figure adopted from original communication VI).

The present evidence for HGT in eukaryotes comes from the study of protists (Top and Springael 2003, Keeling and Palmer 2008). Free-living protozoans mostly live in environments where they are frequently exposed to bacterial genomes. An example is the direct contact of bacterial normal flora and parasites in the digestive system of ruminants. A review of previous literature has indicated some well-established endosymbiotic partnerships between a variety of eukaryotic hosts and prokaryotic or endosymbiotic eukaryotes (Beier, Horn et al. 2002, Skriwan, Fajardo et al. 2002, Greub and Raoult 2004, Wernegreen 2004, Nakajima, Sano et al. 2009,

Smutna, Goncalves et al. 2009, Sun, Noe et al. 2009, Bertelli and Greub 2012, Fujishima and Kodama 2012, Fichorova, Buck et al. 2013, Siegmund, Burmester et al. 2013, Catta-Preta, Brum et al. 2015). As an example, one previous study identified *Caedibacter caryophilus* as bacterial endosymbiont in refractile inclusion bodies of *Paramecium* spp. (Beier, Horn et al. 2002). *Dictyostelium discoideum* is a free-living soil amoeba, which is the host for some human bacterial pathogens and obligate intracellular prokaryotes, such as *Mycobacterium avium, S. typhimurium, Pseudomonas aeruginosa,* and *Legionella pneumophila* (Skriwan, Fajardo et al. 2002). Also, a broad variety of bacterial species from alpha-, beta-, gamma-, and deltaproteobacteria, Chlamydiae, Bacteroidetes, Firmicutes, and Actinobacteria have been reported as prokaryotic endosymbionts of amoeba species, including *Acanthamoeba* spp., *Naegleria* spp., and *Hartmannella* spp. (Greub and Raoult 2004).

Endosymbiosis of prokaryotes is not limited to protozoan species. There are plenty of reports about prokaryotic endosymbionts in insects, such as *Buchnera* spp. in aphids, *Wigglesworthia* and *Sodalis* spp. in tsetse flies, *Blochmannia* spp. in ants, *Tremblaya* spp. in mealybugs, and *Wolbachia* spp. in both insects and nematodes (Wernegreen 2004). In addition to free-living protozoans, prokaryotic endosymbionts, including *Flavobacterium columnare*, and some bacterial species of Rickettsiales and Sphingobacteriales orders have detected from parasitic protozoa *Ichthyophthirius multifiliis* (Sun, Noe et al. 2009). A study revealed that Trypanosomatidae family, such as *Angomonas deanei* and *Strigomonas culicis* obtained their heme synthesis genes from their endosymbiont proteobacteria (Alves, Voegtly et al. 2011). Recently, a novel study identified bacteria *Candidatus* Pandoraea novymonadis as endosymbiont of *Novymonas esmeraldas* from the Trypanosomatidae family (Kostygov, Dobakova et al. 2016). This study suggested that *N. esmeraldas* might be the tentative ancestor of some genes in *Leishmania* spp. An *In vitro* study showed DNA transfer from endosymbiont bacteria to protozoa (Siegmund, Burmester et al. 2013). In this study, *E.coli* and a ciliate *Tetrahymena pyriformis* co-cultured together. *E.coli* could escape from cytoplasmic digestion and reside in the vacuoles of *T. pyriformis*. A PCR method identified DNA from *E. coli* in the *T. pyriformis*. On the other hand, two eminent endosymbiotic processes in eukaryotic evolution resulted in adoption of plastids and mitochondria from cyanobacteria and α-proteobacteria species, respectively (Kjeldsen, Obst et al. 2010). Both pathogenic and endosymbiotic prokaryotes are usually considered as DNA donors to protozoans, insects, and nematodes by HGT (Table 5).

Table 5. Examples of HGT of prokaryotic genes to protozoans, insects, and nematodes.

Prokaryotic gene donors	Protozoan, insect, and nematode gene recipients	Horizontally transferred genes	References
Wolbachia spp.	*Aedes aegypti* (yellow fever mosquito), *Anopheles gambiae* (malaria mosquito), and *Drosophila melanogaster*	Many prokaryotic genes, such as gag-pol, D34 immunodominant antigen, actin, and aminotransferase genes	(Dunning Hotopp, Clark et al. 2007, Baldini, Segata et al. 2014)
Escherichia coli	*Caenorhabditis elegans*	Antibiotic-resistance genes	(Portal-Celhay, Nehrke et al. 2013)
Prokaryotes	Anaerobic protozoans: *Trichomonas vaginalis, Entamoeba histolytica,* and *Naegleria gruberi*	Alcohol dehydrogenase (*adh* gene) and pyruvate:ferredoxin oxidoreductase genes	(Andersson 2005, Loftus, Anderson et al. 2005)
Prokaryotes	*Dictyostelium discoideum* (soil-living amoeba)	18 prokaryotic genes	(Eichinger, Pachebat et al. 2005)
Prokaryotes	Trypanosomatids: *Leishmania* spp., *Angomonas deanei,* and *Strigomonas culicis*	Bacterial amino acid pathways	(Alves, Klein et al. 2013)
α-proteobacteria	*Leishmania* spp.	Mitochondria (initiation point of apoptosis)	(Taylor-Brown and Hurd 2013)
β- and γ-proteobacteria	Trypanosomatids: *Leishmania* spp., *Angomonas deanei,* and *Strigomonas culicis*	Heme synthesis gene	(Alves, Voegtly et al. 2011)
Peptostreptococcus harei	*Trichomonas vaginalis*	Lateral gene transfer fragment (TvLF)	(Strese, Backlund et al. 2014)

3. Aims of the study

One reseach focus of our group has been the expression and physiological role of β-CAs in protozoans, insects, and nematodes. In this work, we have expanded both bioinformatic and experimental studies to cover new β-CAs from parasitic and non-parasitic species. The specific aims of the study have been:

1) To identify β-CA proteins and genomic coding sequences from parasites by bioinformatic approaches through genomics and proteomics databases.
2) To study the role of mobile genetic elements (MGEs) in horizontal gene transfer (HGT) of *β-CA* gene sequences from ancestral prokaryotes to parasites including protozoan, insect, and nematode species.
3) To evaluate whether β-CAs in parasites are suitable vaccine candidates or appropriate drug targets for inhibition by anti-infectious compounds.
4) To produce recombinant parasite β-CAs using bacterial or insect cells and apply them in inhibitory studies to find new potential anti-parasitic agents.
5) To design a new sensitive molecular diagnostic method based on parasite β-CA genomic sequences for detection of parasitic infection in biological samples.

4. Materials and methods

4.1. Bioinformatic methods

The applied bioinformatic methods are described here as a schematic flow chart (Figure 2). The details of the used methods are comprehensively explained in sections 4.1.1 to 4.1.10.

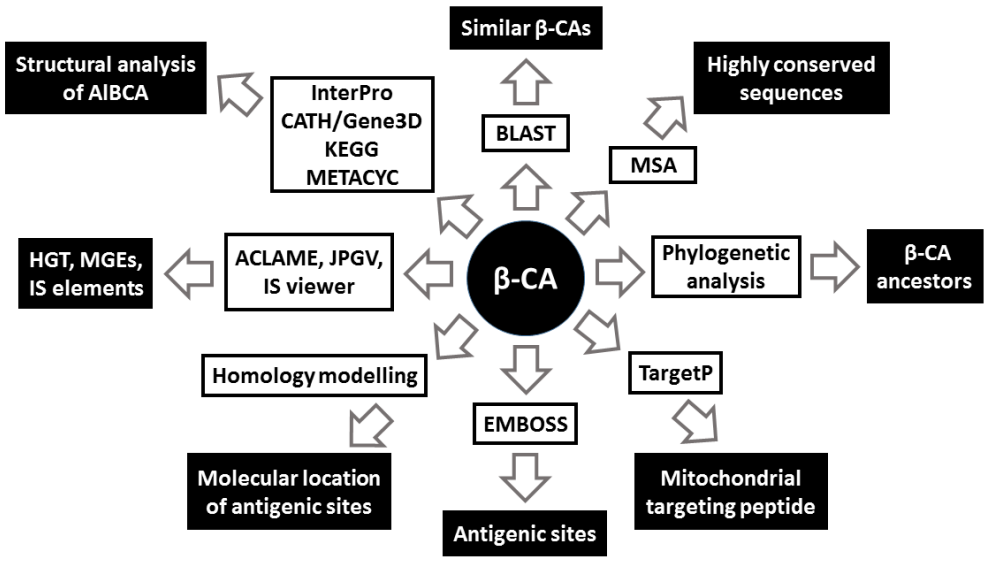


Figure 2. Schematic flow chart of the applied bioinformatic methods.

4.1.1. BLAST (Basic Local Alignment Search Tool) protein homology search

All parasite, bacterial, archaeal, and plant β-CA protein sequences were retrieved from the Universal Protein Resource (UniProt) (http://www.uniprot.org/) (application date: 2014-2016) and National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/) (application date: 2014-2016). We applied β-CA protein sequence IDs from *D. melanogaster* (UniProt ID: Q9VHJ5), *C. elegans* (UniProt IDs: Q22460 and Q2YS41), *L. donovani chagasi* (UniProt ID: E9B8S3), *Arabidopsis thaliana* (UniProt ID: Q9ZUC2), and *Sulfolobus islandicus* (UniProt ID: F0NNH7) as queries in BLAST protein homology search (Alba and Castresana 2007) databases to obtain most β-CA protein homologs from insects, nematodes, protozoans, plants, and archaea. Moreover, a β-CA protein sequence from nematode *Trichinella spiralis* (UniProt ID: E5SH53) was extracted from BLAST protein homology list and used to search Ensembl Metazoa (http://metazoa.ensembl.org/index.html) (application date: March 2016) to obtain corresponding genomic sequence ID of *T. spiralis* (EFV55868) for molecular diagnostics of trichinellosis (I, II, III, IV, V, and VI).

4.1.2. Multiple sequence alignment (MSA)

Clustal omega (Sievers, Wilm et al. 2011) and jalview (version 2.8.ob1) (http://www.jalview.org/) (Waterhouse, Procter et al. 2009) (application date: 2014-2016) were used to create an alignment of β-CA protein sequences from BLAST protein homology searches, including protozoans, metazoans (insects and nematodes), economically important plants, and ten β-CA protein sequences from endosymbiotic bacteria of protozoans, insects, and nematodes, including *Afipia* spp. (UniProt ID: K8NQ88), *Anaeromyxobacter* spp. (UniProt ID: A7HD59), *Campylobacter* spp. (UniProt ID: K0I0K3), *Salmonella* spp. (UniProt ID: Q8ZRS0), *Gardnerella* spp. (UniProt ID: E3D7T4), *Emticicia* spp. (UniProt ID: I2EZ21), *Simkania* spp. (UniProt ID: F8L9G5), *Nostoc* spp. (UniProt ID: Q8YT17), *Exiguobacterium* spp. (UniProt ID: K0ACL8), and *Fusobacterium* spp. (UniProt ID: C6JPI1), as well as β-CA protein sequence encoded from MGEs of bacteria *Acaryochloris marina* (Uniprot ID: A8ZKS7) and archaea *S. islandicus* (UniProt ID: F0NNH7). In addition, some bacterial β-CA protein sequences were used for MSA analysis as outgroup, such as *Pelosinus fermentans* (UniProt ID: EIW34693) and *Desulfosporosinus meridiei* (UniProt ID: J7IY65) (I, II, III, IV, V, and VI). In some evolutionary studies, outgroup organisms can be used for better rooting of phylogeny and clarification of evolutionary relationships. Outgroup organisms should be different from ingroup or analyzed organisms but include the same gene and protein (Wilberg 2015).

4.1.3. Phylogenetic analysis

All β-CA protein sequences were individually analyzed for completeness through UniProt (http://www.uniprot.org/) (application date: 2014-2016) and NCBI (http://www.ncbi.nlm.nih.gov/) (application date: 2014-2016) databases. The protein MSAs, which were created using Clustal Omega algorithm (Sievers, Wilm et al. 2011) were used to derive the corresponding codon-aligned nucleotide sequences using the Pal2Nal program (Suyama, Torrents et al. 2006). All phylogenetic analysis were computed using Mr. Bayes v3.2 (Ronquist, Teslenko et al. 2012) with the generalised time-reversible model of codon substitution and all other parameters set to default. Then all the final phylogenetic trees were visualized in FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/) (Ronquist, Teslenko et al. 2012, Vogeler, Galloway et al. 2014) (application date: 2014-2016). When *β-CA* genes were analyzed for HGT from prokaryotes to protozoans, insects, and nematodes, phylogenetic trees were constructed individually for each β-CA sub-group (A-D). The total numbers of sequences analyzed for each sub-group were: 109(A), 53(B), 36(C), and 22(D). In addition, to evaluate the role of mobilomes, including plasmids and GIs in HGT of *β-CA* genes from prokaryotes to protozoans and metazoans, a total of 107 β-CA protein sequences with chromosomal, plasmid, and GI origins, were retrieved from the UniProt (http://www.uniprot.org/) (application date: 2014-2016) and NCBI (http://www.ncbi.nlm.nih.gov/) (application date: 2014-2016) databases. All the phylogenetic analyses were performed using supercomputer resources provided by the Finnish IT Center for Science (https://www.csc.fi/) (application date: 2014-2016) (I, II, V, and VI).

4.1.4. Prediction of subcellular localization

Subcellular localization of each identified bacterial, protozoan, insect, nematode, and plant β-CA was predicted using the TargetP webserver (http://www.cbs.dtu.dk/services/TargetP/) (Emanuelsson, Nielsen et al. 2000) (application date: 2014-2016). TargetP is built from two layers of neural networks, where the first layer contains one dedicated network for each type of targeting sequences, such as cytoplasmic, mitochondrial, or secretory peptides, and the second layer is an integrating network that outputs the actual prediction (cTP: cytoplasmic, mTP: mitochondrial, SP: secretory, and O: other). It is able to discriminate between cTPs, mTPs, and SPs (I, II, and V).

4.1.5. Prediction of antigenic sites in β-CAs

The protein sequences of 23 agricultural and veterinary relevant parasite and eight economically important plant β-CAs were analyzed with the European Molecular Biology Open Software Suite (EMBOSS) program (http://emboss.bioinformatics.nl/cgi-bin/emboss/antigenic) (Kolaskar and Tongaonkar 1990, Rice, Longden et al. 2000, Mullan and Bleasby 2002, Olson 2002) (application date: August 2014). EMBOSS predicts potentially antigenic regions of a protein sequence using the method of Kolaskar and Tongaonkar (II).

4.1.6. Homology modelling

Homology models of four selected β-CAs from parasites, including *Ancylostoma caninum* (UniProt ID: FC551456), *Ascaris suum* (UniProt ID: F1LE18), *T. spiralis* (UniProt ID: E5SH53), and *E. histolytica* (UniProt ID: C4LXK3) were prepared by first selecting the most suitable template structure. For this purpose, a BLAST protein homology search of the PDB database (http://www.rcsb.org/pdb/home/home.do) (Sussman, Lin et al. 1998) (application date: August 2014) was performed using each of the four β-CA protein sequences. Results for three out of these four searches revealed that PDB structure 1EKJ (β-CA from *P. sativum*) possessed the most similar sequence, while PDB structure 2A5V (β-CA from *M. tuberculosis*) was found to be the most similar to *E. histolytica* (UniProt ID: C4LXK3). Clustal Omega was used to prepare an MSA for those six sequences. The MSA showed nine completely conserved residues within the sequences; the known highly conserved CXDXR (C: Cysteine, D: Aspartic acid, R: Arginine, and X: Any residue) and HXXC (H: Histidine, C: Cysteine, X: Any residue) motifs were among them (data not shown). Homology modelling was performed according to MSA containing β-CA protein sequences from *A. caninum* (UniProt ID: FC551456), *A. suum* (UniProt ID: F1LE18), *T. spiralis* (UniProt ID: E5SH53), and PDB 1EKJ by using the Modeller program (version 9.13) (Eswar, Webb et al. 2006) with PDB model 1EKJ (β-CA from *P. sativum*) as a template. A homology model for β-CA protein sequence from *E. histolytica* (UniProt ID: C4LXK3) was prepared using PDB model 2A5V for pairwise alignment and as a template structure. The resulting models were structurally aligned using the BODIL program (Lehtonen, Still et al. 2004). A figure illustrating the homology models was prepared by using the VMD program (version 1.9.1) (Humphrey, Dalke et al. 1996), and edited within Adobe Photoshop (version 13.0.1). The structural availability of the epitope in the PDB model 1EKJ (β-CA from *P. sativum*) and the homology model based on the β-CA protein sequence from *A. caninum* was studied by preparing the molecular surface with VMD, using a

probe radius of 1.4 Å. The potential epitope residues were excluded from the surface presentation and were shown as Van der Waals (VdW) spheres.

In the study of horizontal transfer of *β-CA* genes from prokaryotes to protozoans, insects, and nematodes, homology models were prepared for selected β-CAs based on the phylogenetic analysis. The most similar prokaryotic and eukaryotic proteins within the group of proteins in 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/) (Sievers and Higgins 2014) (application date: March 2016). 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) (application date: March 2016). For each protein pair (prokaryotic and eukaryotic) analyzed here, the BLAST search resulted in the same template protein as follows: Clade A: *E. coli* β-CA PDB 1I6P; Clade B: *P. sativum* β-CA PDB 1EKJ; Clade C: *M. tuberculosis* β-CA PDB 1YM3; and Clade D: *M. thermoautotrophicum* β-CA PDB 1G5C. The rest of the modelling and visualization of the modelled β-CAs was performed the same as the procedure has been done for *A. caninum* (UniProt ID: FC551456), *A. suum* (UniProt ID: F1LE18), *T. spiralis* (UniProt ID: E5SH53), and *E. histolytica* (UniProt ID: C4LXK3) (II and V).

4.1.7. Identification of β-CA, transposase, integrase, resolvase, and conjugation complex protein (CCP) genes in prokaryotic mobilomes

Identification of *β-CA,* transposase, integrase, resolvase, and CCP genes on the bacterial mobilomes was carried out using the ACLAME (A classification of mobile genetic elements) version 0.4 database (http://aclame.ulb.ac.be/) (Leplae, Lima-Mendez et al. 2010) (application date: March 2016), the plasmid browser from European Molecular Biology Laboratory- European Bioinformatics Institute (EMBL-EBI) (http://www.ebi.ac.uk/genomes/plasmid.html) (application date: March 2016), and the Jena Prokaryote Genome Viewer (JPGV) (http://jpgv.fli-leibniz.de/cgi/index.pl) (Romualdi, Felder et al. 2007) (application date: March 2016). In addition, the IslandViewer version 3 (http://www.pathogenomics.sfu.ca/islandviewer/) (Dhillon, Laird et al. 2015) (application date: June 2016) database was used for identification of both archaeal and bacterial chromosomal GIs (V and VI).

4.1.8. Identification of *β-CA* gene sequences in protozoan, insect, and nematode genomic DNA

Analysis and determination of precise locations of protozoan, insect, and nematode *β-CA* genes on the main genomic DNA were performed using the NCBI database (http://www.ncbi.nlm.nih.gov/) (application date: March 2016). Furthermore, we utilized the *T. vaginalis* genome project database (TrichDB version 1.3) (http://trichdb.org/trichdb/) (Aurrecoechea, Brestelli et al. 2009) (application date: March 2016) and EMBL-EBI database (http://www.ebi.ac.uk/) (application date: March 2016), for detection of *β-CA* genes from *T. vaginalis* (the causative agent of trichomoniasis) and *C. elegans*, respectively. Analysis of mitochondrial coding genes from *A. castellanii* (the most common free-living amoeba in soil and water) was performed using NCBI database (http://www.ncbi.nlm.nih.gov/) (application date: March 2016) (V).

4.1.9. Identification of IS elements from archaeal and bacterial plasmids and chromosomal genomes

The studies defined that IS elements encode the accessory enzymes such as transposases and integrases, which have been applied for HGT of genes from prokaryotes to prokaryotes or eukaryotes (Mahillon and Chandler 1998). Identification of archaeal and bacterial IS elements was performed using the IS Finder database (https://www-is.biotoul.fr/) (Siguier, Perochon et al. 2006) (application date: June 2016) (VI).

4.1.10. Structural and functional predictions of β-CAs based on AlBCA (*A. lumbricoides* β-CA) protein sequence

The AlBCA protein sequence (Uniprot ID: F1LE18) was applied as an enquiry in the integrative protein signature database, InterPro (http://www.ebi.ac.uk/interpro/) (Hunter, Apweiler et al. 2009) (application date: September 2015). InterPro integrates together predictive models of representative protein domains, families, and functional sites from various databases. The resulting AlBCA InterPro ID (IPR001765) was used as an enquiry in the CATH/Gene3D database (http://www.cathdb.info/) (Pearl, Todd et al. 2005) (application date: September 2015). CATH/Gene3D database hierarchically classifies domains into sequence and structure-based families and fold groups, when there is a sufficient

evidence for having diverged from a common ancestor. The CATH/Gene3D database generated a rainbow model for superimposed AlBCA protein sequence and several other close species. In addition, we identified the biochemical pathways and interactions of AlBCA through KEGG (Kyoto Encyclopedia of Genes and Genomes) (http://www.kegg.jp/) (Kanehisa and Goto 2000) (application date: September 2015) and METACYC metabolic pathway databases (http://metacyc.org/) (Caspi, Altman et al. 2014) (application date: September 2015), which have both been linked to InterPro (III).

4.1.11. Primer design

In the study of *T. spiralis* diagnostics based on *β-CA* gene sequence, three sets of primers for genomic sequence of *T. spiralis* β-CA (exonic and intronic) (Ensembl Metazoa ID: EFV55868) were designed automatically by NCBI Primer-BLAST tool (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) (Ye, Coulouris et al. 2012) (application date: March 2016), and one additional primer set was designed manually (Figure 3). The sizes of the PCR products were calculated by Oligo Calc (http://www.basic.northwestern.edu/biotools/oligocalc.html) (Kibbe 2007) (application date: March 2016). For obtaining *T. pseudospiralis* and *T. nativa* β-CA genomic sequences, we first performed PCR amplification using the designed *T. spiralis* primers and *T. pseudospiralis* and *T. nativa* larvae were used as templates. The corresponding bands for *T. pseudospiralis* and *T. nativa* on the agarose gel were cut and sequenced as described in section 4.2.3. The obtained β-CA genomic sequences from *T. pseudospiralis* and *T. nativa* were aligned by the Clustal Omega algorithm (http://www.ebi.ac.uk/Tools/msa/clustalo/) (Sievers, Wilm et al. 2011) (application date: March 2016) within EMBL-EBI database (http://www.ebi.ac.uk/Tools/msa/clustalo/) (application date: March 2016) against *T. spiralis* β-CA genomic sequence, and the most non-identical regions were selected for the second round of primer design. We designed 3 and 4 different forward primers manually for *T. pseudospiralis* (primer set No. 5-7) and *T. nativa* (primer sets No. 8-11), respectively, and used with the same reverse primer No. 2, which was described for the first round (primer set No. 1-4). In the third round of primer design, one new reverse primer was designed for *T. pseudospiralis* (primer set No. 12) and used with its own forward primer No. 7. In addition, four new reverse primers were designed for *T. nativa* and tested with its own forward primers No. 8 and 10 (primer sets No. 13-16). Three other species, *Toxoplasma gondii*, *Toxocara cati*, and *P. equorum* were analyzed for any false-positive reactions with other parasitic species (IV). All of 16 designed primers are presented in Table 6.

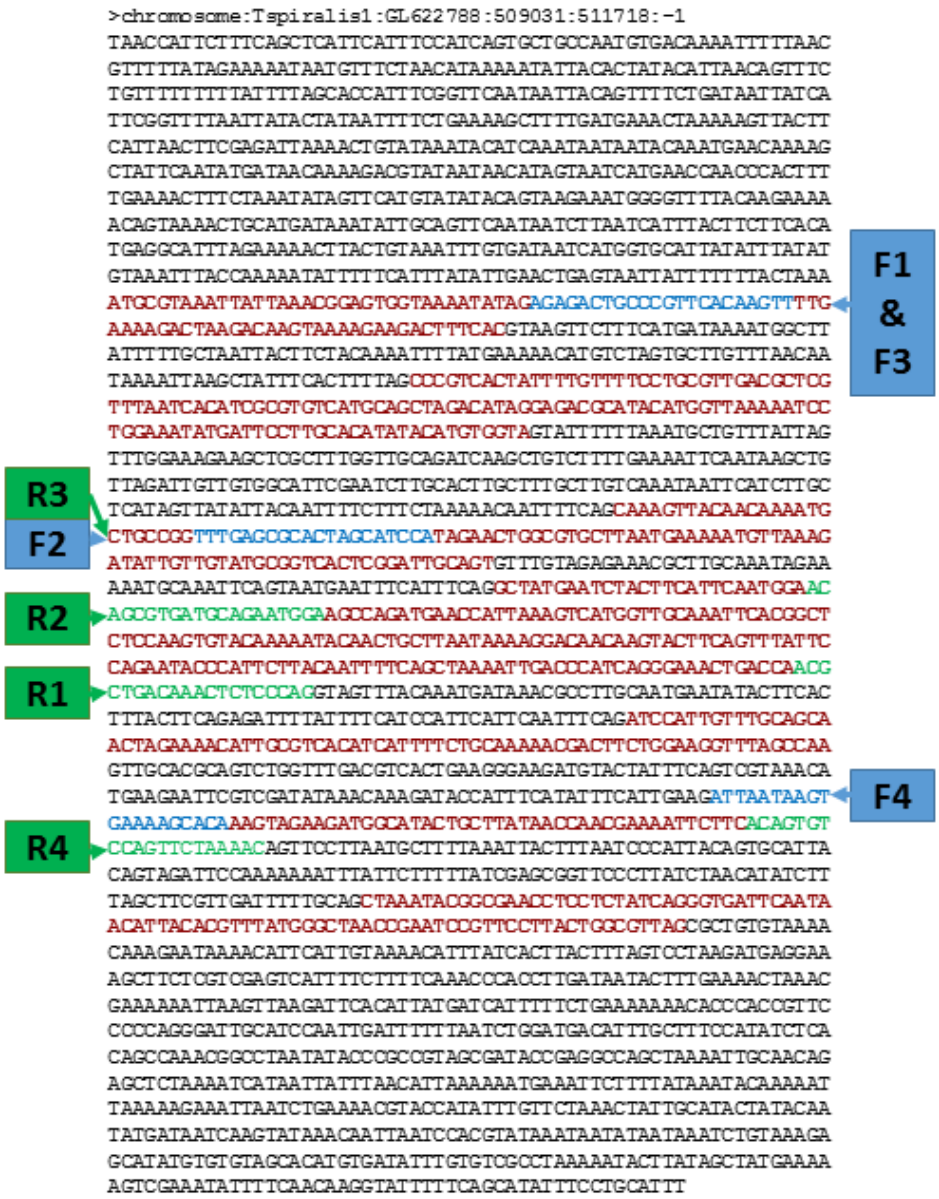


Figure 3. Schematic figure of the first four primers. The full genomic sequence of the *β-CA* gene from *T. spiralis* was used for primer design. Black and red nucleotides are intronic and exonic nucleotides, while blue and green nucleotides are the designed forward (F1-4) and reverse primers (R1-4), respectively. Reverse primers were designed based on reverse complementary conversion of green nucleotide sequences. There are two overlaps for the primers: (1) F1 and F3, and (2) F2 and R3.

Table 6. Designed primers for β-CA genomic sequences from *T. spiralis, T. pseudospiralis* and *T. nativa*.

No.	Designed primers	Species name	Product length (bp)
1	**Forward:** 5′-AGAGACTGCCCGTTCACAAG-3′ **Reverse:** 3′-CTGGGAGAGTTTGTCAGCGT-5′	*T. spiralis*	882
2	**Forward:** 5′-TTTGAGCGCACTAGCATCCA-3′ **Reverse:** 3′-TCCATTCTGCATCACGCTGT-5′	*T. spiralis*	191
3	**Forward:** 5′-AGACTGCCCGTTCACAAGTT-3′ **Reverse:** 3′-TGGATGCTAGTGCGCTCAAA-5′	*T. spiralis*	530
4	**Forward:** 5′-ATTAATAAGTGAAAAGCACA-3′ **Reverse:** 3′-GTTTTAGAACTGGACACTGT-5′	*T. spiralis*	83
5	**Forward:** 5′-CTCACCCATCACCCCGGCTT-3′ **Reverse:** 3′-TCCATTCTGCATCACGCTGT-5′	*T. pseudospiralis*	251
6	**Forward:** 5′-GGCCAGGCCGTTCATCTGGT-3′ **Reverse:** 3′-TCCATTCTGCATCACGCTGT-5′	*T. pseudospiralis*	158
7	**Forward:** 5′-ATCGCGCTCTCGCGATTGGG-3′ **Reverse:** 3′-TCCATTCTGCATCACGCTGT-5′	*T. pseudospiralis*	103
8	**Forward:** 5′-TCCCAGACCAGCGGNAGCAC-3′ **Reverse:** 3′-TCCATTCTGCATCACGCTGT-5′	*T. nativa*	372
9	**Forward:** 5′- GACCCAGCGCGCTTTCGTTG-3′ **Reverse:** 3′-TCCATTCTGCATCACGCTGT-5′	*T. nativa*	317
10	**Forward:** 5′-CGGATACCACGGGCCGATGT-3′ **Reverse:** 3′-TCCATTCTGCATCACGCTGT-5′	*T. nativa*	226
11	**Forward:** 5′-AGTCGCCCAGCTTGATCGCG-3′ **Reverse:** 3′-TCCATTCTGCATCACGCTGT-5′	*T. nativa*	116
12	**Forward:** 5′-ATCGCGCTCTCGCGATTGGG-3′ **Reverse:** 3′-CAACCGATACCGAACGGACC-5′	*T. pseudospiralis*	116
13	**Forward:** 5′-TCCCAGACCAGCGGNAGCAC-3′ **Reverse:** 3′-CGGACCGAACTCTGGTACAG-5′	*T. nativa*	319
14	**Forward:** 5′-CGGATACCACGGGCCGATGT-3′ **Reverse:** 3′-CGGACCGAACTCTGGTACAG-5′	*T. nativa*	171
15	**Forward:** 5′-TCCCAGACCAGCGGNAGCAC-3′ **Reverse:** 3′-ACATCGGCCCGTGGTATCCG-5′	*T. nativa*	168
16	**Forward:** 5′-TCCCAGACCAGCGGNAGCAC-3′ **Reverse:** 3′-GTGAGTCCAGCAGCAACCCG-5′	*T. nativa*	129

4.2. Experimental methods

4.2.1. Preparation of parasites, infected meat samples, and lysed parasite-infected meat samples

Trichinella spp. infected mouse muscle and parasite samples were prepared at the Department of Veterinary Biosciences, University of Helsinki, Finland. The Animal Experiment Board in Finland had approved the study protocol. The tested *Trichinella* spp. involved *T. spiralis*, *T. pseudospiralis*, and *T. nativa*. The blinded samples, including both infected and uninfected meat, were packed and shipped from Helsinki to our laboratory for polymerase chain reaction (PCR). After arriving in the laboratory, the

specimens were stored at -20 °C. To test the specificity of different designed primers, three more parasite samples were analyzed including *T. gondii*, *P. equorum*, and *T. cati* (IV).

The meat and larval samples were handled under the laminar hood. *T. gondii* parasites were separated from the culture medium by centrifugation (Heraeus Biofuge Fresco, Thermo Scientific, Waltham, MA) at 13,000 RPM for 10 minutes (Leroux, et al., 2015). Then the following lysis protocol was performed on all samples: 300 µl lysis buffer (DirectPCR Tail Lysis reagent, Viagen Biotech, LA) was added to 5 mg of samples. Then 6 µl proteinase K (Thermo Scientific) was added to the mixture. The lysis tubes were incubated in a rotating incubator (HYBAID, Thermo Scientific) at 55 °C overnight. It should be noted that no DNA extraction was performed and just the lysis step was done on the samples.

4.2.2. Polymerase chain reaction (PCR)

For PCR reaction of the β-CA genomic sequence from *Trichinella* spp. and other parasites, including *T. gondii*, *P. equorum*, and *T. cati,* 12.5 µl 2X KAPA ReadyMix (KAPA 2G Robust HotStart ReadyMix PCR Kit, Kapa Biosystems, Wilmington, MA) was pipetted in the PCR microtubes. 10 µM forward and reverse primers (Oligomer, Helsinki, Finland) (Table 1, original communication IV) were added to the KAPA ReadyMix. Afterwards, 9 µl dH_2O was mixed with the previous contents of the PCR microtubes. Finally, 1 µl of lysed samples were combined with PCR mixtures. Then, the following temperatures were set and run on the thermocycler (PTC100™, MJ Research Inc., Waltham, MA): 95 °C (3 minutes), [95 °C (10 seconds), 53 °C (10 seconds), 72 °C (10 seconds)] X 36 cycles, 72 °C (5 minutes) (IV). The lysed *T. spiralis* larvae served as a positive control and the negative control was the PCR master mix without any lysed sample.

4.2.3. Sequencing of *T. pseudospiralis* and *T. nativa* β-CA genomic sequences

For sequencing of *T. pseudospiralis* and *T. nativa* β-CA genomic sequences, PCR products of *T. pseudospiralis* and *T. nativa* templates were cut from an agarose gel. The corresponding DNA bands were extracted and purified according to the manufacturer's instructions for Illustra-GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, Buckinghamshire, UK). The DNA samples were dried at room temperature overnight and suspended in 15 µl HiDi™ (Life Technologies Europe, Applied Biosystems®, Finland), which was followed by vortexing and spinning. Thenafter, the samples were heated at 95 °C for 2 minutes, which followed by

transferring to ice for 2 minutes. Finally, the samples were sent to Core Facilities and Research Services, BioMediTech, Tampere, Finland for DNA sequencing. The sequencing was carried out with HITACHI 3130x/Genetic Analyzer (Life Technologies Europe, Applied Biosystems) (IV).

4.2.4. Production of recombinant AlBCA

The CDS sequence of the *AlBCA* gene was obtained from the EMBL database (http://www.ebi.ac.uk/) (application date: September 2015). GeneArt® gene synthesis technology (Life Technologies) was applied to construct the *β-CA* gene sequence for insertion into the cloning vector (pFastBac1) (Notka, Liss et al. 2011). DH10Bac cells (which contain the bacmid baculovirus shuttle vector, and a helper plasmid that produces the proteins needed for transposition), were transformed by pFast-*AlBCA* (Additional file 1: Figure S1, original communication III) (Syrjanen, Tolvanen et al. 2010). Then pFast-*AlBCA* was purified by PureLink™ HiPure Plasmid Purification Kit (Invitrogen).

Primary transfection of sf-9 insect cells for production of *Baculoviruses* was performed using HilyMax transfection reagent (Dojindo) ($0.5x10^7$ cells/ml in 6-well plates), and the cells were incubated at 29°C for three days. Cultured cells were centrifuged, passed through a 0.2 μm filter, and stored in a dark tube at +4°C. For secondary transfection, 20 ml of cultured cells ($2x10^6$/ml) were transfected with a primary stock of *Baculoviruses* and incubated in a 29°C shaker for three days. Cultured cells were centrifuged, passed through 0.2 μm filter, and stored in a dark tube at +4°C. For the expression of recombinant AlBCA, the secondary stocks of *Baculoviruses* were used to infect sf-9 cells ($2x10^6$/ml) (using the same protocol as explained for the production of higher volumes of the secondary stock).

The sf-9 cell culture medium was centrifuged at 5,000 RPM for ten minutes at room temperature. The supernatant containing secreted AlBCA was diluted at a ratio of 1:5 by binding buffer (0.1 M Tris, 0.2 M Na_2SO_4, pH 8). Then Protino™ Ni-NTA Agarose (Macherey-Nagel) was added to the diluted culture medium and put on a magnetic stirrer with a low rotation speed (two hours at room temperature and overnight incubation at +4°C without stirring). The culture medium was poured into the funnel filtration system and passed through Whatman® filter paper by vacuum.

The flow-through was collected in a separate vial. The agarose was washed with the wash buffer (50 mM Na_2HPO_4, 0.5 M NaCl, 20 mM Imidazol, pH 8). Elution buffer (50 mM Na_2HPO_4, 0.5 M NaCl, 0.25 M Imidazol, pH 8) was added into the column to finally elute the recombinant AlBCA. Despite several attempts, thrombin treatment did not cleave the polyhistidine-tag (data not shown) probably due to the specific molecular folding of AlBCA that completely hid the thrombin cleavage site. Therefore, the kinetic measurements were carried out using recombinant AlBCA containing the polyhistidine-tag (III).

4.2.5. Kinetic characterization of AlBCA

The kinetic study of AlBCA was performed at the Neurofarba department, Sezione di Scienze Farmaceutiche e Nutraceutiche, Universita` degli Studi di Firenze, Sesto Fiorentino (Firenze), Italy by Professor Claudiu T. Supuran. An applied photophysics stopped-flow instrument was used for testing the CA catalyzed CO_2 hydration activity (Khalifah 1971). Phenol red (at a concentration of 0.2 mM) was used as an indicator, working at the absorbance maximum of 557 nm, with 20 mM TRIS (pH 8.3) as a buffer, and 20 mM $NaClO_4$ (for maintaining a constant ionic strength; this anion is not inhibitory against AlBCA up to concentrations of 50 mM; data not shown). The CA-catalyzed CO_2 hydration reaction was followed for a period of 10-100 s. The uncatalyzed rates were determined in the same manner and subtracted from the total observed rates.

The inhibition constant of AAZ was obtained by non-linear least-squares methods using PRISM 3 and the Cheng-Prusoff equation, as reported earlier (Maresca, Scozzafava et al. 2013, Maresca, Vullo et al. 2013), and displays the mean from three various designations (III).

5. Results

5.1. Results from bioinformatic studies

5.1.1 BLAST protein homology search and MSA

The Uniprot search of potential β-CA protein sequences and the subsequent MSA analysis identified 75 protozoan and metazoan β-CAs, of which 52 β-CA protein sequences were reported here for the first time (Table 1, original communications I). Some of these β-CAs are present in agricultural and livestock husbandry pests (Table 1, original communication II).

β-CAs from a number of agricultural-veterinary pests and parasites, such as *A. lumbricoides* and *T. spiralis*, and prokaryotic endosymbionts of protozoans were evaluated in the MSA analysis. It showed that all the aligned sequences included both the first (CXDXR) and second (HXXC) highly conserved motifs of the enzyme's active site (Figure 4) (Additional file 1: Figure S1, original communication I; Figure 1, original communication II; Figure 1, original communication III; Figure 1, original communication IV; Additional file 1, original communication V; Figure 1, original communication V; Figure 1, original communication VI).

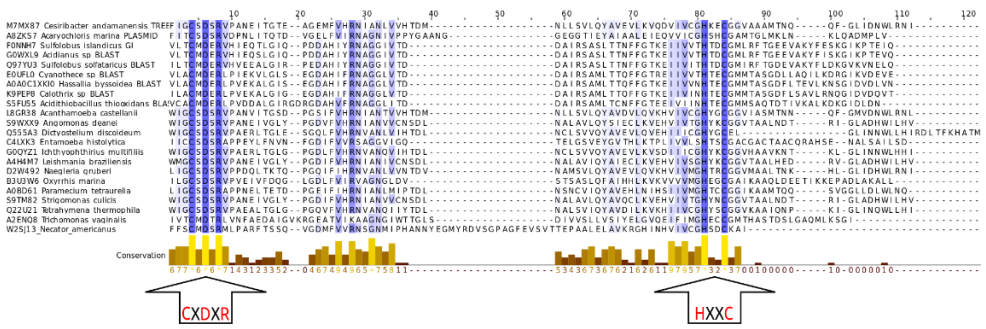


Figure 4. Multiple sequence alignment (MSA) of β-CA protein sequences from defined archaeal, bacterial, protozoan, and nematode species. The aligned 22 β-CA protein sequences show that they all contain the first (CXDXR) and

second (HXXC) highly conserved motifs of β-CAs (Figure adopted from original communication VI).

5.1.2. Phylogenetic analysis

The results of the phylogenetic analysis of β-CAs in prokaryotes, protozoans, metazoans, insects, and nematodes show the evolutionary roots, similarities, and duplications that have occurred. The analysis of phylogenetic trees demonstrates localization of different β-CAs in distinct branches and clades (Figures 5-7) (Figure 2, original communication II; Figure 2, original communication V). These distinct clades (A, B, C, and D) defined separate potential drug and pesticide targets against protozoan and nematode infections and arthropod vectors. In addition, the phylogenetic analysis of these four clades in HGT study revealed a common ancestor of protozoan, insect, and nematode β-CAs within bacterial β-CA protein sequences (Table 7) (Figure 2 and Table 2, original communication V). Evaluation of ten different panels A-J (Figure 7) (Table 1, original communication VI) showed that mobilomes (GIs and plasmids) played a critical role in horizontal transfer of *β-CA* genes within prokaryotes, protozoans, arthropods, and nematodes. We marked a "Golden Box" within the phylogenetic tree to indicate the most feasible evidence on the role of mobilomes in horizontal transfer of *β-CA* genes.

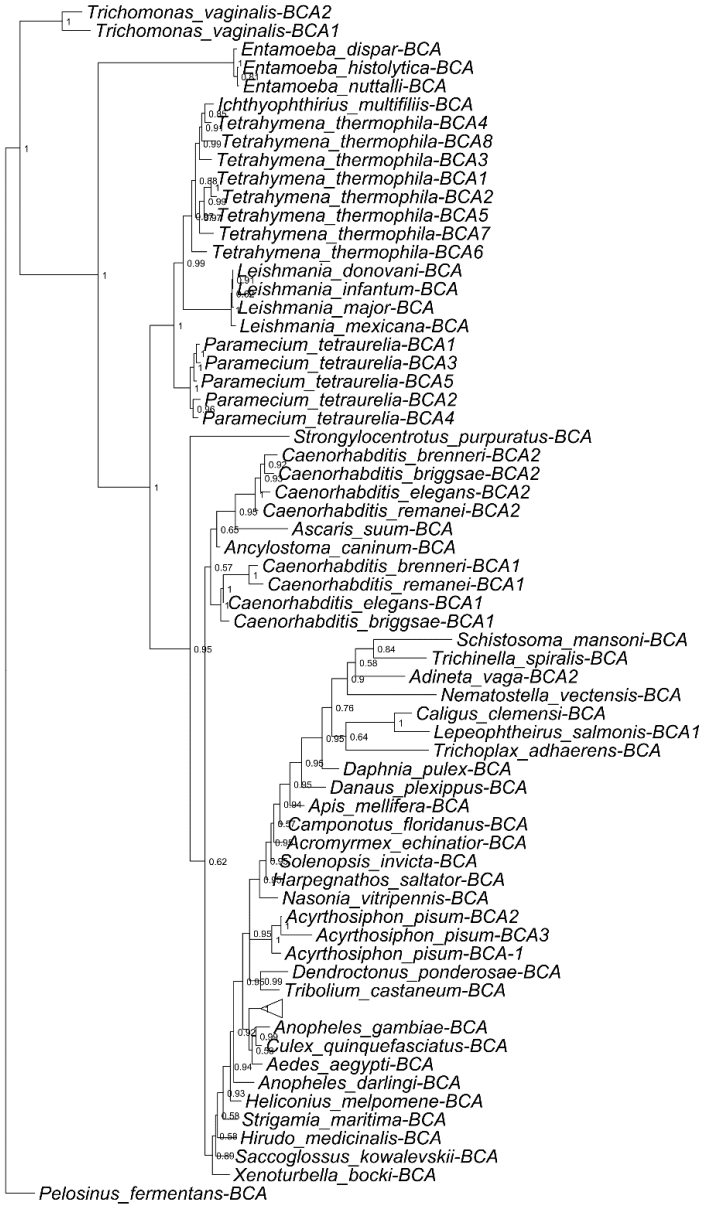


Figure 5. Phylogenetic analysis of 75 metazoan and protozoan β-CAs. The position of β-CAs of *Drosophila* species has been represented at the bottom of the phylogenetic tree by a triangle shape. The details of β-CAs of *Drosophila* spp. in the phylogenetic tree are shown in Figure 6. Some of protozoan and metazoan species include more than one β-CA, which were indicated with consecutative numbers, such as β-CA1, β-CA2, β-CA3, etc. There is no digit with β-CA, if there is no more than one β-CA in the organism (Figure adopted from original communication I).

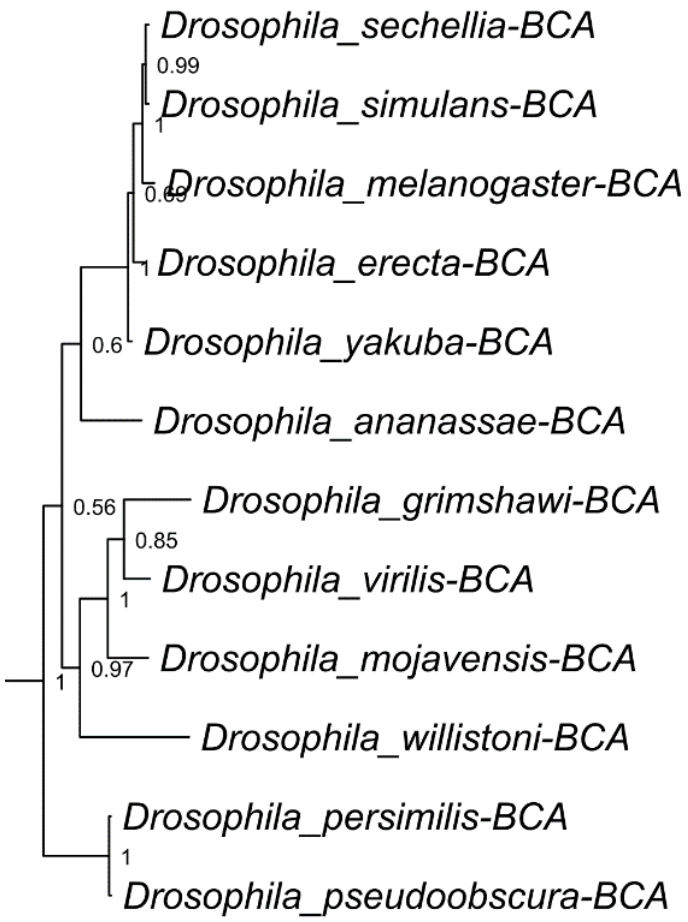


Figure 6. Phylogenetic analysis of β-CAs of *Drosophila* species. This tree represents the expanded view of the triangle located near the bottom of the main phylogenetic tree of β-CAs in Figure 5 (Figure adopted from original communication I).

Table 7. Predicted sources of *β-CA* genes. The tentative prokaryotic endosymbionts and their hosts.

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

*: Figure 2 (A, B, C, and D), original communication V.

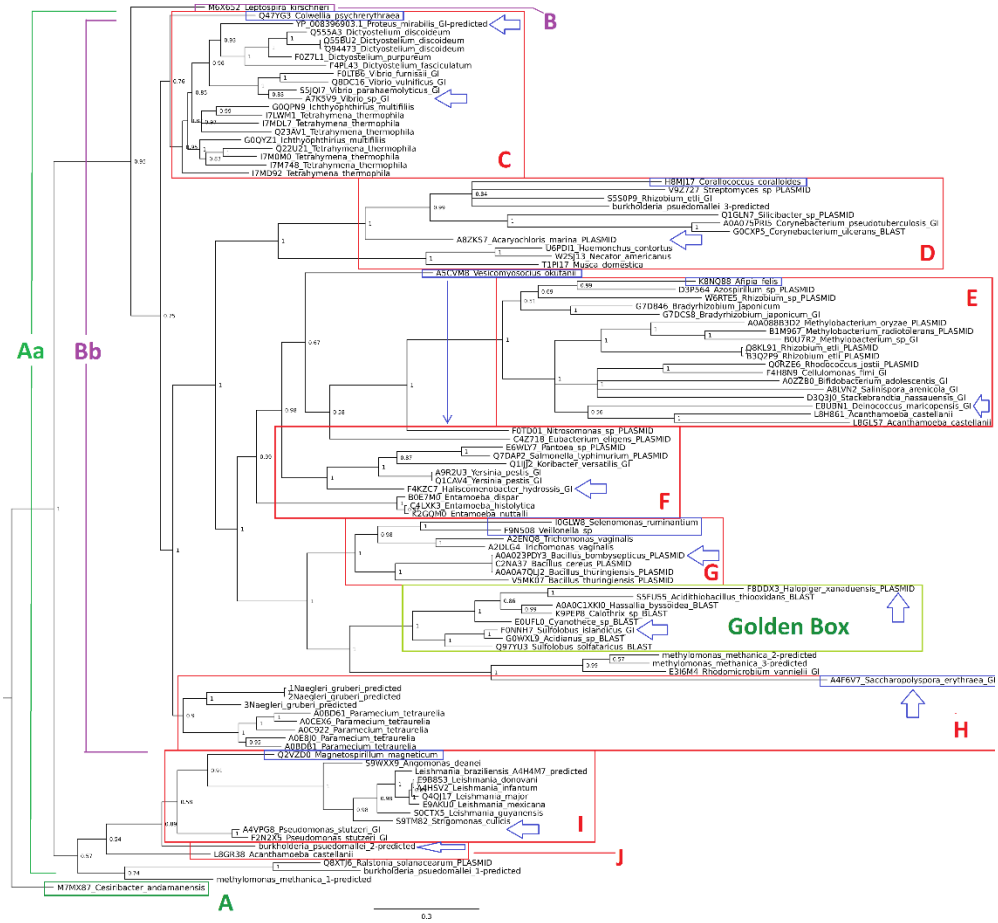


Figure 7- Phylogenetic analysis of β-CA protein sequences from protozoans, metazoans, and prokaryotic mobilomes. We showed boxes (A) and (B) as primitive ancestors, which had common ancestors with "Aa" and "Bb" respectively. In addition, boxes (C) containing *Dictyostelium* spp., *I. multifiliis*, and *T. thermophila*; (D) metazoans including *Haemonchus contortus, Necator americanus,* and *Musca domestica*; (E) *A. castellanii*; (F) *Entamoeba* spp.; (G) *T. vaginalis*; (H) *N. gruberi* and *P. tetraurelia*; (I) *Leishmania* spp.; and (J) *A. castellanii* had common ancestors with defined prokaryotes in the relevant boxes. All of the ancestral prokaryotes included β-CA coding GIs or plasmids. The "Golden Box" showed that β-CA from these archaea, cyanobacteria, and proteobacteria species had a recent common ancestor. Also, this box displayed that GI and plasmid *β-CA* genes from two archaeal species had a common ancestor with each other. The details of the phylogenetic tree were explained in Table 1, original communication VI (Figure adopted from original communication VI).

5.1.3. Prediction of subcellular localization

The results of subcellular localization predictions for defined prokaryotic, metazoan, protozoan, insect, nematode, and plant β-CAs have been shown in several tables of the original publications. The results for protozoan and metazoan β-CAs revealed that 37 are predicted to have a mitochondrial localization, one (*Anopheles darlingi*, Uniprot ID: E3X5Q8) was predicted to be secreted, and the remaining 58 were predicted to have other cellular localizations (Table 2, original communication I; Table 2, original communication II; Additional file 2, original communication V). These results have been pooled and shown in Table 8.

Table 8. Prediction of subcellular localization of *in vitro*-approved prokaryotic endosymbionts of protozoans, insects, and nematodes.

Species name	Entry ID	mTP*	SP*	O*	Result*
Afipia felis	K8NQ88	+	-	-	M
Bradyrhizobium japonicum	G7D846	-	-	+	O
Cesiribacter andamanensis	M7MX87	-	-	+	O
Colwellia psychrerythraea	Q47YG3	-	-	+	O
Corallococcus coralloides	H8MJ17	+	-	-	M
Leptospira kirschneri	M6X652	+	-	-	M
Magnetospirillum magneticum	Q2VZD0	-	-	+	O
Selenomonas ruminantium	I0GLW8	-	-	+	O
Veillonella spp.	F9N508	-	-	+	O
Vesicomyosocius okutanii	A5CVM8	-	-	+	O

***: M:** Mitochondrial, **SP:** Signal peptide (secretory pathway), **mTP:** (mitochondrial targeting peptide), and **O:** other cellular localization.

The prediction results for plant β-CAs defined that six out of eight β-CAs (P17067, Q8LSC8, P27141, D7TWP2, I2FJZ8, and B9GHR1) probably contain a chloroplastic localization signal (Table 2, original communication II). For the HGT study, three bacterial β-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 were positive. The prediction tool provided no definitive result for the other bacterial, metazoan, protozoan, insect, and nematode β-CA proteins.

5.1.4. Prediction of antigenic sites in β-CAs

According to the acceptable 3-85 residue variation in epitope length of an antigen (Singh, Ansari et al. 2013) and default parameters of EMBOSS database, the minimum length of an antigenic region in this set of β-CAs is six amino acid residues. The predictions of antigenic sites in the 31 β-CA proteins are shown in Table 3, original communication II. The highest score belongs to the most antigenic site.

5.1.5. Homology modelling

Homology models of four selected β-CAs from *A. caninum* (UniProt ID: FC551456), *A. suum* (UniProt ID: F1LE18), *T. spiralis* (UniProt ID: E5SH53), and *E. histolytica* (UniProt ID: C4LXK3) verified the predicted localization of conserved residues in the active site (Figure 8).

To study the molecular availability of the predicted main antigenic epitope, surface exposure of the homology model created from PDB model 1EKJ (β-CA from *P. sativum*) and the homology model based on the β-CA sequence from *A. caninum* were studied by visualizing the molecular surface (Figure 9). The analysis revealed that the majority of the antigenic epitopes were mainly buried within the β-CA protein structure.

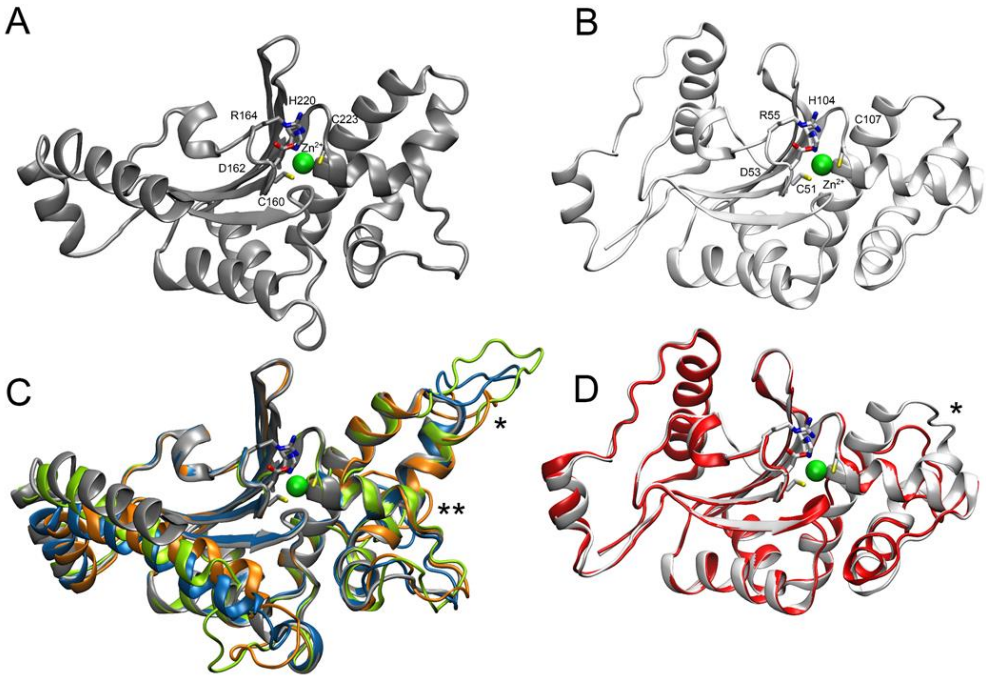


Figure 8. Homology modelling. (A) Cartoon presentation of β-CA from *P. sativum* (PDB 1EKJ). The Zn^{2+}ion is shown as a green sphere and the residues in direct contact with the ion are shown as liquorice models and labeled with residue numbers. (B) β-CA from *M. tuberculosis* (PDB 2A5V, light gray). (C) Aligned homology models of β-CAs from *A. caninum* (green), *A. suum* (blue), and *T. spiralis* (orange) are shown with PDB 1EKJ (gray). (D) Homology model of and *E. histolytica* (red) structurally aligned with PDB 2A5V (light gray). Highly variable loop regions are indicated by stars ("*" and "**") (C, D). The figure was prepared by using VMD (version 1.9.1) (Figure adopted from original communication II).

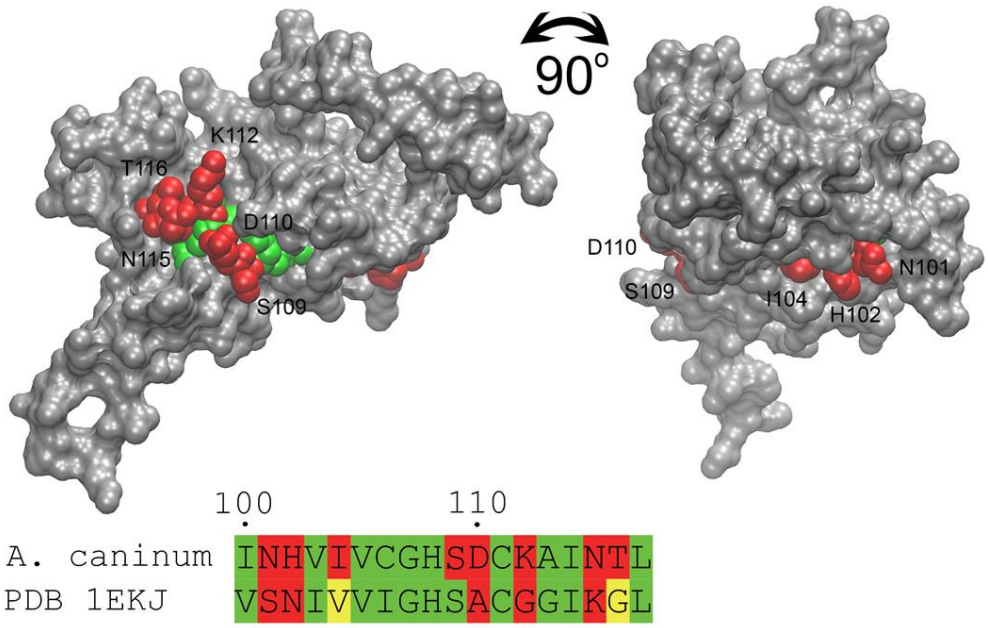


Figure 9. Determination of the localization of the predicted antigenic epitope. The molecular surface of the homology model of β-CA from *A. caninum* is shown as solid grey and the target epitope sequence was excluded from the surface presentation. The epitope residues exposed to solvent are displayed as red VdW spheres and numbered, while buried residues are displayed with green spheres. An alignment containing PDB 1EKJ and the corresponding sequence from *A. caninum* predicted β-CA is shown. The numbering of the residues in the alignment is according to the *A. caninum* sequence. The yellow residues in the alignment indicate partially buried structure (Figure adopted from original communication II).

In the case of horizontal transfer of *β-CA* genes from prokaryotes to protozoans, insects, and nematodes, homology modelling further supported the idea of the high similarity within the analyzed protein groups; no large insertions or deletions were observed. The superimposed homology models created from a pair of proteins from each clade (A, B, C, and D) of the β-CAs are shown in Figure 10.

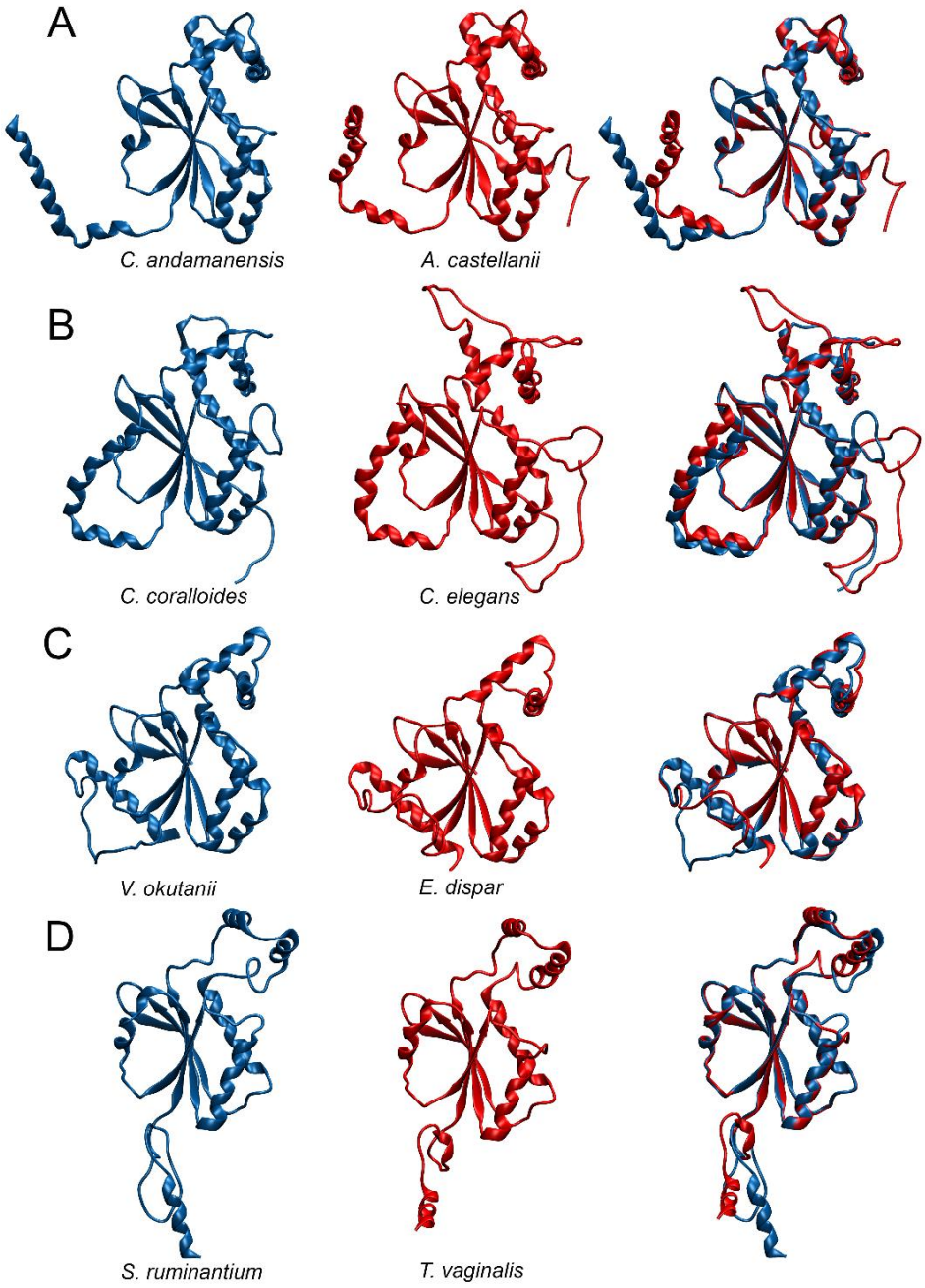


Figure 10. Homology models of representative pairs of β-CAs from clades A, B, C, and D. The blue protein models belong to prokaryotic proteins and the red models to eukaryotic proteins. The superimposed models are shown in the third column at the right side (Figure adopted from original communication V).

5.1.6. Identification of β-CA, transposase, integrase, resolvase, and conjugation complex protein (CCP) genes on the prokaryotic mobilomes

In order to study the genomic context of *β-CA* genes and to understand the molecular mechanisms involved in HGT, we explored the association of prokaryotic *β-CA* genes in mobilomes or MGEs (plasmids and GIs). The ACLAME version 0.4 database (http://aclame.ulb.ac.be/) (Leplae, Lima-Mendez et al. 2010) enabled us to first identify a *β-CA* gene within the pSLT mobile genetic element of *S. 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 the discovery of 40 *β-CA* genes located within plasmids in different prokaryotic species. Each bacterial plasmid contained only one *β-CA* gene sequence and occasionally several transposase, integrase, resolvase, and CCP coding genes. β-CAs, transposase, integrase, resolvase, and CCPs were identified by specific coding IDs from ACLAME and GenBank, and only one representative of each protein is listed (Additional file 3, original communication V) for bacterial species as an example. The locations of β-CA, transposase, integrase, and resolvase gene sequences in plasmid pSLT and pOU1113 from *S. typhimurium* (strain LT2) and *S. enterica*, respectively, are shown in Figure 11.

Through identification of *β-CA* genes from mobilomes, 18 plasmids (Table S1, original communication VI) and 49 chromosomal GIs (Table S2, original communication VI) were detected from archaeal and bacterial species, which contain *β-CA* genes, such as chromosomal GI-encoded β-CA of *M. tuberculosis* (strain H37Rv) (Figure 12). Three out of 18 plasmid located β-CAs were on linear plasmids including pAH1134_566, pRHL1, and pFRL6 from *Bacillus cereus* (strain AH1134) (strain F12), *Rhodococcus jostii* (strain RHA1), and *Streptomyces* spp., respectively. One GI-encoded β-CA, out of 49, was identified in archaea *S. islandicus* (Strain HVE10/4).

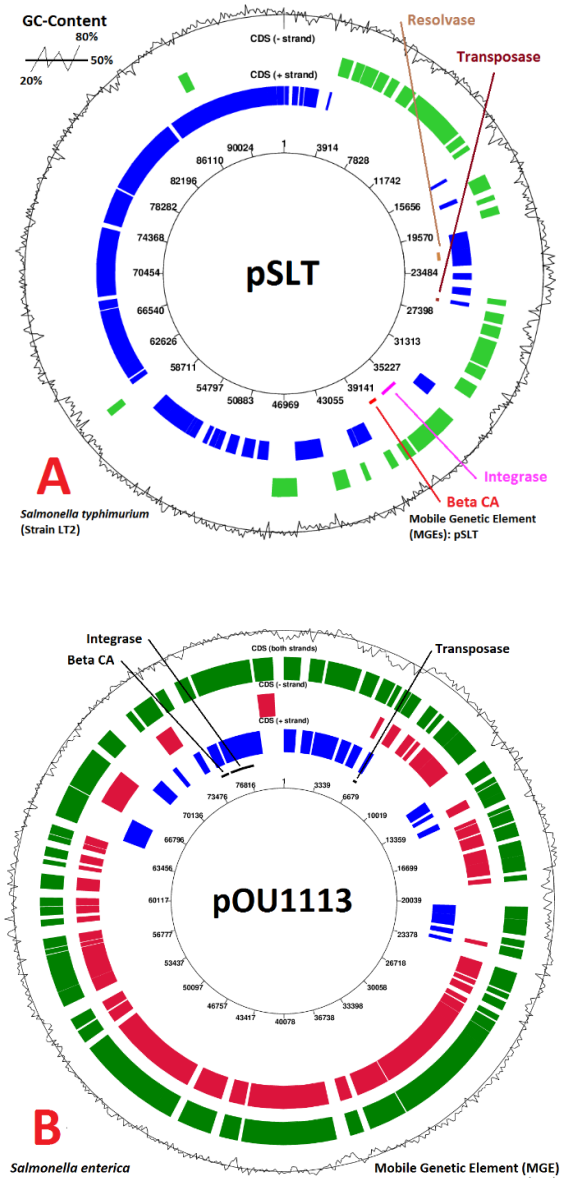


Figure 11. Circular structure of plasmid from *Salmonella* species. (A) The mobile genetic element pSLT from *S. typhimurium* (strain LT2) contains β-CA (37,528-38,268 bp), transposase (25,877-26,140 bp), integrase (35,113-36,777 bp), and resolvase (21,466-22,248 bp) genes. (B) The mobile genetic element pOU1113 from *S. enterica* contains β-CA (74,210-74,950 bp), transposase (6,653-7,003 bp), and integrase (75,157-77,365 bp) genes. The line graph along outer circumference of both plasmid models represents GC-content, which is higher or lower than baseline (50%) (Figure adopted from original communication V and VI).

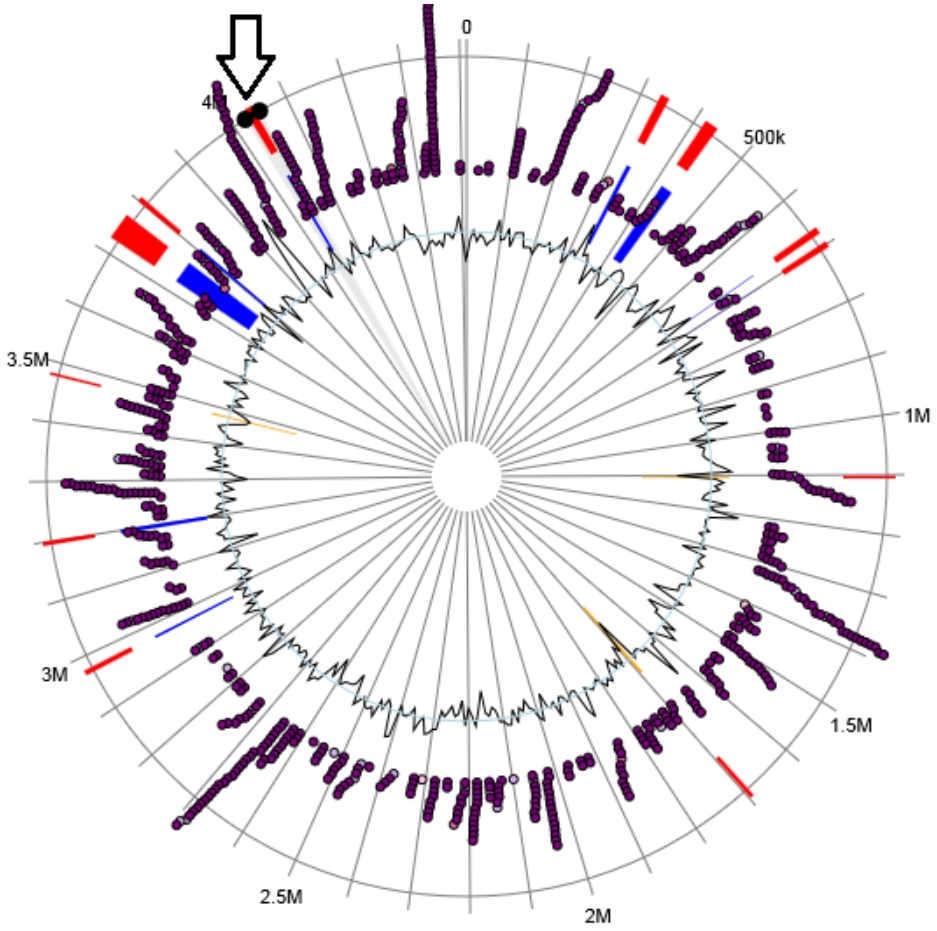


Figure 12. The main chromosome of *M. tuberculosis* (strain H37Rv). The *β-CA* gene (Gene name: *canB*, NCBI ID: NP_218105.1, and sequence: 4,029,871-4,030,494) from *M. tuberculosis* was located on the chromosomal GI, which is shown by an arrow. The GIs were predicted by IslandViewer version 3. In this figure, IslandViewer applied both IslandPath-DIMOB (Hsiao, Wan et al. 2003) and SIGI-HMM (Waack, Keller et al. 2006) methods (blue and yellow color bands, respectively) to predict GIs on the main genome. The final prediction result appears as red color bands. β-CA was predicted as a virulence factor in *M. tuberculosis* (strain H37Rv) using the virulence factor database (VFDB) (Chen, Zheng et al. 2016). The height of purple columns indicate the strength of prediction for the presence of virulence factors (Figure adopted from original communication VI).

5.1.7. Identification of *β-CA* gene sequences in protozoan, insect, and nematode genomic DNA

Analysis of the precise location of *β-CA* gene sequences in protozoan, insect, and nematode genetic structures revealed that all were located in chromosomal DNA. Exon counts, for the group of studied *β-CA* gene sequences, vary in quantity from 1 to 11. The maximum exon counts were eight for *A. castellanii* (Entry ID: L8GR38) and 11 for *P. pacificus* (Entry ID: H3EVA6) for protozoan and nematode species, respectively. Interestingly, some protozoan *β-CA* genes included only one exon (Additional file 4, original communication V). Analysis of the genes on circular mitochondrial DNA from *A. castellanii* revealed that none of the protozoan *β-CA*s were considered mitochondrial coding genes (data not shown).

5.1.8. Identification of IS elements from archaeal and bacterial plasmids and chromosomal genomes

Identification of IS elements revealed that both plasmids and bacterial main chromosomes contain IS elements (Table S2, original communication VI). IS elements are able to express only one of either transposase or integrase. Some IS elements contained two separate starting codons and overlapped ORFs for both transposase and integrase, including IS families IS3 (ISAzs24, ISMra1, ISMra4, ISRel21, ISRhosp5, ISSen3, ISSen4), IS21 (ISAcma26, ISAzs2, ISRel5, ISRel16), and IS630 (ISRel6) from circular plasmids and IS3 (ISYps8), IS6 (ISMtsp1, ISMtsp2, ISMtsp4), and IS21 (ISFK1) from bacterial main chromosomes. The only predicted IS element from a linear plasmid was IS3 (ISRhosp5) from pRHL1 of *R. jostii* (strain RHA1).

5.1.9. Structural and functional predictions of β-CAs based on AlBCA (*A. lumbricoides* β-CA) protein sequence

Analysis of the AlBCA protein sequence by the InterPro database resulted in classification as part of the CA family (InterPro ID IPR001765). In the CATH/Gene3D database, AlBCA was categorized with the CATH superfamily ID 3.40.1050.10 (β-CA; Chain A). The tools of the CATH/Gene3D database were used to generate a rainbow model for superimposition of AlBCA protein with other close relative β-CA proteins (Figure 2, original communication III).

Metabolic pathway analyses of the AlBCA protein sequence in the KEGG and METACYC databases predicted that the enzyme has a critical function in nitrogen metabolism (Figure 13) and gluconeogenesis II pathways (Figure 14).

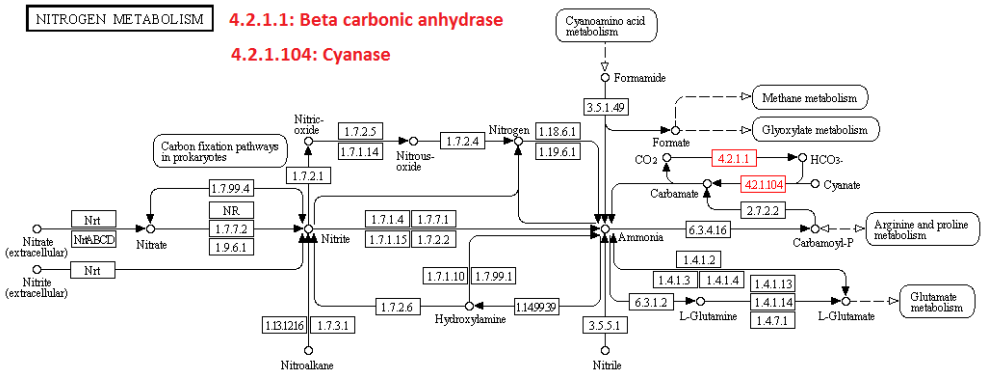


Figure 13. The predicted role of AlBCA in nitrogen metabolism pathway. This pathway has been generated by the KEGG (Kyoto Encyclopedia of Genes and Genomes) database. According to this model, AlBCA catalyzes the conversion of CO_2 and H_2O to HCO_3^- and cyanase converts cyanate to CO_2 and NH_3 in a bicarbonate-dependent reaction (Figure adopted from original communication III).

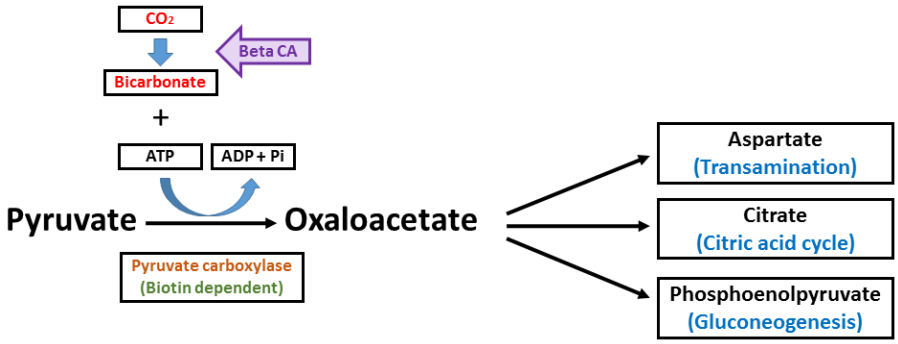


Figure 14. Role of bicarbonate generated by AIBCA in gluconeogenesis II pathway. The conversion of CO_2 to HCO_3^- is catalyzed by β-CA, which is shown with a purple arrow. HCO_3^- is a major element in conversion of pyruvate to oxaloacetate. Oxaloacetate is used to generate aspartate, citrate, and phosphoenolpyruvate through transamination, citric acid cycle, and gluconeogenesis pathways, respectively (Figure adopted from original communication III).

Results from the KEGG database suggested that AlBCA might functionally take part in detoxification of cyanate by providing bicarbonate for cyanase enzyme. The METACYC database also predicted bicarbonate as the product of the β-CA catalytic reaction. This bicarbonate would be needed for the mitochondrial gluconeogenic pathway where pyruvate is converted to oxaloacetate.

5.2. Results from experimental studies

5.2.1. Polymerase chain reaction (PCR) and sequencing of *T. pseudospiralis* and *T. nativa* β-CA genomic sequences

Among the four different pairs of primers (No. 1-4) only the primer set No. 2 (PCR product size 191 bp) showed a clear positive DNA band on agarose gel electrophoresis for detection of *T. spiralis* β-CA genomic sequence in larvae and trichinella-infected meat samples (data not shown). It also produced positive bands on larval samples from *T. pseudospiralis* and *T. nativa* (Figure 15).

The next rounds of PCR analyses were performed to find out whether it was possible to design primers specific for *T. pseudospiralis* and *T. nativa*. We used the same reverse primer (primer set No. 2), which worked well for the first PCR round. The PCR products representing partial sequences of *T. pseudospiralis* and *T. nativa β-CA* gene were sequenced and the forward primers were designed based on the obtained sequences. These primers were included in the sets No. 5-7 for *T. pseudospiralis* and 8-11 for *T. nativa* (Figure 3, original communication IV). Most primers produced multiple bands.

When the most promising primers (No. 6 for *T. pseudospiralis* and No. 10 for *T. nativa*) were tested against the three species of *Trichinella*, they completely cross-amplified because of the high sequence similarity (data not shown). Due to the interspecies cross-amplification, we further designed four new reverse primers, which were used together with the previous forward primers resulting in primer sets No. 12-16.

The results revealed that the primer set No. 13 for *T. nativa* and primer set No. 12 for *T. pseudospiralis* produced the strongest DNA bands on the agarose gel electrophoresis (Figure 4, original communication IV).

For evaluation of cross-amplification, these primers were further tested on all three *Trichinella* spp. (Figure 5, original communication IV), which revealed that there was again a complete cross-amplification between different *Trichinella* spp. by these primers. This finding led us to conclude that finding of species-specific primers for *Trichinella β-CA* genes is probably impossible due to the very high sequence similarity. Based on our results, primer set No. 2 was considered the best potential tool for diagnostic purposes, even though it was not able to discriminate between different

Trichinella spp. Therefore, its specificity was further tested on prepared samples from *T. gondii*, *P. equorum,* and *T. cati*. No false-positive reactions were observed for these parasites (Figure 16).

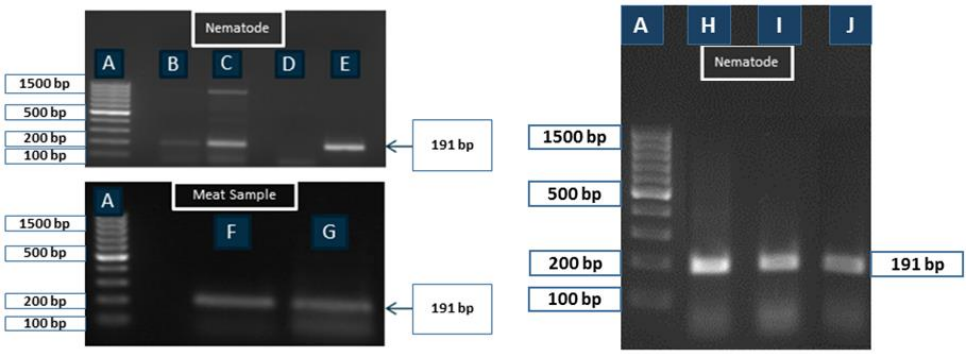


Figure 15. PCR-based detection of genomic β-CA sequences from *Trichinella* larvae and infected meat samples. The primer set No. 2 showed a clear PCR product (size= 191 kb) in all *T. spiralis* -positive samples. (A) 100 bp ladder; (B) one larva; (C) five larvae; (D) negative control (PCR reaction mixture without a digested sample); (E) more than ten larvae; (F and G) 5 mg of *T. spiralis* infected mouse meat samples; (H) *T. nativa* (ten larvae); (I) *T. pseudospiralis* (ten larvae); (J) *T. spiralis* (ten larvae) (Figure adopted from original communication IV).

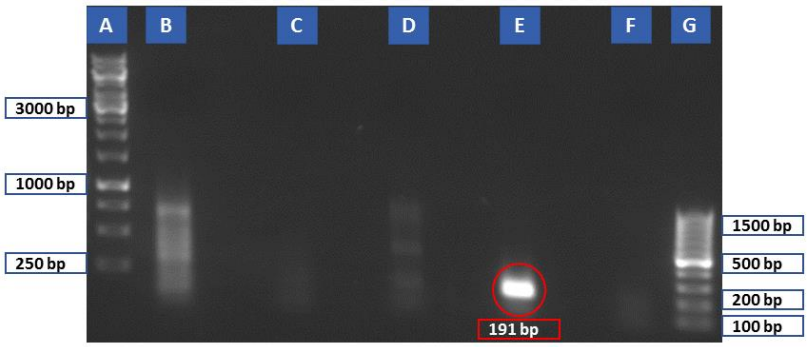


Figure 16. Evaluation of primer set No.2 on different lysed parasites. (A) 1 kb ladder, (B) *T. gondii*, (C) *T. cati*, (D) *P. equorum*, (E) *T. spiralis* as the positive control, (F) negative control, and (G) 100 bp ladder. Red color circle shows 191 bp PCR product of primer set No.2 on *T. spiralis* (Figure adopted from original communication IV).

5.2.2. Production of recombinant AlBCA

A *β-CA* gene insert, including a thrombin cleavage site and a segment encoding a poly-histidine-tag, was produced by GeneArt® technology and integrated into a pFastBac1 plasmid. The purified plasmid was used to produce recombinant AlBCA in sf-9 cells. SDS-PAGE containing the eluted fraction showed three polypeptide bands of 28, 30, and 33 kDa. The 30-kDa band was the major product, which corresponds to the calculated molecular mass of AlBCA (Figure 17).

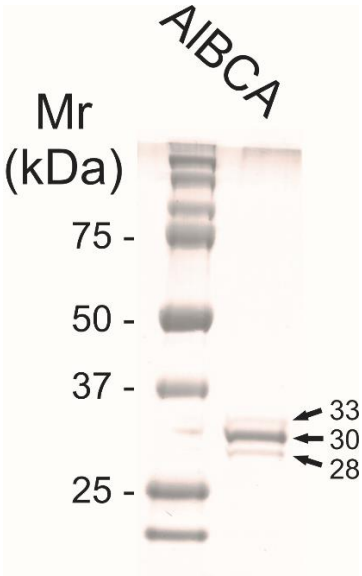


Figure 17. SDS-PAGE of AlBCA. Left lane shows the protein standard. The purified recombinant AlBCA appeared as a triple band (28, 30, and 33 kDa). The predicted molecular mass of His-tagged AlBCA is 30 kDa (Figure adopted from original communication III).

5.2.3. Kinetic characterization of AlBCA

We measured the kinetic properties of AlBCA for the physiologic reaction of CO_2 hydration to bicarbonate and protons at pH 8.3. We compared the activity of the new enzyme with those of the well characterized α- (hCA I and II) and β-CAs from *L. donovani chagasi* (LdcCA), *D. melanogaster* (DmBCA), and *A. gambiae* (AgaCA), characterized earlier by our group (Syrjanen, Tolvanen et al. 2010, Syrjanen, Vermelho et al. 2013, Syrjanen,

Kuuslahti et al. 2015). As shown in Table 9, AlBCA has significant catalytic activity in the physiologic reaction with the following kinetic parameters: k_{cat} of 6.0×10^5 s^{-1} and k_{cat}/K_m of 4.3×10^7 M^{-1} s^{-1}. The measured kinetic parameters show that AlBCA has a moderate activity, comparable to that of hCA I, and in the same range as the enzymes characterized in *A. gambiae* and *L. donovani chagasi* (Syrjanen, Tolvanen et al. 2010, Syrjanen, Vermelho et al. 2013). This activity was effectively inhibited by the classical sulfonamide compound, AAZ, which showed an inhibition constant of 84.1 nM. Although this was the only inhibitor investigated so far for this enzyme, work is in progress to evaluate other classes of inhibitors against this enzyme, i.e., anions, sulfonamides and their isosteres, and dithiocarbamates and their isosteres.

Table 9. Kinetic results for the hydration reaction of CO_2 catalyzed by the β-CAs from *L. donovani chagasi* (LdcCA), *D. melanogaster* (DmBCA), *A. gambiae* (AgaCA), *A. lumbricoides* (AlBCA), and human α-CAs (hCA I and II).

Enzyme	CA class	k_{cat} (s^{-1})	k_{cat}/K_m (M^{-1} s^{-1})	K_i (AAZ) (nM)*	References
hCA I	α	$2.0x10^5$	$5.0x10^7$	250±12	(Vullo, Del Prete et al. 2014)
hCA II	α	$1.4x10^6$	$1.5x10^8$	12±0.8	(Vullo, Del Prete et al. 2014)
LdcCA	β	$9.35x10^5$	$5.9x10^7$	91.7±5.7	(Syrjanen, Vermelho et al. 2013)
DmBCA	β	$9.5x10^5$	$1.1x10^8$	516±24	(Syrjanen, Parkkila et al. 2014)
AgaCA	β	$7.2x10^5$	$5.6x10^7$	27.3±2.0	(Syrjanen, Kuuslahti et al. 2015)
AlBCA	β	$(6.0 \pm 0.1)x10^5$	$(4.3 \pm 0.2)x10^7$	84.1±2.9	-

* Mean ± standard error from three different assays.

6. Discussion

6.1. Identification of β-CAs in parasites and prokaryotic endosymbionts

Our study identified that the β-CA enzyme is present in a variety of parasites and vectors, including protozoans, insects, and nematodes. At the beginning of our studies, a total of 75 β-CA protein sequences were identified from protozoan and metazoan species. During the course of our studies, the number of identified β-CA sequences increased to 96 cases in protozoans or metazoans, including parasites and pests. The MSA results showed that all β-CA protein sequences from parasites contain the first (CXDXR) and second (HXXC) highly conserved residue motifs. Most of the protozoan and metazoan β-CA coding sequences were designated as uncharacterized sequences or CAs with no class specification. Based on our work, these can be now annotated as β-CAs in proteomics and genomics databases.

Our analysis revealed that some of the identified β-CAs are from agricultural pests and livestock parasites (zoonotics), among which the most significant species included: *A. caninum, Ascaris lumbricoides* (*A. suum*), *C. clemensi*, *C. quinquefasciatus*, *Entamoeba* spp., *H. contortus, I. multifiliis*, *L. salmonis, N. americanus, T. spiralis*, and *T. adhaerens*. One was an important pest of food industries (*T. castaneum*). There was also an orchard invasive dipteran fruit fly (*Ceratitis capitata*) and three pests of wood industries, such as *Camponotus floridanus, Dendroctonus ponderosae,* and *Solenopsis invicta.* In addition, *T. spiralis* is considered as a meat-contaminating parasite. Our study defined that endosymbiont prokaryotes from protozoans, insects, and nematodes contained the first (CXDXR) and second (HXXC) highly conserved motifs of β-CA proteins as well as their unicellular or multicellular eukaryotic hosts.

6.2. Prediction of subcellular localization of β-CA protein sequences

Prediction of subcellular localization defined that some bacterial (*A. felis*, *L. kirschneri*, and *C. coralloides*), protozoan (*A. castellanii*, *L. braziliensis*, *L. guyanensis*, *S. culicis*, and *T. thermophila*), insect (*A. aegypti, A. gambiae*, *A. pisum, A. vaga, C. capitata, C. quinquefasciatus, Danaus plexippus, Drosophila* spp., and *Heliconius melpomene*), and nematode (*Caenorhabditis*

spp., *H. contortus, S. kowalevskii,* and *T. spiralis*) species, include mitochondrial signals in their β-CA protein sequences. It is well established that prokaryotes and some anaerobic protozoa, such as *Giardia lamblia*, *E. histolytica, T. vaginalis*, *Cryptosporidium parvum*, *Blastocystis hominis*, *Encephalitozoon cuniculi*, *Sawyeria marylandensis*, *Neocallimastix patriciarum*, and *Mastigamoeba balamuthi* completely lack mitochondria (Makiuchi and Nozaki 2014). Vast numbers of studies have hypothesized that a majority of the anaerobic parasitic protozoan mitochondrial genes have been acquired from α-proteobacterial genomes (Makiuchi and Nozaki 2014), such as horizontal transfer of Monoamine oxidase (an outer membrane enzyme of mitochondria for metabolism of neuromediators) genes from ancestral prokaryote to vertebrate lineages (Lander, Linton et al. 2001).

Therefore, we hypothesize that sequences similar to mitochondrial localization signals emerged in β-CA proteins in prokaryotes, leading to their mitochondrial localization after HGT into protozoans, insects, and nematodes. Supporting this idea, the β-CA of *D. melanogaster* has been experimentally shown to be localized in mitochondria (Syrjanen, Tolvanen et al. 2010).

6.3. Horizontal gene transfer (HGT)

Identification of highly mobile transposase, integrase, and resolvase genes in archaeal and bacterial plasmids and chromosomal genomes indicates that these enzymatic recombination tools were expressed by IS elements and played a major role in horizontal transfer of *β-CA* genes from prokaryotic mobilomes, including plasmids and GIs. Identification of β-CAs with ancillary 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, insect, and nematode species using complex mechanisms and create a conducive environment for integration of *β-CA* gene sequences. During HGT, *β-CA* genes from prokaryotic plasmids are transferred to the genome of another prokaryote (Pok-Pok) or eukaryote (Pok-Euk) (Figure 18). In another HGT possibility, *β-CA* genes are transferred from prokaryotic chromosomal GIs to the different genomic target sites of another prokaryotic or eukaryotic species (Figure 19).

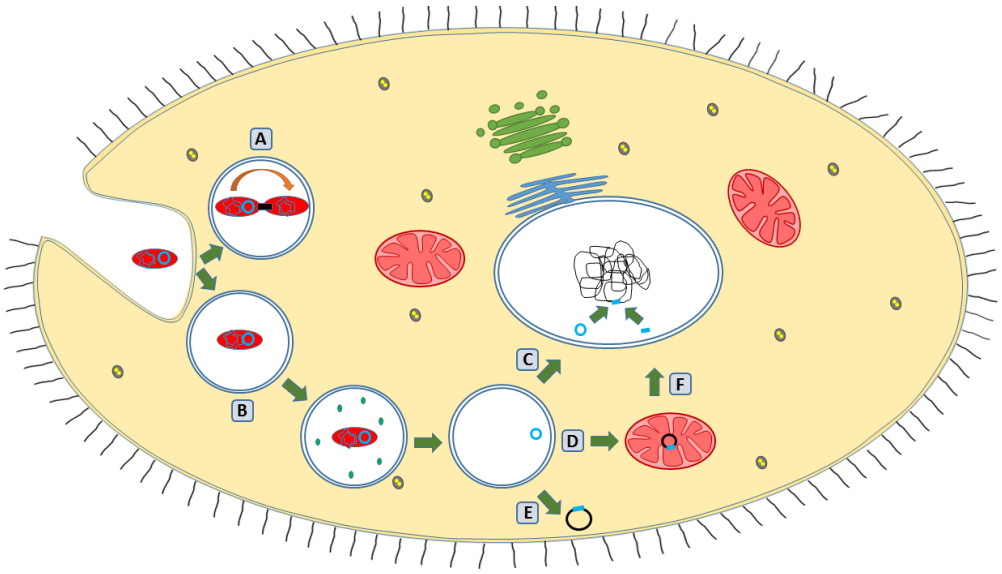


Figure 18. Horizontal transfer of prokaryotic *β-CA* genes from plasmids to other prokaryotic or protozoan species. (A) After endocytosis of bacteria by a protozoan ciliate, a bilayer vesicle or vacuole is formed, which surrounds the bacteria. Bacterial species are able to horizontally transfer *β-CA* genes from plasmids to another recipient bacterial species by conjugation pilus. (B) Bacterial species, which enter into the bilayer vacuole through endocytosis survive as intracellular endosymbionts. Therefore, horizontal transfer of prokaryotic *β-CA* genes to the protozoan genome can follow through four possibilities: (C) Prokaryotic *β-CA* gene from plasmid is integrated into the protozoan main chromosome, (D) mitochondrial DNA, or (E) mobilome. (F) In addition, a *β-CA* gene integrated into circular mitochondrial DNA might have subsequently migrated from mitochondria to the main chromosome (Figure adopted from original communication VI).

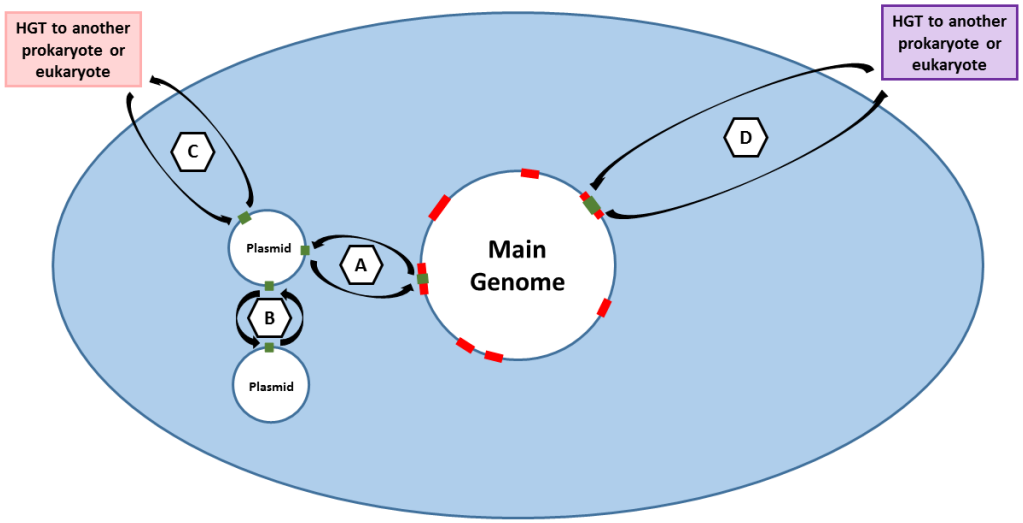


Figure 19. Schematic representation of horizontal transfer of prokaryotic *β-CA* genes from chromosomal GIs to other hosts. (A) Horizontal transfer of a *β-CA* gene from a plasmid to the GI region of the main chromosome of the same prokaryote. (B) Horizontal transfer of a *β-CA* gene between two different plasmids from the same prokaryote. (C) Horizontal transfer of a *β-CA* gene from a prokaryotic plasmid to another prokaryote or eukaryote, and (D) Horizontal transfer of a β-CA gene from the GI region of the main genome to another prokaryote or eukaryote (Figure adopted from original communication VI).

Our current findings may shed some light into the question of why *β-CA* gene sequences are completely absent in the genomes of vertebrates. In protozoan and invertebrate metazoans, including insect and nematode species, *β-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 parasitic host. There are some possible reasons for the lack of subsequent horizontal transfer of *β-CA* genes to the vertebrate genomes. First, there may not be a specific transposable element insertion site for these enzymatic cutters within vertebrate genomes. Second, vertebrates are complex multicellular organisms in which evolutionarily stable integration of *β-CA* gene sequences would need to have taken place in the germ cells that give rise to egg and sperm cells (Andersson, Doolittle et al. 2001). Hence, the lack of *β-CA* genes from vertebrate's genomes is understandable, especially because there is no evolutionary pressure for the aquisition of another CA class due to the presence of several efficient α-CAs in all vertebrates. Hence, the main reason for the absence of *β-CA* gene in the vertebrate genomes might be related to β-CA-coding mobilomes, which were not able

to detect an integration site on the vertebrate egg or sperm genomes to insert the *β-CA* genes as exonic sequences.

During analysis of β-CA evolution, one would expect that high similarity between β-CA 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 β-CAs, but not in the β-CAs used as a template in homology modelling. 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 through HGT, due to their origins in a relatively recent identical genetic source, not because of selection pressure towards the particular residue observed in each position.

6.4. Phylogenetic analysis

The phylogenetic analysis results revealed that significant duplications of *β-CA* genes have occurred in the protozoan *T. thermophila* and *P. tetraurelia*, resulting in eight and five distinct proteins, respectively. The presence of β-CA proteins in various pests and parasites and their absence in vertebrates suggests that these enzymes could be potential targets for the development of novel anti-parasitic agents or pesticides with minimal adverse effects on vertebrates. A key requirement for such novel β-CA inhibitors would be the high isoform specificity. The distinction among β-CA proteins elucidated in the phylogenetic tree suggests that inhibitors can probably be designed, which would target β-CAs specific to different groups of species; thus leaving those in other species, such as plants, unaffected. Unfortunately, design of highly specific inhibitors will require proper structural data of proteins based on crystallography. Thus far, β-CA crystal structures are available only from a few different species in the PDB database, including some algae, bacteria, archaea, yeast, and a plant *P. sativum* (Kimber, Coleman et al. 2000).

Our phylogenetic analysis of HGT of *β-CA* genes from prokaryotes to protozoans, insects, and nematodes, revealed that *A. castellanii* possesses two *β*-CA genes, one from clade A and one from clade C (Table 7) (Figure 2 and Table 2, original communication V). The *β-CA* gene of *A. castellanii* (L8GLS7) was potentially horizontally transferred from a prokaryotic species, which probably was a common ancestor of *B. japonicum* (G7D846) and *A. felis* (K8NQ88). Moreover, previous studies have shown that *B. japonicum* (La Scola, Mezi et al. 2002) and *A. felis* (La Scola and Raoult 1999) are prokaryotic endosymbionts of *A. castellanii*.

Phylogenetic analysis of clade A (Table 7) of β-CAs, showed that all β-CA enzymes in *N. gruberi* and *P. tetraurelia* protozoa have a common source with the single β-CA from spirochaetes bacteria, *L. kirschneri* (M6X652). Potentially, after HGT of a *β-CA* gene from the common source to these two protozoan hosts, the gene duplicated and created five different *β-CA* genes for *P. tetraurelia* (A0BD61, A0CEX6, A0C922, A0BDB1, A0E8I0) and three for *N. gruberi* (Predicted 1, 2, 3).

Among the various prokaryotic endosymbionts, it is proposed that *Dictyostelium* spp., *I. multifiliis*, and *T. thermophila*, 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. Thereafter, gene duplication in these protozoans led to multiple copies of *β-CA* in *Dictyostelium* spp. (Q555A3, Q55BU2, Q94473, F0Z7L1, F4PL43), *I. multifiliis* (G0QYZ1, G0QPN9), and *T. thermophila* (Q22U21, Q22U16, I7M0M0, I7M748, I7LWM1, I7MDL7, Q23AV1, I7MD92). Indeed, previous studies have defined some major similarities between *I. multifiliis* and *T. thermophila,* which also suggested a common evolutionary origin for these species. These similarities included: (1) encoding similar small subunit rRNAs from rDNA of *I. multifiliis* and chromosome of *T. thermophila* (Butler, Yasuda et al. 1995, Coyne, Hannick et al. 2011), (2) the identical genome size and gene contents of mitochondrial DNA and having five identical tRNA genes with the same locations and orientations in both protozoan species (Brunk, Lee et al. 2003), (3) the identical ancestral origin for three basal body and centriole proteins (Carvalho-Santos, Machado et al. 2010), and (4) the same ciliopathy genes in both protozoan genomes (Hodges, Scheumann et al. 2010).

Our phylogenetic results identified that *β-CA* genes in Trypanosomatidae, including *A. daenei* (S9WXX9), *S. culicis* (S9TM82), and *Leishmania* spp. (A4H4M7, E9B8S3, A4HSV2, Q4QJ17, E9AKU0, and S0CTX5), have a common source with an alphaproteobacterium similar to *M. magneticum* (Q2VZD0).

The phylogenetic analysis showed that insect and nematode β-CAs belong to clade B and also suggested that they may have a common source with myxobacterial β-CAs. The various myxobacteria *Corallococcus, Enhygromyxa, Stigmatella,* and *Myxococcus*, are part of the same subtree that contains insect and nematode β-CAs. However, a larger analysis with more insect, nematode, and plant β-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 a *β-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 phyologenetic trees do not provide conclusive evidence for this, and it is thus possible to speculate that the β-CAs of clade B, which we see in insects and nematodes, have been retained from an ancestral eukaryote.

Phylogenetic analysis of clade C showed that *β-CA* genes from *Entamoeba* spp. (B0E7M0, 1C4LXK3, K2GQM0) have a common source with the *β-CA* gene of gammaproteobacterium *V. okutanii* (A5CVM8). Based on the obtained result, we propose that *β-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 revealed that *β-CA* genes in *T. vaginalis* (A2ENQ8, A2DLG4) have a common source with *β-CA* genes from firmicutes bacteria *S. ruminantium* (I0GLW8) and *Veillonella* spp. (F9N508). Prior results have shown that *Clostridium sordellii* and *Veillonella* spp. from firmicutes phylum and *T. vaginalis* have a symbiotic interaction in sexual organs of animals (Smutna, Goncalves et al. 2009, Fichorova, Buck et al. 2013), providing the environment in which a transfer of firmicutes bacteria *β-CA* genes into the *T. vaginalis* genome is possible. However, it is tempting to assume that β-CAs of all four clades in protozoans, insects, and nematodes were derived by HGT from prokaryotes.

More detailed analysis of the phylogenetic information, especially focusing on the "Golden Box", identified that both GIs and plasmids played a critical role in horizontal transfer of *β-CA* genes within archaea and bacteria (both cyanobacteria and proteobacteria). The "Golden Box" clearly showed that *β-CA* genes from proteobacteria *Acidithiobacillus thiooxidans* (S5FU55), and cyanobacteria *Hassallia byssoidea* (A0A0C1XKI0), *Calothrix* spp. (K9PEP8), and *Cyanothece* spp. (E0UFL0) have a common ancestor with archaea *Halopiger xanaduensis* (F8DDX3), which is located on circular pHALXA01. In addition, we could determine that pHALXA01 coding β-CA from archaea *H. xanaduensis* (F8DDX3) has a common ancestor with *β-CA* gene from chromosomal GI of archaea *S. islandicus* (F0NNH7). In addition, we could identify no plasmids for HGT of *β-CA* genes. Therefore, the results suggest that archaeal species mostly employed chromosomal GIs for HGT of *β-CA* genes. This prediction seems logical, because after the origin of archaeal species around 3.5 billion years ago (Doolittle 1997), they probably utilized chromosomal GIs more than plasmids for HGT. Rate of evolutionary change in this period of life could have increased because bacteria were equipped with more accessory tools, including chromosomal GIs, plasmids, and IS elements, in the presence of atmospheric oxygen (Falkowski and Godfrey 2008, Sobecky and Hazen 2009).

6.5. Inhibitory and kinetic studies on β-CAs from parasites

Since β-CAs are present in parasites (protozoans, helminths, and arthropods) and absent in vertebrates, this makes β-CAs good potential parasitic-specific targets for therapeutic and diagnostic purposes. The kinetic study on the purified recombinant AlBCA defined that the recombinant β-CA from *A. lumbricoides* has moderate CA enzymatic activity. Nanomolar concentration of AAZ (84.1 nM) was enough to inhibit AlBCA. MSA analysis of the AlBCA sequence with β-CAs from *C. elegans* revealed that β-CA2 (isoform c, Uniprot ID: Q2YS41) is most similar to AlBCA. The expression pattern study of *C. elegans* β-CA2 reveals that this enzyme is expressed in all larval (EE, LE, L1, L2, L3, L4) and adult (including male and hermaphrodites) stages. The highest expression levels were detected in the body wall muscles of L2 stage. This enzyme is also present in larval

neurons, muscles, coelomocytes, hypodermis layer, intestine, excretory cells, whole body of adult male, and hermaphrodite gonads. In addition, *D. melanogaster* β-CA has shown the highest upregulated values of expression in the spermatheca (female), fat body, and heart of adults; as well as early larval stages and late in metamorphosis. The knockdown studies of *D. melanogaster* β-CA showed that females were unable to produce eggs, which resulted in infertility of female flies.

Figure 20 presents 14 categories of known α- and/or β-CA inhibitors, which are able to inhibit catalytic activity of these enzyme families (Supuran 2010, McKenna and Supuran 2014). Inhibition of CA activity would slow down some cellular biochemical pathways in pests and parasites, such as gluconeogenesis, nucleotide biosynthesis, fatty acid synthesis, gastrointestinal function, neuronal signaling, respiration, and reproduction. It is known that β-CAs are required for CO_2 sequestration within chloroplasts of plants and algae, and hence CA inhibition would affect the rate of photosynthesis (Badger 2003). Importantly, β-CA inhibition in fungi and *D. melanogaster* showed completely different inhibition profiles (Syrjanen, Tolvanen et al. 2010), suggesting that β-CAs of parasites can be inhibited with higher affinity than plant CAs by applying the right inhibitors and concentrations. On the other hand, the structural and functional predictions of β-CAs from parasites based on AlBCA revealed that β-CAs from parasites might play a crucial role in detoxification of cyanate in nitrogen metabolism and mitochondrial gluconeogenic pathways, where pyruvate is converted to oxaloacetate. Therefore, β-CAs from parasites are considered as good potential targets for inhibition to control or treat parasitic infections, such as ascariasis, by efficient inhibitor molecules, such as AAZ, to block cyanate detoxification, transamination, citric acid cycle, and gluconeogenesis pathways.

Another important goal is to discover proper inhibitors that are β-CA-specific and do not inhibit α-CAs. This would primarily require detailed structural information on selected parasitic CAs. The resolved structures would then allow high-throughput screening of chemical compounds, identification of the most promising inhibitors, and testing of potential inhibitors *in vitro* and *in vivo*.

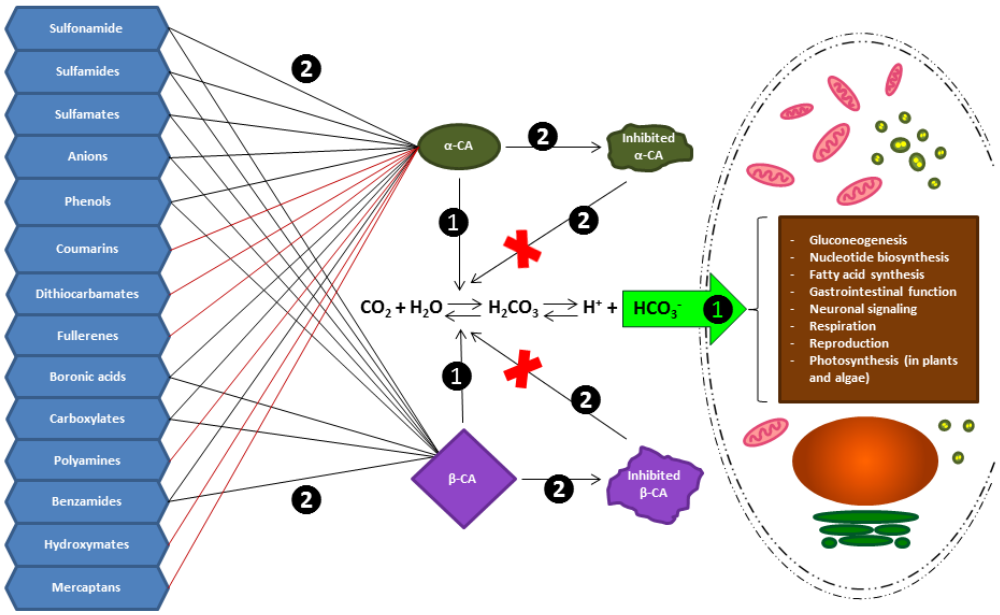


Figure 20. Effects of 14 CA inhibitors on α- and β-CAs of pests and parasites. Some compounds inhibit members of both α- and β-CA enzyme families. The brown box shows biological pathways where bicarbonate plays a critical role as a biochemical substrate. The ultimate goal of future research should be the creation of inhibitors specific to both enzyme families and to each isozyme. Ideally, the specific inhibitors would cause tissue- and organ-specific effects in pests and parasites with minimal adverse effects on other species. Number 1 shows α- and β-CA catalytic pathways and number 2 shows the inhibitory effects of α- and β-CA inhibitors (Figure adopted from original communication II).

6.6. Molecular diagnosis of parasitic infections based on β-CA genomic sequences

Our study on detection of β-CA genomic sequences of *Trichinella* spp. in meat samples, demonstrated that the method is robust enough for screening of *Trichinella* infection in routine meat samples. The specificity of the method was further tested on prepared samples from *T. gondii*, *T. cati*, and *P. equorum*, and no false-positive reactions were detectable. Other advantages of the present method include both ease of use and rapidity of detection. The meat samples were lysed overnight, after which the samples were ready for the PCR assay. The chosen reagents allowed PCR amplification without further need for time-consuming DNA isolation. Duration of both the PCR assay and gel

electrophoresis took about three hours, after which the positive bands became clearly visible for recording. Even though there are several diagnostic methods available for *Trichinella* infection, most of them are either expensive, time-consuming, or require special expertise. Therefore, our results revealed that the PCR amplification of β-CA genomic sequences seems to be *Trichinella* genus-specific and robust enough for routine screening purposes in comparison with the conventional direct microscopic detection. Based on these encouraging results, we are in the process of developing analogous PCR-based methods for several human parasites, which could be present in biological samples, such as stool or blood.

7. Summary and conclusions

In conclusion, this study presents the identification and partial characterization of some β-CAs from parasites, and discusses potential applications based on β-CAs in medicine, veterinary medicine, food industries, and agriculture. This study provides novel information about the evolution of β-CAs from parasites and predicts their subcellular localization. Hence, the major conclusions are:

1) Parasites and vectors including protozoans, insects, and nematodes contain *β-CA* genes and proteins in their genomes and proteomes, respectively.
2) The second highly conserved sequence (HXXC) of β-CAs is predicted to be the most antigenic site of the protein. This protein is intracellular in most organisms and this epitope is also mostly buried within the protein molecule. Therefore, it may not be accessible to activate the immune system of vertebrates. The main reason for the intramolecular location could be related to the hydrophobicity of the adjacent amino acids of "HXXC", such as alanine, isoleucine, and valine.
3) Recombinant AlBCA from *A. lumbricoides* showed moderate CA enzymatic activity and was efficiently inhibited by AAZ. Inhibition of β-CAs in *A. lumbricoides* or other parasites could represent a new chemotherapy strategy for treatment of parasitic diseases.
4) Genomic *β-CA* gene sequences from *Trichinella* species were defined as novel targets for innovative, robust, and straightforward diagnostic methods to detect *Trichinella* contamination in meat samples.
5) Prokaryotic endosymbionts of protozoans and metazoans (parasitic and non-parasitic insects and nematodes) contain mobilomes (plasmids and GIs), which have been equipped with some accessory enzymes, such as transposase, integrase, and resolvase as well as *β-CA* genes. These enzymatic tools might have prepared an environment suitable for horizontal transfer of *β-CA* genes from genomes of prokaryotic endosymbionts to protozoan, insect, and nematode hosts.

8. Acknowledgements

This study was performed at BioMediTech and the School of Medicine, University of Tampere between 2012 and 2016. At the beginning, I wish to say my special thanks to the deans of BioMediTech and the School of Medicine, Dr. Hannu Hanhijärvi (DDS, PhD) and Professor Matti Lehto (MD, PhD), respectively. Thank you for accepting me as a PhD researcher at your institutes, preparing a high-quality scientific atmosphere and providing advanced equipments and facilities, as well as expert laboratory technicians for scientific collaboration.

Dear Professor Seppo Parkkila (MD, PhD), you were more than a thesis supervisor for me. You granted a PhD position to me and prepared a quiet and pleasant scientific environment, where I could conduct my study. You acted as an elder and caring brother for my family and I, and helped us to adapt to Finnish society and culture. Your office was open and you always answered all of my questions, doubts, and requests honestly with shining eyes and open arms. I wish you the best in your future endeavors, success in your scientific plans, and good health in your life. Also, I hope that we will continue our scientific collaboration in the future.

In addition, I shared brilliant and memorable moments with my other supervisor, Emeritus Professor Markku Kulomaa (PhD). Dear Professor Markku Kulomaa, please accept my ebullient appreciation for all of your creative and friendly supervision and comments.

I very much appreciate the external reviewers of my thesis, Dr. Jukka Hytönen (MD, PhD) and Dr. Lotta-Riina Sundberg (PhD), for their superlative, transcendent, and valuable comments.

I spent many creative periods with my thesis follow-up committee members, Dr. Marko Pesu (MD, PhD) and Dr. Vesa P. Hytönen (PhD). I would like to say my special thanks to both of them for their participation in my thesis annual report meetings, expression of their scientific comments, and participation in writing reliable and valuable articles for publication.

I had a great opportunity to have a very valuable scientific collaboration with Professor Claudiu T. Supuran (PhD). He and the other Florence team performed the inhibitory studies of our CAs, which was invaluable for this work. In addition, he always answered my doubts and questions honestly and played a crucial role for my thesis. I want to say, dear Professor Supuran, thank you for your valuable scientific collaboration in my thesis and I hope to continue our research together in the future.

At this moment, I would like to acknowledge Professor Antti Sukura (DVM, PhD) and Dr. Anu Näreaho (DVM, PhD), who participated in the parasite diagnostics project. They prepared meat samples with parasite *Trichinella* species as well as provided their valuable comments for the publication of our *Trichinella* research article.

Meanwhile, I would like to say my special thanks to Professor Pekka Karhunen (MD, PhD) and Professor Olli Silvennoinen (MD, PhD), who prepared the opportunity for me to have suitable access to their laboratory facilities for my thesis.

Another important person for my thesis was Dr. Martti Tolvanen (PhD). I learned much new information about bioinformatics, genomics, proteomics, and transcriptomics from him, and he very much helped me to improve my thesis. I would like to express my heartfelt appreciation to him for his reliable and creative comments and helps.

Dr. Csaba Ortutay (PhD) played a major role in teaching phylogenetics to me, especially during my first research article publication. I want to say my thanks for his help during my thesis work.

It is the time to acknowledge one of my great colleagues whom I missed very soon after starting my thesis. Dr. Peiwen Pan (PhD) taught me many important technical issues for molecular biology work. I want to express my honest appreciation for her, as she greatly helped my family and me at the beginning of our stay in Tampere. I wish all the best in her new position abroad.

Harlan Barker (MSc) gave me several creative comments on bioinformatics and phylogenetic analysis methods. He also performed many English revisions on my manuscripts and thesis. I would like to say my hearty appreciation to him for his reliable and critical role in my thesis project.

I want to express my particular appreciation for the expert assistance provided by Ms. Marianne Kuuslahti and Ms. Aulikki Lehmus in conducting the laboratory experiments of my thesis. They explained all of the details and rules to me honestly and clearly.

Special thanks belong to Ms. Arja Ahola. She has helped me and my family in many ways since my arrival to Finland in 2012. She guided and explained everything warmly to my family and me as a caring mother.

My doctoral study involved many administrative tasks during all these years. I would like to say my special thanks to Professor Jorma Isola (MD, PhD), Dr. Henna Mattila (PhD), Ms. Marjatta Viilo, and Ms. Nastasia Pråhl who have been in charge of the Doctoral Program in Biomedicine and Biotechnology at BioMediTech, University of Tampere.

During my PhD study, I participated in several courses. I would like to express my appreciation to all teachers, including Professor Anne Kallioniemi (MD, PhD), Dr. Eric Dufour (PhD), Dr. Alberto Sanz Montero (PhD), Dr. Mataleena Parikka (DDS, PhD), Dr. Susanna Valanne (PhD), and Dr. Teemu Ihalainen (PhD).

Several other faculty members, colleagues, students, laboratory technicians, and employees gave me a lot of favors that greatly helped me during these years. I want to say my thanks to Professor Timo Ylikomi (MD, PhD), Dr. Fatemeh Bootorabi (PhD), Dr. Ashok Aspatwar (PhD), Dr. Leo Syrjänen (MD, PhD), Dr. Sami Purmonen (PhD), Dr. Pasi Pennanen (PhD), Dr. Heimo Syvälä (PhD), Dr. Riina Sarkanen (PhD), Dr. Tenho Hietanen (MD), Maarit Patrikainen (MSc), Linda Urbański (MSc, Tech.), Nirmal Iyer (MSc), Prajwol Manandhar (MSc), Rawnak Jahan Hoque (MSc), Ms. Niina Ikonen, Ms. Hilkka Mäkinen, Ms. Alma Yrjänäinen, Ms. Maarit Pehkonen, and Ms. Carita Vartiainen.

I would like to express my great appreciation for the Ministry of Science, Research, and Technology (MSRT), Student Affairs Organization (SAORG), and National Institute of Genetic Engineering and Biotechnology (NIGEB) of the Islamic Republic of Iran, which prepared a PhD scholarship position for me at NIGEB. Also, I want to acknowledge both former and current heads, general directors, deputies, and employees of MSRT, SAORG, and NIGEB, including Professor Abbas Sahebghadam Lotfi (PhD), Professor Mohammad Reza Zamani (PhD), Professor Mahmoud Mollabashi (PhD), Professor Mojtaba Sedighi (PhD), Professor Hassan Moslemi Naeini (PhD), Professor Ibrahim Hajizadeh (PhD), Ms. Mahshid Atashi , Ms. Sedigheh Khodami, Ms. Majidzadeh, and Mr. Abdolhossein Daneshfar.

Also, I want to express my gratitude to Professor Kambiz Akbari (PhD, faculty member of NIGEB) who gave me a vast amount of valuable scientific advice before and during my thesis.

I would like to express my greatest gratitude to Fimlab Laboratories Ltd. and Finnish Cultural Foundation (Pirkanmaa Regional Fund), which granted me the funds for conducting my thesis.

In addition, my special thanks go to the Academy of Finland, Sigrid Juselius Foundation, Jane and Aatos Erkko Foundation, Tampere Tuberculosis Foundation, Competitive Research Funding of the Tampere University Hospital, and Tampere University Foundation, which supported my thesis and attendance at conferences.

In this section, I am going to express my sincerest appreciation to my family. I remember that my father and mother tried to prepare the best suitable conditions for my education in Iran and if they did not support me, it would not have been possible for me to come to Finland to get higher education. I would like to kiss your hands and tell you: I love you! In addition, I would like to say my greatest appreciation to my wife (Fatemeh) and my lovely daughter (Setareh). Dear Fatemeh and Setareh, if you did not accept to come to Finland with me, if you did not accept to live far from your other Iranian family members, and if you did not tolerate all the difficulties during our life in Tampere, I would not have been successful in my PhD education. I would like to dedicate my doctoral dissertation to my wife and daughter, Fatemeh and Setareh.

Tampere, Finland/November 2016

Reza Zolfaghari Emameh

تقدیر و تشکر

روزها، ماهها و سالها گذشت تا چنین لحظه تاریخی و به یاد ماندنی یعنی دفاع در مقطع دکتری تخصصی فرا رسد. خواست و تقدیر حق تعالی، این بار گران و پر مسؤولیت را بر دوش این بنده حقیر گذاشت تا در معرض امتحانی سخت قرار گیرم. امتحانی سخت از آن جهت که پاسخگویی در پیشگاه حضرت ولیعصر مهدی موعود (عج)، مقام عظمای ولایت حضرت آیت الله خامنه ای (مد ظله العالی)، روح پر فتوح حضرت امام خمینی (ره) و خون پاک شهدای گرانقدر، بند بند وجودم را هر لحظه می فشرد تا پایان نامه ای در خور و شأن ایران اسلامی ارائه دهم و پرچم کشورم را در عرصه علم و فناوری برافرازم. ایزد منان را شاکرم که یار و مددکار اینجانب در طی دوران تحصیل بود تا این مرحله را نیز همچون گذشته با سربلندی سپری نمایم.

بنابراین، بر خود وظیفه میدانم از تمام مسؤولین وزارت علوم، تحقیقات و فناوری، سازمان امور دانشجویان و پژوهشگاه ملی مهندسی ژنتیک و زیست فناوری تشکر نمایم که اینجانب را لایق این خدمتگزاری برگزیدند. جا دارد که به نمایندگی از همه عزیزان شاغل حال و گذشته در این وزارت، سازمان و پژوهشگاه، از جناب آقایان: دکتر مجتبی صدیقی، دکتر محمود ملا باشی، دکتر ابراهیم حاجی زاده، دکتر حسن مسلمی نائینی، مهندس علي پرهیزکار ابیانه، دکتر محمد حسین مجلس آرا، دکتر محمد رضا زمانی، دکتر عباس صاحبقدم لطفی و عبدالحسین دانشفر و سرکار خانمها: آتشی، خدامی، فراهانی و مجید زاده تشکر نمایم. همچنین از زحمات جناب آقای دکتر کامبیز اکبری (عضو محترم هیئت علمی پژوهشگاه ملی مهندسی ژنتیک و زیست فناوری) بسیار سپاسگزاری مینمایم که عهده دار مسؤولیت استاد ناظر در این پایان نامه بودند. نظرات ارزشمند این استاد فرهیخته در تمام مراحل اخذ پذیرش و تحصیل برای اینجانب از جایگاه والایی برخوردار بوده است.

بر کسی پوشیده نیست که موفقیت فرزندان در تمام عرصه های زندگی، نشأت گرفته از حضور پدر و مادری دلسوز و مهربان در زنگی آنهاست. در این لحظه حساس و آکنده از شادمانی، بر دستان پر تلاش و سرشار از محبت پدر و مادرم بوسه میزنم و از صمیم قلب به این عزیزانم میگویم که دوستتان دارم.

زبانم از ابراز کمال تشکراتم از همسر عزیزم فاطمه و فرزند دلبندم ستاره قاصر است. اگر همراهی ایندو عزیز با اینجانب در این هجرت طولانی مدت علمی نبود، قطعاً چنین لحظه خاطره انگیزی در دفاع از پایان نامه دکتری تخصصی رخ نمی داد. تحمل تمام سختی ها، مشقتها و تنگناهای روحی و اقتصادی در دوران تحصیلم، جایگاه ویژه ای برای آنان در قلبم فراهم کرده است. به راستی که همسرم الگوی واقعی استقامت و وفاداری است. در این لحظه، از تمام وجود دوست دارم که به پاس تشکر، این پایان نامه را به همسرم فاطمه و ستاره آسمان زندگیم تقدیم نمایم.

علم دریاییست بی حد و کنار　　طالب علمست غواص بحار

گر هزاران سال باشد عمر او　　او نگردد سیر خود از جست و جو

کان رسول حق بگفت اندر بیان　　اینک منهومان هما لا یشبعان

"مولانا"

با آرزوی موفقیت و سلامتی
التماس دعا
رضا ذوالفقاری امامه

آبان ۱۳۹۵

9. References

Abuaita, B. H. and J. H. Withey (2009). "Bicarbonate Induces Vibrio cholerae virulence gene expression by enhancing ToxT activity." Infect Immun 77(9): 4111-4120.

Alba, M. M. and J. Castresana (2007). "On homology searches by protein Blast and the characterization of the age of genes." BMC Evol Biol 7: 53.

Ali, M. Y., A. Pavasovic, P. B. Mather and P. J. Prentis (2015). "Analysis, characterisation and expression of gill-expressed carbonic anhydrase genes in the freshwater crayfish Cherax quadricarinatus." Gene 564(2): 176-187.

Alterio, V., R. M. Vitale, S. M. Monti, C. Pedone, A. Scozzafava, A. Cecchi, G. De Simone and C. T. Supuran (2006). "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 128(25): 8329-8335.

Alves, J. M., C. C. Klein, F. M. da Silva, A. G. Costa-Martins, M. G. Serrano, G. A. Buck, A. T. Vasconcelos, M. F. Sagot, M. M. Teixeira, M. C. Motta and E. P. Camargo (2013). "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 13: 190.

Alves, J. M., L. Voegtly, A. V. Matveyev, A. M. Lara, F. M. da Silva, M. G. Serrano, G. A. Buck, M. M. Teixeira and E. P. Camargo (2011). "Identification and phylogenetic analysis of heme synthesis genes in trypanosomatids and their bacterial endosymbionts." PLoS One 6(8): e23518.

Amoroso, G., L. Morell-Avrahov, D. Muller, K. Klug and D. Sultemeyer (2005). "The gene NCE103 (YNL036w) from Saccharomyces cerevisiae encodes a functional carbonic anhydrase and its transcription is regulated by the concentration of inorganic carbon in the medium." Mol Microbiol 56(2): 549-558.

Andersson, B., P. O. Nyman and L. Strid (1972). "Amino acid sequence of human erythrocyte carbonic anhydrase B." Biochem Biophys Res Commun 48(3): 670-677.

Andersson, J. O. (2005). "Lateral gene transfer in eukaryotes." Cell Mol Life Sci 62(11): 1182-1197.

Andersson, J. O., W. F. Doolittle and C. L. Nesbo (2001). "Genomics. Are there bugs in our genome?" Science 292(5523): 1848-1850.

Andersson, M., G. Chen, M. Otikovs, M. Landreh, K. Nordling, N. Kronqvist, P. Westermark, H. Jornvall, S. Knight, Y. Ridderstrale, L. Holm, Q. Meng, K. Jaudzems, M. Chesler, J. Johansson and A. Rising (2014). "Carbonic anhydrase generates CO2 and H+ that drive spider silk formation via opposite effects on the terminal domains." PLoS Biol 12(8): e1001921.

Armentia, A., F. J. Martin-Gil, C. Pascual, M. Martin-Esteban, A. Callejo and C. Martinez (2006). "Anisakis simplex allergy after eating chicken meat." J Investig Allergol Clin Immunol 16(4): 258-263.

Aurrecoechea, C., J. Brestelli, B. P. Brunk, J. M. Carlton, J. Dommer, S. Fischer, B. Gajria, X. Gao, A. Gingle, G. Grant, O. S. Harb, M. Heiges, F. Innamorato, J. Iodice, J. C. Kissinger, E. Kraemer, W. Li, J. A. Miller, H. G. Morrison, V. Nayak, C. Pennington, D. F. Pinney, D. S. Roos, C. Ross, C. J. Stoeckert, Jr., S. Sullivan, C. Treatman and H. Wang (2009). "GiardiaDB and TrichDB: integrated genomic resources for the eukaryotic protist pathogens Giardia lamblia and Trichomonas vaginalis." Nucleic Acids Res 37(Database issue): D526-530.

Badger, M. (2003). "The roles of carbonic anhydrases in photosynthetic CO(2) concentrating mechanisms." Photosynth Res 77(2-3): 83-94.

Bahn, Y. S., G. M. Cox, J. R. Perfect and J. Heitman (2005). "Carbonic anhydrase and CO2 sensing during Cryptococcus neoformans growth, differentiation, and virulence." Curr Biol 15(22): 2013-2020.

Baldini, F., N. Segata, J. Pompon, P. Marcenac, W. Robert Shaw, R. K. Dabire, A. Diabate, E. A. Levashina and F. Catteruccia (2014). "Evidence of natural Wolbachia infections in field populations of Anopheles gambiae." Nat Commun 5: 3985.

Banerjee, S. and P. A. Deshpande (2016). "On origin and evolution of carbonic anhydrase isozymes: A phylogenetic analysis from whole-enzyme to active site." Comput Biol Chem 61: 121-129.

Beier, C. L., M. Horn, R. Michel, M. Schweikert, H. D. Gortz and M. Wagner (2002). "The genus Caedibacter comprises endosymbionts of Paramecium spp. related to the Rickettsiales (Alphaproteobacteria) and to Francisella tularensis (Gammaproteobacteria)." Appl Environ Microbiol 68(12): 6043-6050.

Bertelli, C. and G. Greub (2012). "Lateral gene exchanges shape the genomes of amoeba-resisting microorganisms." Front Cell Infect Microbiol 2: 110.

Betson, M., P. Nejsum, R. P. Bendall, R. M. Deb and J. R. Stothard (2014). "Molecular epidemiology of ascariasis: a global perspective on the transmission dynamics of Ascaris in people and pigs." J Infect Dis 210(6): 932-941.

Bockarie, M. J., L. A. Kelly-Hope, M. Rebollo and D. H. Molyneux (2013). "Preventive chemotherapy as a strategy for elimination of neglected tropical parasitic diseases: endgame challenges." Philos Trans R Soc Lond B Biol Sci 368(1623): 20120144.

Boucher, Y., C. J. Douady, R. T. Papke, D. A. Walsh, M. E. Boudreau, C. L. Nesbo, R. J. Case and W. F. Doolittle (2003). "Lateral gene transfer and the origins of prokaryotic groups." Annu Rev Genet 37: 283-328.

Bowes, G. W. (1969). "Carbonic anhydrase in marine algae." Plant Physiol 44(5): 726-732.

Brunk, C. F., L. C. Lee, A. B. Tran and J. Li (2003). "Complete sequence of the mitochondrial genome of Tetrahymena thermophila and comparative methods for identifying highly divergent genes." Nucleic Acids Res 31(6): 1673-1682.

Brussow, H., C. Canchaya and W. D. Hardt (2004). "Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion." Microbiol Mol Biol Rev 68(3): 560-602, table of contents.

Bunyaratvej, S., P. Unpunyo and A. Pongtippan (2007). "The Sarcocystis-cyst containing beef and pork as the sources of natural intestinal sarcocystosis in Thai people." J Med Assoc Thai 90(10): 2128-2135.

Burghout, P., D. Vullo, A. Scozzafava, P. W. Hermans and C. T. Supuran (2011). "Inhibition of the beta-carbonic anhydrase from Streptococcus pneumoniae by inorganic anions and small molecules: Toward innovative drug design of antiinfectives?" Bioorg Med Chem 19(1): 243-248.

Burnell, J. N., M. J. Gibbs and J. G. Mason (1990). "Spinach chloroplastic carbonic anhydrase: nucleotide sequence analysis of cDNA." Plant Physiol 92(1): 37-40.

Butler, D. K., L. E. Yasuda and M. C. Yao (1995). "An intramolecular recombination mechanism for the formation of the rRNA gene palindrome of Tetrahymena thermophila." Mol Cell Biol 15(12): 7117-7126.

Cameron, C. B. and C. D. Bishop (2012). "Biomineral ultrastructure, elemental constitution and genomic analysis of biomineralization-related proteins in hemichordates." Proc Biol Sci 279(1740): 3041-3048.

Cantagalli, L. B., C. A. Mangolin and M. C. Ruvolo-Takasusuki (2010). "Isoenzymatic polymorphism in the leaf-cutting ant Atta capiguara Goncalves (Hymenoptera: Formicidae)." Neotrop Entomol 39(1): 46-49.

Carter, N., A. Shiels and R. Tashian (1978). "Carbonic anhydrase III isoenzyme from human and bovine muscle [proceedings]." Biochem Soc Trans 6(3): 552-553.

Carvalho-Santos, Z., P. Machado, P. Branco, F. Tavares-Cadete, A. Rodrigues-Martins, J. B. Pereira-Leal and M. Bettencourt-Dias (2010). "Stepwise evolution of the centriole-assembly pathway." J Cell Sci 123(Pt 9): 1414-1426.

Caspi, R., T. Altman, R. Billington, K. Dreher, H. Foerster, C. A. Fulcher, T. A. Holland, I. M. Keseler, A. Kothari, A. Kubo, M. Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, P. Subhraveti, D. S. Weaver, D. Weerasinghe, P. Zhang and P. D. Karp (2014). "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases." Nucleic Acids Res 42(Database issue): D459-471.

Castro-Borges, W., A. Dowle, R. S. Curwen, J. Thomas-Oates and R. A. Wilson (2011). "Enzymatic shaving of the tegument surface of live schistosomes for proteomic analysis: a rational approach to select vaccine candidates." PLoS Negl Trop Dis 5(3): e993.

Catala, M. (1997). "Carbonic anhydrase activity during development of the choroid plexus in the human fetus." Childs Nerv Syst 13(7): 364-368.

Catta-Preta, C. M., F. L. Brum, C. C. da Silva, A. A. Zuma, M. C. Elias, W. de Souza, S. Schenkman and M. C. Motta (2015). "Endosymbiosis in trypanosomatid protozoa: the bacterium division is controlled during the host cell cycle." Front Microbiol 6: 520.

Chang, X., Y. Zheng, Q. Yang, L. Wang, J. Pan, Y. Xia, X. Yan and J. Han (2012). "Carbonic anhydrase I (CA1) is involved in the process of bone formation and is susceptible to ankylosing spondylitis." Arthritis Res Ther 14(4): R176.

Chen, L., D. Zheng, B. Liu, J. Yang and Q. Jin (2016). "VFDB 2016: hierarchical and refined dataset for big data analysis-10 years on." Nucleic Acids Res 44(D1): D694-697.

Cheng, L., Y. Du, J. Hu, D. Jiao, J. Li, Z. Zhou, Q. Xu and F. Li (2015). "PROTEOMIC ANALYSIS OF UBIQUITINATED PROTEINS FROM DELTAMETHRIN-

RESISTANT AND SUSCEPTIBLE STRAINS OF THE DIAMONDBACK MOTH, Plutella Xylostella L." Arch Insect Biochem Physiol.

Cleves, A. E., D. N. Cooper, S. H. Barondes and R. B. Kelly (1996). "A new pathway for protein export in Saccharomyces cerevisiae." J Cell Biol 133(5): 1017-1026.

Colbourne, J. K., M. E. Pfrender, D. Gilbert, W. K. Thomas, A. Tucker, T. H. Oakley, S. Tokishita, A. Aerts, G. J. Arnold, M. K. Basu, D. J. Bauer, C. E. Caceres, L. Carmel, C. Casola, J. H. Choi, J. C. Detter, Q. Dong, S. Dusheyko, B. D. Eads, T. Frohlich, K. A. Geiler-Samerotte, D. Gerlach, P. Hatcher, S. Jogdeo, J. Krijgsveld, E. V. Kriventseva, D. Kultz, C. Laforsch, E. Lindquist, J. Lopez, J. R. Manak, J. Muller, J. Pangilinan, R. P. Patwardhan, S. Pitluck, E. J. Pritham, A. Rechtsteiner, M. Rho, I. B. Rogozin, O. Sakarya, A. Salamov, S. Schaack, H. Shapiro, Y. Shiga, C. Skalitzky, Z. Smith, A. Souvorov, W. Sung, Z. Tang, D. Tsuchiya, H. Tu, H. Vos, M. Wang, Y. I. Wolf, H. Yamagata, T. Yamada, Y. Ye, J. R. Shaw, J. Andrews, T. J. Crease, H. Tang, S. M. Lucas, H. M. Robertson, P. Bork, E. V. Koonin, E. M. Zdobnov, I. V. Grigoriev, M. Lynch and J. L. Boore (2011). "The ecoresponsive genome of Daphnia pulex." Science 331(6017): 555-561.

Coyne, R. S., L. Hannick, D. Shanmugam, J. B. Hostetler, D. Brami, V. S. Joardar, J. Johnson, D. Radune, I. Singh, J. H. Badger, U. Kumar, M. Saier, Y. Wang, H. Cai, J. Gu, M. W. Mather, A. B. Vaidya, D. E. Wilkes, V. Rajagopalan, D. J. Asai, C. G. Pearson, R. C. Findly, H. W. Dickerson, M. Wu, C. Martens, Y. Van de Peer, D. S. Roos, D. M. Cassidy-Hanley and T. G. Clark (2011). "Comparative genomics of the pathogenic ciliate Ichthyophthirius multifiliis, its free-living relatives and a host species provide insights into adoption of a parasitic lifestyle and prospects for disease control." Genome Biol 12(10): R100.

Culver, B. W. and P. K. Morton (2015). "The Evolutionary History of Daphniid alpha-Carbonic Anhydrase within Animalia." Int J Evol Biol 2015: 538918.

de Morais, C. G., A. K. Castro Lima, R. Terra, R. F. dos Santos, S. A. Da-Silva and P. M. Dutra (2015). "The Dialogue of the Host-Parasite Relationship: Leishmania spp. and Trypanosoma cruzi Infection." Biomed Res Int 2015: 324915.

De Simone, G., A. Di Fiore, C. Capasso and C. T. Supuran (2015). "The zinc coordination pattern in the eta-carbonic anhydrase from Plasmodium falciparum is different from all other carbonic anhydrase genetic families." Bioorg Med Chem Lett 25(7): 1385-1389.

Del Prete, S., D. Vullo, G. M. Fisher, K. T. Andrews, S. A. Poulsen, C. Capasso and C. T. Supuran (2014). "Discovery of a new family of carbonic anhydrases in the malaria pathogen Plasmodium falciparum--the eta-carbonic anhydrases." Bioorg Med Chem Lett 24(18): 4389-4396.

DeRosa, A. A., S. R. Chirgwin, J. C. Williams and T. R. Klei (2008). "Isolation and characterization of a gene encoding carbonic anhydrase from Ostertagia ostertagi and quantitative measurement of expression during in vivo exsheathment." Vet Parasitol 154(1-2): 58-66.

Dhillon, B. K., M. R. Laird, J. A. Shay, G. L. Winsor, R. Lo, F. Nizam, S. K. Pereira, N. Waglechner, A. G. McArthur, M. G. Langille and F. S. Brinkman (2015). "IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis." Nucleic Acids Res.

Doolittle, W. F. (1997). "Fun with genealogy." Proc Natl Acad Sci U S A 94(24): 12751-12753.

Dunning Hotopp, J. C., M. E. Clark, D. C. Oliveira, J. M. Foster, P. Fischer, M. C. Munoz Torres, J. D. Giebel, N. Kumar, N. Ishmael, S. Wang, J. Ingram, R. V. Nene, J. Shepard, J. Tomkins, S. Richards, D. J. Spiro, E. Ghedin, B. E. Slatko, H. Tettelin and J. H. Werren (2007). "Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes." Science 317(5845): 1753-1756.

Eckert, J. and P. Deplazes (2004). "Biological, epidemiological, and clinical aspects of echinococcosis, a zoonosis of increasing concern." Clin Microbiol Rev 17(1): 107-135.

Eichinger, L., J. A. Pachebat, G. Glockner, M. A. Rajandream, R. Sucgang, M. Berriman, J. Song, R. Olsen, K. Szafranski, Q. Xu, B. Tunggal, S. Kummerfeld, M. Madera, B. A. Konfortov, F. Rivero, A. T. Bankier, R. Lehmann, N. Hamlin, R. Davies, P. Gaudet, P. Fey, K. Pilcher, G. Chen, D. Saunders, E. Sodergren, P. Davis, A. Kerhornou, X. Nie, N. Hall, C. Anjard, L. Hemphill, N. Bason, P. Farbrother, B. Desany, E. Just, T. Morio, R. Rost, C. Churcher, J. Cooper, S. Haydock, N. van Driessche, A. Cronin, I. Goodhead, D. Muzny, T. Mourier, A. Pain, M. Lu, D. Harper, R. Lindsay, H. Hauser, K. James, M. Quiles, M. Madan Babu, T. Saito, C. Buchrieser, A. Wardroper, M. Felder, M. Thangavelu, D. Johnson, A. Knights, H. Loulseged, K. Mungall, K. Oliver, C. Price, M. A. Quail, H. Urushihara, J. Hernandez, E. Rabbinowitsch, D. Steffen, M. Sanders, J. Ma, Y. Kohara, S. Sharp, M. Simmonds, S. Spiegler, A. Tivey, S. Sugano, B. White, D. Walker, J. Woodward, T. Winckler, Y. Tanaka, G. Shaulsky, M. Schleicher, G. Weinstock, A. Rosenthal, E. C. Cox, R. L. Chisholm, R. Gibbs, W. F. Loomis, M. Platzer, R. R. Kay, J. Williams, P. H. Dear, A. A. Noegel, B. Barrell and A. Kuspa (2005). "The genome of the social amoeba Dictyostelium discoideum." Nature 435(7038): 43-57.

Elleuche, S. and S. Poggeler (2008). "A cyanase is transcriptionally regulated by arginine and involved in cyanate decomposition in Sordaria macrospora." Fungal Genet Biol 45(11): 1458-1469.

Elleuche, S. and S. Poggeler (2009). "Evolution of carbonic anhydrases in fungi." Curr Genet 55(2): 211-222.

Elleuche, S. and S. Poggeler (2010). "Carbonic anhydrases in fungi." Microbiology 156(Pt 1): 23-29.

Emanuelsson, O., H. Nielsen, S. Brunak and G. von Heijne (2000). "Predicting subcellular localization of proteins based on their N-terminal amino acid sequence." J Mol Biol 300(4): 1005-1016.

Eriksson, M., J. Karlsson, Z. Ramazanov, P. Gardestrom and G. Samuelsson (1996). "Discovery of an algal mitochondrial carbonic anhydrase: molecular cloning and characterization of a low-CO2-induced polypeptide in Chlamydomonas reinhardtii." Proc Natl Acad Sci U S A 93(21): 12031-12034.

Eriksson, M., P. Villand, P. Gardestrom and G. Samuelsson (1998). "Induction and Regulation of Expression of a Low-CO2-Induced Mitochondrial Carbonic Anhydrase in Chlamydomonas reinhardtii." Plant Physiol 116(2): 637-641.

Esbaugh, A. J. and B. L. Tufts (2006). "The structure and function of carbonic anhydrase isozymes in the respiratory system of vertebrates." Respir Physiol Neurobiol 154(1-2): 185-198.

Eswar, N., B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M. Y. Shen, U. Pieper and A. Sali (2006). "Comparative protein structure modeling using Modeller." Curr Protoc Bioinformatics Chapter 5: Unit 5 6.

Falkowski, P. G. and L. V. Godfrey (2008). "Electrons, life and the evolution of Earth's oxygen cycle." Philos Trans R Soc Lond B Biol Sci 363(1504): 2705-2716.

Fasseas, M. K., D. Tsikou, E. Flemetakis and P. Katinakis (2010). "Molecular and biochemical analysis of the beta class carbonic anhydrases in Caenorhabditis elegans." Mol Biol Rep 37(6): 2941-2950.

Feldshtein, M., S. Elkrinawi, B. Yerushalmi, B. Marcus, D. Vullo, H. Romi, R. Ofir, D. Landau, S. Sivan, C. T. Supuran and O. S. Birk (2010). "Hyperchlorhidrosis caused by homozygous mutation in CA12, encoding carbonic anhydrase XII." Am J Hum Genet 87(5): 713-720.

Ferry, J. G. (2010). "The gamma class of carbonic anhydrases." Biochim Biophys Acta 1804(2): 374-381.

Fichorova, R. N., O. R. Buck, H. S. Yamamoto, T. Fashemi, H. Y. Dawood, B. Fashemi, G. R. Hayes, D. H. Beach, Y. Takagi, M. L. Delaney, M. L. Nibert, B. N. Singh and A. B. Onderdonk (2013). "The villain team-up or how Trichomonas vaginalis and bacterial vaginosis alter innate immunity in concert." Sex Transm Infect 89(6): 460-466.

Fisher, S. Z., I. Tariku, N. M. Case, C. Tu, T. Seron, D. N. Silverman, P. J. Linser and R. McKenna (2006). "Expression, purification, kinetic, and structural characterization of an alpha-class carbonic anhydrase from Aedes aegypti (AaCA1)." Biochim Biophys Acta 1764(8): 1413-1419.

Friedley, N. J. and S. Rosen (1975). "Carbonic anhydrase activity in the mammalian ovary, fallopian tube, and uterus: histochemical and biochemical studies." Biol Reprod 12(2): 293-304.

Frost, L. S., R. Leplae, A. O. Summers and A. Toussaint (2005). "Mobile genetic elements: the agents of open source evolution." Nat Rev Microbiol 3(9): 722-732.

Fujikawa-Adachi, K., I. Nishimori, S. Sakamoto, M. Morita, S. Onishi, S. Yonezawa and M. A. Hollingsworth (1999). "Identification of carbonic anhydrase IV and VI mRNA expression in human pancreas and salivary glands." Pancreas 18(4): 329-335.

Fujikawa-Adachi, K., I. Nishimori, T. Taguchi and S. Onishi (1999). "Human carbonic anhydrase XIV (CA14): cDNA cloning, mRNA expression, and mapping to chromosome 1." Genomics 61(1): 74-81.

Fujikawa-Adachi, K., I. Nishimori, T. Taguchi and S. Onishi (1999). "Human mitochondrial carbonic anhydrase VB. cDNA cloning, mRNA expression, subcellular localization, and mapping to chromosome x." J Biol Chem 274(30): 21228-21233.

Fujishima, M. and Y. Kodama (2012). "Endosymbionts in paramecium." Eur J Protistol 48(2): 124-137.

Fukuzawa, H., E. Suzuki, Y. Komukai and S. Miyachi (1992). "A gene homologous to chloroplast carbonic anhydrase (icfA) is essential to photosynthetic carbon dioxide fixation by Synechococcus PCC7942." Proc Natl Acad Sci U S A 89(10): 4437-4441.

Gawryluk, R. M. and M. W. Gray (2010). "Evidence for an early evolutionary emergence of gamma-type carbonic anhydrases as components of mitochondrial respiratory complex I." BMC Evol Biol 10: 176.

Geers, C. and G. Gros (2000). "Carbon dioxide transport and carbonic anhydrase in blood and muscle." Physiol Rev 80(2): 681-715.

Ghigo, J. M. (2001). "Natural conjugative plasmids induce bacterial biofilm development." Nature 412(6845): 442-445.

Giordano, M., A. Norici, M. Forssen, M. Eriksson and J. A. Raven (2003). "An anaplerotic role for mitochondrial carbonic anhydrase in Chlamydomonas reinhardtii." Plant Physiol 132(4): 2126-2134.

Gotz, R., A. Gnann and F. K. Zimmermann (1999). "Deletion of the carbonic anhydrase-like gene NCE103 of the yeast Saccharomyces cerevisiae causes an oxygen-sensitive growth defect." Yeast 15(10A): 855-864.

Graeff-Teixeira, C., A. C. da Silva and K. Yoshimura (2009). "Update on eosinophilic meningoencephalitis and its clinical relevance." Clin Microbiol Rev 22(2): 322-348, Table of Contents.

Greub, G. and D. Raoult (2004). "Microorganisms resistant to free-living amoebae." Clin Microbiol Rev 17(2): 413-433.

Groth, A. C. and M. P. Calos (2004). "Phage integrases: biology and applications." J Mol Biol 335(3): 667-678.

Guilloton, M. B., A. F. Lamblin, E. I. Kozliak, M. Gerami-Nejad, C. Tu, D. Silverman, P. M. Anderson and J. A. Fuchs (1993). "A physiological role for cyanate-induced carbonic anhydrase in Escherichia coli." J Bacteriol 175(5): 1443-1451.

Haritos, V. S. and G. Dojchinov (2005). "Carbonic anhydrase metabolism is a key factor in the toxicity of CO2 and COS but not CS2 toward the flour beetle Tribolium castaneum [Coleoptera: Tenebrionidae]." Comp Biochem Physiol C Toxicol Pharmacol 140(1): 139-147.

Harju, A. K., F. Bootorabi, M. Kuuslahti, C. T. Supuran and S. Parkkila (2013). "Carbonic anhydrase III: a neglected isozyme is stepping into the limelight." J Enzyme Inhib Med Chem 28(2): 231-239.

Hilvo, M., A. M. Salzano, A. Innocenti, M. S. Kulomaa, A. Scozzafava, A. Scaloni, S. Parkkila and C. T. Supuran (2009). "Cloning, expression, post-translational modifications and inhibition studies on the latest mammalian carbonic anhydrase isoform, CA XV." J Med Chem 52(3): 646-654.

Hilvo, M., M. Tolvanen, A. Clark, B. Shen, G. N. Shah, A. Waheed, P. Halmi, M. Hanninen, J. M. Hamalainen, M. Vihinen, W. S. Sly and S. Parkkila (2005). "Characterization of CA XV, a new GPI-anchored form of carbonic anhydrase." Biochem J 392(Pt 1): 83-92.

Hodges, M. E., N. Scheumann, B. Wickstead, J. A. Langdale and K. Gull (2010). "Reconstructing the evolutionary history of the centriole from protein components." J Cell Sci 123(Pt 9): 1407-1413.

Hofreuter, D., S. Odenbreit, G. Henke and R. Haas (1998). "Natural competence for DNA transformation in Helicobacter pylori: identification and genetic characterization of the comB locus." Mol Microbiol 28(5): 1027-1038.
Hong, S. T. and Y. Fang (2012). "Clonorchis sinensis and clonorchiasis, an update." Parasitol Int 61(1): 17-24.
Hori, H., K. Yoshida, H. Fukazawa, Y. Kiso, N. Sayama, K. Mori, Y. Aizawa, J. I. Tani and S. Ito (1998). "Effects of thyroid hormone on carbonic anhydrase I gene expression in human erythroid cells." Thyroid 8(6): 525-531.
Hotez, P. J. (2010). "Nuclear weapons and neglected diseases: the "ten-thousand-to-one gap"." PLoS Negl Trop Dis 4(4): e680.
Hotez, P. J., D. H. Molyneux, A. Fenwick, L. Savioli and T. Takeuchi (2008). "A Global Fund to Fight Neglected Tropical Diseases: is the G8 Hokkaido Toyako 2008 Summit ready?" PLoS Negl Trop Dis 2(3): e220.
Hsiao, W., I. Wan, S. J. Jones and F. S. Brinkman (2003). "IslandPath: aiding detection of genomic islands in prokaryotes." Bioinformatics 19(3): 418-420.
Huang, S., T. Hainzl, C. Grundstrom, C. Forsman, G. Samuelsson and A. E. Sauer-Eriksson (2011). "Structural studies of beta-carbonic anhydrase from the green alga Coccomyxa: inhibitor complexes with anions and acetazolamide." PLoS One 6(12): e28458.
Humphrey, W., A. Dalke and K. Schulten (1996). "VMD: visual molecular dynamics." J Mol Graph 14(1): 33-38, 27-38.
Hunter, S., R. Apweiler, T. K. Attwood, A. Bairoch, A. Bateman, D. Binns, P. Bork, U. Das, L. Daugherty, L. Duquenne, R. D. Finn, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, A. Laugraud, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, J. Mistry, A. Mitchell, N. Mulder, D. Natale, C. Orengo, A. F. Quinn, J. D. Selengut, C. J. Sigrist, M. Thimma, P. D. Thomas, F. Valentin, D. Wilson, C. H. Wu and C. Yeats (2009). "InterPro: the integrative protein signature database." Nucleic Acids Res 37(Database issue): D211-215.
Inada, A., C. Nienaber, H. Katsuta, Y. Fujitani, J. Levine, R. Morita, A. Sharma and S. Bonner-Weir (2008). "Carbonic anhydrase II-positive pancreatic cells are progenitors for both endocrine and exocrine pancreas after birth." Proc Natl Acad Sci U S A 105(50): 19915-19919.
Innocenti, A., W. Leewattanapasuk, F. A. Muhlschlegel, A. Mastrolorenzo and C. T. Supuran (2009). "Carbonic anhydrase inhibitors. Inhibition of the beta-class enzyme from the pathogenic yeast Candida glabrata with anions." Bioorg Med Chem Lett 19(16): 4802-4805.
Iverson, T. M., B. E. Alber, C. Kisker, J. G. Ferry and D. C. Rees (2000). "A closer look at the active site of gamma-class carbonic anhydrases: high-resolution crystallographic studies of the carbonic anhydrase from Methanosarcina thermophila." Biochemistry 39(31): 9222-9231.
Jaiswal, P., R. Prasanna and A. K. Kashyap (2005). "Modulation of carbonic anhydrase activity in two nitrogen fixing cyanobacteria, Nostoc calcicola and Anabaena sp." J Plant Physiol 162(10): 1087-1094.
Jenkins, E. J., L. J. Castrodale, S. J. de Rosemond, B. R. Dixon, S. A. Elmore, K. M. Gesy, E. P. Hoberg, L. Polley, J. M. Schurer, M. Simard and R. C. Thompson (2013).

"Tradition and transition: parasitic zoonoses of people and animals in Alaska, northern Canada, and Greenland." Adv Parasitol 82: 33-204.
Johnston, S. P., M. M. Ballard, M. J. Beach, L. Causer and P. P. Wilkins (2003). "Evaluation of three commercial assays for detection of Giardia and Cryptosporidium organisms in fecal specimens." J Clin Microbiol 41(2): 623-626.
Joseph, P., F. Turtaut, S. Ouahrani-Bettache, J. L. Montero, I. Nishimori, T. Minakuchi, D. Vullo, A. Scozzafava, S. Kohler, J. Y. Winum and C. T. Supuran (2010). "Cloning, characterization, and inhibition studies of a beta-carbonic anhydrase from Brucella suis." J Med Chem 53(5): 2277-2285.
Just, F. and B. Walz (1994). "Localization of carbonic anhydrase in the salivary glands of the cockroach, Periplaneta americana." Histochemistry 102(4): 271-277.
Kanehisa, M. and S. Goto (2000). "KEGG: kyoto encyclopedia of genes and genomes." Nucleic Acids Res 28(1): 27-30.
Keeling, P. J. and J. D. Palmer (2008). "Horizontal gene transfer in eukaryotic evolution." Nat Rev Genet 9(8): 605-618.
Khademvatan, S., R. Abdizadeh and M. Tavalla (2014). "Molecular characterization of Toxocara spp. from soil of public areas in Ahvaz southwestern Iran." Acta Trop 135: 50-54.
Khalifah, R. G. (1971). "The carbon dioxide hydration activity of carbonic anhydrase. I. Stop-flow kinetic studies on the native human isoenzymes B and C." J Biol Chem 246(8): 2561-2573.
Kibbe, W. A. (2007). "OligoCalc: an online oligonucleotide properties calculator." Nucleic Acids Res 35(Web Server issue): W43-46.
Kida, E., S. Palminiello, A. A. Golabek, M. Walus, T. Wierzba-Bobrowicz, A. Rabe, G. Albertini and K. E. Wisniewski (2006). "Carbonic anhydrase II in the developing and adult human brain." J Neuropathol Exp Neurol 65(7): 664-674.
Kimber, M. S., J. R. Coleman and E. F. Pai (2000). "beta-carbonic anhydrase from Pisum sativum: crystallization and preliminary X-ray analysis." Acta Crystallogr D Biol Crystallogr 56(Pt 7): 927-929.
Kjeldsen, K. U., M. Obst, H. Nakano, P. Funch and A. Schramm (2010). "Two types of endosymbiotic bacteria in the enigmatic marine worm Xenoturbella bocki." Appl Environ Microbiol 76(8): 2657-2662.
Klengel, T., W. J. Liang, J. Chaloupka, C. Ruoff, K. Schroppel, J. R. Naglik, S. E. Eckert, E. G. Mogensen, K. Haynes, M. F. Tuite, L. R. Levin, J. Buck and F. A. Muhlschlegel (2005). "Fungal adenylyl cyclase integrates CO2 sensing with cAMP signaling and virulence." Curr Biol 15(22): 2021-2026.
Koenig, S. H. and R. D. Brown, 3rd (1972). "H 2 CO 3 as substrate for carbonic anhydrase in the dehydration of HCO 3." Proc Natl Acad Sci U S A 69(9): 2422-2425.
Kolaskar, A. S. and P. C. Tongaonkar (1990). "A semi-empirical method for prediction of antigenic determinants on protein antigens." FEBS Lett 276(1-2): 172-174.
Kostygov, A. Y., E. Dobakova, A. Grybchuk-Ieremenko, D. Vahala, D. A. Maslov, J. Votypka, J. Lukes and V. Yurchenko (2016). "Novel Trypanosomatid-Bacterium Association: Evolution of Endosymbiosis in Action." MBio 7(2): e01985.

Kotwica, J., M. A. Ciuk, E. Joachimiak, S. Rowinski, B. Cymborowski and P. Bebas (2006). "Carbonic anhydrase activity in the vas deferens of the cotton leafworm - Spodoptera littoralis (Lepidoptera: Noctuidae) controlled by circadian clock." J Physiol Pharmacol 57 Suppl 8: 107-123.

Kovacikova, G., W. Lin and K. Skorupski (2010). "The LysR-type virulence activator AphB regulates the expression of genes in Vibrio cholerae in response to low pH and anaerobiosis." J Bacteriol 192(16): 4181-4191.

Krungkrai, J., A. Scozzafava, S. Reungprapavut, S. R. Krungkrai, R. Rattanajak, S. Kamchonwongpaisan and C. T. Supuran (2005). "Carbonic anhydrase inhibitors. Inhibition of Plasmodium falciparum carbonic anhydrase with aromatic sulfonamides: towards antimalarials with a novel mechanism of action?" Bioorg Med Chem 13(2): 483-489.

Kumar, R. S. and J. G. Ferry (2014). "Prokaryotic carbonic anhydrases of Earth's environment." Subcell Biochem 75: 77-87.

Kumpulainen, T. and P. Jalovaara (1981). "Immunohistochemical localization of carbonic anhydrase isoenzymes in the human pancreas." Gastroenterology 80(4): 796-799.

Kumpulainen, T. and S. H. Nystrom (1981). "Immunohistochemical localization of carbonic anhydrase isoenzyme C in human brain." Brain Res 220(1): 220-225.

La Scola, B., L. Mezi, J. P. Auffray, Y. Berland and D. Raoult (2002). "Patients in the intensive care unit are exposed to amoeba-associated pathogens." Infect Control Hosp Epidemiol 23(8): 462-465.

La Scola, B. and D. Raoult (1999). "Afipia felis in hospital water supply in association with free-living amoebae." Lancet 353(9161): 1330.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K.

Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen and C. International Human Genome Sequencing (2001). "Initial sequencing and analysis of the human genome." Nature 409(6822): 860-921.

Lane, T. W. and F. M. Morel (2000). "A biological function for cadmium in marine diatoms." Proc Natl Acad Sci U S A 97(9): 4627-4631.

Lane, T. W., M. A. Saito, G. N. George, I. J. Pickering, R. C. Prince and F. M. Morel (2005). "Biochemistry: a cadmium enzyme from a marine diatom." Nature 435(7038): 42.

Langille, M. G., W. W. Hsiao and F. S. Brinkman (2010). "Detecting genomic islands using bioinformatics approaches." Nat Rev Microbiol 8(5): 373-382.

Lehtonen, J., B. Shen, M. Vihinen, A. Casini, A. Scozzafava, C. T. Supuran, A. K. Parkkila, J. Saarnio, A. J. Kivela, A. Waheed, W. S. Sly and S. Parkkila (2004). "Characterization of CA XIII, a novel member of the carbonic anhydrase isozyme family." J Biol Chem 279(4): 2719-2727.

Lehtonen, J. V., D. J. Still, V. V. Rantanen, J. Ekholm, D. Bjorklund, Z. Iftikhar, M. Huhtala, S. Repo, A. Jussila, J. Jaakkola, O. Pentikainen, T. Nyronen, T. Salminen, M. Gyllenberg and M. S. Johnson (2004). "BODIL: a molecular modeling environment for structure-function analysis and drug design." J Comput Aided Mol Des 18(6): 401-419.

Leplae, R., G. Lima-Mendez and A. Toussaint (2010). "ACLAME: a classification of Mobile genetic Elements, update 2010." Nucleic Acids Res 38(Database issue): D57-61.

Liljas, A., K. K. Kannan, P. C. Bergsten, I. Waara, K. Fridborg, B. Strandberg, U. Carlbom, L. Jarup, S. Lovgren and M. Petef (1972). "Crystal structure of human carbonic anhydrase C." Nat New Biol 235(57): 131-137.

Loftus, B., I. Anderson, R. Davies, U. C. Alsmark, J. Samuelson, P. Amedeo, P. Roncaglia, M. Berriman, R. P. Hirt, B. J. Mann, T. Nozaki, B. Suh, M. Pop, M. Duchene, J. Ackers, E. Tannich, M. Leippe, M. Hofer, I. Bruchhaus, U. Willhoeft, A.

Bhattacharya, T. Chillingworth, C. Churcher, Z. Hance, B. Harris, D. Harris, K. Jagels, S. Moule, K. Mungall, D. Ormond, R. Squares, S. Whitehead, M. A. Quail, E. Rabbinowitsch, H. Norbertczak, C. Price, Z. Wang, N. Guillen, C. Gilchrist, S. E. Stroup, S. Bhattacharya, A. Lohia, P. G. Foster, T. Sicheritz-Ponten, C. Weber, U. Singh, C. Mukherjee, N. M. El-Sayed, W. A. Petri, Jr., C. G. Clark, T. M. Embley, B. Barrell, C. M. Fraser and N. Hall (2005). "The genome of the protist parasite Entamoeba histolytica." Nature 433(7028): 865-868.

Ludwig, M. (2011). "The molecular evolution of beta-carbonic anhydrase in Flaveria." J Exp Bot 62(9): 3071-3081.

Mahillon, J. and M. Chandler (1998). "Insertion sequences." Microbiol Mol Biol Rev 62(3): 725-774.

Majeau, N. and J. R. Coleman (1996). "Effect of CO2 Concentration on Carbonic Anhydrase and Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Expression in Pea." Plant Physiol 112(2): 569-574.

Makiuchi, T. and T. Nozaki (2014). "Highly divergent mitochondrion-related organelles in anaerobic parasitic protozoa." Biochimie 100: 3-17.

Mann, K., F. H. Wilt and A. J. Poustka (2010). "Proteomic analysis of sea urchin (Strongylocentrotus purpuratus) spicule matrix." Proteome Sci 8: 33.

Marchesini, N., F. A. Ruiz, M. Vieira and R. Docampo (2002). "Acidocalcisomes are functionally linked to the contractile vacuole of Dictyostelium discoideum." J Biol Chem 277(10): 8146-8153.

Maresca, A., A. Scozzafava, D. Vullo and C. T. Supuran (2013). "Dihalogenated sulfanilamides and benzolamides are effective inhibitors of the three beta-class carbonic anhydrases from Mycobacterium tuberculosis." J Enzyme Inhib Med Chem 28(2): 384-387.

Maresca, A., D. Vullo, A. Scozzafava, G. Manole and C. T. Supuran (2013). "Inhibition of the beta-class carbonic anhydrases from Mycobacterium tuberculosis with carboxylic acids." J Enzyme Inhib Med Chem 28(2): 392-396.

Mastrolorenzo, A., G. Zuccati, D. Massi, M. G. Gabrielli, A. Casini, A. Scozzafava and C. T. Supuran (2003). "Immunohistochemical study of carbonic anhydrase isozymes in human skin." Eur J Dermatol 13(5): 440-444.

Matsuda, Y., K. Nakajima and M. Tachibana (2011). "Recent progresses on the genetic basis of the regulation of CO2 acquisition systems in response to CO2 concentration." Photosynth Res 109(1-3): 191-203.

Mayta, H., R. H. Gilman, E. Prendergast, J. P. Castillo, Y. O. Tinoco, H. H. Garcia, A. E. Gonzalez, C. R. Sterling and P. Cysticercosis Working Group in (2008). "Nested PCR for specific diagnosis of Taenia solium taeniasis." J Clin Microbiol 46(1): 286-289.

McKenna, R. and C. T. Supuran (2014). "Carbonic anhydrase inhibitors drug design." Subcell Biochem 75: 291-323.

Merlin, C., M. Masters, S. McAteer and A. Coulson (2003). "Why is carbonic anhydrase essential to Escherichia coli?" J Bacteriol 185(21): 6415-6424.

Mezquita, P., C. Mezquita and J. Mezquita (1999). "Novel transcripts of carbonic anhydrase II in mouse and human testis." Mol Hum Reprod 5(3): 199-205.

Mitsuhashi, S., T. Mizushima, E. Yamashita, S. Miyachi and T. Tsukihara (2000). "Crystallization and preliminary X-ray diffraction studies of a beta-carbonic anhydrase from the red alga Porphyridium purpureum." Acta Crystallogr D Biol Crystallogr 56(Pt 2): 210-211.

Mitsuhashi, S., T. Mizushima, E. Yamashita, M. Yamamoto, T. Kumasaka, H. Moriyama, T. Ueki, S. Miyachi and T. Tsukihara (2000). "X-ray structure of beta-carbonic anhydrase from the red alga, Porphyridium purpureum, reveals a novel catalytic site for CO(2) hydration." J Biol Chem 275(8): 5521-5526.

Mitsuhashi, S., J. Ohnishi, M. Hayashi and M. Ikeda (2004). "A gene homologous to beta-type carbonic anhydrase is essential for the growth of Corynebacterium glutamicum under atmospheric conditions." Appl Microbiol Biotechnol 63(5): 592-601.

Montgomery, J. C., P. J. Venta, R. L. Eddy, Y. S. Fukushima, T. B. Shows and R. E. Tashian (1991). "Characterization of the human gene for a newly discovered carbonic anhydrase, CA VII, and its localization to chromosome 16." Genomics 11(4): 835-848.

Moustafa, A., B. Beszteri, U. G. Maier, C. Bowler, K. Valentin and D. Bhattacharya (2009). "Genomic footprints of a cryptic plastid endosymbiosis in diatoms." Science 324(5935): 1724-1726.

Mullan, L. J. and A. J. Bleasby (2002). "Short EMBOSS User Guide. European Molecular Biology Open Software Suite." Brief Bioinform 3(1): 92-94.

Munoz-Lopez, M. and J. L. Garcia-Perez (2010). "DNA transposons: nature and applications in genomics." Curr Genomics 11(2): 115-128.

Murakami, H. and W. S. Sly (1987). "Purification and characterization of human salivary carbonic anhydrase." J Biol Chem 262(3): 1382-1388.

Nagao, Y., J. S. Platero, A. Waheed and W. S. Sly (1993). "Human mitochondrial carbonic anhydrase: cDNA cloning, expression, subcellular localization, and mapping to chromosome 16." Proc Natl Acad Sci U S A 90(16): 7623-7627.

Nakajima, T., A. Sano and H. Matsuoka (2009). "Auto-/heterotrophic endosymbiosis evolves in a mature stage of ecosystem development in a microcosm composed of an alga, a bacterium and a ciliate." Biosystems 96(2): 127-135.

Nedenskov-Sorensen, P., G. Bukholm and K. Bovre (1990). "Natural competence for genetic transformation in Campylobacter pylori." J Infect Dis 161(2): 365-366.

Neish, A. C. (1939). "Studies on chloroplasts: Their chemical composition and the distribution of certain metabolites between the chloroplasts and the remainder of the leaf." Biochem J 33(3): 300-308.

Nishimori, I., T. Minakuchi, S. Onishi, D. Vullo, A. Scozzafava and C. T. Supuran (2007). "Carbonic anhydrase inhibitors. DNA cloning, characterization, and inhibition studies of the human secretory isoform VI, a new target for sulfonamide and sulfamate inhibitors." J Med Chem 50(2): 381-388.

Nishimori, I., S. Onishi, H. Takeuchi and C. T. Supuran (2008). "The alpha and beta classes carbonic anhydrases from Helicobacter pylori as novel drug targets." Curr Pharm Des 14(7): 622-630.

Nockler, K., E. Pozio, W. P. Voigt and J. Heidrich (2000). "Detection of Trichinella infection in food animals." Vet Parasitol 93(3-4): 335-350.

Noda, Y., S. Sumitomo, N. Hikosaka and M. Mori (1986). "Immunohistochemical observations on carbonic anhydrase I and II in human salivary glands and submandibular obstructive adenitis." J Oral Pathol 15(4): 187-190.

Notka, F., M. Liss and R. Wagner (2011). "Industrial scale gene synthesis." Methods Enzymol 498: 247-275.

Ogilvie, J. M., K. K. Ohlemiller, G. N. Shah, B. Ulmasov, T. A. Becker, A. Waheed, A. K. Hennig, P. D. Lukasiewicz and W. S. Sly (2007). "Carbonic anhydrase XIV deficiency produces a functional defect in the retinal light response." Proc Natl Acad Sci U S A 104(20): 8514-8519.

Olson, S. A. (2002). "EMBOSS opens up sequence analysis. European Molecular Biology Open Software Suite." Brief Bioinform 3(1): 87-91.

Opavsky, R., S. Pastorekova, V. Zelnik, A. Gibadulinova, E. J. Stanbridge, J. Zavada, R. Kettmann and J. Pastorek (1996). "Human MN/CA9 gene, a novel member of the carbonic anhydrase family: structure and exon to protein domain relationships." Genomics 33(3): 480-487.

Osborn, A. M. and D. Boltner (2002). "When phage, plasmids, and transposons collide: genomic islands, and conjugative- and mobilizable-transposons as a mosaic continuum." Plasmid 48(3): 202-212.

Otranto, D. and D. Traversa (2002). "A review of dicrocoeliosis of ruminants including recent advances in the diagnosis and treatment." Vet Parasitol 107(4): 317-335.

Pan, P., A. B. Vermelho, G. Capaci Rodrigues, A. Scozzafava, M. E. Tolvanen, S. Parkkila, C. Capasso and C. T. Supuran (2013). "Cloning, characterization, and sulfonamide and thiol inhibition studies of an alpha-carbonic anhydrase from Trypanosoma cruzi, the causative agent of Chagas disease." J Med Chem 56(4): 1761-1771.

Parisi, G., M. Perales, M. S. Fornasari, A. Colaneri, N. Gonzalez-Schain, D. Gomez-Casati, S. Zimmermann, A. Brennicke, A. Araya, J. G. Ferry, J. Echave and E. Zabaleta (2004). "Gamma carbonic anhydrases in plant mitochondria." Plant Mol Biol 55(2): 193-207.

Parkkila, S., K. Kaunisto, L. Rajaniemi, T. Kumpulainen, K. Jokinen and H. Rajaniemi (1990). "Immunohistochemical localization of carbonic anhydrase isoenzymes VI, II, and I in human parotid and submandibular glands." J Histochem Cytochem 38(7): 941-947.

Parkkila, S., A. K. Parkkila, T. Juvonen and H. Rajaniemi (1994). "Distribution of the carbonic anhydrase isoenzymes I, II, and VI in the human alimentary tract." Gut 35(5): 646-650.

Parkkila, S., A. K. Parkkila, H. Rajaniemi, G. N. Shah, J. H. Grubb, A. Waheed and W. S. Sly (2001). "Expression of membrane-associated carbonic anhydrase XIV on neurons and axons in mouse and human brain." Proc Natl Acad Sci U S A 98(4): 1918-1923.

Parkkila, S., A. K. Parkkila, J. Saarnio, J. Kivela, T. J. Karttunen, K. Kaunisto, A. Waheed, W. S. Sly, O. Tureci, I. Virtanen and H. Rajaniemi (2000). "Expression of the membrane-associated carbonic anhydrase isozyme XII in the human kidney and renal tumors." J Histochem Cytochem 48(12): 1601-1608.

Paulus, J. K., D. Schlieper and G. Groth (2013). "Greater efficiency of photosynthetic carbon fixation due to single amino-acid substitution." Nat Commun 4: 1518.
Pearl, F., A. Todd, I. Sillitoe, M. Dibley, O. Redfern, T. Lewis, C. Bennett, R. Marsden, A. Grant, D. Lee, A. Akpor, M. Maibaum, A. Harrison, T. Dallman, G. Reeves, I. Diboun, S. Addou, S. Lise, C. Johnston, A. Sillero, J. Thornton and C. Orengo (2005). "The CATH Domain Structure Database and related resources Gene3D and DHS provide comprehensive domain family information for genome analysis." Nucleic Acids Res 33(Database issue): D247-251.
Pederzoli, A., M. Mandrioli and L. Mola (2014). "Expression of carbonic anhydrase, cystic fibrosis transmembrane regulator (CFTR) and V-H(+)-ATPase in the lancelet Branchiostoma lanceolatum (Pallas, 1774)." Acta Histochem 116(3): 487-492.
Pedrozo, H. A., Z. Schwartz, H. Nakaya, J. L. Harrison, D. D. Dean, M. L. Wiederhold and B. D. Boyan (1995). "Carbonic anhydrase is required for statoconia homeostasis in organ cultures of statocysts from Aplysia californica." J Comp Physiol A 177(4): 415-425.
Portal-Celhay, C., K. Nehrke and M. J. Blaser (2013). "Effect of Caenorhabditis elegans age and genotype on horizontal gene transfer in intestinal bacteria." FASEB J 27(2): 760-768.
Premkumar, L., U. K. Bageshwar, I. Gokhman, A. Zamir and J. L. Sussman (2003). "An unusual halotolerant alpha-type carbonic anhydrase from the alga Dunaliella salina functionally expressed in Escherichia coli." Protein Expr Purif 28(1): 151-157.
Purkerson, J. M. and G. J. Schwartz (2007). "The role of carbonic anhydrases in renal physiology." Kidney Int 71(2): 103-115.
Rapsch, C., G. Schweizer, F. Grimm, L. Kohler, C. Bauer, P. Deplazes, U. Braun and P. R. Torgerson (2006). "Estimating the true prevalence of Fasciola hepatica in cattle slaughtered in Switzerland in the absence of an absolute diagnostic test." Int J Parasitol 36(10-11): 1153-1158.
Rebello, G., R. Ramesar, A. Vorster, L. Roberts, L. Ehrenreich, E. Oppon, D. Gama, S. Bardien, J. Greenberg, G. Bonapace, A. Waheed, G. N. Shah and W. S. Sly (2004). "Apoptosis-inducing signal sequence mutation in carbonic anhydrase IV identified in patients with the RP17 form of retinitis pigmentosa." Proc Natl Acad Sci U S A 101(17): 6617-6622.
Reed, R. R. and C. D. Moser (1984). "Resolvase-mediated recombination intermediates contain a serine residue covalently linked to DNA." Cold Spring Harb Symp Quant Biol 49: 245-249.
Rice, P., I. Longden and A. Bleasby (2000). "EMBOSS: the European Molecular Biology Open Software Suite." Trends Genet 16(6): 276-277.
Riehl, B. and W. R. Schlue (1990). "Localization of carbonic anhydrase in identified glial cells of the leech central nervous system: application of histochemical and immunocytochemical methods." J Histochem Cytochem 38(8): 1173-1178.
Romualdi, A., M. Felder, D. Rose, U. Gausmann, M. Schilhabel, G. Glockner, M. Platzer and J. Suhnel (2007). "GenColors: annotation and comparative genomics of prokaryotes made easy." Methods Mol Biol 395: 75-96.
Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Hohna, B. Larget, L. Liu, M. A. Suchard and J. P. Huelsenbeck (2012). "MrBayes 3.2: efficient Bayesian

phylogenetic inference and model choice across a large model space." Syst Biol 61(3): 539-542.
Rowlett, R. S. (2010). "Structure and catalytic mechanism of the beta-carbonic anhydrases." Biochim Biophys Acta 1804(2): 362-373.
Rowlett, R. S., C. Tu, J. Lee, A. G. Herman, D. A. Chapnick, S. H. Shah and P. C. Gareiss (2009). "Allosteric site variants of Haemophilus influenzae beta-carbonic anhydrase." Biochemistry 48(26): 6146-6156.
Satoh, D., Y. Hiraoka, B. Colman and Y. Matsuda (2001). "Physiological and molecular biological characterization of intracellular carbonic anhydrase from the marine diatom Phaeodactylum tricornutum." Plant Physiol 126(4): 1459-1470.
Sawaya, M. R., G. C. Cannon, S. Heinhorst, S. Tanaka, E. B. Williams, T. O. Yeates and C. A. Kerfeld (2006). "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 281(11): 7546-7555.
Schneider, R., A. Obwaller and H. Auer (2015). "Immunoblot for the detection of Ascaris suum-specific antibodies in patients with visceral larva migrans (VLM) syndrome." Parasitol Res 114(1): 305-310.
Sender, S., B. Decker, C. D. Fenske, W. S. Sly, N. D. Carter and G. Gros (1998). "Localization of carbonic anhydrase IV in rat and human heart muscle." J Histochem Cytochem 46(7): 855-861.
Sherman, T. A., S. C. Rongali, T. A. Matthews, J. Pfeiffer and K. Nehrke (2012). "Identification of a nuclear carbonic anhydrase in Caenorhabditis elegans." Biochim Biophys Acta 1823(4): 808-817.
Siegmund, L., A. Burmester, M. S. Fischer and J. Wostemeyer (2013). "A model for endosymbiosis: interaction between Tetrahymena pyriformis and Escherichia coli." Eur J Protistol 49(4): 552-563.
Sievers, F. and D. G. Higgins (2014). "Clustal omega." Curr Protoc Bioinformatics 48: 3 13 11-13 13 16.
Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson and D. G. Higgins (2011). "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega." Mol Syst Biol 7: 539.
Siguier, P., J. Perochon, L. Lestrade, J. Mahillon and M. Chandler (2006). "ISfinder: the reference centre for bacterial insertion sequences." Nucleic Acids Res 34(Database issue): D32-36.
Singh, H., H. R. Ansari and G. P. Raghava (2013). "Improved method for linear B-cell epitope prediction using antigen's primary sequence." PLoS One 8(5): e62216.
Skriwan, C., M. Fajardo, S. Hagele, M. Horn, M. Wagner, R. Michel, G. Krohne, M. Schleicher, J. Hacker and M. Steinert (2002). "Various bacterial pathogens and symbionts infect the amoeba Dictyostelium discoideum." Int J Med Microbiol 291(8): 615-624.
Slesak, G., S. Inthalad, P. Basy, D. Keomanivong, O. Phoutsavath, S. Khampoui, A. Grosrenaud, V. Amstutz, H. Barennes, Y. Buisson and P. Odermatt (2011). "Ziehl-Neelsen staining technique can diagnose paragonimiasis." PLoS Negl Trop Dis 5(5): e1048.

Sly, W. S., D. Hewett-Emmett, M. P. Whyte, Y. S. Yu and R. E. Tashian (1983). "Carbonic anhydrase II deficiency identified as the primary defect in the autosomal recessive syndrome of osteopetrosis with renal tubular acidosis and cerebral calcification." Proc Natl Acad Sci U S A 80(9): 2752-2756.

Smith, K. S. and J. G. Ferry (1999). "A plant-type (beta-class) carbonic anhydrase in the thermophilic methanoarchaeon Methanobacterium thermoautotrophicum." J Bacteriol 181(20): 6247-6253.

Smith, M. C. and H. M. Thorpe (2002). "Diversity in the serine recombinases." Mol Microbiol 44(2): 299-307.

Smutna, T., V. L. Goncalves, L. M. Saraiva, J. Tachezy, M. Teixeira and I. Hrdy (2009). "Flavodiiron protein from Trichomonas vaginalis hydrogenosomes: the terminal oxygen reductase." Eukaryot Cell 8(1): 47-55.

So, A. K. and G. S. Espie (1998). "Cloning, characterization and expression of carbonic anhydrase from the cyanobacterium Synechocystis PCC6803." Plant Mol Biol 37(2): 205-215.

Sobecky, P. A. and T. H. Hazen (2009). "Horizontal gene transfer and mobile genetic elements in marine systems." Methods Mol Biol 532: 435-453.

Solomon, J. M. and A. D. Grossman (1996). "Who's competent and when: regulation of natural genetic competence in bacteria." Trends Genet 12(4): 150-155.

Spicer, S. S., M. A. Sens and R. E. Tashian (1982). "Immunocytochemical demonstration of carbonic anhydrase in human epithelial cells." J Histochem Cytochem 30(9): 864-873.

Steinmann, P., G. Cringoli, F. Bruschi, B. Matthys, L. K. Lohourignon, B. Castagna, M. P. Maurelli, M. E. Morgoglione, J. Utzinger and L. Rinaldi (2012). "FLOTAC for the diagnosis of Hymenolepis spp. infection: proof-of-concept and comparing diagnostic accuracy with other methods." Parasitol Res 111(2): 749-754.

Strese, A., A. Backlund and C. Alsmark (2014). "A recently transferred cluster of bacterial genes in Trichomonas vaginalis--lateral gene transfer and the fate of acquired genes." BMC Evol Biol 14: 119.

Sun, H. Y., J. Noe, J. Barber, R. S. Coyne, D. Cassidy-Hanley, T. G. Clark, R. C. Findly and H. W. Dickerson (2009). "Endosymbiotic bacteria in the parasitic ciliate Ichthyophthirius multifiliis." Appl Environ Microbiol 75(23): 7445-7452.

Supuran, C. T. (2008). "Carbonic anhydrases: novel therapeutic applications for inhibitors and activators." Nat Rev Drug Discov 7(2): 168-181.

Supuran, C. T. (2010). "Carbonic anhydrase inhibition/activation: trip of a scientist around the world in the search of novel chemotypes and drug targets." Curr Pharm Des 16(29): 3233-3245.

Supuran, C. T. (2011). "Bacterial carbonic anhydrases as drug targets: toward novel antibiotics?" Front Pharmacol 2: 34.

Supuran, C. T. (2016). "Structure and function of carbonic anhydrases." Biochem J 473(14): 2023-2032.

Sussman, J. L., D. Lin, J. Jiang, N. O. Manning, J. Prilusky, O. Ritter and E. E. Abola (1998). "Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules." Acta Crystallogr D Biol Crystallogr 54(Pt 6 Pt 1): 1078-1084.

Suyama, M., D. Torrents and P. Bork (2006). "PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments." Nucleic Acids Res 34(Web Server issue): W609-612.

Syrjanen, L., M. Kuuslahti, M. Tolvanen, D. Vullo, S. Parkkila and C. T. Supuran (2015). "The beta-carbonic anhydrase from the malaria mosquito Anopheles gambiae is highly inhibited by sulfonamides." Bioorg Med Chem 23(10): 2303-2309.

Syrjanen, L., S. Parkkila, A. Scozzafava and C. T. Supuran (2014). "Sulfonamide inhibition studies of the beta carbonic anhydrase from Drosophila melanogaster." Bioorg Med Chem Lett 24(13): 2797-2801.

Syrjanen, L., M. Tolvanen, M. Hilvo, A. Olatubosun, A. Innocenti, A. Scozzafava, J. Leppiniemi, B. Niederhauser, V. P. Hytonen, T. A. Gorr, S. Parkkila and C. T. Supuran (2010). "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 11: 28.

Syrjanen, L., S. Valanne, M. Kuuslahti, T. Tuomela, A. Sriram, A. Sanz, H. T. Jacobs, M. Ramet and S. Parkkila (2015). "beta carbonic anhydrase is required for female fertility in Drosophila melanogaster." Front Zool 12: 19.

Syrjanen, L., A. B. Vermelho, A. Rodrigues Ide, S. Corte-Real, T. Salonen, P. Pan, D. Vullo, S. Parkkila, C. Capasso and C. T. Supuran (2013). "Cloning, characterization, and inhibition studies of a beta-carbonic anhydrase from Leishmania donovani chagasi, the protozoan parasite responsible for leishmaniasis." J Med Chem 56(18): 7372-7381.

Szuplewska, M., J. Czarnecki and D. Bartosik (2014). "Autonomous and non-autonomous Tn3-family transposons and their role in the evolution of mobile genetic elements." Mob Genet Elements 4(6): 1-4.

Tabaran, F., A. D. Sandor, M. Marinov, C. Catoi and A. D. Mihalca (2013). "Alaria alata Infection in European mink." Emerg Infect Dis 19(9): 1547-1549.

Tachibana, M., A. E. Allen, S. Kikutani, Y. Endo, C. Bowler and Y. Matsuda (2011). "Localization of putative carbonic anhydrases in two marine diatoms, Phaeodactylum tricornutum and Thalassiosira pseudonana." Photosynth Res 109(1-3): 205-221.

Takeuchi-Storm, N., M. N. Al-Sabi, S. M. Thamsborg and H. L. Enemark (2015). "Alaria alata Mesocercariae among Feral Cats and Badgers, Denmark." Emerg Infect Dis 21(10): 1872-1874.

Takeuchi, N., K. Kaneko and E. V. Koonin (2014). "Horizontal gene transfer can rescue prokaryotes from Muller's ratchet: benefit of DNA from dead cells and population subdivision." G3 (Bethesda) 4(2): 325-339.

Tanaka, Y., D. Nakatsuma, H. Harada, M. Ishida and Y. Matsuda (2005). "Localization of soluble beta-carbonic anhydrase in the marine diatom Phaeodactylum tricornutum. Sorting to the chloroplast and cluster formation on the girdle lamellae." Plant Physiol 138(1): 207-217.

Tanyuksel, M. and W. A. Petri, Jr. (2003). "Laboratory diagnosis of amebiasis." Clin Microbiol Rev 16(4): 713-729.

Tappe, D. and D. W. Buttner (2009). "Diagnosis of human visceral pentastomiasis." PLoS Negl Trop Dis 3(2): e320.

Taylor-Brown, E. and H. Hurd (2013). "The first suicides: a legacy inherited by parasitic protozoans from prokaryote ancestors." Parasit Vectors 6: 108.
Tetu, S. G., S. K. Tanz, N. Vella, J. N. Burnell and M. Ludwig (2007). "The Flaveria bidentis beta-carbonic anhydrase gene family encodes cytosolic and chloroplastic isoforms demonstrating distinct organ-specific expression patterns." Plant Physiol 144(3): 1316-1327.
Thorpe, H. M. and M. C. Smith (1998). "In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family." Proc Natl Acad Sci U S A 95(10): 5505-5510.
Tobal, J. M. and M. E. Balieiro (2014). "Role of carbonic anhydrases in pathogenic micro-organisms: a focus on Aspergillus fumigatus." J Med Microbiol 63(Pt 1): 15-27.
Top, E. M. and D. Springael (2003). "The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds." Curr Opin Biotechnol 14(3): 262-269.
Toussaint, A. and C. Merlin (2002). "Mobile elements as a combination of functional modules." Plasmid 47(1): 26-35.
Tripp, B. C., K. Smith and J. G. Ferry (2001). "Carbonic anhydrase: new insights for an ancient enzyme." J Biol Chem 276(52): 48615-48618.
Tureci, O., U. Sahin, E. Vollmar, S. Siemer, E. Gottert, G. Seitz, A. K. Parkkila, G. N. Shah, J. H. Grubb, M. Pfreundschuh and W. S. Sly (1998). "Human carbonic anhydrase XII: cDNA cloning, expression, and chromosomal localization of a carbonic anhydrase gene that is overexpressed in some renal cell cancers." Proc Natl Acad Sci U S A 95(13): 7608-7613.
Udonsom, R., R. Buddhirongawatr and Y. Sukthana (2010). "Is Sabin-Feldman dye test using T. gondii tachyzoites from animal inoculation still the best method for detecting Toxoplasma gondii antibodies?" Southeast Asian J Trop Med Public Health 41(5): 1059-1064.
Vaananen, H. K. (1984). "Immunohistochemical localization of carbonic anhydrase isoenzymes I and II in human bone, cartilage and giant cell tumor." Histochemistry 81(5): 485-487.
Vaananen, H. K., M. Paloniemi and J. Vuori (1985). "Purification and localization of human carbonic anhydrase. III. Typing of skeletal muscle fibers in paraffin embedded sections." Histochemistry 83(3): 231-235.
van Karnebeek, C. D., W. S. Sly, C. J. Ross, R. Salvarinova, J. Yaplito-Lee, S. Santra, C. Shyr, G. A. Horvath, P. Eydoux, A. M. Lehman, V. Bernard, T. Newlove, H. Ukpeh, A. Chakrapani, M. A. Preece, S. Ball, J. Pitt, H. D. Vallance, M. Coulter-Mackie, H. Nguyen, L. H. Zhang, A. P. Bhavsar, G. Sinclair, A. Waheed, W. W. Wasserman and S. Stockler-Ipsiroglu (2014). "Mitochondrial carbonic anhydrase VA deficiency resulting from CA5A alterations presents with hyperammonemia in early childhood." Am J Hum Genet 94(3): 453-461.
Veitch, F. P. and L. C. Blankenship (1963). "Carbonic anhydrase in bacteria." Nature 197: 76-77.
Vogeler, S., T. S. Galloway, B. P. Lyons and T. P. Bean (2014). "The nuclear receptor gene family in the Pacific oyster, Crassostrea gigas, contains a novel subfamily group." BMC Genomics 15: 369.

Vullo, D., S. Del Prete, S. M. Osman, A. Scozzafava, Z. Alothman, C. T. Supuran and C. Capasso (2014). "Anion inhibition study of the beta-class carbonic anhydrase (PgiCAb) from the oral pathogen Porphyromonas gingivalis." Bioorg Med Chem Lett 24(18): 4402-4406.

Vullo, D., M. Franchi, E. Gallori, J. Antel, A. Scozzafava and C. T. Supuran (2004). "Carbonic anhydrase inhibitors. Inhibition of mitochondrial isozyme V with aromatic and heterocyclic sulfonamides." J Med Chem 47(5): 1272-1279.

Vullo, D., A. Innocenti, I. Nishimori, J. Pastorek, A. Scozzafava, S. Pastorekova and C. T. Supuran (2005). "Carbonic anhydrase inhibitors. Inhibition of the transmembrane isozyme XII with sulfonamides-a new target for the design of antitumor and antiglaucoma drugs?" Bioorg Med Chem Lett 15(4): 963-969.

Vullo, D., I. Nishimori, T. Minakuchi, A. Scozzafava and C. T. Supuran (2011). "Inhibition studies with anions and small molecules of two novel beta-carbonic anhydrases from the bacterial pathogen Salmonella enterica serovar Typhimurium." Bioorg Med Chem Lett 21(12): 3591-3595.

Waack, S., O. Keller, R. Asper, T. Brodag, C. Damm, W. F. Fricke, K. Surovcik, P. Meinicke and R. Merkl (2006). "Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models." BMC Bioinformatics 7: 142.

Waki, K., H. Oi, S. Takahashi, T. Nakabayashi and T. Kitani (1986). "Successful treatment of Diphyllobothrium latum and Taenia saginata infection by intraduodenal 'Gastrografin' injection." Lancet 2(8516): 1124-1126.

Walz, B. (1988). "Brain but not retinal glial cells have carbonic anhydrase activity in the honeybee drone." Neurosci Lett 85(1): 47-50.

Waterhouse, A. M., J. B. Procter, D. M. Martin, M. Clamp and G. J. Barton (2009). "Jalview Version 2--a multiple sequence alignment editor and analysis workbench." Bioinformatics 25(9): 1189-1191.

Weber, P. C., C. D. Cunningham, 3rd and B. A. Schulte (2001). "Potassium recycling pathways in the human cochlea." Laryngoscope 111(7): 1156-1165.

Wernegreen, J. J. (2004). "Endosymbiosis: lessons in conflict resolution." PLoS Biol 2(3): E68.

Wilberg, E. W. (2015). "What's in an Outgroup? The Impact of Outgroup Choice on the Phylogenetic Position of Thalattosuchia (Crocodylomorpha) and the Origin of Crocodyliformes." Syst Biol 64(4): 621-637.

Wistrand, P. J. and K. G. Knuuttila (1989). "Renal membrane-bound carbonic anhydrase. Purification and properties." Kidney Int 35(3): 851-859.

Wolfensberger, T. J., I. Mahieu, J. Jarvis-Evans, M. Boulton, N. D. Carter, A. Nogradi, E. Hollande and A. C. Bird (1994). "Membrane-bound carbonic anhydrase in human retinal pigment epithelium." Invest Ophthalmol Vis Sci 35(9): 3401-3407.

Worasith, C., C. Kamamia, A. Yakovleva, K. Duenngai, C. Wangboon, J. Sithithaworn, N. Watwiengkam, N. Namwat, A. Techasen, W. Loilome, P. Yongvanit, A. Loukas, P. Sithithaworn and J. M. Bethony (2015). "Advances in the Diagnosis of Human Opisthorchiasis: Development of Opisthorchis viverrini Antigen Detection in Urine." PLoS Negl Trop Dis 9(10): e0004157.

Xia, J., S. Cai, H. Zhang, W. Lin, Y. Fan, J. Qiu, L. Sun, B. Chang, Z. Zhang and S. Nie (2015). "Spatial, temporal, and spatiotemporal analysis of malaria in Hubei Province, China from 2004-2011." Malar J 14: 145.

Xu, Y., L. Feng, P. D. Jeffrey, Y. Shi and F. M. Morel (2008). "Structure and metal exchange in the cadmium carbonic anhydrase of marine diatoms." Nature 452(7183): 56-61.

Yamasaki, H., T. Nakamura, P. M. Intapan, W. Maleewong, Y. Morishima, H. Sugiyama, H. Matsuoka, K. Kobayashi, K. Takayama and Y. Kobayashi (2014). "Development of a rapid diagnostic kit that uses an immunochromatographic device to detect antibodies in human sparganosis." Clin Vaccine Immunol 21(9): 1360-1363.

Ye, J., G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen and T. L. Madden (2012). "Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction." BMC Bioinformatics 13: 134.

Yones, D. A., L. A. Galal, A. M. Abdallah and K. S. Zaghlol (2015). "Effect of enteric parasitic infection on serum trace elements and nutritional status in upper Egyptian children." Trop Parasitol 5(1): 29-35.

Zabaleta, E., M. V. Martin and H. P. Braun (2012). "A basal carbon concentrating mechanism in plants?" Plant Sci 187: 97-104.

Zolfaghari Emameh, R., L. Syrjanen, H. Barker, C. T. Supuran and S. Parkkila (2015). "Drosophila melanogaster: a model organism for controlling Dipteran vectors and pests." J Enzyme Inhib Med Chem 30(3): 505-513.

10. Original communications

Parasites & Vectors

RESEARCH Open Access

Bioinformatic analysis of beta carbonic anhydrase sequences from protozoans and metazoans

Reza Zolfaghari Emameh[1,2*], Harlan Barker[1,2], Martti E E Tolvanen[2,3], Csaba Ortutay[2] and Seppo Parkkila[1,2,4]

Abstract

Background: Despite the high prevalence of parasitic infections, and their impact on global health and economy, the number of drugs available to treat them is extremely limited. As a result, the potential consequences of large-scale resistance to any existing drugs are a major concern. A number of recent investigations have focused on the effects of potential chemical inhibitors on bacterial and fungal carbonic anhydrases. Among the five classes of carbonic anhydrases (alpha, beta, gamma, delta and zeta), beta carbonic anhydrases have been reported in most species of bacteria, yeasts, algae, plants, and particular invertebrates (nematodes and insects). To date, there has been a lack of knowledge on the expression and molecular structure of beta carbonic anhydrases in metazoan (nematodes and arthropods) and protozoan species.

Methods: Here, the identification of novel beta carbonic anhydrases was based on the presence of the highly-conserved amino acid sequence patterns of the active site. A phylogenetic tree was constructed based on codon-aligned DNA sequences. Subcellular localization prediction for each identified invertebrate beta carbonic anhydrase was performed using the TargetP webserver.

Results: We verified a total of 75 beta carbonic anhydrase sequences in metazoan and protozoan species by proteome-wide searches and multiple sequence alignment. Of these, 52 were novel, and contained highly conserved amino acid residues, which are inferred to form the active site in beta carbonic anhydrases. Mitochondrial targeting peptide analysis revealed that 31 enzymes are predicted with mitochondrial localization; one was predicted to be a secretory enzyme, and the other 43 were predicted to have other undefined cellular localizations.

Conclusions: These investigations identified 75 beta carbonic anhydrases in metazoan and protozoan species, and among them there were 52 novel sequences that were not previously annotated as beta carbonic anhydrases. Our results will not only change the current information in proteomics and genomics databases, but will also suggest novel targets for drugs against parasites.

Keywords: Beta carbonic anhydrase, Inhibitor, Metazoa, Mitochondrial targeting peptide, Multiple sequence alignment, Protozoa

Background

Carbonic anhydrases (CAs) are ubiquitous metalloenzymes. They are encoded by five evolutionary divergent gene families and the corresponding enzymes are designated α, β, γ, δ and ζ-CAs. α-CAs are present in animals, some fungi, bacteria, algae, and cytoplasm of green plants. β-CAs are expressed mainly in fungi, bacteria, archaea, algae, and chloroplasts of monocotyledons and dicotyledons. γ-CAs are expressed in plants, archaea, and some bacteria. δ- and ζ-CAs are present in several classes of marine phytoplankton [1-6]. A total of 13 enzymatically active α-CAs have been reported in mammals: CA I, CA II, CA III, CA VII, and CA XIII are cytosolic enzymes; CA IV, CA IX, CA XII, CA XIV, and CA XV are membrane-bound; CA VA and CA VB are mitochondrial; CA VI is secreted and CA VIII, CA X, and CA XI are acatalytic CA-related proteins [3,7]. The active site of CA contains a zinc ion (Zn^{2+}) which has a critical role in the catalytic activity of the enzyme. ζ-and γ-CAs represent exceptions to this rule since they can use

* Correspondence: reza.zolfaghari.emameh@uta.fi
[1]School of Medicine, University of Tampere, Medisiinarinkatu 3, 33520 Tampere, Finland
[2]Institute of Biomedical Technology and BioMediTech, University of Tampere, 33520 Tampere, Finland
Full list of author information is available at the end of the article

cadmium (ζ), iron (γ), or cobalt (γ) as cofactors [8-10]. CAs are involved in many biological processes, such as respiration involving transport of CO_2 and bicarbonate between metabolizing tissues, pH homeostasis, electrolyte transfer, bone resorption, calcification, and tumor progression. They also participate in some biosynthetic reactions, such as gluconeogenesis, lipogenesis, and ureagenesis [3,11-14].

The first β-CA was serendipitously discovered by Neish in 1939 [15]. In 1990, the cDNA sequence of spinach (*Spinacea oleracea*) chloroplast CA was determined, and found to be non-homologous to animal α-CA [16,17]. Thereafter, cDNA sequences of β-CA from pea (*Pisium sativum*) and *Arabidopsis thaliana* were determined [17-19]. It is believed that the plant β-CAs are distributed in the chloroplastic stroma, thylakoid space, and cytoplasm of plant cells [17]. Many putative β-CAs have been discovered since 1990, not only in photosynthetic organisms, but also in eubacteria, yeast, and archaea [17].

The first bacterial β-CA gene was named *CynT* and recognized in *Escherichia coli* [20,21]. Later, β-CA was identified in some other pathogenic bacteria, such as *Helicobacter pylori, Mycobacterium tuberculosis, Salmonella typhimurium* [17,22], *Haemophilus influenzae* [23,24], *Brucella suis* [24,25], *Streptococcus pneumoniae* [24,26], *Salmonella enterica* [24,27], and *Vibrio cholerae* [24,28,29]. β-CAs have also been identified in fungi, such as *Candida albicans* [1,30], *Candida glabrata* [1,31], *Cryptococcus neoformans* [1,32], and *Sordaria macrospora* [6,33]. This class of enzyme has also been discovered in a wide range of taxa, such as yeast (*Saccharomyces cerevisiae*) [34-36], cyanobacteria (*Synechocystis* sp. PCC6803) [37], carboxysomes of chemoautotrophic bacteria (*Halothiobacillus neapolitanus*) [38], green algae (*Chlamydomonas reinhardtii*) [39], red algae (*Porphyridium purpureum*) [40], nematodes (*Caenorhabditis elegans)* [41], and insects (*Drosophila melanogaster)* [4]. While β-CAs were initially thought to be expressed only in plants, this enzyme family is indeed present in a wide variety of species – from bacteria and archaea to invertebrate animals, missing only from vertebrates and most chordates, making it an attractive target for evolutionary studies [5].

β-CA is an important accessory enzyme for many CO_2 or HCO_3^- utilizing enzymes (e.g. RuBisCO in chloroplasts, cyanase in *E. coli* [42], urease in *H. pylori* [43], and carboxylases in *Corynebacterium glutamicum* [44]). In cyanobacteria, β-CA is an essential component of the CO_2-concentrating carboxysome organelle [17,45]. β-CA activity is required for growth of *E. coli* bacteria in air [46]; it is also indispensable if the atmospheric partial pressure of CO_2 is high or during anaerobic growth in a closed vessel at low pH, where copious CO_2 is generated endogenously. β-CA is also needed for growth of *C. glutamicum* [44,47] and some yeasts, such as *S. cerevisiae* [40]. In higher plants, the *Flaveria bidentis* genome contains at least three β-CA genes, named *CA1*, *CA2*, and *CA3* [48]. The functional roles of β-CAs in plants are not yet fully understood, even though a lot of new data has emerged in recent years. C_3 and C_4 plants have different mechanisms for carbon fixation and photosynthesis and, thus, β-CAs might possess different roles, depending on the location of the enzyme and the type of plant [49]. In plants, the highest CA activity has been found within the chloroplast stroma, but there is also some CA activity in the cytosol of mesophyll cells [50]. Carbon dioxide coming from the external environment must be rapidly hydrated by β-CA and converted into HCO_3^- for the phosphoenolpyruvate carboxylase enzyme [49]. Additionally, CAs play a role in photosynthesis by facilitating diffusion into and across the chloroplast, and by catalyzing HCO_3^- dehydration to supply CO_2 for RuBisCO. Interestingly, both RuBisCO and β-CA expression levels increase together when *P. sativum* is transferred from an environment with high levels of CO_2 to one with low levels [47].

Crystal structures of β-CAs reveal that a zinc ion (Zn^{2+}) is ligated by two conserved cysteines and one conserved histidine [5]. Until now, the only X-ray crystallography structure defined for β-CAs in plants belongs to *P. sativum* [51]. *E. coli* was the first bacteria in which the β-CA crystal structure was determined [20]. β-CA can adopt a variety of oligomeric states with molecular masses ranging from 45 to 200 kDa [52].

The first metazoan β-CAs were reported in 2010 [41]. In one of the studies [4,41], two genes encoding β-CAs (y116a8c.28 and bca-1) were identified in *Caenorhabditis elegans*. Another study reported a novel β-CA gene identified from FlyBase, which was named DmBCA (short for *Drosophila melanogaster* β-CA) [4]. Additionally, orthologs were retrieved from sequence databases, and reconstructed when necessary. The results confirmed the presence of β-CA sequences in 55 metazoan species, such as *Aedes aegypti, Culex quinquefasciatus, Anopheles gambiae, Drosophila virilis, Tribolium castaneum, Nasonia vitripennis, Apis mellifera, Acyrthosiphon pisum, Daphnia pulex, Caenorhabditis elegans, Pristionchus pacificus, Trichoplax adhaerens, Caligus clemensi, Lepeophtheirus salmonis, Nematostella vectensis, Strongylocentrotus purpuratus,* and *Saccoglossus kowalevskii*. The DmBCA enzyme was produced as a recombinant protein in Sf9 insect cells, and its kinetic and inhibition profiles were determined. The enzyme showed high CO_2 hydratase activity, with a k_{cat} of 9.5×10^5 s^{-1} and a k_{cat}/K_M of 1.1×10^8 M^{-1} s^{-1}. DmBCA was inhibited by the clinically-used sulfonamide, acetazolamide, with an inhibition constant of 49 nM. Subcellular localization studies have indicated that DmBCA is probably a mitochondrial enzyme, as is also suggested by sequence analysis.

In this study, using bioinformatics tools, we discovered and verified the presence of β-CA in various other

metazoan species, and, for the first time, in protozoa. Previously, most β-CA proteins have been identified in protein databases as 'unknown' proteins or 'putative' CAs, without a specific reference to β-CAs. Based on the present findings, new avenues will be opened to biochemically characterize β-CAs and their inhibitors in arthropods, nematodes and protozoans.

Methods

Identification of putative β-CA enzymes in protozoan and metazoan species and multiple sequence alignment

Identification of novel β-CAs was based on the presence of the highly-conserved amino acid sequence patterns of the active site, namely Cys-Xaa-Asp-Xaa-Arg and His-Xaa-Xaa-Cys also marked in Additional file 1: Figure S1. Alignment was visualized in Jalview [53]. In total, 75 invertebrate β-CA sequences were retrieved from Uniprot (http://www.uniprot.org/) for alignment analysis, and one bacterial sequence (*Pelosinus fermentans*) was included as an outgroup. All protein sequences were aligned using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) [54]. The sequences were manually curated to remove residues associated with an incorrect starting methionine. A total of 90 residues were removed from the N-terminal end of Uniprot IDs D4NWE5_ADIVA, G0QPN9_ICHMG, D6WK56_TRICA, I7LWM1_TETTS and I7M0M0_TETTS. The modified protein sequences were then re-aligned. This protein alignment then served as the template for codon alignment of corresponding nucleotide sequences using the Pal2Nal program (http://www.bork.embl.de/pal2nal/) [55].

Phylogenetic analysis

The phylogenetic analysis was computed using Mr. Bayes v3.2 [56]. After 8 million generations using the GTR codon substitution model, with all other parameters as default, the standard deviation of split frequencies was 1.39×10^{-3}. The final output tree was produced using 50% majority rule consensus. FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/) [56] was used to visualize the phylogenetic tree and the *Pelosinus fermentans* [57] sequence set as outgroup. Additional trees were constructed for comparison using maximum likelihood (PhyML)[58], UPGMA, and neighbor-joining methods within Geneious version 7.0.5 from Biomatters (Auckland, New Zealand) (http://www.geneious.com/).

Prediction of subcellular localization

Subcellular localization prediction of each identified invertebrate β-CA was performed using the TargetP webserver (http://www.cbs.dtu.dk/services/TargetP/). TargetP is built from two layers of neural networks, where the first layer contains one dedicated network for each type of pre-sequence [cTP (cytoplasmic targeting peptide), mTP (mitochondrial targeting peptide, or SP (secretory signal peptide)], and the second is an integrating network that outputs the actual prediction (cTP, mTP, SP, other). It is able to discriminate between cTPs, mTPs, and SPs with sensitivities and specificities higher than what has been obtained with other available subcellular localization predictors [59].

Results

Multiple sequence alignment

The Uniprot search of potential β-CA sequences, and the subsequent multiple sequence alignment, identified 75 β-CAs in metazoan and protozoan species, of which 23 sequences were reported as β-CAs previously [4]. Thus, 52 metazoan and protozoan β-CA sequences were novel and reported here for the first time. All 75 β-CAs in metazoan and protozoan species are shown in Table 1. The multiple sequence alignment results of these 75 β-CAs, plus a bacterial β-CA sequence from *Pelosinus fermentans*, are shown as Additional file 1: Figure S1. Multiple sequence alignment of all animal β-CAs confirmed conservation of the known active site motifs CxDxR and HxxC in all identified enzymes. Several other key residues were also highly conserved. Notably, all β-CA sequences from *Leishmania* species (*Leishmania donovani*, *Leishmania infantum*, *Leishmania major*, and *Leishmania mexicana*) contained a 71 residue N-terminal extension not present in any other sequences.

Phylogenetic analysis

The results of the phylogenetic analysis of 75 β-CAs in metazoan and protozoan species are shown in Figure 1. A β-CA sequence from the *Pelosinus fermentans* bacterium was used as an outgroup [60]. The phylogenetic results represent the evolutionary root of β-CAs in metazoan and protozoan species, the similarity between them, and duplications that have occurred. The branching pattern and branch lengths reveal interesting evolutionary relationships of β-CAs in various invertebrate species. There is a close relationship between our bacterial outgroup and *Trichomonas vaginalis* β-CAs, both having originated well before the other species within the tree. β-CAs of nematodes and arthropods are located in the lower evolutionary branches. In the protozoan *Tetrahymena thermophilia* and *Paramecium tetraurelia* clades significant duplications of β-CA have occurred, with 8 and 5 distinct proteins respectively. Meanwhile, metazoan and nematode species tend to have just one or two β-CAs. Surprisingly, β-CAs of the nematode *Trichinella spiralis* and trematode *Schistosoma mansoni* appear more closely related to arthropod than to nematode enzymes. The triangle located near the bottom of Figure 1 represents the clade of β-CAs in different *Drosophila* species. The details of the phylogenetic tree of β-CAs in *Drosophila* species are shown in Figure 2. The likely presence of inaccuracies in some of the database

Table 1 Identified β-CAs in protozoan and metazoan species

Species	β- CA ID	Entry ID	Gene name	Protein name
Acromyrmex echinatior	BCA	F4WAG3	G5I_02499	Beta carbonic anhydrase 1
Acyrthosiphon pisum	BCA1	J9K706	Uncharacterized	Uncharacterized
	BCA2	C4WVD8	ACYPI006033	ACYPI006033
	BCA3	J9JZY3	XM_001950078.1	Uncharacterized
Adineta vaga	BCA	D4NWE5	Uncharacterized	Putative uncharacterized protein
Aedes aegypti	BCA	Q17N64	AAEL000816	AAEL000816-PA
Ancylostoma caninum	BCA	FC551456	Uncharacterized	Uncharacterized protein
Anopheles darlingi	BCA	E3X5Q8	AND_14274	Uncharacterized protein
Anopheles gambiae	BCA	Q5TU56	AGAP002992 AgaP_AGAP002992	AGAP002992-PA
Apis mellifera	BCA	H9KS29	Uncharacterized	Uncharacterized protein
Ascaris suum	BCA	F1LE18	Uncharacterized	Beta carbonic anhydrase 1
Caenorhabditis brenneri	BCA1	G0MSW4	Cbn-bca-1 CAEBREN_17105	CBN-BCA-1 protein
	BCA2	G0MRG1	Cbn-bca-2 CAEBREN_06024	CBN-BCA-2 protein
Caenorhabditis briggsae	BCA1	A8XKV0	bca-1 CBG14861	Beta carbonic anhydrase 1
	BCA2	A8WN21	bca-2 Cbr-bca-2 cbr-bca-2 CBG00424 CBG_00424	Protein CBR-BCA-2
Caenorhabditis elegans	BCA1	Q22460	bca-1 T13C5.5	Beta carbonic anhydrase 1
	BCA2	Q2YS41	bca-2 Y116A8C.28	Protein BCA-2
Caenorhabditis remanei	BCA1	E3LDN3	Cre-bca-1 CRE_00190	CRE-BCA-1 protein
	BCA2	E3MK96	Cre-bca-2 CRE_28742	CRE-BCA-2 protein
Caligus clemensi	BCA	C1C2M7	CYNT	Carbonic anhydrase
Camponotus floridanus	BCA	E2ANQ9	EAG_05651	Carbonic anhydrase
Culex quinquefasciatus	BCA	B0WKV7	CpipJ_CPIJ007527	Carbonic anhydrase
Danaus plexippus	BCA	G6D7Z4	Uncharacterized	Putative carbonic anhydrase
Daphnia pulex	BCA	E9GLB5	CAB	Beta-carbonic anhydrase
Dendroctonus ponderosae	BCA	J3JTM9	Uncharacterized	Uncharacterized protein
Drosophila ananassae	BCA	B3LZ10	GF17694 Dana\GF17694 Dana_GF17694	GF17694
Drosophila erecta	BCA	B3P1V8	GG13874 Dere\GG13874 Dere_GG13874	GG13874
Drosophila grimshawi	BCA	B4JHY1	GH19010 Dgri\GH19010 Dgri_GH19010	GH19010
Drosophila melanogaster	BCA	Q9VHJ5	CAHbeta CG11967 Dmel_CG11967	CG11967
Drosophila mojavensis	BCA	B4KDC1	GI23065 Dmoj\GI23065 Dmoj_GI23065	GI23065
Drosophila persimilis	BCA	B4GFA1	GL22171 Dper\GL22171 Dper_GL22171	GL22171
Drosophila pseudoobscura	BCA	Q296E4	GA11301 Dpse\GA11301 Dpse_GA11301	GA11301
Drosophila sechellia	BCA	B4HKY7	GM23772 Dsec\GM23772 Dsec_GM23772	GM23772
Drosophila simulans	BCA	B4QXC5	GD18582 Dsim\GD18582 Dsim_GD18582	GD18582
Drosophila virilis	BCA	B4LZE7	CAHbeta Dvir\GJ24578 GJ24578 Dvir_GJ24578	GJ24578
Drosophila willistoni	BCA	B4NBB9	GK11865 Dwil\GK11865 Dwil_GK11865	GK11865
Drosophila yakuba	BCA	B4PTY0	GE25916 Dyak\GE25916 Dyak_GE25916	GE25916
Entamoeba dispar	BCA	B0E7M0	EDI_275880	Carbonic anhydrase
Entamoeba histolytica	BCA	C4LXK3	EHI_073380	Carbonic anhydrase
Entamoeba nuttalli	BCA	K2GQM0	ENU1_204230	Carbonate dehydratase domain containing protein
Harpegnathos saltator	BCA	E2B2Q1	EAI_05019	Carbonic anhydrase
Heliconius melpomene	BCA	HMEL015257	Uncharacterized	Uncharacterized protein
Hirudo medicinalis	BCA	EY481200	Uncharacterized	Uncharacterized protein

Table 1 Identified β-CAs in protozoan and metazoan species *(Continued)*

Ichthyophthirius multifiliis	BCA	G0QPN9	IMG5_069900	Carbonic anhydrase
Leishmania donovani	BCA	E9B8S3	LDBPK_060630	Carbonic anhydrase
Leishmania infantum	BCA	A4HSV2	LINJ_06_0630	Carbonic anhydrase
Leishmania major	BCA	Q4QJ17	LMJF_06_0610	Carbonic anhydrase
Leishmania mexicana	BCA	E9AKU0	LMXM_06_0610	Carbonic anhydrase
Lepeophtheirus salmonis	BCA	D3PI48	BCA1	Beta carbonic anhydrase 1
Nasonia vitripennis	BCA	K7IWK8	Uncharacterized	Uncharacterized protein
Nematostella vectensis	BCA	A7S717	v1g186479	Predicted protein
Paramecium tetraurelia	BCA1	A0BD61	GSPATT00004572001	Carbonic anhydrase
	BCA2	A0E8J0	GSPATT00024336001	Carbonic anhydrase
	BCA3	A0CEX6	GSPATT00037782001	Carbonic anhydrase
	BCA4	A0BDB1	GSPATT00004622001	Carbonic anhydrase
	BCA5	A0C922	GSPATT00006595001	Carbonic anhydrase
Saccoglossus kowalevskii	BCA	187043763	Uncharacterized	Uncharacterized protein
Schistosoma mansoni	BCA	G4V6B2	Smp_004070	Putative carbonic anhydrase
Solenopsis invicta	BCA	E9IP13	SINV_09652	Putative carbonic anhydrase
Strigamia maritima	BCA	SMAR006741	Uncharacterized	Uncharacterized protein
Strongylocentrotus purpuratus	BCA	H3I177	Uncharacterized	Uncharacterized protein
Tetrahymena thermophila	BCA1	Q22U21	TTHERM_00263620	Carbonic anhydrase
	BCA2	Q22U16	TTHERM_00263670	Carbonic anhydrase
	BCA3	I7MDL7	TTHERM_00373840	Carbonic anhydrase
	BCA4	I7LWM1	TTHERM_00558270	Carbonic anhydrase
	BCA5	I7M0M0	TTHERM_00374880	Carbonic anhydrase
	BCA6	I7MD92	TTHERM_00541480	Carbonic anhydrase
	BCA7	I7M748	TTHERM_00374870	Carbonic anhydrase
	BCA8	Q23AV1	TTHERM_00654260	Carbonic anhydrase
Tribolium castaneum	BCA	D6WK56	TcasGA2_TC014816	Putative uncharacterized protein
Trichinella spiralis	BCA	E5SH53	Uncharacterized	Carbonic anhydrase
Trichomonas vaginalis	BCA1	A2ENQ8	TVAG_005270	Carbonic anhydrase
	BCA2	A2DLG4	TVAG_268150	Carbonic anhydrase
Trichoplax adhaerens	BCA	B3S5Y1	TRIADDRAFT_29634	Putative uncharacterized protein
Xenoturbella bocki	BCA	117195962	Uncharacterized	Uncharacterized protein

sequences, and inherent limitations of Bayesian inference, prompted use of additional phylogenetic methods. These analyses generally supported the major features of the final tree achieved via Bayesian inference.

Subcellular localization of β-CAs

The predictions for subcellular localization of the 75 β-CAs are shown in Table 2. The results reveal that 31 are predicted to have a mitochondrial localization, one (*Anopheles darlingi*, Uniprot ID: E3X5Q8) was predicted to be secreted, and the remaining 43 were predicted to have other cellular localizations. The predictions were based on the analysis of 175 N-terminal amino acids of each sequence. In the **Name** column, there are both IDs of the β-CAs in Uniprot database and scientific name of the metazoan and protozoan species.

Discussion

This study shows that the β-CA enzyme is present in a range of protozoans and metazoans. A total of 75 sequences were identified and a phylogenetic tree constructed. The multiple sequence alignment results revealed that all 75 sequences have the highly conserved residues (Cysteine, Aspartic acid, Arginine, and Histidine) consistent with a β-CA enzyme (Additional file 1: Figure S1). Most of the metazoan and protozoan β-CAs, and corresponding coding sequences, were designated as uncharacterized sequences or CAs with no class specification. These

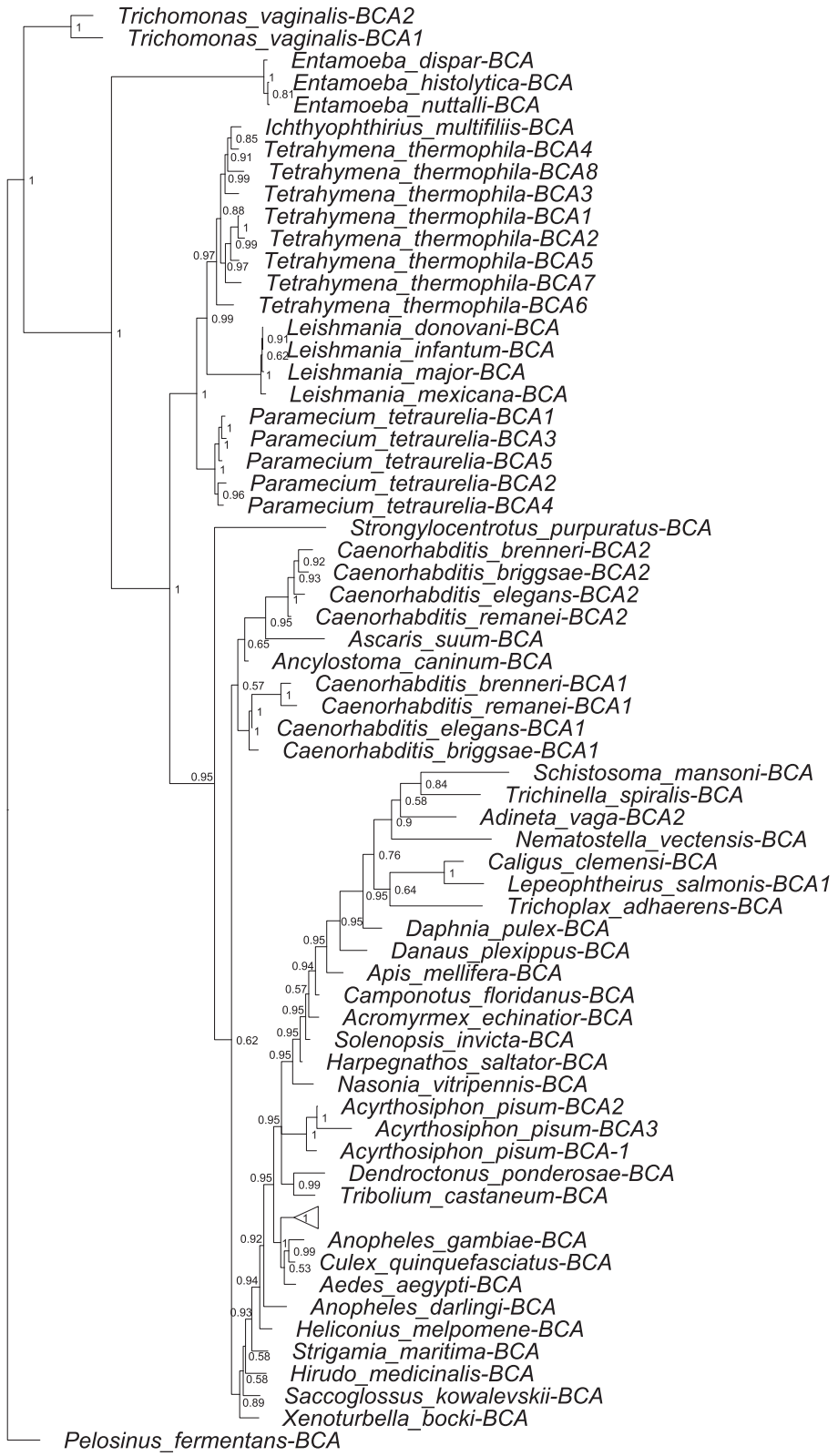


Figure 1 Phylogenetic analysis of 75 metazoan and protozoan β-CAs. The position of β-CAs of *Drosophila* species has been represented at the bottom of the phylogenetic tree by a triangle shape. The details of β-CAs of *Drosophila* species in the phylogenetic tree are shown in Figure 2.

can be now assigned to β-CAs in proteomics and genomics databases.

β-CAs have been identified in the mitochondria of a variety of different organisms, such as plants [61], green algae [62], fungi [1,63], and *Drosophila melanogaster* [4]. Our results of subcellular localization prediction (Table 2) suggested that 31 of the β-CAs are targeted to mitochondria. In mitochondrial targeting peptides (mTPs), Arginine, Alanine and Serine are over-represented, while negatively charged amino acid residues (Aspartic acid and Glutamic acid) are rare. Furthermore, mTPs are believed to form an amphiphilic α-helix, which is important for the import of the nascent protein into the mitochondrion [59]. The successful construction of the TargetP predictor demonstrates that protein sorting signals can be recognized with reasonable reliability from amino acid sequence data alone, thus, to some extent, mimicking the cellular recognition processes [59]. The prediction of the mitochondrial localization for many of the proteins studied is also supported by the previous experimental data, showing that recombinant DmBCA protein is indeed located in mitochondria of insect cells [4]. As mitochondrial proteins the β-CAs may contribute to key metabolic functions. Among the mammalian α-CAs, CA VA and CA VB are the only enzymes that have been exclusively located to mitochondria. Functional studies, summarized in [64], have indicated them in several metabolic processes, such as gluconeogenesis, urea synthesis, and fatty acid synthesis. It has been shown previously that the gluconeogenic enzyme, pyruvate carboxylase, is expressed in protozoan (*Toxoplasma gondii*) mitochondria [65]. This enzyme utilizes bicarbonate to convert pyruvate to oxaloacetate. Mitochondrial CA V is also involved in lipid synthesis through pyruvate carboxylation reaction [66]. Importantly, lipid metabolism is of crucial importance for parasites. Lipids serve as cellular building blocks, signaling

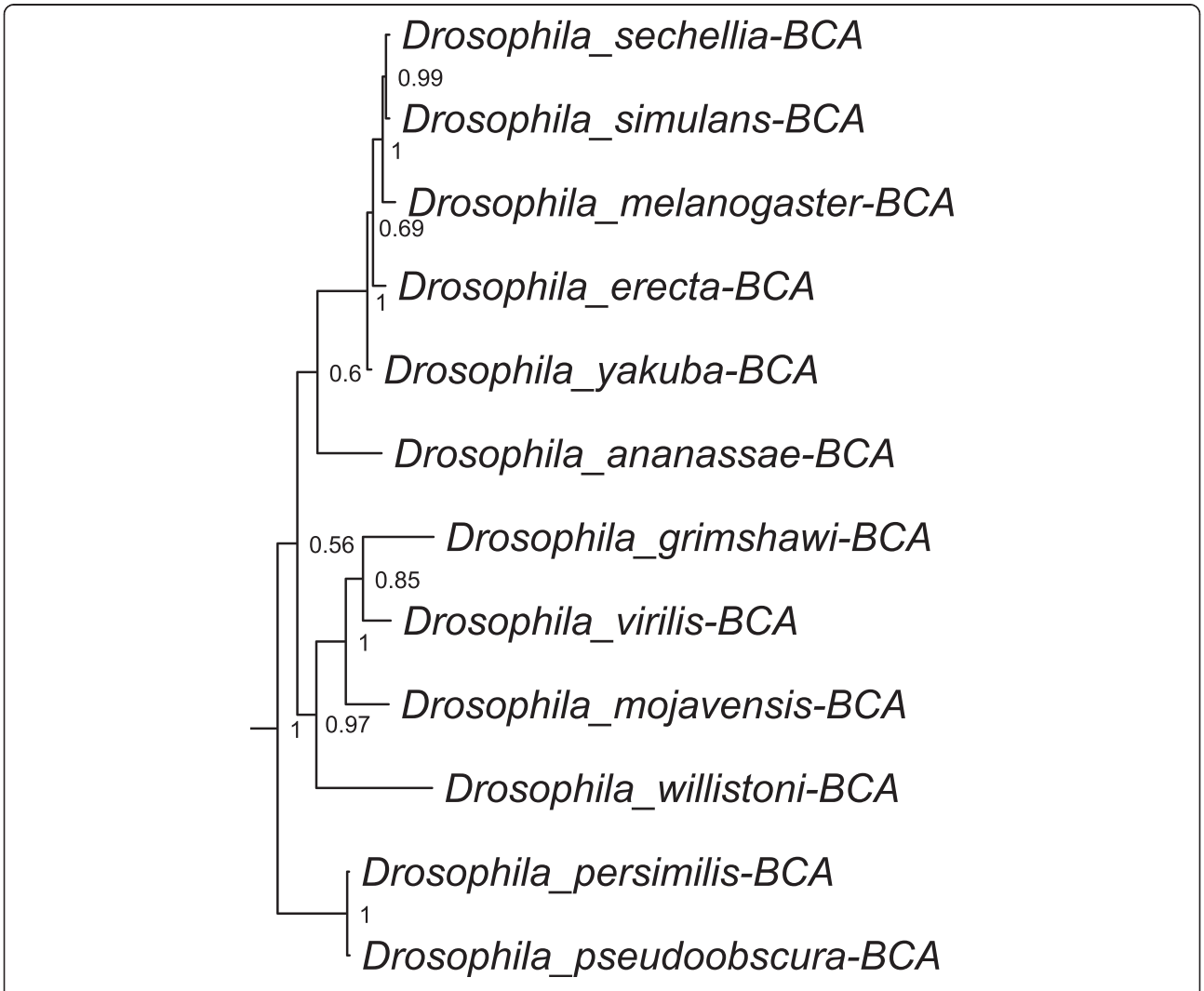


Figure 2 Phylogenetic analysis of β-CAs of *Drosophila* species. This tree represents the expanded view of the triangle located near the bottom of the main phylogenetic tree of β-CAs in Figure 1.

Table 2 Prediction of the subcellular localization of 75 β-CAs of metazoan and protozoan species

Species	β- CA ID	Entry ID	mTP	SP	Other	Loc	RC
Acromyrmex echinatior	BCA	F4WAG3	0.199	0.054	0.86	-	2
Acyrthosiphon pisum	BCA1	J9K706	0.473	0.05	0.631	-	5
	BCA2	C4WVD8	0.579	0.043	0.536	M	5
	BCA3	J9JZY3	0.579	0.043	0.534	M	5
Adineta vaga	BCA	D4NWE5	0.509	0.102	0.375	M	5
Aedes aegypti	BCA	Q17N64	0.589	0.029	0.491	M	5
Ancylostoma caninum	BCA	FC551456	0.466	0.046	0.514	-	5
Anopheles darlingi	BCA	E3X5Q8	0.044	0.836	0.144	S	2
Anopheles gambiae	BCA	Q5TU56	0.713	0.03	0.34	M	4
Apis mellifera	BCA	H9KS29	0.126	0.08	0.875	-	2
Ascaris suum	BCA	F1LE18	0.388	0.079	0.406	-	5
Caenorhabditis brenneri	BCA1	G0MSW4	0.522	0.036	0.518	M	5
	BCA2	G0MRG1	0.52	0.051	0.473	M	5
Caenorhabditis briggsae	BCA1	A8XKV0	0.392	0.047	0.615	-	4
	BCA2	A8WN21	0.546	0.048	0.466	M	5
Caenorhabditis elegans	BCA1	Q22460	0.475	0.039	0.549	-	5
	BCA2	Q2YS41	0.465	0.05	0.529	-	5
Caenorhabditis remanei	BCA1	E3LDN3	0.327	0.045	0.69	-	4
	BCA2	E3MK96	0.51	0.051	0.48	M	5
Caligus clemensi	BCA	C1C2M7	0.21	0.04	0.873	-	2
Camponotus floridanus	BCA	E2ANQ9	0.325	0.051	0.735	-	3
Culex quinquefasciatus	BCA	B0WKV7	0.573	0.032	0.507	M	5
Danaus plexippus	BCA	G6D7Z4	0.793	0.032	0.273	M	3
Daphnia pulex	BCA	E9GLB5	0.157	0.055	0.843	-	2
Dendroctonus ponderosae	BCA	J3JTM9	0.27	0.064	0.742	-	3
Drosophila ananassae	BCA	B3LZ10	0.537	0.041	0.518	M	5
Drosophila erecta	BCA	B3P1V8	0.531	0.04	0.53	M	5
Drosophila grimshawi	BCA	B4JHY1	0.605	0.037	0.454	M	5
Drosophila melanogaster	BCA	Q9VHJ5	0.531	0.04	0.53	M	5
Drosophila mojavensis	BCA	B4KDC1	0.556	0.039	0.511	M	5
Drosophila persimilis	BCA	B4GFA1	0.595	0.037	0.466	M	5
Drosophila pseudoobscura	BCA	Q296E4	0.595	0.037	0.466	M	5
Drosophila sechellia	BCA	B4HKY7	0.531	0.04	0.53	M	5
Drosophila simulans	BCA	B4QXC5	0.531	0.04	0.53	M	5
Drosophila virilis	BCA	B4LZE7	0.531	0.04	0.53	M	5
Drosophila willistoni	BCA	B4NBB9	0.531	0.04	0.53	M	5
Drosophila yakuba	BCA	B4PTY0	0.531	0.04	0.53	M	5
Entamoeba dispar	BCA	B0E7M0	0.114	0.158	0.766	-	2
Entamoeba histolytica	BCA	C4LXK3	0.113	0.151	0.779	-	2
Entamoeba nuttalli	BCA	K2GQM0	0.132	0.142	0.763	-	2
Harpegnathos saltator	BCA	E2B2Q1	0.248	0.055	0.801	-	3
Heliconius melpomene	BCA	HMEL015257	0.77	0.032	0.302	M	3
Hirudo medicinalis	BCA	EY481200	0.121	0.098	0.778	-	2
Ichthyophthirius multifiliis	BCA	G0QPN9	0.181	0.04	0.872	-	2

Table 2 Prediction of the subcellular localization of 75 β-CAs of metazoan and protozoan species *(Continued)*

Leishmania donovani	BCA	E9B8S3	0.106	0.13	0.826	-	2
Leishmania infantum	BCA	A4HSV2	0.106	0.13	0.826	-	2
Leishmania major	BCA	Q4QJ17	0.108	0.124	0.822	-	2
Leishmania mexicana	BCA	E9AKU0	0.109	0.135	0.82	-	2
Lepeophtheirus salmonis	BCA	D3PI48	0.126	0.068	0.889	-	2
Nasonia vitripennis	BCA	K7IWK8	0.388	0.046	0.713	-	4
Nematostella vectensis	BCA	A7S717	0.775	0.052	0.211	M	3
Paramecium tetraurelia	BCA1	A0BD61	0.196	0.045	0.843	-	2
	BCA2	A0E8J0	0.107	0.056	0.909	-	1
	BCA3	A0CEX6	0.28	0.045	0.725	-	3
	BCA4	A0BDB1	0.073	0.065	0.938	-	1
	BCA5	A0C922	0.178	0.056	0.826	-	2
Saccoglossus kowalevskii	BCA	187043763	0.565	0.049	0.463	M	5
Schistosoma mansoni	BCA	G4V6B2	0.388	0.064	0.605	-	4
Solenopsis invicta	BCA	E9IP13	0.326	0.052	0.756	-	3
Strigamia maritima	BCA	SMAR006741	0.683	0.046	0.28	M	3
Strongylocentrotus purpuratus	BCA	H3I177	0.804	0.047	0.16	M	2
Tetrahymena thermophila	BCA1	Q22U21	0.092	0.064	0.92	-	1
	BCA2	Q22U16	0.087	0.075	0.918	-	1
	BCA3	I7MDL7	0.659	0.067	0.203	M	3
	BCA4	I7LWM1	0.115	0.058	0.871	-	2
	BCA5	I7M0M0	0.087	0.034	0.947	-	1
	BCA6	I7MD92	0.058	0.069	0.941	-	1
	BCA7	I7M748	0.09	0.047	0.933	-	1
	BCA8	Q23AV1	0.187	0.123	0.758	-	3
Tribolium castaneum	BCA	D6WK56	0.054	0.097	0.938	-	1
Trichinella spiralis	BCA	E5SH53	0.876	0.028	0.177	M	2
Trichomonas vaginalis	BCA1	A2ENQ8	0.043	0.137	0.933	-	2
	BCA2	A2DLG4	0.073	0.061	0.937	-	1
Trichoplax adhaerens	BCA	B3S5Y1	0.582	0.038	0.459	M	5
Xenoturbella bocki	BCA	117195962	0.222	0.056	0.78	-	3

molecules, energy stores, posttranslational modifiers, and pathogenesis factors [67]. Parasites rely on complex metabolic systems to satisfy their lipid needs. The present findings open a new avenue to investigate whether mitochondrial β-CAs are functionally involved in these processes.

The single β-CA of *Anopheles darlingi* is the first predicted secretory β-CA. Among the various α-CAs, the first secreted form (CA VI) was identified in human saliva in 1987 [68], and in 2011 another α-CA was identified in the salivary gland of *Aedes aegypti* [69]. Complementary research, such as morphological, biochemical, and spatial mapping of gene expression in *Anopheles darlingi* will clarify the exact expression pattern of β-CA in this mosquito [69,70].

The TargetP predictor defined 43 β-CAs with 'other' cellular localizations. Although it is possible that β-CAs are truly located in different subcellular compartments depending on the species, these results should be interpreted with caution. Both the common errors in full genomic DNA, cDNA, or protein sequences in databases, and the potential inaccuracy of TargetP predictor could contribute to the observed deviations of the results. The highest prediction accuracy, with appropriate selection of specificity and sensitivity, is 90% [59].

Among the species mentioned in Table 1, some have important medical relevance, such as *Aedes aegypti, Anopheles darlingi, Anopheles gambiae, Ascaris suum (Ascaris lumbricoides), Culex quinquefasciatus, Entamoeba histolytica, Hirudo medicinalis, Leishmania* species,

Schistosoma mansoni, *Trichinella spiralis*, and *Trichomonas vaginalis*. In the past decade, inhibition profiles of β-CAs of bacteria [24,31,71] and fungi [72-75] have been investigated with various inhibitors. Our results suggest that various protozoans and metazoans express β-CAs and that these molecules represent protein targets appropriate for inhibitor development. These proteins are not restricted to nematodes, insects, or protozoa causing human diseases, but are also present in many species with relevance to agriculture or veterinary medicine. These species include: *Acyrthosiphon pisum*, *Ancylostoma caninum*, *Ascaris suum*, *Caligus clemensi*, *Camponotus floridanus*, *Culex quinquefasciatus*, *Dendroctonus ponderosae*, *Entamoeba* species, *Ichthyophthirius multifiliis*, *Solenopsis invicta*, *Tribolium castaneum*, *Trichinella spiralis*, and *Trichoplax adhaerens*. Therefore, our findings also suggest that it might be possible to develop specific β-CA inhibitors as pesticides for the protection of crops and other natural resources against pathogens and pests.

Conclusions

The present data identifies β-CA enzymes that are expressed in a number of protozoans and metazoans. Metazoan and protozoan β-CAs represent promising diagnostic and therapeutic targets for parasitic infections, because this CA family is absent from mammalian proteomes. Many of these enzymes are predicted to be present in mitochondria where they might contribute to cell metabolism by providing bicarbonate for biosynthetic reactions and regulating intra-mitochondrial pH.

Additonal file

Additional file 1: Figure S1. Multiple sequence alignment of all 75 β-CAs in metazoan and protozoan species with β-CA of *Pelosinus fermentans* (a bacterial out group). β-CAs contain two highly conserved active site motifs, CxDxR as well as HxxC (C=Cysteine, D=Aspartic acid, R=Arginine, H=Histidine, C=Cysteine) which are indicated by arrows. Alignment was visualized in Jalview [53].

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

RZE, HB, MEET, CO carried out the bioinformatics searches on metazoan and protozoan species. RZE and HB participated in the sequence alignment and made the phylogenetic analysis. RZE performed the mitochondrial targeting peptide prediction. All authors participated in the design of the study. RZE and HB drafted the first version of the manuscript. All authors read and approved the final manuscript.

Acknowledgement

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. This study was also funded by Finnish Cultural Foundation (HB), Academy of Finland (SP), Sigrid Juselius Foundation (SP), Jane and Aatos Erkko Foundation (SP), Tampere Tuberculosis Foundation (SP), and Competitive Research Funding of the Tampere University Hospital (SP; 9 N054).

Author details

[1]School of Medicine, University of Tampere, Medisiinarinkatu 3, 33520 Tampere, Finland. [2]Institute of Biomedical Technology and BioMediTech, University of Tampere, 33520 Tampere, Finland. [3]Department of Information Technology, University of Turku, 20520 Turku, Finland. [4]Fimlab Ltd and Tampere University Hospital, Biokatu 4, 33520 Tampere, Finland.

Received: 9 October 2013 Accepted: 10 January 2014
Published: 21 January 2014

References

1. Elleuche S, Poggeler S: **Carbonic anhydrases in fungi.** *Microbiology* 2010, **156**(Pt 1):23–29.
2. Huang S, Hainzl T, Grundstrom C, Forsman C, Samuelsson G, Sauer-Eriksson AE: **Structural studies of beta-carbonic anhydrase from the green alga Coccomyxa: inhibitor complexes with anions and acetazolamide.** *PLoS One* 2011, **6**(12):e28458.
3. Supuran CT: **Carbonic anhydrases: novel therapeutic applications for inhibitors and activators.** *Nat Rev Drug Discov* 2008, **7**(2):168–181.
4. Syrjanen L, Tolvanen M, Hilvo M, Olatubosun A, Innocenti A, Scozzafava A, Leppiniemi J, Niederhauser B, Hytonen VP, Gorr TA, *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.
5. Tripp BC, Smith K, Ferry JG: **Carbonic anhydrase: new insights for an ancient enzyme.** *J Biol Chem* 2001, **276**(52):48615–48618.
6. Elleuche S, Poggeler S: **Evolution of carbonic anhydrases in fungi.** *Curr Genet* 2009, **55**(2):211–222.
7. Esbaugh AJ, Tufts BL: **The structure and function of carbonic anhydrase isozymes in the respiratory system of vertebrates.** *Respir Physiol Neurobiol* 2006, **154**(1–2):185–198.
8. Lane TW, Saito MA, George GN, Pickering IJ, Prince RC, Morel FM: **Biochemistry: a cadmium enzyme from a marine diatom.** *Nature* 2005, **435**(7038):42.
9. Xu Y, Feng L, Jeffrey PD, Shi Y, Morel FM: **Structure and metal exchange in the cadmium carbonic anhydrase of marine diatoms.** *Nature* 2008, **452**(7183):56–61.
10. Ferry JG: **The gamma class of carbonic anhydrases.** *Biochim Biophys Acta* 2010, **1804**(2):374–381.
11. Alterio V, Vitale RM, Monti SM, Pedone C, Scozzafava A, Cecchi A, De Simone G, Supuran CT: **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):8329–8335.
12. 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 isoform VI, a new target for sulfonamide and sulfamate inhibitors.** *J Med Chem* 2007, **50**(2):381–388.
13. 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):1272–1279.
14. Vullo D, Innocenti A, Nishimori I, Pastorek J, Scozzafava A, Pastorekova S, Supuran CT: **Carbonic anhydrase inhibitors. Inhibition of the transmembrane isozyme XII with sulfonamides-a new target for the design of antitumor and antiglaucoma drugs?** *Bioorg Med Chem Lett* 2005, **15**(4):963–969.
15. Neish AC: **Studies on chloroplasts: Their chemical composition and the distribution of certain metabolites between the chloroplasts and the remainder of the leaf.** *Biochem J* 1939, **33**(3):300–308.
16. Burnell JN, Gibbs MJ, Mason JG: **Spinach chloroplastic carbonic anhydrase: nucleotide sequence analysis of cDNA.** *Plant Physiol* 1990, **92**(1):37–40.
17. Rowlett RS: **Structure and catalytic mechanism of the beta-carbonic anhydrases.** *Biochim Biophys Acta* 2010, **1804**(2):362–373.
18. Fett JP, Coleman JR: **Characterization and expression of two cDNAs encoding carbonic anhydrase in Arabidopsis thaliana.** *Plant Physiol* 1994, **105**(2):707–713.
19. Majeau N, Coleman JR: **Nucleotide sequence of a complementary DNA encoding tobacco chloroplastic carbonic anhydrase.** *Plant Physiol* 1992, **100**(2):1077–1078.
20. Cronk JD, Endrizzi JA, Cronk MR, O'Neill JW, Zhang KY: **Crystal structure of E. coli beta-carbonic anhydrase, an enzyme with an unusual pH-dependent activity.** *Protein Sci* 2001, **10**(5):911–922.

21. Guilloton MB, Korte JJ, Lamblin AF, Fuchs JA, Anderson PM: **Carbonic anhydrase in Escherichia coli. A product of the cyn operon.** *J Biol Chem* 1992, **267**(6):3731–3734.
22. Smith KS, Ferry JG: **A plant-type (beta-class) carbonic anhydrase in the thermophilic methanoarchaeon Methanobacterium thermoautotrophicum.** *J Bacteriol* 1999, **181**(20):6247–6253.
23. Rowlett RS, Tu C, Lee J, Herman AG, Chapnick DA, Shah SH, Gareiss PC: **Allosteric site variants of Haemophilus influenzae beta-carbonic anhydrase.** *Biochemistry* 2009, **48**(26):6146–6156.
24. Supuran CT: **Bacterial carbonic anhydrases as drug targets: toward novel antibiotics?** *Front Pharmacol* 2011, **2**:34.
25. Joseph P, Turtaut F, Ouahrani-Bettache S, Montero JL, Nishimori I, Minakuchi T, Vullo D, Scozzafava A, Kohler S, Winum JY, *et al*: **Cloning, characterization, and inhibition studies of a beta-carbonic anhydrase from Brucella suis.** *J Med Chem* 2010, **53**(5):2277–2285.
26. Burghout P, Vullo D, Scozzafava A, Hermans PW, Supuran CT: **Inhibition of the beta-carbonic anhydrase from Streptococcus pneumoniae by inorganic anions and small molecules: Toward innovative drug design of antiinfectives?** *Bioorg Med Chem* 2011, **19**(1):243–248.
27. Vullo D, Nishimori I, Minakuchi T, Scozzafava A, Supuran CT: **Inhibition studies with anions and small molecules of two novel beta-carbonic anhydrases from the bacterial pathogen Salmonella enterica serovar Typhimurium.** *Bioorg Med Chem Lett* 2011, **21**(12):3591–3595.
28. Abuaita BH, Withey JH: **Bicarbonate Induces Vibrio cholerae virulence gene expression by enhancing ToxT activity.** *Infect Immun* 2009, **77**(9):4111–4120.
29. Kovacikova G, Lin W, Skorupski K: **The LysR-type virulence activator AphB regulates the expression of genes in Vibrio cholerae in response to low pH and anaerobiosis.** *J Bacteriol* 2010, **192**(16):4181–4191.
30. Klengel T, Liang WJ, Chaloupka J, Ruoff C, Schroppel K, Naglik JR, Eckert SE, Mogensen EG, Haynes K, Tuite MF, *et al*: **Fungal adenylyl cyclase integrates CO2 sensing with cAMP signaling and virulence.** *Curr Biol* 2005, **15**(22):2021–2026.
31. Innocenti A, Leewattanapasuk W, Muhlschlegel FA, Mastrolorenzo A, Supuran CT: **Carbonic anhydrase inhibitors. Inhibition of the beta-class enzyme from the pathogenic yeast Candida glabrata with anions.** *Bioorg Med Chem Lett* 2009, **19**(16):4802–4805.
32. Bahn YS, Cox GM, Perfect JR, Heitman J: **Carbonic anhydrase and CO2 sensing during Cryptococcus neoformans growth, differentiation, and virulence.** *Curr Biol* 2005, **15**(22):2013–2020.
33. Elleuche S, Poggeler S: **A cyanase is transcriptionally regulated by arginine and involved in cyanate decomposition in Sordaria macrospora.** *Fungal Genet Biol* 2008, **45**(11):1458–1469.
34. Amoroso G, Morell-Avrahov L, Muller D, Klug K, Sultemeyer D: **The gene NCE103 (YNL036w) from Saccharomyces cerevisiae encodes a functional carbonic anhydrase and its transcription is regulated by the concentration of inorganic carbon in the medium.** *Mol Microbiol* 2005, **56**(2):549–558.
35. Cleves AE, Cooper DN, Barondes SH, Kelly RB: **A new pathway for protein export in Saccharomyces cerevisiae.** *J Cell Biol* 1996, **133**(5):1017–1026.
36. Gotz R, Gnann A, Zimmermann FK: **Deletion of the carbonic anhydrase-like gene NCE103 of the yeast Saccharomyces cerevisiae causes an oxygen-sensitive growth defect.** *Yeast* 1999, **15**(10A):855–864.
37. So AK, Espie GS: **Cloning, characterization and expression of carbonic anhydrase from the cyanobacterium Synechocystis PCC6803.** *Plant Mol Biol* 1998, **37**(2):205–215.
38. Sawaya MR, Cannon GC, Heinhorst S, Tanaka S, Williams EB, Yeates TO, Kerfeld CA: **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):7546–7555.
39. Eriksson M, Karlsson J, Ramazanov Z, Gardestrom P, Samuelsson G: **Discovery of an algal mitochondrial carbonic anhydrase: molecular cloning and characterization of a low-CO2-induced polypeptide in Chlamydomonas reinhardtii.** *Proc Natl Acad Sci U S A* 1996, **93**(21):12031–12034.
40. Mitsuhashi S, Mizushima T, Yamashita E, Yamamoto M, Kumasaka T, Moriyama H, Ueki T, Miyachi S, Tsukihara T: **X-ray structure of beta-carbonic anhydrase from the red alga, Porphyridium purpureum, reveals a novel catalytic site for CO(2) hydration.** *J Biol Chem* 2000, **275**(8):5521–5526.
41. Fasseas MK, Tsikou D, Flemetakis E, Katinakis P: **Molecular and biochemical analysis of the beta class carbonic anhydrases in Caenorhabditis elegans.** *Mol Biol Rep* 2010, **37**(6):2941–2950.
42. Guilloton MB, Lamblin AF, Kozliak EI, Gerami-Nejad M, Tu C, Silverman D, Anderson PM, Fuchs JA: **A physiological role for cyanate-induced carbonic anhydrase in Escherichia coli.** *J Bacteriol* 1993, **175**(5):1443–1451.
43. Nishimori I, Onishi S, Takeuchi H, Supuran CT: **The alpha and beta classes carbonic anhydrases from Helicobacter pylori as novel drug targets.** *Curr Pharm Des* 2008, **14**(7):622–630.
44. Mitsuhashi S, Ohnishi J, Hayashi M, Ikeda M: **A gene homologous to beta-type carbonic anhydrase is essential for the growth of Corynebacterium glutamicum under atmospheric conditions.** *Appl Microbiol Biotechnol* 2004, **63**(5):592–601.
45. Fukuzawa H, Suzuki E, Komukai Y, Miyachi S: **A gene homologous to chloroplast carbonic anhydrase (icfA) is essential to photosynthetic carbon dioxide fixation by Synechococcus PCC7942.** *Proc Natl Acad Sci U S A* 1992, **89**(10):4437–4441.
46. Merlin C, Masters M, McAteer S, Coulson A: **Why is carbonic anhydrase essential to Escherichia coli?** *J Bacteriol* 2003, **185**(21):6415–6424.
47. Majeau N, Coleman JR: **Effect of CO2 Concentration on Carbonic Anhydrase and Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Expression in Pea.** *Plant Physiol* 1996, **112**(2):569–574.
48. Tetu SG, Tanz SK, Vella N, Burnell JN, Ludwig M: **The Flaveria bidentis beta-carbonic anhydrase gene family encodes cytosolic and chloroplastic isoforms demonstrating distinct organ-specific expression patterns.** *Plant Physiol* 2007, **144**(3):1316–1327.
49. Ludwig M: **The molecular evolution of beta-carbonic anhydrase in Flaveria.** *J Exp Bot* 2011, **62**(9):3071–3081.
50. Zabaleta E, Martin MV, Braun HP: **A basal carbon concentrating mechanism in plants?** *Plant Sci: Int J Exp Plant Biol* 2012, **187:**97–104.
51. Kimber MS, Pai EF: **The active site architecture of Pisum sativum beta-carbonic anhydrase is a mirror image of that of alpha-carbonic anhydrases.** *EMBO J* 2000, **19**(7):1407–1418.
52. Mitsuhashi S, Mizushima T, Yamashita E, Miyachi S, Tsukihara T: **Crystallization and preliminary X-ray diffraction studies of a beta-carbonic anhydrase from the red alga Porphyridium purpureum.** *Acta Crystallogr D Biol Crystallogr* 2000, **56**(Pt 2):210–211.
53. 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):1189–1191.
54. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, *et al*: **Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.** *Mol Syst Biol* 2011, **7:**539.
55. Suyama M, Torrents D, Bork P: **PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments.** *Nucleic Acids Res* 2006, **34:**W609–W612. eb Server issue.
56. Ronquist F, Teslenko M, Van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP: **MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space.** *Syst Biol* 2012, **61**(3):539–542.
57. Brown SD, Podar M, Klingeman DM, Johnson CM, Yang ZK, Utturkar SM, Land ML, Mosher JJ, Hurt RA Jr, Phelps TJ, *et al*: **Draft genome sequences for two metal-reducing Pelosinus fermentans strains isolated from a Cr (VI)-contaminated site and for type strain R7.** *J Bacteriol* 2012, **194**(18):5147–5148.
58. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O: **New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.** *Systematic biology* 2010, **59**(3):307–321.
59. Emanuelsson O, Nielsen H, Brunak S, von Heijne G: **Predicting subcellular localization of proteins based on their N-terminal amino acid sequence.** *J Mol Biol* 2000, **300**(4):1005–1016.
60. Woese CR: **Interpreting the universal phylogenetic tree.** *Proc Natl Acad Sci U S A* 2000, **97**(15):8392–8396.
61. Fabre N, Reiter IM, Becuwe-Linka N, Genty B, Rumeau D: **Characterization and expression analysis of genes encoding alpha and beta carbonic anhydrases in Arabidopsis.** *Plant Cell Environ* 2007, **30**(5):617–629.
62. Mitra M, Lato SM, Ynalvez RA, Xiao Y, Moroney JV: **Identification of a new chloroplast carbonic anhydrase in Chlamydomonas reinhardtii.** *Plant Physiol* 2004, **135**(1):173–182.
63. Bahn YS, Muhlschlegel FA: **CO2 sensing in fungi and beyond.** *Curr Opin Microbiol* 2006, **9**(6):572–578.

64. Shah GN, Rubbelke TS, Hendin J, Nguyen H, Waheed A, Shoemaker JD, Sly WS: **Targeted mutagenesis of mitochondrial carbonic anhydrases VA and VB implicates both enzymes in ammonia detoxification and glucose metabolism.** *Proc Natl Acad Sci U S A* 2013, **110**(18):7423–7428.
65. Fleige T, Pfaff N, Gross U, Bohne W: **Localisation of gluconeogenesis and tricarboxylic acid (TCA)-cycle enzymes and first functional analysis of the TCA cycle in Toxoplasma gondii.** *Int J Parasit* 2008, **38**(10):1121–1132.
66. Hazen SA, Waheed A, Sly WS, LaNoue KF, Lynch CJ: **Differentiation-dependent expression of CA V and the role of carbonic anhydrase isozymes in pyruvate carboxylation in adipocytes.** *FASEB J* 1996, **10**(4):481–490.
67. Ramakrishnan S, Serricchio M, Striepen B, Butikofer P: **Lipid synthesis in protozoan parasites: A comparison between kinetoplastids and apicomplexans.** *Prog Lipid Res* 2013, **52**(4):488–512.
68. Murakami H, Sly WS: **Purification and characterization of human salivary carbonic anhydrase.** *J Biol Chem* 1987, **262**(3):1382–1388.
69. Juhn J, Naeem-Ullah U, Maciel Guedes BA, Majid A, Coleman J, Paolucci Pimenta PF, Akram W, James AA, Marinotti O: **Spatial mapping of gene expression in the salivary glands of the dengue vector mosquito, Aedes aegypti.** *Parasit Vectors* 2011, **4**:1.
70. Moreira-Ferro CK, Marinotti O, Bijovsky AT: **Morphological and biochemical analyses of the salivary glands of the malaria vector, Anopheles darlingi.** *Tissue Cell* 1999, **31**(3):264–273.
71. Maresca A, Scozzafava A, Kohler S, Winum JY, Supuran CT: **Inhibition of beta-carbonic anhydrases from the bacterial pathogen Brucella suis with inorganic anions.** *J Inorg Biochem* 2012, **110**:36–39.
72. Isik S, Guler OO, Kockar F, Aydin M, Arslan O, Supuran CT: **Saccharomyces cerevisiae beta-carbonic anhydrase: inhibition and activation studies.** *Curr Pharm Des* 2010, **16**(29):3327–3336.
73. Isik S, Kockar F, Aydin M, Arslan O, Guler OO, Innocenti A, Scozzafava A, Supuran CT: **Carbonic anhydrase inhibitors: inhibition of the beta-class enzyme from the yeast Saccharomyces cerevisiae with sulfonamides and sulfamates.** *Bioorg Med Chem* 2009, **17**(3):1158–1163.
74. Monti SM, Maresca A, Viparelli F, Carta F, De Simone G, Muhlschlegel FA, Scozzafava A, Supuran CT: **Dithiocarbamates are strong inhibitors of the beta-class fungal carbonic anhydrases from Cryptococcus neoformans, Candida albicans and Candida glabrata.** *Bioorg Med Chem Lett* 2012, **22**(2):859–862.
75. Guzel O, Maresca A, Hall RA, Scozzafava A, Mastrolorenzo A, Muhlschlegel FA, Supuran CT: **Carbonic anhydrase inhibitors. The beta-carbonic anhydrases from the fungal pathogens Cryptococcus neoformans and Candida albicans are strongly inhibited by substituted-phenyl-1H-indole-5-sulfonamides.** *Bioorg Med Chem Lett* 2010, **20**(8):2508–2511.

doi:10.1186/1756-3305-7-38
Cite this article as: Zolfaghari Emameh *et al.*: Bioinformatic analysis of beta carbonic anhydrase sequences from protozoans and metazoans. *Parasites & Vectors* 2014 **7**:38.

Parasites & Vectors

RESEARCH Open Access

Beta carbonic anhydrases: novel targets for pesticides and anti-parasitic agents in agriculture and livestock husbandry

Reza Zolfaghari Emameh[1,2,3*], Harlan Barker[1,2], Vesa P Hytönen[2,3], Martti E E Tolvanen[4,2] and Seppo Parkkila[1,2,3]

Abstract

Background: The genomes of many insect and parasite species contain beta carbonic anhydrase (β-CA) protein coding sequences. The lack of β-CA proteins in mammals makes them interesting target proteins for inhibition in treatment of some infectious diseases and pests. Many insects and parasites represent important pests for agriculture and cause enormous economic damage worldwide. Meanwhile, pollution of the environment by old pesticides, emergence of strains resistant to them, and their off-target effects are major challenges for agriculture and society.

Methods: In this study, we analyzed a multiple sequence alignment of 31 β-CAs from insects, some parasites, and selected plant species relevant to agriculture and livestock husbandry. Using bioinformatics tools a phylogenetic tree was generated and the subcellular localizations and antigenic sites of each protein were predicted. Structural models for β-CAs of *Ancylostoma caninum*, *Ascaris suum*, *Trichinella spiralis*, and *Entamoeba histolytica*, were built using *Pisum sativum* and *Mycobacterium tuberculosis* β-CAs as templates.

Results: Six β-CAs of insects and parasites and six β-CAs of plants are predicted to be mitochondrial and chloroplastic, respectively, and thus may be involved in important metabolic functions. All 31 sequences showed the presence of the highly conserved β-CA active site sequence motifs, CXDXR and HXXC (C: cysteine, D: aspartic acid, R: arginine, H: histidine, X: any residue). We discovered that these two motifs are more antigenic than others. Homology models suggested that these motifs are mostly buried and thus not well accessible for recognition by antibodies.

Conclusions: The predicted mitochondrial localization of several β-CAs and hidden antigenic epitopes within the protein molecule, suggest that they may not be considered major targets for vaccines. Instead, they are promising candidate enzymes for small-molecule inhibitors which can easily penetrate the cell membrane. Based on current knowledge, we conclude that β-CAs are potential targets for development of small molecule pesticides or anti-parasitic agents with minimal side effects on vertebrates.

Keywords: Beta carbonic anhydrase, Inhibitors, Insecticides, Pesticides, Anti-parasitic agents, Agriculture, Livestock husbandry

Background

Various pests, including weeds, insects, and plant viruses, often reduce crop production by 25-50% [1,2]. On the other hand, the widespread use of synthetic insecticides for controlling pests produces many negative consequences (e.g. insecticide resistance, toxicity to mammals and other non-target animals, residue problems, and environmental pollution). High risk groups exposed to pesticides include: production workers, formulators, sprayers, mixers, loaders, and agricultural farm workers. Residual pesticides represent a real threat for human health. When 9700 samples of fruits and vegetables were analyzed for seven pesticides (Acephate, Chlopyriphos, Chlopyriphos-methyl, Methamidophos, Iprodione, Procymidone, and Chlorothalonil), 5.2% of the samples were found to contain residues. Pesticides can contaminate soil, water, and turf. In addition to killing insects or weeds, pesticides can be toxic to other organisms including birds,

* Correspondence: reza.zolfaghari.emameh@uta.fi
[1]School of Medicine, University of Tampere, 33520 Tampere, Finland
[2]BioMediTech, University of Tampere, 33520 Tampere, Finland
Full list of author information is available at the end of the article

fish, beneficial insects, and non-target plants [3]. The extensive use of pesticides, such as Dichlorodiphenyltrichloroethane (DDT), in recent decades has led to their recurrent detection in many surface and ground waters [4]. As a result of these negative consequences, natural products have become popular among consumers [5].

As of the 1960s pesticide resistance had already evolved in some key greenhouse pests, prompting the development of alternative methods of management. The pressure to reduce insecticide usage was reinforced by the adoption of bumble-bees for pollination within greenhouses [6]. Biological control plays a central role in the production of many greenhouse crops. The term "Biopesticide" encompasses a broad array of microbial pesticides, including biochemicals derived from micro-organisms and other natural sources, and those resulting from the incorporation of DNA into various agricultural commodities [7]. Bacteria, fungi, viruses, entomopathogenic nematodes (ENPs), and herbal essential oils are often used as biopesticides [8]. Novel approaches to control pests involve targeting of specific insect and parasite enzymes. This can be done using either chemical or biological compounds. Acetylcholinesterase (AChE) of the malaria mosquito (*Anopheles gambiae*) has been reported as a target site for pesticides [9]. Three pesticides, Atrazine, DDT, and Chlorpyrifos, have been determined to affect the esterase (GE), glutathione S-transferase (GST), cytochrome P450 monooxygenase (P450), and acetylcholinesterase (AChE) activities of *Chironomus tentans* (an aquatic midge) [4]. Proteinases serving as insect digestive enzymes are defined targets in pest control [10]. Enzyme inhibitors, such as: piperonyl butoxide (PB), a mixed-function oxidase (MFO) inhibitor; triphenyl phosphate (TPP), a carboxyesterase (CarE) inhibitor; and diethyl maleate (DEM), a glutathione S-transferase (GST) inhibitor, have been used to inhibit insect enzymes [11]. Inhibition of *Plasmodium falciparum* carbonic anhydrase (CA) with aromatic heterocyclic sulfonamides was investigated in 2011 [12]. In another study, a thiabendazole sulfonamide showed a potent inhibitory activity against both mammalian and nematode α-CAs [13].

Five independently evolved classes of CAs (α, β, γ, δ, and ζ) have been identified, of which one or more are found in nearly every cell type, underscoring the general importance of this ubiquitous enzyme in nature [14]. The CAs are involved in several important biological processes, such as respiration and transportation of CO_2 and bicarbonate between metabolizing tissues, pH and CO_2 homeostasis, electrolyte secretion in different organs, bone resorption, calcification, tumorigenicity, and some biosynthetic reactions including gluconeogenesis, lipogenesis, and ureagenesis [15]. Since 1990, many demonstrated and putative β-CAs have been discovered not only in photosynthetic organisms, but also in eubacteria, yeast, archaeal species [16] and 18 metazoan species [17]. Recently, we reported 52 β-CAs in metazoan and protozoan species [18]. At least one study has shown the effects of β-CA inhibitors as anti-infective agents on different bacterial and fungal pathogens [19], yet this approach has not been tested *in vivo* in metazoans or protozoans. In this article, we introduce β-CAs as novel potential target enzymes to control agricultural and veterinary insects and parasites which cause enormous economic losses worldwide.

Methods

Identification of putative β-CA enzymes and multiple sequence alignment (MSA)

In total, 23 parasite and 8 plant β-CA sequences relevant to agriculture and livestock husbandry, or as model organisms, and one bacterial sequence (*Desulfosporosinus meridiei*) were retrieved from UniProt (http://www.uniprot.org/) and NCBI (http://www.ncbi.nlm.nih.gov/). The full list of agriculture and livestock husbandry pests and plants containing β-CA addressed in this research are shown in Table 1. We focused on 98 amino acid residues around the catalytic active site of all tested β-CAs, starting 7 amino acid residues prior to the first highly conserved sequence (CXDXR). The Clustal Omega algorithm [20] within the Jalview program (version 2.8.ob1) (http://www.jalview.org/) was used to create a multiple sequence alignment (MSA) [21].

Phylogenetic analysis

All sequences were individually analyzed for completeness and quality. The β-CA sequence for *Solenopsis invicta* (UniProt ID: E9IP13) was determined to have a spurious exon when the genomic sequence was analyzed by the Exonerate program using the other β-CA proteins as query sequences, and subsequently 17 amino acids were removed [49]. Similarly, the full genome of *Acyrthosiphon pisum* was analyzed. Of the three *Acyrthosiphon pisum* β-CA sequences identified in UniProt, two were incomplete (UniProt IDs: C4WVD8 and J9JZY3) and found to be fragments of the same complete protein predicted in our analysis (*Acyrthosiphon pisum* BCA-2). Finally, the full genome of *Ichthyophthirius multifiliis* was scanned for β-CA proteins using the same method, and two new putative β-CA proteins were identified (*Ichthyophthirius multifiliis* BCA-3 and BCA-4).

A protein sequence alignment was created using Clustal Omega [20] based on which the corresponding nucleotide sequences were then codon-aligned by the Pal2Nal program [50]. Using the *Desulfosporosinus meridiei* bacterial sequence as an outgroup, a phylogenetic analysis was computed using Mr. Bayes v3.2 [51] with the GTR model of codon substitution and all other parameters set to default. In total, 200,000 generations were computed with a final standard deviation of split frequencies of 3.33×10^{-4}.

Table 1 Agriculture and livestock husbandry pests, and plants containing β-CA which applied in this research

Species name	General name	Parasitic Features	Main concerns
Acyrthosiphon pisum	pea aphid	Sap-sucking in forage crops, such as peas, clover, alfalfa, and broad beans	Food canning industry [22]
Ancylostoma caninum	A species of phylum *Nematoda*	Infection of the small intestine of dogs and human (zoonosis)	Dog breeding [23]
***Ascaris suum* (*Ascaris lumbricoides*)**	large roundworm of pigs	Ascariasis in pig and human (zoonosis)	Pig breeding [24]
Caligus clemensi	Plural sea lice	Major ectoparasites of farmed and wild Atlantic salmon	Fishing and fish farming [25]
Camponotus floridanus	Carpenter ant	Nest in live or dead trees, rotting logs and stumps, buildings, telephone poles, and other wooden structures	Wooden instrument industries and consumers [26]
Ceratitis capitata	Mediterranean fruit fly (Medfly)	Causing extensive damage to a wide range of fruit crops	Invasion to orchards [27]
Culex quinquefasciatus	Southern house mosquito	Vector of West Nile virus (WNV), St. Louis encephalitis virus and other arboviruses, lymphatic filariasis, *Wuchereria bancrofti*, and *Plasmodium relictum* (avian malaria)	Zoonotic diseases which affect both humans and animals health [28]
Dendroctonus ponderosae	Mountain pine beetle (MPB)	Attacks to old or weakened trees, and speeds to younger forests	Wooden instrument industries and consumers [29]
Entamoeba histolytica Entamoeba nuttalli Entamoeba dispar	A genus of phylum *Amoebozoa*	Causative agent of amoebiasis in animals and human (zoonosis)	Humans and animals health [30]
Haemonchus contortus	Trichostrongyloid nematode (Red stomach worm, wire worm or barber's pole worm)	Causative agent of Haemonchosis by blood feeding through attachment to abomasal mucosa of ruminants	Sheep and goat farming [31]
Ichthyophthirius multifiliis	Freshwater ich, or freshwater ick	White spot disease in freshwater fishes and rarely in human (zoonosis)	Fish and fish farming [32]
Lepeophtheirus salmonis	Salmon louse	Parasite living on wild salmon and fish farming	Fish and fish farming [25]
Necator americanus	New World hookworm	Necatoriasis in dog, cat, and human (zoonosis)	Humans and animals health [33]
Solenopsis invicta	Red imported fire ant (RIFA)	Mound-building activity, Damage plant roots which leads to loss of crops, and interfere with mechanical cultivation	Wooden instrument industries and consumers, and gardening [34]
Tribolium castaneum	Red flour beetle	Pest of stored grain products, carcinogenic by secretion of quinones, causative agent of occupational IgE-mediated allergy and some other diseases	Wheat, flour, cereal and nut based food industries [35-38]
Trichinella spiralis	Pork worm	Trichinosis in rat, pig, bear and human (zoonosis)	Pig breeding [39]
Trichoplax adhaerens	Adherent hairy plate	Adherence to the wall of a marine aquariums	Aquarium and ornamental fishing industry [40]
Arabidopsis thaliana	Mouse-ear cress	-	A popular model organism in plant biology and genetics [41]
Pisum sativum	Pea	-	Pea is most commonly the small spherical seed or the seed-pod [42]
Gossypium hirsutum	Upland cotton	-	Upland cotton is the most widely planted species of cotton [43]
Nicotiana tabacum	Tobacco	-	Its leaves are commercially processed into tobacco [44]
Vitis vinifera	Grape vine	-	Commercial significance for wine and table grape production [45]
Solanum tuberosum	Potato	-	The world's fourth-largest food crop, following maize, wheat and rice [46]
Populus trichocarpa	Black cottonwood or California poplar	-	A model organism in plant biology [47]
Capsella rubella	A genus from Mustard family	-	A member of Mustard family [48]

The final phylogenetic tree was visualized in FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

Prediction of subcellular localization

Subcellular localization of each identified invertebrate β-CA was predicted using the TargetP webserver (http://www.cbs.dtu.dk/services/TargetP/). TargetP is built from two layers of neural networks, where the first layer contains one dedicated network for each type of targeting sequences, such as cytoplasmic, mitochondrial, or secretory peptides, and the second layer is an integrating network that outputs the actual prediction (cTP = cytoplasmic, mTP = mitochondrial, SP = secretory, or other). It is able to discriminate between cTPs, mTPs, and SPs with sensitivities and specificities higher than what has been obtained with other available subcellular localization predictors [52].

Prediction of antigenic sites in β-CA

The protein sequences of 23 parasite and 8 plant β-CAs were analyzed with the European Molecular Biology Open Software Suite (EMBOSS) program Antigenic (http://emboss.bioinformatics.nl/cgi-bin/emboss/antigenic). EMBOSS Antigenic predicts potentially antigenic regions of a protein sequence, using the method of Kolaskar and Tongaonkar [53]. Application of this method to a large number of proteins has shown that their accuracy is better than most of the known methods [54-56].

Homology modelling

Homology models of four selected β-CAs, including FC551456 (*Ancylostoma caninum*), F1LE18 (*Ascaris suum*), E5SH53 (*Trichinella spiralis*), and C4LXK3 (*Entamoeba histolytica*) were prepared by first selecting the most suitable template structure. For this purpose, a BLAST search of the PDB database (http://www.rcsb.org/pdb/home/home.do) was performed using each of the four sequences. Results for three out of these four searches revealed that PDB structure 1EKJ (β-CA from *Pisum sativum*) possessed the most similar sequence, while PDB id 2A5V (β-CA from *Mycobacterium tuberculosis*) was found to be the most similar to C4LXK3 (*Entamoeba histolytica*). Clustal Omega was used to prepare a multiple sequence alignment for those six sequences.

The multiple sequence alignment showed nine completely conserved residues within the sequences; the known highly conserved CXDXR and HXXC motifs were among them (data not shown). Homology modelling was performed according to multiple sequence alignment containing FC551456 (*Ancylostoma caninum*), F1LE18 (*Ascaris suum*), E5SH53 (*Trichinella spiralis*), and PDB 1EKJ by using the Modeller program (version 9.13) [57] with PDB model 1EKJ (β-CA from *Pisum sativum*) as a template. A homology model for C4LXK3 (*Entamoeba histolytica*) was prepared using PDB 2A5V for pairwise alignment and as a template structure. The resulting models were structurally aligned using the BODIL program [58]. A figure illustrating the homology models was prepared by using the VMD program (version 1.9.1) [59], and edited within Adobe Photoshop (version 13.0.1).

The structural availability of the epitope in the PDB model 1EKJ (β-CA from *Pisum sativum*) and the homology model based on the β-CA sequence from *Ancylostoma caninum* was studied by preparing the molecular surface with VMD, using a probe radius of 1.4 Å. The potential epitope residues were excluded from the surface presentation and were shown as Van der Waals (VdW) spheres.

Results

Multiple sequence alignment (MSA)

The MSA of 23 parasite and 8 plant β-CA sequences revealed the presence of the highly characteristic conserved sequence motifs CXDXR and HXXC (C: cysteine, D: aspartic acid, R: arginine, H: histidine, X: any residue) in all sequences. These results verify the presence of the β-CA enzyme in several insects and parasites which are pathogenic to various species of plants and animals and are thus considered relevant to agriculture and livestock husbandry (Figure 1).

Phylogenetic analysis

The results of the phylogenetic analysis of DNA sequences encoding 23 parasite and 8 plant β-CAs are shown in Figure 2. From the resulting tree we see four distinct clades, three of which represent distinct potential β-CA targets. From the top, the first clade represents β-CAs of invertebrate pests, the second clade are plant model organisms, the third clade is entirely represented by the four β-CAs of *Ichthyophthirius multifiliis*, and the final clade represents three species of amoeba. The *Entamoeba spp*. sequences occupy a midpoint between our outgroup bacteria species and the others.

Prediction of subcellular localization

The results of subcellular localization prediction of β-CAs in selected parasite and plant species are shown in Table 2. The predictions were based on the analysis of full-length β-CA protein sequences. In the *Name* column, there are both the UniProt ID and species scientific name. The results reveal that 6 of 23 β-CAs from parasites (XP_004537221.1, B0WKV7, U6PDI1, E5SH53, B3S5Y1, and predicted BCA2 in *A. pisum*) were predicted to have a mitochondrial localization signal; 6 of 8 β-CAs of plants (P17067, Q8LSC8, P27141, D7TWP2, I2FJZ8, and B9GHR1) were predicted to have a chloroplastic localization.

Prediction of antigenic sites in β-CA

According to the acceptable 3–85 residue variation in epitope length of an antigen [60] and default parameters

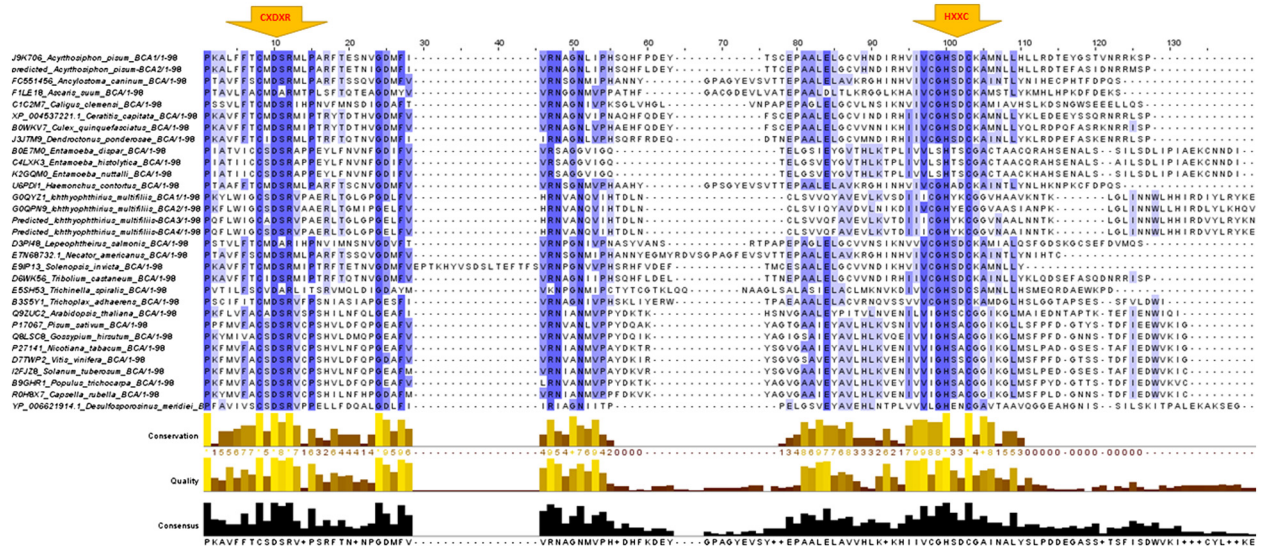


Figure 1 Multiple sequence alignment of 23 parasite, 8 plant, and one bacterial (*Desulfosporosinus meridiei*) β-CA sequences showing the most conserved region of the active site. The first (CXDXR) and second (HXXC) highly conserved sequences which are involved in zinc atom binding in catalytic active sites of the enzyme are marked with arrows at the top of the figure.

of EMBOSS Antigenic database, the minimum length of an antigenic region in this set of β-CAs is 6 amino acid residues. The predictions of antigenic sites in the 31 β-CA proteins are shown in Table 3; the highest score belongs to the most antigenic site.

Homology modelling

Homology models of four selected β-CAs verified the predicted localization of conserved residues in the active site. Two loop regions showed high variability in the sequence length which is apparent in the Figure 3C, D and indicated by "*" and "**". In addition, homology modelling suggested insertion located within the longest α-helix in case of homology models based on 1EKJ (Figure 3C, indicated by "***").

To study the molecular availability of the predicted main antigenic epitope, surface exposure of the homology model created from PDB model 1EKJ (β-CA from *Pisum sativum*) and the homology model based on the β-CA sequence from *Ancylostoma caninum* were studied

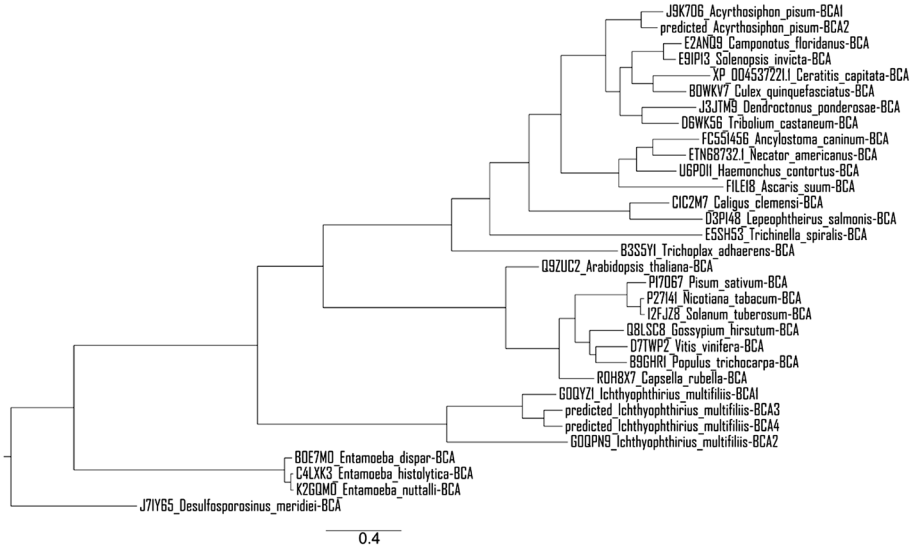


Figure 2 Phylogenetic analysis of β-CAs from 23 parasite and 8 plant species. β-CA from *Desulfosporosinus meridiei* was used as a bacterial outgroup.

Table 2 Prediction of subcellular localization of 23 pest and 8 plant β-CAs

Species name	Entry ID	β-CA ID	cTP	mTP	SP	Other	RC	Loc
Acyrthosiphon pisum	J9K706	BCA1	-	0.473	0.050	0.631	5	-
Acyrthosiphon pisum	Predicted	BCA2	-	0.579	0.043	0.536	5	M
Ancylostoma caninum	FC551456	BCA	-	0.466	0.046	0.514	5	-
Ascaris suum	F1LE18	BCA	-	0.388	0.079	0.406	5	-
Caligus clemensi	C1C2M7	BCA	-	0.210	0.040	0.873	2	-
Camponotus floridanus	E2ANQ9	BCA	-	0.325	0.051	0.735	3	-
Ceratitis capitata	XP_004537221.1	BCA	-	0.549	0.039	0.512	5	M
Culex quinquefasciatus	B0WKV7	BCA	-	0.573	0.032	0.507	5	M
Dendroctonus ponderosae	J3JTM9	BCA	-	0.270	0.064	0.742	3	-
Entamoeba dispar	B0E7M0	BCA	-	0.114	0.158	0.766	2	-
Entamoeba histolytica	C4LXK3	BCA	-	0.113	0.151	0.779	2	-
Entamoeba nuttalli	K2GQM0	BCA	-	0.132	0.142	0.763	2	-
Haemonchus contortus	U6PDI1	BCA	-	0.587	0.057	0.403	5	M
Ichthyophthirius multifiliis	G0QYZ1	BCA1	-	0.071	0.046	0.946	1	-
Ichthyophthirius multifiliis	G0QPN9	BCA2	-	0.181	0.040	0.872	2	-
Ichthyophthirius multifiliis	Predicted	BCA3	-	0.059	0.078	0.954	1	-
Ichthyophthirius multifiliis	Predicted	BCA4	-	0.050	0.178	0.868	2	-
Lepeophtheirus salmonis	D3PI48	BCA	-	0.126	0.068	0.889	2	-
Necator americanus	ETN68732.1	BCA	-	0.379	0.036	0.604	4	-
Solenopsis invicta	E9IP13	BCA	-	0.326	0.052	0.756	3	-
Tribolium castaneum	D6WK56	BCA	-	0.054	0.097	0.938	1	-
Trichinella spiralis	E5SH53	BCA	-	0.876	0.028	0.177	2	M
Trichoplax adhaerens	B3S5Y1	BCA	-	0.582	0.038	0.459	5	M
Arabidopsis thaliana	Q9ZUC2	BCA	0.043	0.171	0.108	0.923	2	-
Pisum sativum	P17067	BCA	0.969	0.050	0.014	0.023	1	C
Gossypium hirsutum	Q8LSC8	BCA	0.947	0.154	0.008	0.019	2	C
Nicotiana tabacum	P27141	BCA	0.956	0.059	0.019	0.039	1	C
Vitis vinifera	D7TWP2	BCA	0.902	0.183	0.016	0.034	2	C
Solanum tuberosum	I2FJZ8	BCA	0.954	0.051	0.024	0.045	1	C
Populus trichocarpa	B9GHR1	BCA	0.931	0.231	0.021	0.012	2	C
Capsella rubella	R0H8X7	BCA	0.040	0.208	0.176	0.907	2	-

cTP = a chloroplast transit peptide, *mTP* = a mitochondrial targeting peptide, *SP* = secretory pathway, *Loc* (predicted localization) where *C* = chloroplastic, *M* = mitochondrial, *S* = secretory, *−* = other, *RC* = reliability class, from 1 to 5, where 1 indicates the strongest prediction. *RC* is a measure of the difference between the highest and the second highest output scores. There are 5 reliability classes, defined as follows: 1: diff ≥ 0.800, 2: 0.800 > diff ≥ 0.600, 3: 0.600 > diff ≥ 0.400, 4: 0.400 > diff ≥ 0.200 and 5: 0.200 > diff. Thus, the lower the value of RC the safer the prediction.

by visualizing the molecular surface (Figure 4). The analysis revealed that the majority of the epitope was buried within the structure. The residues considered to be mainly buried in the structure are shown in green, while solvent-exposed residues are shown with red colour. Two residues in PDB model 1EKJ (β-CA from *Pisum sativum*) appear considerably smaller than their complements in the *Ancylostoma caninum*-based homology model, and those residues can be considered to be only partially exposed (Figure 4, indicated by yellow colour in the alignment). Taken together, these results indicate that the predicted epitope sequence is mainly buried in β-CA sequences.

Discussion

Several insect, parasite, and plant genomes contain genes which encode β-CA enzymes. Some of these parasites and insects are either causative agents or vectors of important veterinary, fish farming, and zoonotic diseases (Table 1). For this analysis we selected 31 β-CAs, 23 from parasites and 8 from plants. These sequences were retrieved from protein databases, or predicted from their genomes,

Table 3 Predicted antigenic sites of 23 pest and 8 plant β-CA primary sequences

Species name	Entry ID	β-CA ID	Pest or plant	HitCount*	The most antigenic epitope
Acyrthosiphon pisum	J9K706	BCA1	Pest	14	77 YTSCEPAALELGCVHNDIRHVIVCG***HSDC*** 105
Acyrthosiphon pisum	Predicted	BCA2	Pest	14	79 TCEPAALELGCVHNDIRHVIVCG***HSDC*** 105
Ancylostoma caninum	FC551456	BCA	Pest	11	101 INHVIVCG***HSDC***KAINTLYNIHECPHTFDP 130
Ascaris suum	F1LE18	BCA	Pest	15	102 KHAIVCG***HSDC***KAMST 117
Caligus clemensi	C1C2M7	BCA	Pest	10	84 EPAGLELGCVLNSIKNVIVCG***HSDC***KAMIAVHSL 117
Camponotus floridanus	E2ANQ9	BCA	Pest	11	80 CESAALELGCVVNDIRHVIVCG***HSDC*** 105
Ceratitis capitata	XP_004537221.1	BCA	Pest	13	72 HFQDEYFSCEPAALELGCVINDIRHIIVCGHSD 104
Culex quinquefasciatus	B0WKV7	BCA	Pest	14	75 DEYFSCEPAALELGCVVNNIKHIIVCG***HSDC*** 105
Dendroctonus ponderosae	J3JTM9	BCA	Pest	13	95 RHIIVCG***HSDC***KAINLLYKL 114
Entamoeba dispar	B0E7M0	BCA	Pest	8	85 SIEYGVTHLKTPLIVVLS***HTSC***GACTAACQRA 116
Entamoeba histolytica	C4LXK3	BCA	Pest	8	83 LGSVEYGVTHLKTPLIVVLS***HTSC***GACTAACQRA 116
Entamoeba nuttalli	K2GQM0	BCA	Pest	7	83 LGSVEYGVTHLKTPLIVVLS***HTSC***GACTAACKHA 116
Haemonchus contortus	U6PDI1	BCA	Pest	13	101 HINHVIVCGHADCKAINTLYNL 122
Ichthyophthirius multifiliis	G0QYZ1	BCA1	Pest	13	193 ANQVIHTDLNCLSVVQYAVEVLKVSDIIICG***HYKC***GGVHAAVKNT 237
Ichthyophthirius multifiliis	G0QPN9	BCA2	Pest	9	86 ANQVIHTDLNCLSVIQYAVDVLNIKDIIVCG***HYEC***GGVAASIANPKLGL 134
Ichthyophthirius multifiliis	Predicted	BCA3	Pest	7	65 ANQVIHTDLNCLSVVQFAVEVLKVTDIIICG***HYKC***GGVNAA 105
Ichthyophthirius multifiliis	Predicted	BCA4	Pest	6	62 ANQVIHTDLNCLSVVQFAVEVLKVTDIIICG***HYKC***GGVNA 101
Lepeophtheirus salmonis	D3PI48	BCA	Pest	10	82 PEPAGLELGCVVNSIKNVVVCG***HSDC***KAMIALQSF 116
Necator americanus	ETN68732.1	BCA	Pest	10	108 HINHVIVCGHSDCKAINTLYNIHTCPQ 134
Solenopsis invicta	E9IP13	BCA	Pest	14	97 CESAALELGCVVNDIKHVIVCG***HSDC*** 122
Tribolium castaneum	D6WK56	BCA	Pest	13	116 ALELGCVVNDIRHIIVCG***HSDC***KAINLLYKLQDS 149
Trichinella spiralis	E5SH53	BCA	Pest	11	100 KDIVVCG***HSDC*** 110
Trichoplax adhaerens	B3S5Y1	BCA	Pest	13	82 EAAALELACVRNQVSSVVVCG***HSDC*** 106
Arabidopsis thaliana	Q9ZUC2	BCA	Plant	13	80 PKFLVFA***CADSR***VSPSHILNFQ 101
Pisum sativum	P17067	BCA	Plant	16	153 PFMVFA***CSDSR***VCPSHVLDFQ 173
Gossypium hirsutum	Q8LSC8	BCA	Plant	14	151 KYMIVA***CSDSR***VCPSHVLDM 170
Nicotiana tabacum	P27141	BCA	Plant	15	146 KFMVFA***CSDSR***VCPSHVLNF 165
Vitis vinifera	D7TWP2	BCA	Plant	13	149 KFMVFA***CSDSR***VCPSHVLDFQ 169
Solanum tuberosum	I2FJZ8	BCA	Plant	15	146 KFMVFA***CSDSR***VCPSHVLNF 165
Populus trichocarpa	B9GHR1	BCA	Plant	13	146 KFMVFA***CSDSR***VCPSHVLDFQ 166
Capsella rubella	R0H8X7	BCA	Plant	11	84 KYMVFA***CSDSR***VCPSHILNFH 104

The italic and bolded residues represent the first (CXDXR) and second (HXXC) highly conserved sequences in the catalytic active sites of the enzyme whenever present in the predicted epitope.
*:HitCount means the total number of antigenic residues in the whole sequence of one protein or antigen.

and all selected host or vector species are considered important for agriculture or livestock husbandry, or represent model organisms. The most significant species included *Ancylostoma caninum*, *Ascaris suum* (*Ascaris lumbricoides*), *Caligus clemensi*, *Culex quinquefasciatus*, *Entamoeba spp*, *Haemonchus contortus*, *Ichthyophthirius multifiliis*, *Lepeophtheirus salmonis*, *Necator americanus*, *Trichinella spiralis*, and *Trichoplax adhaerens*. One was an important pest in food industries (*Tribolium castaneum*). There was also an orchard invasive dipteran fruit fly (*Ceratitis capitata*) and three pests of wood industries, such as *Camponotus floridanus*, *Dendroctonus ponderosae*, and *Solenopsis invicta*.

Our MSA of β-CAs in plants, parasites, and insects showed that they all contain the first (CXDXR) and second (HXXC) highly conserved sequences of β-CA. The presence of β-CA proteins in various insects and parasites and their absence in mammals suggests that these enzymes could be potential targets for the development of novel pesticides or anti-parasitic drugs with minimal side effects on vertebrates. A key requirement for such novel β-CA inhibitors is the high isoform specificity. The

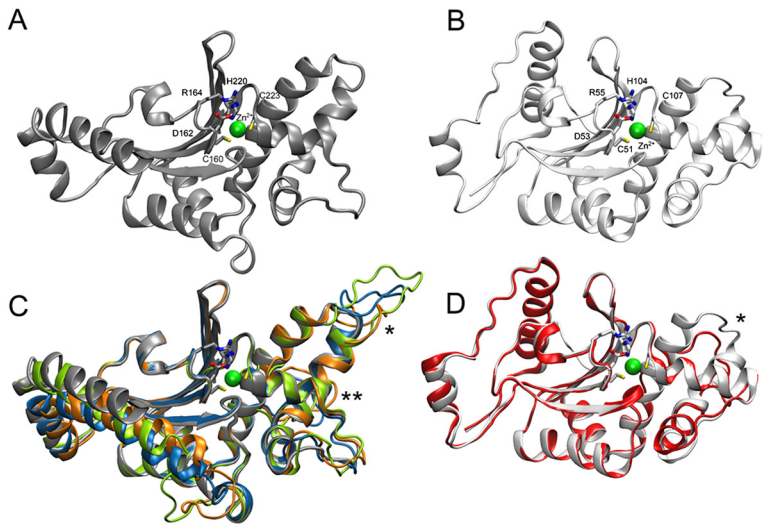


Figure 3 Homology modelling. (A) Cartoon presentation of β-CA from *Pisum sativum* (PDB 1EKJ). The Zn^{2+}ion is shown as green sphere and the residues in direct contact with the ion are shown as liquorice models and labeled with residue numbers. **(B)** β-CA from *Mycobacterium tuberculosis* (PDB 2A5V, light gray). **(C)** Aligned homology models of β-CAs from *Ancylostoma caninum* (green), *Ascaris suum* (blue), and *Trichinella spiralis* (orange) are shown with PDB 1EKJ (gray). **(D)** Homology model of and *Entamoeba histolytica* (red) structurally aligned with PDB 2A5V (light gray). Highly variable loop regions are indicated by stars ("*" and "**") **(C, D)**. Insertion suggested by homology models of *Ancylostoma caninum*, *Ascaris suum* and *Trichinella spiralis* is indicated by three stars **(C)**. The figure was prepared by using VMD (version 1.9.1).

distinction among β-CA proteins elucidated in the phylogenetic tree indicates that inhibitors can be created which would target β-CAs specific to different groups of species, leaving those in other species, such as plants, unaffected. Unfortunately, design of highly specific inhibitors will require proper structural data based on protein crystallography. Thus far, β-CA crystal structures from only a few different species are available in PDB database (http://www.rcsb.org/pdb/home/home.do), including some algae, bacteria, archaea, yeast, and a plant *Pisum sativum* [61].

Our prediction results on the subcellular localization of β-CAs showed that 6 of 23 β-CAs from parasites (XP_004537221.1, B0WKV7, U6PDI1, E5SH53, B3S5Y1, and predicted BCA2 in *A. pisum*) are probably mitochondrial enzymes. It is well known that several pesticides have unwanted side effects because of their off-target impacts on mitochondria [62]. Blocking of β-CAs in insect and parasitic cells can affect mitochondrial metabolic cycles and possibly eradicate the pathogens. Figure 5 presents 14 categories of known α- and/or β-CA inhibitors, which

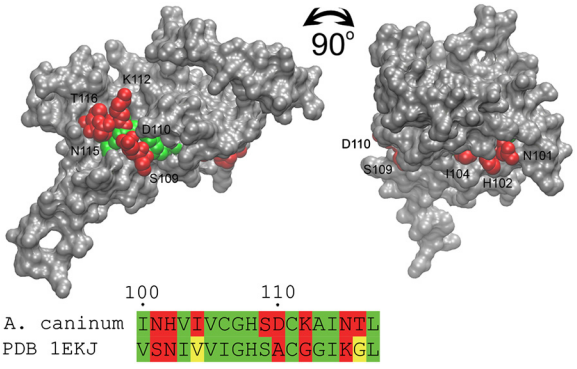


Figure 4 Determination of the availability of the predicted epitope. The molecular surface of the homology model of β-CA from *Ancylostoma caninum* is shown as solid grey and the target epitope sequence was excluded from the surface presentation. The epitope residues exposed to solvent are shown as red VdW spheres and numbered, while buried residues are shown with green spheres. An alignment containing PDB 1EKJ and the corresponding sequence from *Ancylostoma caninum* predicted β-CA is shown. The numbering of the residues in the alignment is according to the *Ancylostoma caninum* sequence. The yellow residues in the alignment indicate partially buried structure.

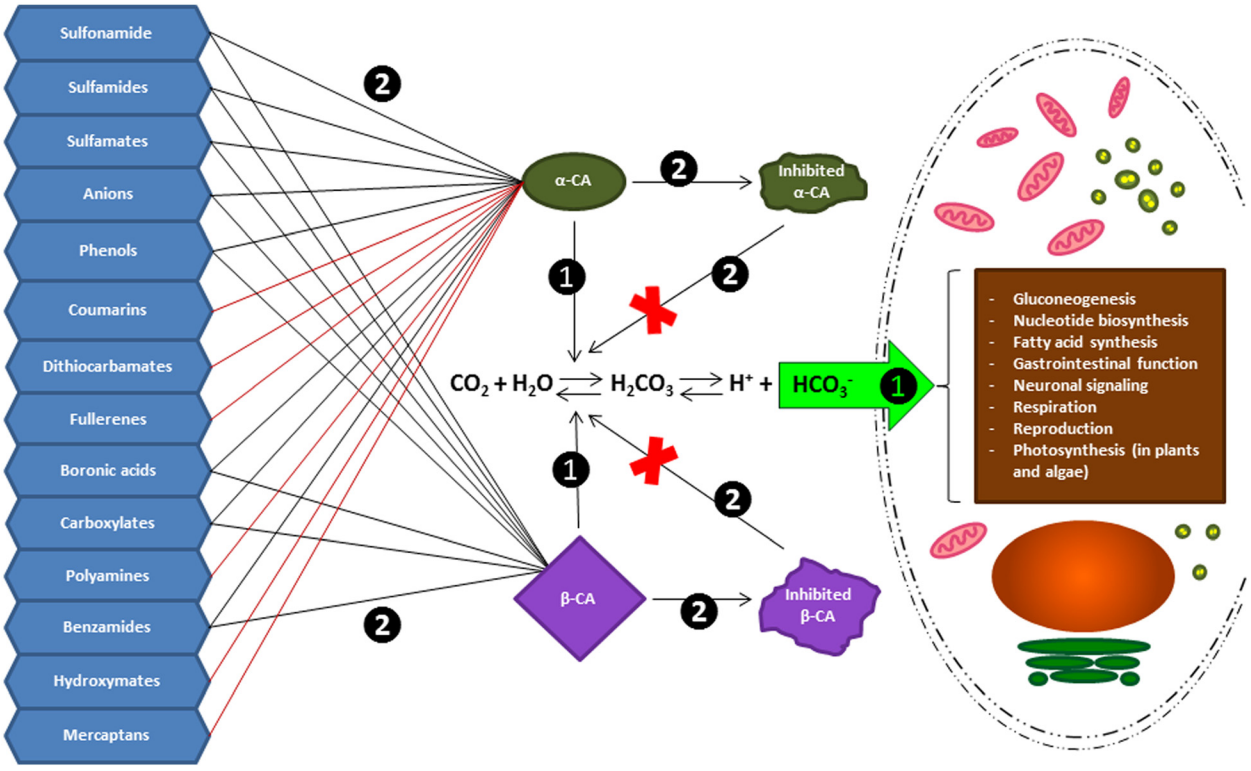


Figure 5 Effects of 14 CA inhibitors on α- and β-CAs of parasites and insects. Some compounds inhibit members of both α- and β- CA enzyme families. The brown box shows physiological processes where bicarbonate plays a role as a biochemical substrate. The ultimate goal of future research should be the creation of inhibitors specific to both enzyme families and to each isozyme. Ideally, the specific inhibitors would cause tissue- and organ-specific effects in parasites and vectors with minimal off-target effects on other species. Number 1 shows the catalytic pathway of α- and β-CA and number 2 shows the inhibitory effects of α- and β-CA inhibitors.

are able to inhibit catalytic activity of these enzyme families [63,64]. As the result, inhibition of CA activity would slow down some cellular biochemical pathways in parasites and insects, such as gluconeogenesis, nucleotide biosynthesis, fatty acid synthesis, gastrointestinal function, neuronal signaling, respiration, and reproduction. In plants and algae, it is known that β-CAs are required for CO_2 sequestration within chloroplast, and therefore CA inhibition would affect the rate of photosynthesis [65]. Importantly, β-CA inhibition in fungi and *Drosophila melanogaster* revealed completely different inhibition profiles [17], suggesting that β-CAs of parasites and insects can be inhibited with higher affinity than plant CAs by applying the right inhibitors and concentrations.

Another important goal is to find inhibitors that are specific for β-CAs and do not affect α-CAs at all. This would first require detailed structural data on selected parasite and insect CAs. The resolved structures would then allow high throughput screening of chemical compounds, identification of the most promising inhibitor molecules, and testing of potential compounds *in vitro* and *in vivo*.

Vaccination would offer another option to develop antiparasitic treatments based on β-CAs. In our study we used computational antigen prediction tools, which have been developed to reduce the laboratory work required to identify important antigenic epitopes in pathogenic proteins [66]. The Protegen database (http://www.violinet.org/protegen/) has been used to identify a number of predicted antigens from bacteria, viruses, parasites and fungi, which are involved in immune responses against various infectious and non-infectious diseases [67]. Antigenic site prediction of β-CA of parasites and plants revealed that the first and second highly conserved sequences (CXDXR and HXXC) represent the most plausible antigenic sites of β-CAs. Because these epitopes are located in the region of the active site and are mainly buried (Figure 4), they show very limited promise as vaccine targets. Furthermore, most β-CAs are intracellular proteins which are not readily accessible for immunological recognition. Taking all of these results together, small molecule inhibitors should still be considered the first option when β-CAs are investigated as therapeutic target proteins.

Conclusions

Our present work is the first study that discusses the potential role of β-CAs as target proteins for pesticides and anti-parasitic agents in agriculture and livestock husbandry. Our results could potentially have significant impacts on development of novel pesticides, which would directly benefit both food and forest industries. This is important as pests cause significant costs for agricultural, horticultural, and livestock husbandry products due to production losses [68]. Since β-CA sequences are not present in the genomes of vertebrates, the possible off-target effects in human and vertebrate animals should be minimal if high isozyme specificity is achieved. Discovery and validation of a new generation of β-CA inhibitors as pesticides and anti-parasitic agents would be a novel research field for chemical and pharmaceutical industries to improve safe nutrition and general health in societies.

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 parasites and plant species. RZE and HB participated in the sequence alignment. HB made protein sequence corrections and predictions and performed the phylogenetic analysis. RZE performed the mitochondrial targeting peptide and antigenic site prediction. RZE and VH participated in the homology modelling. RZE, HB and VH drafted the first version of the manuscript. All authors participated in writing further versions and read and approved the final manuscript.

Acknowledgment

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. This study was also funded by Finnish Cultural Foundation (HB), Sigrid Juselius Foundation (SP), Jane and Aatos Erkko Foundation (SP), Academy of Finland (VH) and Competitive Research Funding of the Tampere University Hospital (SP, VH).

Author details

[1]School of Medicine, University of Tampere, 33520 Tampere, Finland. [2]BioMediTech, University of Tampere, 33520 Tampere, Finland. [3]Fimlab Laboratories Ltd and Tampere University Hospital, Biokatu 4, 33520 Tampere, Finland. [4]Department of Information Technology, University of Turku, 20014 Turku, Finland.

Received: 21 May 2014 Accepted: 20 August 2014
Published: 29 August 2014

References

1. Zalucki MP, Shabbir A, Silva R, Adamson D, Shu-Sheng L, Furlong MJ: **Estimating the economic cost of one of the world's major insect pests, Plutella xylostella (Lepidoptera: Plutellidae): just how long is a piece of string?** *J Econ Entomol* 2012, **105**(4):1115–1129.
2. Way MJ: **Crop losses by insects and the problem of control.** *Proc Nutr Soc* 1961, **20:**5–11.
3. Aktar MW, Sengupta D, Chowdhury A: **Impact of pesticides use in agriculture: their benefits and hazards.** *Interdiscip Toxicol* 2009, **2**(1):1–12.
4. Rakotondravelo ML, Anderson TD, Charlton RE, Zhu KY: **Sublethal effects of three pesticides on activities of selected target and detoxification enzymes in the aquatic midge, Chironomus tentans (diptera: chironomidae).** *Archiv Environ Contam Toxicol* 2006, **51**(3):360–366.
5. Ayvaz A, Sagdic O, Karaborklu S, Ozturk I: **Insecticidal activity of the essential oils from different plants against three stored-product insects.** *J Insect Sci* 2010, **10:**21.
6. Chandler D, Bailey AS, Tatchell GM, Davidson G, Greaves J, Grant WP: **The development, regulation and use of biopesticides for integrated pest management.** *Philos Trans R Soc Lond B Biol Sci* 2011, **366**(1573):1987–1998.
7. Sudakin DL: **Biopesticides.** *Toxicol Rev* 2003, **22**(2):83–90.
8. Glare T, Caradus J, Gelernter W, Jackson T, Keyhani N, Kohl J, Marrone P, Morin L, Stewart A: **Have biopesticides come of age?** *Trends Biotechnol* 2012, **30**(5):250–258.
9. Pang YP: **Novel acetylcholinesterase target site for malaria mosquito control.** *PLoS One* 2006, **1:**e58.
10. Haq SK, Atif SM, Khan RH: **Protein proteinase inhibitor genes in combat against insects, pests, and pathogens: natural and engineered phytoprotection.** *Arch Biochem Biophys* 2004, **431**(1):145–159.
11. Wu G, Miyata T, Kang CY, Xie LH: **Insecticide toxicity and synergism by enzyme inhibitors in 18 species of pest insect and natural enemies in crucifer vegetable crops.** *Pest Manag Sci* 2007, **63**(5):500–510.
12. Krungkrai SR, Krungkrai J: **Malaria parasite carbonic anhydrase: inhibition of aromatic/heterocyclic sulfonamides and its therapeutic potential.** *Asian Pac J Trop Biomed* 2011, **1**(3):233–242.
13. Crocetti L, Maresca A, Temperini C, Hall RA, Scozzafava A, Muhlschlegel FA, Supuran CT: **A thiabendazole sulfonamide shows potent inhibitory activity against mammalian and nematode alpha-carbonic anhydrases.** *Bioorg Med Chem Lett* 2009, **19**(5):1371–1375.
14. Kumar RS, Ferry JG: **Prokaryotic carbonic anhydrases of Earth's environment.** *Subcell Biochem* 2014, **75:**77–87.
15. Alterio V, Vitale RM, Monti SM, Pedone C, Scozzafava A, Cecchi A, De Simone G, Supuran CT: **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):8329–8335.
16. Rowlett RS: **Structure and catalytic mechanism of the beta-carbonic anhydrases.** *Biochim Biophys Acta* 2010, **1804**(2):362–373.
17. Syrjanen L, Tolvanen M, Hilvo M, Olatubosun A, Innocenti A, Scozzafava A, Leppiniemi J, Niederhauser B, Hytonen VP, Gorr TA, Parkkila S, Supuran CT: **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.
18. 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**(1):38.
19. Tobal JM, Balieiro ME: **Role of carbonic anhydrases in pathogenic micro-organisms: a focus on Aspergillus fumigatus.** *J Med Microbiol* 2014, **63**(Pt 1):15–27.
20. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG: **Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.** *Mol Syst Biol* 2011, **7:**539.
21. 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):1189–1191.
22. Stavrinides J, McCloskey JK, Ochman H: **Pea aphid as both host and vector for the phytopathogenic bacterium Pseudomonas syringae.** *Appl Environ Microbiol* 2009, **75**(7):2230–2235.
23. Prociv P, Croese J: **Human enteric infection with Ancylostoma caninum: hookworms reappraised in the light of a "new" zoonosis.** *Acta Trop* 1996, **62**(1):23–44.
24. Leles D, Gardner SL, Reinhard K, Iniguez A, Araujo A: **Are Ascaris lumbricoides and Ascaris suum a single species?** *Parasit Vectors* 2012, **5:**42.
25. Jones SR, Prosperi-Porta G, Kim E, Callow P, Hargreaves NB: **The occurrence of Lepeophtheirus salmonis and Caligus clemensi (Copepoda: Caligidae) on three-spine stickleback Gasterosteus aculeatus in coastal British Columbia.** *J Parasitol* 2006, **92**(3):473–480.
26. Kafil M, Bandani AR, Kaltenpoth M, Goldansaz SH, Alavi SM: **Role of Symbiotic Bacteria in the Growth and Development of the Sunn Pest, Eurygaster integriceps.** *J Insect Sci* 2013, **13:**99.
27. Tabilio MR, Fiorini D, Marcantoni E, Materazzi S, Delfini M, De Salvador FR, Musmeci S: **Impact of the Mediterranean fruit fly (medfly) Ceratitis capitata on different peach cultivars: the possible role of peach volatile compounds.** *Food Chem* 2013, **140**(1–2):375–381.
28. Linthicum KJ: **Introduction to the symposium Global Perspective on the Culex pipiens Complex in the 21st century: The Interrelationship of**

Culex pipiens, quinquefasciatus, molestus and others. *J Am Mosq Control Assoc* 2012, **28**(4 Suppl):4–9.
29. Knapp PA, Soule PT, Maxwell JT: **Mountain pine beetle selectivity in old-growth ponderosa pine forests, Montana, USA.** *Ecol Evol* 2013, **3**(5):1141–1148.
30. Feng M, Yang B, Yang L, Fu Y, Zhuang Y, Liang L, Xu Q, Cheng X, Tachibana H: **High prevalence of Entamoeba infections in captive long-tailed macaques in China.** *Parasitol Res* 2011, **109**(4):1093–1097.
31. Peter JW, Chandrawathani P: **Haemonchus contortus: parasite problem No. 1 from tropics - Polar Circle. Problems and prospects for control based on epidemiology.** *Trop Biomed* 2005, **22**(2):131–137.
32. Buchmann K, Sigh J, Nielsen CV, Dalgaard M: **Host responses against the fish parasitizing ciliate Ichthyophthirius multifiliis.** *Vet Parasitol* 2001, **100**(1–2):105–116.
33. Schneider B, Jariwala AR, Periago MV, Gazzinelli MF, Bose SN, Hotez PJ, Diemert DJ, Bethony JM: **A history of hookworm vaccine development.** *Hum Vaccin* 2011, **7**(11):1234–1244.
34. Sun Q, Zhou X: **Corpse management in social insects.** *Int J Biol Sci* 2013, **9**(3):313–321.
35. Toews MD, Campbell JF, Arthur FH: **The presence of flour affects the efficacy of aerosolized insecticides used to treat the red flour beetle, Tribolium castaneum.** *J Insect Sci* 2010, **10:**196.
36. Lis LB, Bakula T, Baranowski M, Czarnewicz A: **The carcinogenic effects of benzoquinones produced by the flour beetle.** *Pol J Vet Sci* 2011, **14**(1):159–164.
37. El-Mofty MM, Osman SI, Sakr SA, Toulan BA: **Carcinogenicity of flour infested with Tribolium castaneum using the Egyptian toad Bufo regularis as a biological test animal.** *Oncology* 1988, **45**(1):65–67.
38. Alanko K, Tuomi T, Vanhanen M, Pajari-Backas M, Kanerva L, Havu K, Saarinen K, Bruynzeel DP: **Occupational IgE-mediated allergy to Tribolium confusum (confused flour beetle).** *Allergy* 2000, **55**(9):879–882.
39. Gottstein B, Pozio E, Nockler K: **Epidemiology, diagnosis, treatment, and control of trichinellosis.** *Clin Microbiol Rev* 2009, **22**(1):127–145. Table of Contents.
40. Schierwater B: **My favorite animal, Trichoplax adhaerens.** *Bioessays* 2005, **27**(12):1294–1302.
41. Theologis A, Ecker JR, Palm CJ, Federspiel NA, Kaul S, White O, Alonso J, Altafi H, Araujo R, Bowman CL, Brooks SY, Buehler E, Chan A, Chao Q, Chen H, Cheuk RF, Chin CW, Chung MK, Conn L, Conway AB, Conway AR, Creasy TH, Dewar K, Dunn P, Etgu P, Feldblyum TV, Feng J, Fong B, Fujii CY, Gill JE, *et al*: **Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana.** *Nature* 2000, **408**(6814):816–820.
42. Majeau N, Coleman JR: **Isolation and characterization of a cDNA coding for pea chloroplastic carbonic anhydrase.** *Plant Physiol* 1991, **95**(1):264–268.
43. Hoang CV, Wessler HG, Local A, Turley RB, Benjamin RC, Chapman KD: **Identification and expression of cotton (Gossypium hirsutum L.) plastidial carbonic anhydrase.** *Plant Cell Physiol* 1999, **40**(12):1262–1270.
44. Majeau N, Arnoldo MA, Coleman JR: **Modification of carbonic anhydrase activity by antisense and over-expression constructs in transgenic tobacco.** *Plant Mol Biol* 1994, **25**(3):377–385.
45. Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyère C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, *et al*: **The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla.** *Nature* 2007, **449**(7161):463–467.
46. Rumeau D, Cuine S, Fina L, Gault N, Nicole M, Peltier G: **Subcellular distribution of carbonic anhydrase in Solanum tuberosum L. leaves: characterization of two compartment-specific isoforms.** *Planta* 1996, **199**(1):79–88.
47. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, *et al*: **The genome of black cottonwood, Populus trichocarpa (Torr. & Gray).** *Science* 2006, **313**(5793):1596–1604.
48. Moyroud E, Minguet EG, Ott F, Yant L, Pose D, Monniaux M, Blanchet S, Bastien O, Thevenon E, Weigel D, Schmid M, Parcy F: **Prediction of regulatory interactions from genome sequences using a biophysical model for the Arabidopsis LEAFY transcription factor.** *Plant Cell* 2011, **23**(4):1293–1306.
49. Slater GS, Birney E: **Automated generation of heuristics for biological sequence comparison.** *BMC Bioinform* 2005, **6:**31.
50. Suyama M, Torrents D, Bork P: **PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments.** *Nucleic Acids Res* 2006, **34:**W609–W612. Web Server issue.
51. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP: **MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space.** *Syst Biol* 2012, **61**(3):539–542.
52. Emanuelsson O, Nielsen H, Brunak S, von Heijne G: **Predicting subcellular localization of proteins based on their N-terminal amino acid sequence.** *J Mol Biol* 2000, **300**(4):1005–1016.
53. Kolaskar AS, Tongaonkar PC: **A semi-empirical method for prediction of antigenic determinants on protein antigens.** *FEBS Lett* 1990, **276**(1–2):172–174.
54. Rice P, Longden I, Bleasby A: **EMBOSS: the European Molecular Biology Open Software Suite.** *Trends Genet* 2000, **16**(6):276–277.
55. Olson SA: **EMBOSS opens up sequence analysis. European Molecular Biology Open Software Suite.** *Brief Bioinform* 2002, **3**(1):87–91.
56. Mullan LJ, Bleasby AJ: **Short EMBOSS User Guide. European Molecular Biology Open Software Suite.** *Brief Bioinform* 2002, **3**(1):92–94.
57. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A: **Comparative protein structure modeling using Modeller.** In *Curr Protoc Bioinformatics.* Edited by Baxevanis AD. Hoboken, NJ: John Wiley and Sons, Inc; 2006. Chapter 5:Unit 5 6.
58. Lehtonen JV, Still DJ, Rantanen VV, Ekholm J, Bjorklund D, Iftikhar Z, Huhtala M, Repo S, Jussila A, Jaakkola J, Pentikäinen O, Nyrönen T, Salminen T, Gyllenberg M, Johnson MS: **BODIL: a molecular modeling environment for structure-function analysis and drug design.** *J Comput Aided Mol Des* 2004, **18**(6):401–419.
59. Humphrey W, Dalke A, Schulten K: **VMD: visual molecular dynamics.** *J Mol Graph* 1996, **14**(1):33–38. 27–38.
60. Singh H, Ansari HR, Raghava GP: **Improved method for linear B-cell epitope prediction using antigen's primary sequence.** *PLoS One* 2013, **8**(5):e62216.
61. Kimber MS, Coleman JR, Pai EF: **Beta-carbonic anhydrase from Pisum sativum: crystallization and preliminary X-ray analysis.** *Acta Crystallogr D Biol Crystallogr* 2000, **56**(Pt 7):927–929.
62. Yamano T, Morita S: **Effects of pesticides on isolated rat hepatocytes, mitochondria, and microsomes II.** *Arch Environ Contam Toxicol* 1995, **28**(1):1–7.
63. Supuran CT: **Carbonic anhydrase inhibition/activation: trip of a scientist around the world in the search of novel chemotypes and drug targets.** *Curr Pharm Des* 2010, **16**(29):3233–3245.
64. McKenna R, Supuran CT: **Carbonic anhydrase inhibitors drug design.** *Subcell Biochem* 2014, **75:**291–323.
65. Badger M: **The roles of carbonic anhydrases in photosynthetic CO_2 concentrating mechanisms.** *Photosynth Res* 2003, **77**(2–3):83–94.
66. Flower DR, Macdonald IK, Ramakrishnan K, Davies MN, Doytchinova IA: **Computer aided selection of candidate vaccine antigens.** *Immunome Res* 2010, **6**(Suppl 2):S1.
67. Yang B, Sayers S, Xiang Z, He Y: **Protegen: a web-based protective antigen database and analysis system.** *Nucleic Acids Res* 2011, **39**(Database issue): D1073–D1078.
68. Gustafson DJ: **Rising food costs & global food security: key issues & relevance for India.** *Indian J Med Res* 2013, **138**(3):398–410.

doi:10.1186/1756-3305-7-403
Cite this article as: Zolfaghari Emameh *et al.*: Beta carbonic anhydrases: novel targets for pesticides and anti-parasitic agents in agriculture and livestock husbandry. *Parasites & Vectors* 2014 **7**:403.

Zolfaghari Emameh et al. Parasites & Vectors (2015) 8:479
DOI 10.1186/s13071-015-1098-5

Parasites & Vectors

RESEARCH Open Access

Ascaris lumbricoides β carbonic anhydrase: a potential target enzyme for treatment of ascariasis

Reza Zolfaghari Emameh[1,2,3*], Marianne Kuuslahti[1], Daniela Vullo[4,5], Harlan R. Barker[1], Claudiu T. Supuran[4,5] and Seppo Parkkila[1,3]

Abstract

Background: A parasitic roundworm, *Ascaris lumbricoides*, is the causative agent of ascariasis, with approximately 760 million cases around the world. Helminthic infections occur with a high prevalence mostly in tropical and developing xcountries. Therefore, design of affordable broad-spectrum anti-helminthic agents against a variety of pathogens, including not only *A. lumbricoides* but also hookworms and whipworms, is desirable. Beta carbonic anhydrases (β-CAs) are considered promising targets of novel anthelminthics because these enzymes are present in various parasites, while completely absent in vertebrates.

Methods: In this study, we identified an *A. lumbricoides* β-CA (AlBCA) protein from protein sequence data using bioinformatics tools. We used computational biology resources and methods (including InterPro, CATH/Gene3D, KEGG, and METACYC) to analyze AlBCA and define potential roles of this enzyme in biological pathways. The *AlBCA* gene was cloned into pFastBac1, and recombinant AlBCA was produced in sf-9 insect cells. Kinetics of AlBCA were analyzed by a stopped-flow method.

Results: Multiple sequence alignment revealed that AlBCA contains the two sequence motifs, CXDXR and HXXC, typical for β-CAs. Recombinant AlBCA showed significant CA catalytic activity with k_{cat} of 6.0×10^5 s^{-1} and k_{cat}/K_M of 4.3×10^7 M^{-1} s^{-1}. The classical CA inhibitor, acetazolamide, showed an inhibition constant of 84.1 nM. Computational modeling suggests that the molecular architecture of AlBCA is highly similar to several other known β-CA structures. Functional predictions suggest that AlBCA might play a role in bicarbonate-mediated metabolic pathways, such as gluconeogenesis and removal of metabolically produced cyanate.

Conclusions: These results open new avenues to further investigate the precise functions of β-CAs in parasites and suggest that novel β-CA specific inhibitors should be developed and tested against helminthic diseases.

Keywords: *Ascaris lumbricoides*, Beta carbonic anhydrase, Enzyme inhibition, Sulfonamide, Acetazolamide, Bioinformatics, Computational biology, Ascariasis

Background

Two parasitic worms, *Ascaris lumbricoides* and *Ascaris suum* were independently nominated by Linnaeus in 1758 and Goeze in 1782, respectively. Recent genetic and paleoparasitological evidence has suggested that these strains are, in fact, a single species [1]. Therefore, the original name, *A. lumbricoides*, should be used upon priority on taxonomic nomination.

Around 760 million people worldwide are infected with *A. lumbricoides*, mainly in Southeast Asia [2]. The human ascariasis infection is normally caused by feces contamination in water, vegetables, and other food. The eggs of the worm hatch into larvae within the small intestine. The larvae spread through the blood stream to different organs and finally arrive in the lung. From the lungs they eventually enter the throat and are swallowed. In the intestinal tract, the larvae complete development into adult worms. A female *A. lumbricoides* worm can produce 240,000 eggs daily, which pass within feces to the environment to begin the cycle anew. The eggs are resistant to cold weather and disinfectants and can

* Correspondence: zolfaghari.emameh.reza.x@student.uta.fi
[1]Department of Anatomy, School of Medicine, University of Tampere, Tampere, Finland
[2]BioMediTech, University of Tampere, Tampere, Finland
Full list of author information is available at the end of the article

remain viable for more than 10 years. Because of the high load of nematodes in ascariasis, there are also severe complications including intestinal obstruction, peritonitis, and acute pancreatitis [3]. There are different treatment strategies for ascariasis, such as surgery (in case of bowel obstruction) and application of anthelminthic drugs including albendazole, mebendazole, and pyrantel pamoate. At present and in the future, access to new broad-spectrum anthelmintics against *A. lumbricoides*, as well as hookworms and whipworms, are needed in countries where these infections are endemic [4].

Carbonic anhydrases (CAs) have been recently identified as potential targets for novel anti-infective drugs. CAs are encoded by six evolutionary divergent gene families: α, β, γ, δ, ζ, and η CAs [5–7]. All members of these gene families are metalloenzymes, which usually contain zinc ion in their catalytic active site [8]. Certain ζ- and γ-CAs contain cadmium (II), iron (II) or cobalt (II) as alternative metal ion cofactors [9–11]. α-CAs are the most intensively studied family, which contains 13 catalytically active members in mammals: cytosolic enzymes (CA I, CA II, CA III, CA VII, and CA XIII), membrane-bound (CA IV, CA IX, CA XII, CA XIV, and CA XV), mitochondrial CAs (VA and CA VB), and secreted CA (VI) [12]. β-CAs are found in plants, algae, fungi, bacteria, protozoans, arthropods, and nematodes [6, 13, 14]; γ-CAs in algae, plants, bacteria, and archaea [15]; δ-CAs in free-living marine dinoflagellates [16]; ζ-CAs in marine diatoms [9]; and η-CAs in *Plasmodium* parasites [5]. CAs play a critical role in many biochemical pathways, including respiration, pH homeostasis, electrolyte transfer, bone resorption, calcification, gluconeogenesis, lipogenesis, and ureagenesis [12, 17]. Because *β-CA* genes are absent in vertebrate genomes, while present in many parasite genomes, they are considered potential candidate target enzymes for novel anti-infectives [6, 7, 18, 19]. Literature on CA inhibition reveals that many inhibitors, such as sulfonamide, sulfamides, sulfamates, anions, phenols, coumarins, dithiocarbamates, fullerenes, boronic acids, carboxylates, polyamines, benzamides, hydroxymates, and mercaptans have been tested against β-CAs to control infectious organisms, such as *Candida albicans*, *Cryptococcus neoformans*, *Leishmania donovani, Salmonella typhimurium*, *Porphyromonas gingivalis*, *Helicobacter pylori*, *Streptococcus pneumoniae*, *Mycobacterium tuberculosis*, and *Brucella suis* [20–27]. Meanwhile, inhibitory studies have been also carried out on β-CAs from non-pathogenic model organisms, including *Saccharomyces cerevisiae* and *Drosophila melanogaster* [12, 28–34].

In this study, we analyzed properties of *A. lumbricoides* β-CA (AlBCA) using bioinformatics tools, produced AlBCA as a recombinant protein in insect cells, and tested its kinetic and inhibition properties. These investigations represent the first experimental study on a β-CA protein from a parasitic nematode.

Methods

Identification of AlBCA protein sequence

A β-CA protein sequence from *Caenorhabditis elegans* (Uniprot ID: Q2YS41) [35] was used for the initial NCBI BLAST protein homology search (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Ten nematode β-CA protein sequences, including AlBCA, were aligned with the Clustal Omega algorithm to create a multiple sequence alignment (MSA) within the Jalview program (version 2.8.ob1) (http://www.jalview.org/).

Structural and functional predictions based on AlBCA sequence

The AlBCA protein sequence (Uniprot ID: F1LE18) was used as a query in the integrative protein signature database, InterPro (http://www.ebi.ac.uk/interpro/). This database integrates together predictive models of representative protein domains, families, and functional sites from multiple and diverse databases, such as Gene3D, PANTHER, Pfam, PIRSF, PRINTS, ProDom, PROSITE, SMART, SUPERFAMILY, and TIGRFAMs [36]. The resulting InterPro ID (IPR001765) for AlBCA protein sequence was used as a query in the CATH/Gene3D database (http://www.cathdb.info/) [37]. This database hierarchically classifies domains into sequence and structure-based families and fold groups, when there is a sufficient evidence for having diverged from a common ancestor. The CATH/Gene3D database generated a rainbow model for superimposed AlBCA protein sequence and several other close species. We also identified the biochemical pathways and interactions of AlBCA through KEGG (Kyoto Encyclopedia of Genes and Genomes) (http://www.kegg.jp/) [38] and METACYC metabolic pathway databases (http://metacyc.org/) [39], which have both been linked to InterPro.

Production of recombinant AlBCA

The CDS sequence of *AlBCA* gene was retrieved from the EMBL database (http://www.ebi.ac.uk/). GeneArt® gene synthesis technology (Life Technologies) was used to construct the *β-CA* gene sequence for insertion into the cloning vector (pFastBac1) [40]. DH10Bac cells (which contain the bacmid baculovirus shuttle vector, and a helper plasmid that produces the proteins needed for transposition), were transformed by pFast-*AlBCA* [7]. Then pFast-*AlBCA* was purified by PureLink™ HiPure Plasmid Purification Kit (Invitrogen).

Primary transfection of *Spodoptera frugiperda* (sf-9) insect cells for production of *Baculoviruses* was performed by HilyMax transfection reagent (Dojindo) (0.5×10^7 cells/ml in 6-well plates), and the cells were

incubated for 3 days at 29 °C. Cultured cells were centrifuged, passed through a 0.2 μm filter, and stored in a dark tube at +4 °C. For secondary transfection, 20 ml of cultured cells (2×10^6/ml) were transfected with a primary stock of *Baculoviruses* and incubated for 3 days in a 29 °C shaker. Cultured cells were centrifuged, passed through 0.2 μm filter, and stored in a dark tube at +4 °C. For the expression of recombinant AlBCA, the secondary stocks of *Baculoviruses* were used to infect sf-9 cells (2×10^6/ml) (using the same procedure as described for the production of the secondary stock, but at higher volumes).

The sf-9 cell culture medium was centrifuged at 5000 RPM for 10 min at room temperature. The supernatant containing secreted AlBCA was diluted at a ratio of 1:5 by binding buffer (0.1 M Tris, 0.2 M Na_2SO_4, pH 8). Then Protino™ Ni-NTA Agarose (Macherey-Nagel) was added to the diluted culture medium and put on a magnetic stirrer with a low rotation speed (2 h at room temperature and overnight incubation at +4 °C without stirring). The culture medium was poured into the funnel filtration system and passed through Whatman® filter paper by vacuum. The flow-through was collected in a separate vial. The agarose was washed with the wash buffer (50 mM Na_2HPO_4, 0.5 M NaCl, 20 mM Imidazol, pH 8). Elution buffer (50 mM Na_2HPO_4, 0.5 M NaCl, 0.25 M Imidazol, pH 8) was added into the column to finally elute the recombinant AlBCA. Thrombin treatment did not cleave the polyhistidine tag in spite of several attempts (data not shown) probably due to the specific molecular folding of AlBCA that completely hid the thrombin cleaving site. Therefore, the kinetic measurements were carried out using recombinant AlBCA containing the polyhistidine tag.

Kinetic characterization of AlBCA

An Applied Photophysics stopped-flow instrument has been used for assaying the CA catalyzed CO_2 hydration activity [41]. Phenol red (at a concentration of 0.2 mM) has been used as indicator, working at the absorbance maximum of 557 nm, with 20 mM TRIS (pH 8.3) as buffer, and 20 mM $NaClO_4$ (for maintaining a constant ionic strength; this anion is not inhibitory against AlBCA up to concentrations of 50 mM, data not shown), following the initial rates of the CA-catalyzed CO_2 hydration reaction for a period of 10–100 s. The CO_2 concentrations ranged from 1.7 to 17 mM for the determination of the kinetic parameters (by Lineweaver-Burk plots) and inhibition constants. For each measurement at least six traces of the initial 5-10 % of the reaction have been used for determining the initial velocity. The uncatalyzed rates were determined in the same manner and subtracted from the total observed rates. The inhibition constant of acetazolamide (AAZ, 5-acetamido-1,3,4-thiadiazole-2-sulfonamide) was obtained by non-linear least-squares methods using PRISM 3 and the Cheng-Prusoff equation, as reported earlier [42, 43], and represents the mean from at least three different determinations.

Results

Identification of AlBCA protein sequence

Multiple sequence alignment (MSA) revealed that all 10 β-CA protein sequences from nematodes contained the first (CXDXR; C: cysteine, D: aspartic acid, R: arginine, X: any residue) and second (HXXC; H: histidine, C: cysteine, X: any residue) highly conserved sequence motifs of the catalytic site, which are the hallmark residues for β-CAs (Table 1, Fig. 1). In addition, the neighbor residues present within or close to the active site were almost identical.

Structural and functional predictions based on AlBCA sequence

Analysis of the AlBCA protein sequence by the InterPro database resulted in classification as part of the carbonic anhydrase family InterPro ID IPR001765. In the CATH/Gene3D database, AlBCA is categorized with the CATH superfamily ID 3.40.1050.10 (Beta-carbonic Anhydrase; Chain A). The tools of CATH/Gene3D database were used to generate a rainbow model for superimposition of AlBCA protein with other close relative β-CA proteins (Fig. 2). Metabolic pathway analyses of the AlBCA protein sequence in the KEGG and METACYC databases predict that the enzyme plays a major role in nitrogen metabolism (Fig. 3) and gluconeogenesis ll pathways (Figs. 4 and 5). Results from the KEGG database suggested that AlBCA might functionally participate in detoxification of cyanate by providing bicarbonate for cyanase enzyme. The METACYC database also predicted

Table 1 IDs for ten β-CA protein sequences from nematodes

Nematode name	β-CA protein IDs[a]
Ascaris lumbricoides	F1LE18
Caenorhabditis brenneri	G0MRG1
Caenorhabditis briggsae	A8WN21
Caenorhabditis elegans	Q22460 (bca-1)
	Q2YS41 (bca-2, isoform c)
	D3NQA9 (bca-2, isoform d)
Caenorhabditis remanei	E3MK96
Haemonchus contortus	U6PDI1
Necator americanus	W2SJ13
Saccoglossus kowalevskii	NP_001171747.1[b]
Strongyloides ratti	A0A090LV46
Toxocara canis	KHN75516.1

[a]Uniprot β-CA protein IDs
[b]NCBI β-CA protein ID

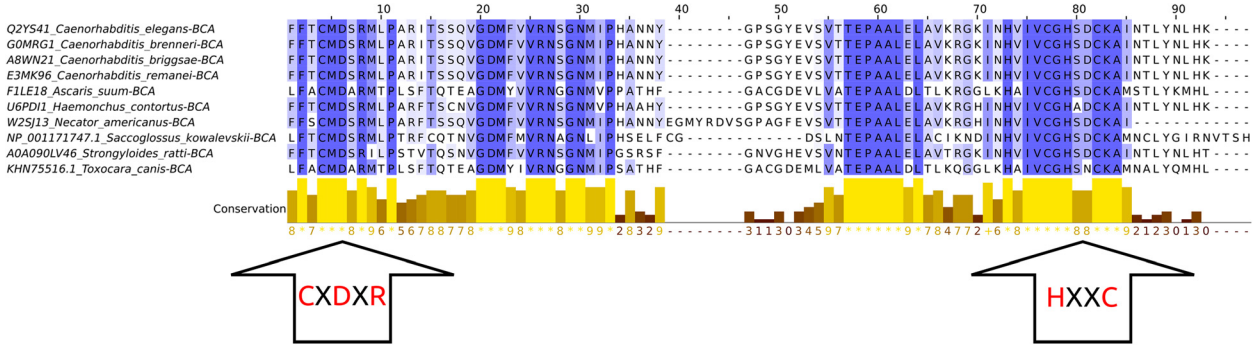


Fig. 1 Multiple sequence alignment (MSA) of β-CA protein sequences from 10 nematodes. Only the central region of the sequences (85 amino acids, starting 3 amino acids prior to first highly conserved motif) are shown in this alignment. MSA revealed that all of them including AlBCA contain the first (CXDXR; C: cysteine, D: aspartic acid, R: arginine, X: any residue) and second (HXXC; H: histidine, C: cysteine, X: any residue) highly conserved sequences, which have been indicated by two black arrows at the bottom. Zinc ion (Zn^{2+}) is the metal cofactor in catalytic active site of β-CAs, which binds to Cys from the first motif and His and Cys from the second motif [55]

bicarbonate as the final product of the β-CA catalytic reaction. This bicarbonate would be needed for the mitochondrial gluconeogenic pathway where pyruvate is converted to oxaloacetate.

Production and characterization of recombinant AlBCA

A *β-CA* gene insert, including a thrombin cleavage site and a segment encoding a poly-histidine tag, was produced by GeneArt® technology and integrated into a pFastBac1 plasmid (Additional file 1: Figure S1). The purified plasmid was used to produce recombinant AlBCA in sf-9 cells. SDS-PAGE containing the eluted fraction showed three polypeptide bands of 28, 30, and 33 kDa. The 30-kDa band was the major product, which corresponds to the calculated molecular mass of AlBCA (Fig. 6).

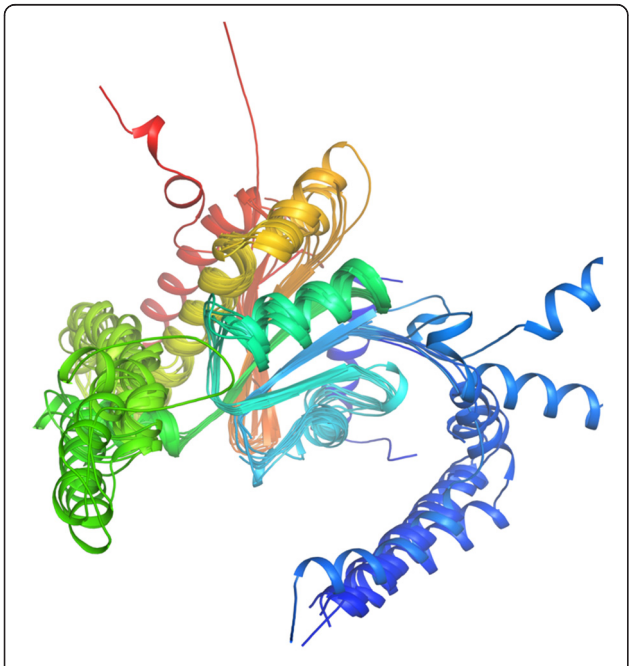


Fig. 2 A rainbow model for superimposition of AlBCA and β-CAs of close relative species. This image has been generated from a superposition of nine representative domains within this superfamily. The domain positions and protein structures appear to be highly similar in this model

Kinetic properties of AlBCA

We have measured the kinetic properties of AlBCA for the physiologic reaction of CO_2 hydration to bicarbonate and protons at pH 8.3 where, irrespective if this protein is a type I or type II β-CA, the active site should be open. The Zn (II) ion is coordinated by two Cys and one His residues as well as by a water molecule/hydroxide ion, acting as a nucleophile in the catalyzed reaction. We compared the activity of the new enzyme with those of well characterized α-class (human CA I and II) and β-CAs from *Leishmania donovani chagasi* (LdcCA), *Drosophila melanogaster* (DmBCA), and *Anopheles gambiae* (AgaCA), characterized earlier by our group [7, 21, 44]. As displayed in Table 2, AlBCA has significant catalytic activity in the physiologic reaction with the following kinetic parameters: k_{cat} of 6.0×10^5 s^{-1} and k_{cat}/K_m of 4.3×10^7 M^{-1} s^{-1}. The measured kinetic parameters show that AlBCA has a moderate activity, comparable to that of the relatively slower human carbonic anhydrase isoform hCA I, and in the same range as the enzymes characterized in *Anopheles gambiae* and *Leishamania donovani chagasi* [7, 21]. Furthermore, this activity has been effectively inhibited by the sulfonamide compound acetazolamide, which showed an inhibition constant of 84.1 nM. Although this was the only inhibitor investigated so far for this enzyme, work is in progress to understand in detail the inhibitory profiles of the main classes of activity modulators against this enzyme,

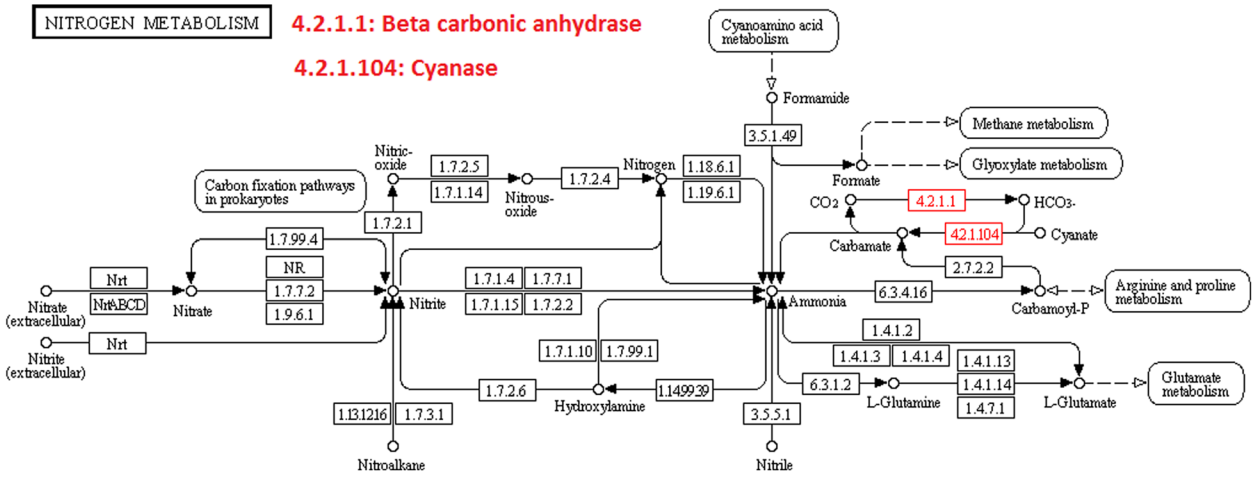


Fig. 3 The predicted role of AlBCA in nitrogen metabolism pathway. This pathway has been generated by the KEGG (Kyoto Encyclopedia of Genes and Genomes) database. According to this model AlBCA catalyzes the conversion of CO_2 and H_2O to HCO_3^-. Cyanase converts cyanate to CO and NH in a bicarbonate-dependent reaction

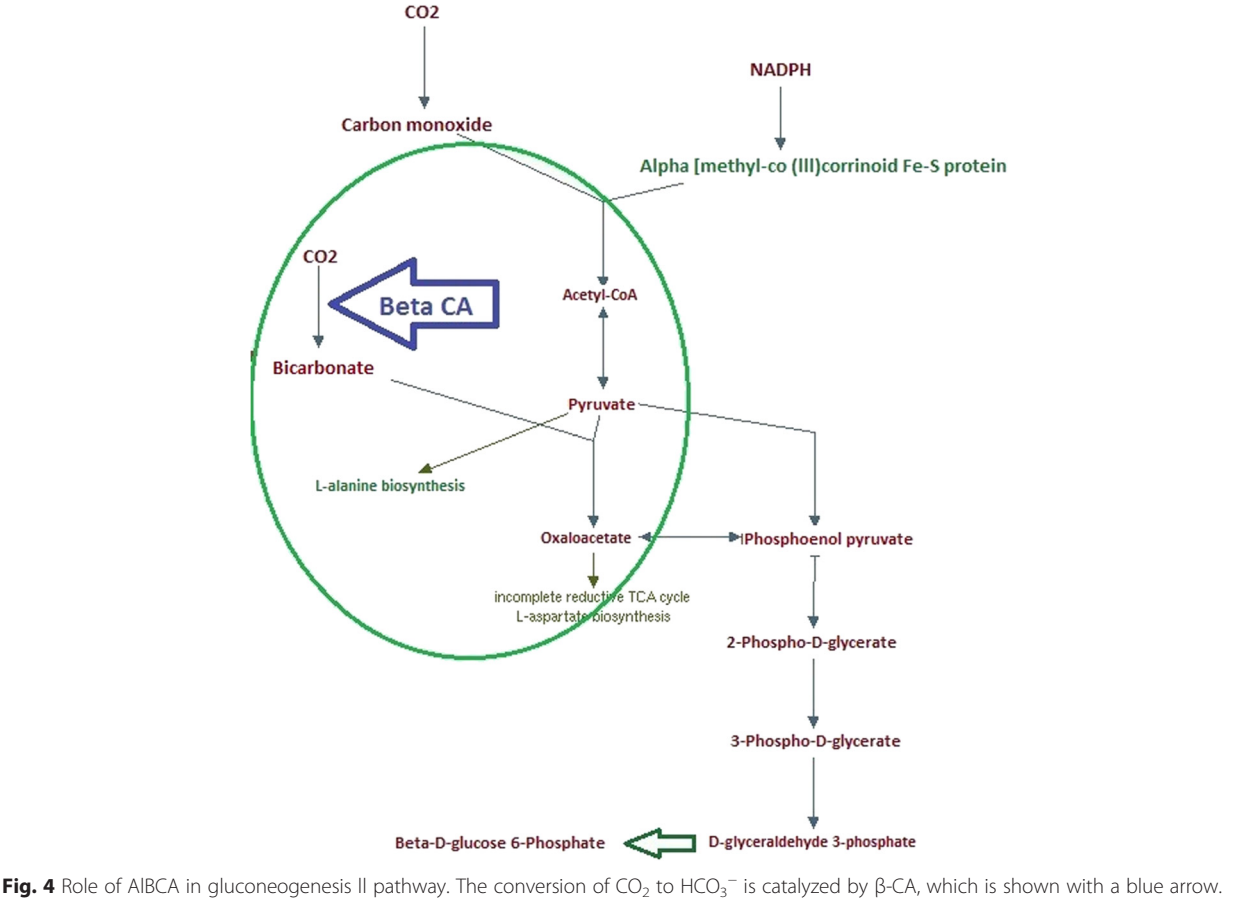


Fig. 4 Role of AlBCA in gluconeogenesis II pathway. The conversion of CO_2 to HCO_3^- is catalyzed by β-CA, which is shown with a blue arrow. Bicarbonate is then used for conversion of pyruvate to oxaloacetate

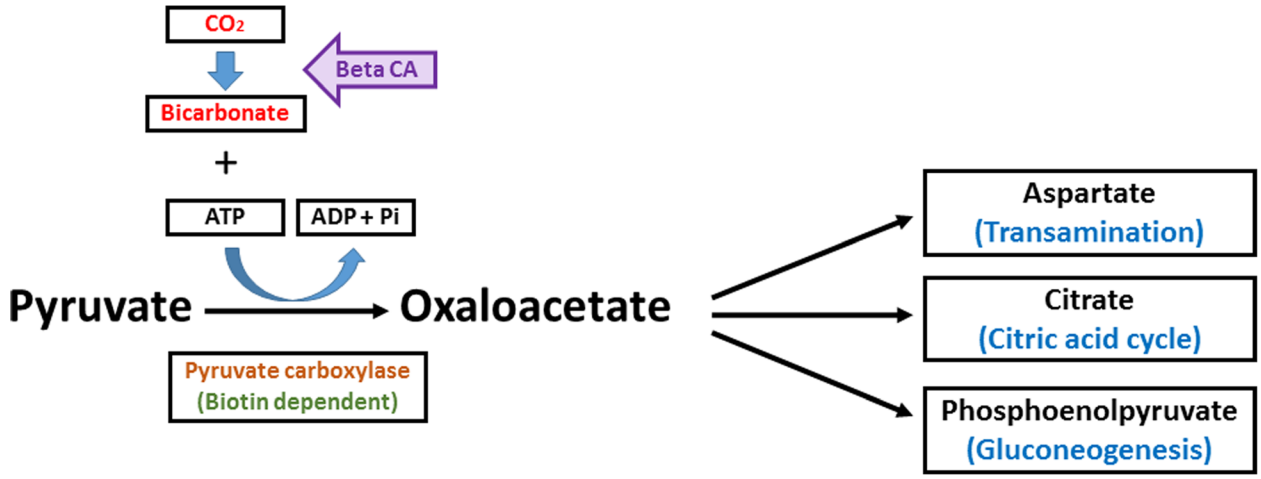


Fig. 5 Role of bicarbonate generated by AlBCA in gluconeogenesis II pathway. The conversion of CO_2 to HCO_3^- is catalyzed by β-CA, which is shown with a purple arrow. HCO_3^- is a major element in conversion of pyruvate to oxaloacetate. Oxaloacetate is used to generate aspartate, citrate, and phosphoenolpyruvate through transamination, citric acid cycle, and gluconeogenesis pathways, respectively

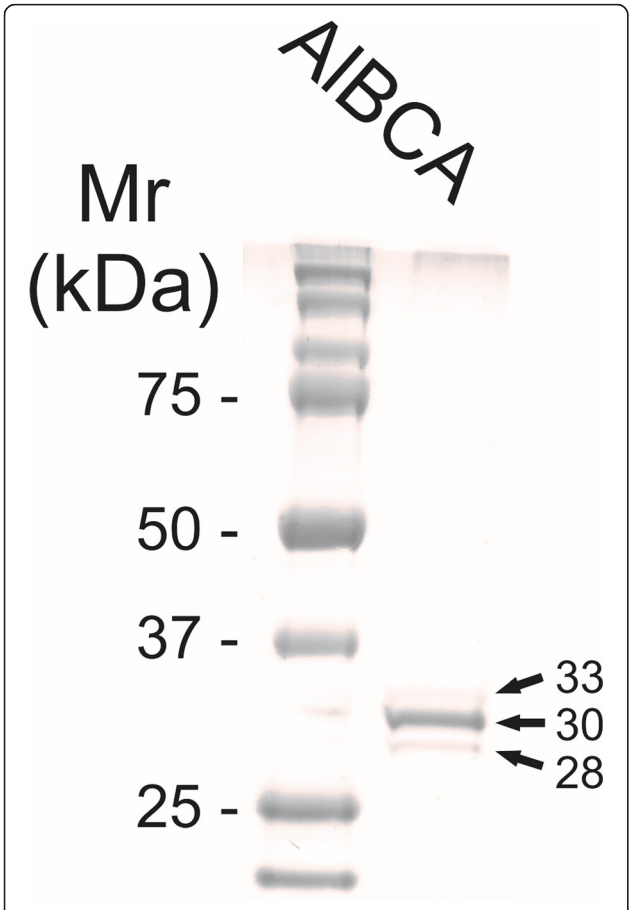


Fig. 6 SDS-PAGE of AlBCA. Left lane shows the protein standard. The purified recombinant AlBCA appeared as a triple band (28, 30, and 33 kDa). The predicted molecular mass of His-tagged AlBCA was 30 kDa

i.e., anions, sulfonamides and their isosteres, and dithiocarbamates and their isosteres.

Discussion

Multiple sequence alignment (MSA) has confirmed the presence of a β-CA enzyme in the proteome of *A. lumbricoides*, an important pathogen which is the causative agent of the parasitic roundworm disease, ascariasis. After this discovery our aim was focused on investigation of the putative functions of this enzyme by bioinformatics prediction tools, production as a recombinant protein, and characterization of its kinetic properties. We determined that AlBCA protein contains the first (CXDXR) and second (HXXC) hallmark sequences of β-CAs, and most of the neighbor residues present within or near the active site of the enzyme were also highly homologous compared to other aligned nine helminthic β-CAs. This result suggested that AlBCA is a catalytically active enzyme. The rainbow model of AlBCA protein was generated by superimposing its nine protein domains with β-CA proteins from close species. α-helices and β-sheets structures showed a high similarity in general domain architecture. We successfully produced recombinant AlBCA protein in Sf-9 insect cells. The SDS-PAGE analysis of the purified recombinant protein showed three polypeptide bands with molecular masses of 28, 30, and 33 kDa. Among them the middle band was clearly strongest and probably represents mature recombinant AlBCA protein. The lower 28-kDa polypeptide band variably appeared in SDS-PAGE analyses, and it probably represents a partially degraded protein. The upper 33-kDa polypeptide might represent a premature form of the enzyme. Even though our previous predictions for subcellular localization did not provide any definitive result [6], the

Table 2 Kinetic parameters for the CO_2 hydration reaction catalysed by the human isozymes hCA I, II, III, IV, VA, VB, VI, VII, IX, XII, XIII, and XIV (α-class CAs) and the β-CAs from *Leishmania donovani chagasi* (LdcCA), *Drosophila melanogaster* (DmBCA), *Anopheles gambiae* (AgaCA) and *Ascaris lumbricoides* (AlBCA)

Enzyme	Class	kcat (s^{-1})	kcat/Km (M^{-1}x s^{-1})	Ki (acetazolamide) (nM)[a]
hCA I	α	2.0×10^5	5.0×10^7	250 ± 12
hCA II	α	1.4×10^6	1.5×10^8	12 ± 0.8
hCA III	α	1.3×10^4	2.5×10^5	240000 ± 25000
hCA IV	α	1.1×10^6	5.1×10^7	74 ± 5.5
hCA VA	α	2.9×10^5	2.9×10^7	63 ± 2.1
hCA VB	α	9.5×10^5	9.8×10^7	54 ± 3.0
hCA VI	α	3.4×10^5	4.9×10^7	11 ± 0.7
hCA VII	α	9.5×10^5	8.3×10^7	2.5 ± 0.11
hCA IX	α	1.1×10^6	1.5×10^8	16 ± 0.8
hCA XII	α	4.2×10^5	3.5×10^7	5.7 ± 0.04
hCA XIII	α	1.5×10^5	1.1×10^7	16 ± 0.3
hCA XIV	α	3.1×10^5	3.9×10^7	41 ± 2.2
LdcCA	β	9.35×10^5	5.9×10^7	91.7 ± 5.7
DmBCA	β	9.5×10^5	1.1×10^8	516 ± 24
AgaCA	β	7.2×10^5	5.6×10^7	27.3 ± 2.0
AlBCA[a]	β	$(6.0 \pm 0.1)\times10^5$	$(4.3 \pm 0.2)\times10^7$	84.1 ± 2.9

[a]Mean ± standard error from three different assays

AlBCA protein is probably either a mitochondrial or secretory protein like the other parasite β-CAs defined so far. Our previous studies have shown that there are examples of metazoan β-CAs in both subcellular locations [6]. Several species, such as *Saccoglossus kowalevskii, Trichinella spiralis,* and *Strigamia maritima*, possess mitochondrial β-CAs. The highest score for a secretory signal peptide was predicted to the β-CA of malaria mosquito *Anopheles darlingi*. In our previous analysis using the SignalP server, the β-CA of malaria mosquito *Anopheles darlingi* had the highest score in likelihood to be a secreted protein [6]. In the functional predictions performed by computational tools, AlBCA associated to conversion of CO_2 and H_2O to bicarbonate, as expected. Then bicarbonate was functionally linked to detoxification of cyanate, which is a toxic byproduct of some metabolites, such as urea and carbamoylphosphate. Cyanase catalyzes the decomposition of cyanate into CO_2 and ammonia. Bicarbonate serves as a nucleophilic reactant that attacks and breaks down the cyanate, with carbamate as an unstable intermediate. Therefore, the role of CA in recycling of CO_2 into bicarbonate, and the importance of bicarbonate in the nitrogen metabolism pathway whole metabolic process, are evident [45]. Furthermore, it was predicted that β-CA plays a role in gluconeogenesis ll pathway. If AlBCA locates in mitochondria, it would generate bicarbonate as the key element for conversion of pyruvate to other final products, such as aspartate, citrate, and phosphoenolpyruvate through transamination, citric acid cycle, and gluconeogenesis, respectively. Indeed, our results have indicated that AlBCA shows a significant catalytic activity for the conversion of CO_2 into bicarbonate, as demonstrated *in vitro* by stopped-flow kinetic measurements.

Identification of β-CA from *C. elegans* in the Ensembl Metazoa database (http://metazoa.ensembl.org/index.html) [46] revealed that this model nematode contains three full-length β-CA protein sequences (Additional file 2: Table S1). An MSA of these β-CAs with AlBCA sequence, created with the Clustal Omega algorithm (http://www.ebi.ac.uk/Tools/msa/clustalo/) [47], showed that β-CA2 (isoform c, Uniprot ID: Q2YS41) from *C. elegans* is most similar to AlBCA. Thus, there is the possibility that the expression pattern of AlBCA is similar to *C. elegans* β-CA2 (isoform c). The expression of *C. elegans* β-CA2 (isoform c, Ensembl gene ID WBGene00013805) in the WormViz expression database (http://www.vanderbilt.edu/wormdoc/wormmap/WormViz.html) of WormBase (https://www.wormbase.org/#01-23-6) [48] showed that β-CA2 (isoform c) is expressed in all larval (EE, LE, L1, L2, L3, L4) and adult (including male and hermaphrodites) stages. The highest expression levels were detected in the body wall muscles of L2 stage. The results defined that β-CA is also present in larval neurons, muscles, coelomocytes, hypodermis layer, intestine, and excretory cells. In addition, β-CA2 (isoform c) is detectable in the whole body of adult male and hermaphrodite gonads of *C. elegans*. Previously, Fasseas *et al.* investigated the function of *C. elegans* β-CA using an RNAi technique [35]. They did not find any significant phenotypic change, which might be due to several reasons. First, other CA isoforms might compensate the lack of one β-CA. Second, the efficiency of gene silencing might have been inadequate due to challenges with RNAi technique. In another model organism, *D. melanogaster*, the highest upregulated values of β-CA mRNA were observed in the spermatheca (female), fat body, and heart of adults; as well as early larval stages and late in metamorphosis [7]. The knockdown studies of *D. melanogaster* β-CA showed a clear phenotypic change. Surprisingly, females were sterile and unable to produce eggs. An apparent defect was shown in migration of border cells, which probably contributed to infertility of female *D. melanogaster* [49].

For *in vivo* inhibitory studies, acetazolamide has been tested on live *C. elegans* [50] and *Plasmodium falciparum* (malaria parasite) [51]. These studies showed, however, that acetazolamide could not penetrate through the nematode cuticle or protozoan surface. A BLAST search using 14 human α-CAs as queries recently identified six α-CAs (CAH-1 to 6) in *C. elegans* [52]. Güzel *et al.* [53] carried out an inhibition study on *C. elegans* CAH-4 and 13 mammalian CAs using

different sulfonamide derivatives including 2-(hydrazinocarbonyl)-3-phenyl-1H-indole-5-sulfonamides of type 1 and 2, and compared their inhibition efficacy with traditional CA inhibitors including acetazolamide (AZA) and ethoxzolamide (EZA). Their study defined that some of the new derivatives displayed excellent inhibitory action on CAH-4 with K_i as low as 6 nM, which was 5.83 times better value compared to AZA. Within this context, AZA and EZA showed moderate inhibitory effects on CAH-4. In another study, Giacomotto *et al.* [52] realized that *cah-4* null mutant *C. elegans* nematodes were nonviable, and knockdown experiments, using the RNAi technique, showed reduced muscle degeneration in dystrophin deficient muscle. Furthermore, *cah-4* RNAi treatment caused a significant shift in the dose response curves of CA inhibitors, methazolamide and dichlorphenamide. Giacomotto and coworkers concluded that their studies confirmed the suitability of *C. elegans* as a model organism for screening, identifying and characterizing potential lead pharmacological agents [52].

Albendazole and Mebendazole are drugs quite well-tolerated when used against ascariasis and resistance has not yet been a major issue. Novel anthelmintic approaches are needed, however, because global healthcare will likely face these challenges at some point, since millions of people require treatment [2, 54]. Inhibition of AlBCA by single dose and broad-spectrum inhibitors, which are effective against various helminthic infections, would be a novel strategy for treatment of ascariasis. It could potentially disrupt the normal detoxification of cyanate, which would in turn increase the intracellular cyanate concentration to a toxic level, leading to the death of the parasite.

Conclusions

β-CAs represent promising targets for novel antiparasitic drug design. In the future, new broad-spectrum, and preferably single dose β-CA inhibitors, should be designed against AlBCA and corresponding enzymes of whipworms and hookworms. For the moment, the clinically used sulfonamide acetazolamide, the only inhibitor tested so far, showed a promising *in vitro* inhibitory power, with an inhibition constant of 84.1 nM on AlBCA. Acetazolamide's inability to penetrate the nematode is an obvious problem. Therefore, further studies should be planned to improve the penetration efficacy of CA inhibitors through biological membranes and cuticles of worms. The new sulfonamide derivatives, which were recently shown to inhibit *C. elegans* α-CAs, could represent useful leads for design of novel compounds having higher efficiency, better penetration, and minimal side effects on human CAs.

Additional files

Additional file 1: Figure S1. Construction of pFast-AlBCA cloning vector for production of recombinant AlBCA. The construct contained the restriction sites for EcoR1 and Xho1, thrombin-cutting sequence, and 6× His-tag sequences. (TIFF 102 kb)

Additional file 2: Table S1. Full-length β-CA protein sequences from *Caenorhabditis elegans*. (DOC 28 kb)

Abbreviations

EE: Early embryo; LE: Late Embryo; L1: Larval stage 1; L2: Larval stage 2; L3: Larval stage 3; L4: Larval stage 4.

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 most bioinformatics and computational biology studies on AlBCA. RZE and HRB participated in the sequence alignment. RZE and MK designed the expression vector and produced recombinant AlBCA. CTS and DV participated in the kinetic and inhibitory studies of AlBCA. RZE and MK drafted the first version of the manuscript. All authors participated in revision of the manuscript and approved the final version.

Acknowledgments

The authors thank Aulikki Lehmus for skillful technical assistance. 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. This study was also funded 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 (SP).

Author details

[1]Department of Anatomy, School of Medicine, University of Tampere, Tampere, Finland. [2]BioMediTech, University of Tampere, Tampere, Finland. [3]Fimlab Laboratories Ltd and Tampere University Hospital, Tampere, Finland. [4]Dipartimento di Chimica, Laboratorio di Chimica Bioinorganica, Universita' degli Studi di Firenze, Sesto Fiorentino, Firenze, Italy. [5]Neurofarba Department, Sezione di Scienze Farmaceutiche e Nutraceutiche, Universita' degli Studi di Firenze, Sesto Fiorentino, Firenze, Italy.

Received: 4 August 2015 Accepted: 15 September 2015
Published online: 18 September 2015

References

1. Leles D, Gardner SL, Reinhard K, Iniguez A, Araujo A. Are Ascaris lumbricoides and Ascaris suum a single species? Parasit Vectors. 2012;5:42.
2. Betson M, Nejsum P, Bendall RP, Deb RM, Stothard JR. Molecular epidemiology of ascariasis: a global perspective on the transmission dynamics of Ascaris in people and pigs. J Infect Dis. 2014;210(6):932–41.
3. Das AK. Hepatic and biliary ascariasis. J Glob Infect Dis. 2014;6(2):65–72.
4. Kappagoda S, Singh U, Blackburn BG. Antiparasitic therapy. Mayo Clin Proc. 2011;86(6):561–83.
5. 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 Plasmodium falciparum–the eta-carbonic anhydrases. Bioorg Med Chem Lett. 2014;24(18):4389–96.
6. 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.
7. 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.

8. Bracey MH, Christiansen J, Tovar P, Cramer SP, Bartlett SG. Spinach carbonic anhydrase: investigation of the zinc-binding ligands by site-directed mutagenesis, elemental analysis, and EXAFS. Biochemistry. 1994;33(44):13126–31.
9. Lane TW, Saito MA, George GN, Pickering IJ, Prince RC, Morel FM. Biochemistry: a cadmium enzyme from a marine diatom. Nature. 2005;435(7038):42.
10. Xu Y, Feng L, Jeffrey PD, Shi Y, Morel FM. Structure and metal exchange in the cadmium carbonic anhydrase of marine diatoms. Nature. 2008;452(7183):56–61.
11. Ferry JG. The gamma class of carbonic anhydrases. Biochim Biophys Acta. 2010;1804(2):374–81.
12. Supuran CT. Carbonic anhydrases: novel therapeutic applications for inhibitors and activators. Nat Rev Drug Discov. 2008;7(2):168–81.
13. Burnell JN, Gibbs MJ, Mason JG. Spinach chloroplastic carbonic anhydrase: nucleotide sequence analysis of cDNA. Plant Physiol. 1990;92(1):37–40.
14. Rowlett RS. Structure and catalytic mechanism of the beta-carbonic anhydrases. Biochim Biophys Acta. 2010;1804(2):362–73.
15. Capasso C, Supuran CT. An overview of the alpha-, beta- and gamma-carbonic anhydrases from Bacteria: can bacterial carbonic anhydrases shed new light on evolution of bacteria? J Enzyme Inhib Med Chem. 2015;30(2):325–32.
16. Lapointe M, Mackenzie TD, Morse D. An external delta-carbonic anhydrase in a free-living marine dinoflagellate may circumvent diffusion-limited carbon acquisition. Plant Physiol. 2008;147(3):1427–36.
17. Aggarwal M, McKenna R. Update on carbonic anhydrase inhibitors: a patent review (2008–2011). Expert Opin Ther Pat. 2012;22(8):903–15.
18. 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.
19. 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.
20. Innocenti A, Hall RA, Schlicker C, Muhlschlegel FA, Supuran CT. Carbonic anhydrase inhibitors. Inhibition of the beta-class enzymes from the fungal pathogens Candida albicans and Cryptococcus neoformans with aliphatic and aromatic carboxylates. Bioorg Med Chem. 2009;17(7):2654–7.
21. Syrjanen L, Vermelho AB, Rodrigues Ide A, Corte-Real S, Salonen T, Pan P, et al. Cloning, characterization, and inhibition studies of a beta-carbonic anhydrase from Leishmania donovani chagasi, the protozoan parasite responsible for leishmaniasis. J Med Chem. 2013;56(18):7372–81.
22. Nishimori I, Minakuchi T, Vullo D, Scozzafava A, Supuran CT. Inhibition studies of the beta-carbonic anhydrases from the bacterial pathogen Salmonella enterica serovar Typhimurium with sulfonamides and sulfamates. Bioorg Med Chem. 2011;19(16):5023–30.
23. Prete SD, Vullo D, Osman SM, Scozzafava A, AlOthman Z, Capasso C, et al. Sulfonamide inhibition study of the carbonic anhydrases from the bacterial pathogen Porphyromonas gingivalis: the beta-class (PgiCAb) versus the gamma-class (PgiCA) enzymes. Bioorg Med Chem. 2014;22(17):4537–43.
24. Nishimori I, Minakuchi T, Kohsaki T, Onishi S, Takeuchi H, Vullo D, et al. Carbonic anhydrase inhibitors: the beta-carbonic anhydrase from Helicobacter pylori is a new target for sulfonamide and sulfamate inhibitors. Bioorg Med Chem Lett. 2007;17(13):3585–94.
25. Burghout P, Vullo D, Scozzafava A, Hermans PW, Supuran CT. Inhibition of the beta-carbonic anhydrase from Streptococcus pneumoniae by inorganic anions and small molecules: Toward innovative drug design of antiinfectives? Bioorg Med Chem. 2011;19(1):243–8.
26. Buchieri MV, Riafrecha LE, Rodriguez OM, Vullo D, Morbidoni HR, Supuran CT, et al. Inhibition of the beta-carbonic anhydrases from Mycobacterium tuberculosis with C-cinnamoyl glycosides: identification of the first inhibitor with anti-mycobacterial activity. Bioorg Med Chem Lett. 2013;23(3):740–3.
27. Joseph P, Turtaut F, Ouahrani-Bettache S, Montero JL, Nishimori I, Minakuchi T, et al. Cloning, characterization, and inhibition studies of a beta-carbonic anhydrase from Brucella suis. J Med Chem. 2010;53(5):2277–85.
28. Isik S, Guler OO, Kockar F, Aydin M, Arslan O, Supuran CT. Saccharomyces cerevisiae beta-carbonic anhydrase: inhibition and activation studies. Curr Pharm Des. 2010;16(29):3327–36.
29. Syrjanen L, Parkkila S, Scozzafava A, Supuran CT. Sulfonamide inhibition studies of the beta carbonic anhydrase from Drosophila melanogaster. Bioorg Med Chem Lett. 2014;24(13):2797–801.
30. Supuran CT, Casini A, Scozzafava A. Protease inhibitors of the sulfonamide type: anticancer, antiinflammatory, and antiviral agents. Med Res Rev. 2003;23(5):535–58.
31. Alterio V, Di Fiore A, D'Ambrosio K, Supuran CT, De Simone G. Multiple binding modes of inhibitors to carbonic anhydrases: how to design specific drugs targeting 15 different isoforms? Chem Rev. 2012;112(8):4421–68.
32. Vullo D, Del Prete S, Osman SM, De Luca V, Scozzafava A, Alothman Z, et al. Sulfonamide inhibition studies of the gamma-carbonic anhydrase from the oral pathogen Porphyromonas gingivalis. Bioorg Med Chem Lett. 2014;24(1):240–4.
33. Supuran CT. Structure-based drug discovery of carbonic anhydrase inhibitors. J Enzyme Inhib Med Chem. 2012;27(6):759–72.
34. Capasso C, Supuran CT. Anti-infective carbonic anhydrase inhibitors: a patent and literature review. Expert Opin Ther Pat. 2013;23(6):693–704.
35. Fasseas MK, Tsikou D, Flemetakis E, Katinakis P. Molecular and biochemical analysis of the beta class carbonic anhydrases in Caenorhabditis elegans. Mol Biol Rep. 2010;37(6):2941–50.
36. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. InterPro: the integrative protein signature database. Nucleic Acids Res. 2009;37(Database issue):D211–5.
37. Pearl F, Todd A, Sillitoe I, Dibley M, Redfern O, Lewis T, et al. The CATH Domain structure database and related resources Gene3D and DHS provide comprehensive domain family information for genome analysis. Nucleic Acids Res. 2005;33(Database issue):D247–51.
38. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.
39. Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. 2014;42(Database issue):D459–71.
40. Notka F, Liss M, Wagner R. Industrial scale gene synthesis. Methods Enzymol. 2011;498:247–75.
41. Khalifah RG. The carbon dioxide hydration activity of carbonic anhydrase. I. Stop-flow kinetic studies on the native human isoenzymes B and C. J Biol Chem. 1971;246(8):2561–73.
42. Maresca A, Vullo D, Scozzafava A, Manole G, Supuran CT. Inhibition of the beta-class carbonic anhydrases from Mycobacterium tuberculosis with carboxylic acids. J Enzyme Inhib Med Chem. 2013;28(2):392–6.
43. Maresca A, Scozzafava A, Vullo D, Supuran CT. Dihalogenated sulfanilamides and benzolamides are effective inhibitors of the three beta-class carbonic anhydrases from Mycobacterium tuberculosis. J Enzyme Inhib Med Chem. 2013;28(2):384–7.
44. Syrjanen L, Kuuslahti M, Tolvanen M, Vullo D, Parkkila S, Supuran CT. The beta-carbonic anhydrase from the malaria mosquito Anopheles gambiae is highly inhibited by sulfonamides. Bioorg Med Chem. 2015;23(10):2303–9.
45. Luque-Almagro VM, Huertas MJ, Saez LP, Luque-Romero MM, Moreno-Vivian C, Castillo F, et al. Characterization of the Pseudomonas pseudoalcaligenes CECT5344 Cyanase, an enzyme that is not essential for cyanide assimilation. Appl Environ Microbiol. 2008;74(20):6280–8.
46. Kersey PJ, Staines DM, Lawson D, Kulesha E, Derwent P, Humphrey JC, et al. Ensembl Genomes: an integrative resource for genome-scale data from non-vertebrate species. Nucleic Acids Res. 2012;40(Database issue):D91–7.
47. Sievers F, Higgins DG. Clustal omega. Curr Protoc Bioinformatics / Editoral Board Andreas D Baxevanis [et al]. 2014;48:3 13 11–13 13 16.
48. Yook K, Harris TW, Bieri T, Cabunoc A, Chan J, Chen WJ, et al. WormBase 2012: more genomes, more data, new website. Nucleic Acids Res. 2012;40(Database issue):D735–41.
49. Syrjanen L, Valanne S, Kuuslahti M, Tuomela T, Sriram A, Sanz A, et al. beta carbonic anhydrase is required for female fertility in Drosophila melanogaster. Front Zool. 2015;12:19.
50. Hall RA, Vullo D, Innocenti A, Scozzafava A, Supuran CT, Klappa P, et al. External pH influences the transcriptional profile of the carbonic anhydrase, CAH-4b in Caenorhabditis elegans. Mol Biochem Parasitol. 2008;161(2):140–9.
51. Krungkrai J, Scozzafava A, Reungprapavut S, Krungkrai SR, Rattanajak R, Kamchonwongpaisan S, et al. Carbonic anhydrase inhibitors. Inhibition of Plasmodium falciparum carbonic anhydrase with aromatic sulfonamides: towards antimalarials with a novel mechanism of action? Bioorg Med Chem. 2005;13(2):483–9.
52. Giacomotto J, Pertl C, Borrel C, Walter MC, Bulst S, Johnsen B, et al. Evaluation of the therapeutic potential of carbonic anhydrase inhibitors in two animal models of dystrophin deficient muscular dystrophy. Hum Mol Genet. 2009;18(21):4089–101.
53. Guzel O, Innocenti A, Hall RA, Scozzafava A, Muhlschlegel FA, Supuran CT. Carbonic anhydrase inhibitors. The nematode alpha-carbonic anhydrase of

Caenorhabditis elegans CAH-4b is highly inhibited by 2-(hydrazinocarbonyl)-3-substituted-phenyl-1H-indole-5-sulfonamides. Bioorg Med Chem. 2009;17(8):3212–5.
54. Hagel I, Giusti T. Ascaris lumbricoides: an overview of therapeutic targets. Infect Disord Drug Targets. 2010;10(5):349–67.
55. Kimber MS, Pai EF. The active site architecture of Pisum sativum beta-carbonic anhydrase is a mirror image of that of alpha-carbonic anhydrases. EMBO J. 2000;19(7):1407–18.

microbial biotechnology

Open Access

Innovative molecular diagnosis of *Trichinella* species based on β-carbonic anhydrase genomic sequence

Reza Zolfaghari Emameh,[1,2,3]* Marianne Kuuslahti,[1] Anu Näreaho,[4] Antti Sukura[4] and Seppo Parkkila[1,3]

[1]*School of Medicine, University of Tampere, FI-33520 Tampere, Finland*

[2]*BioMediTech, University of Tampere, FI-33520 Tampere, Finland*

[3]*Fimlab Laboratories Ltd and Tampere University Hospital, Biokatu 4, FI-33520 Tampere, Finland*

[4]*Department of Veterinary Biosciences, Faculty of Veterinary Medicine, University of Helsinki, FI-00014 Helsinki, Finland*

Summary

Trichinellosis is a helminthic infection where different species of *Trichinella* nematodes are the causative agents. Several molecular assays have been designed to aid diagnostics of trichinellosis. These assays are mostly complex and expensive. The genomes of *Trichinella* species contain certain parasite-specific genes, which can be detected by polymerase chain reaction (PCR) methods. We selected β-carbonic anhydrase (*β-CA*) gene as a target, because it is present in many parasites genomes but absent in vertebrates. We developed a novel β-CA gene-based method for detection of *Trichinella* larvae in biological samples. We first identified a β-CA protein sequence from *Trichinella spiralis* by bioinformatic tools using β-CAs from *Caenorhabditis elegans* and *Drosophila melanogaster*. Thereafter, 16 sets of designed primers were tested to detect β-CA genomic sequences from three species of *Trichinella*, including *T. spiralis*, *Trichinella pseudospiralis* and *Trichinella nativa*. Among all 16 sets of designed primers, the primer set No. 2 efficiently amplified β-CA genomic sequences from *T. spiralis*, *T. pseudospiralis* and *T. nativa* without any false-positive amplicons from other parasite samples including *Toxoplasma gondii*, *Toxocara cati* and *Parascaris equorum*. This robust and straightforward method could be useful for meat inspection in slaughterhouses, quality control by food authorities and medical laboratories.

Received 21 July, 2015; revised 3 September, 2015; accepted 18 September, 2015. *For correspondence. E-mail zolfaghari.emameh.reza.x@student.uta.fi; Tel. +358-3-355111.

doi:10.1111/1751-7915.12327

Funding Information 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. This study was also funded by Academy of Finland, Finnish Cultural Foundation (Pirkanmaa Regional Fund), Sigrid Juselius Foundation, Jane and Aatos Erkko Foundation, Tampere Tuberculosis Foundation, and Competitive Research Funding of the Tampere University Hospital.

Introduction

Trichinella spp. are the causative agents of a zoonotic helminthic disease, which is called trichinellosis (Gottstein *et al*., 2009). Within the main sources of *Trichinella* spp. [including pigs, horses, wild boars, dogs, walruses, foxes, bears and birds (Alban *et al*., 2011)], the most important source of human infection worldwide is the domestic pig, e.g. in Europe, whereas horse and wild boar meat have played a significant role during outbreaks within the past three decades (Gottstein *et al*., 2009). Infection in humans occurs with the ingestion of insufficiently cooked meat contaminated with *Trichinella* larvae, which are encysted in muscle tissues of the host animals (Clausen *et al*., 1996; Pozio, 2007; Gottstein *et al*., 2009). Due to the re-emerging problem also in Europe, the European Union and some associated non-European Union member countries still run *Trichinella* monitoring programmes (Gottstein *et al*., 2009). Following delivery by the gravid female worm, which lives within the intestinal mucosa of the host, newborn larvae (NBL) predominantly migrate into lymphatic and blood vessels of the host and thereafter to the ultimate target organ, i.e. highly oxygenated muscle. NBL enter the striated muscle cells by the aid of their stylet (Gottstein *et al*., 2009). NBL survive in muscle nurse cells for years (up to 40 years in humans and over 20 years, e.g. in polar bears) (Froscher *et al*., 1988; Kumar *et al*., 1990). Among 12 different genetically detected *Trichinella* spp., *Trichinella spiralis* is the most common species in domestic and wild swine and is also the most important aetiological agent that causes wide and global distribution of trichinellosis in humans (Pozio and Darwin Murrell, 2006). Other *Trichinella* spp. include *Trichinella nativa* (La Rosa *et al*., 2001; Nelson *et al*., 2003), *Trichinella pseudospiralis* (La Rosa *et al*., 2001; Pozio and Darwin Murrell, 2006), *Trichinella britovi* (Gottstein *et al*., 2009), *Trichinella murelli* (Ancelle, 1998), *Trichinella nelsoni* (Marucci *et al*., 2009), *Trichinella papuae* (Pozio *et al*., 2005), *Trichinella zimbabwensis*

(Pozio *et al*., 2002), *Trichinella* genotype T8 (Marucci *et al*., 2009), *Trichinella* genotype T9 (Zarlenga *et al*., 2006) and *Trichinella* genotype T12 (Krivokapich *et al*., 2012). Seven percent of infected animals carry mixed infections (Airas *et al*., 2010).

Diagnosis of trichinellosis can be made by several methods. The classical method includes meat inspection and sampling for direct microscopic detection of larvae (Nockler *et al*., 2000). The other choices are detection of anti-*Trichinella* immunoglobulin G (IgG) in animal serum or meat juice by serological methods including enzyme-linked immunosorbent assay (Gamble *et al*., 2000; Korinkova *et al*., 2008), proteomics (Wang *et al*., 2014), and molecular methods including conventional polymerase chain reaction (PCR), real-time PCR, loop-mediated isothermal amplification (Tantrawatpan *et al*., 2012; Lin *et al*., 2013) and reverse line blot hybridization (Rombout *et al*., 2001). Many studies have been focused on detection of novel molecular markers, because the classical methods have limitations, such as labour-intensiveness, necessity of high level expertise for direct microscopic diagnosis, and high costs due to the required manpower. The potential molecular markers have included Internal transcribed spacer (ITS1 and ITS2) (Zarlenga *et al*., 1999; Lin *et al*., 2013), mitochondrial cytochrome C oxidase subunit III (cox3) gene (Van De *et al*., 2012), mitochondrial large subunit ribosomal RNA (rrnL) (Borsuk *et al*., 2003; Guenther *et al*., 2008; Blaga *et al*., 2009; Li *et al*., 2011), mitochondrial small subunit ribosomal RNA (rrnS) (Blaga *et al*., 2009), DNA sequence of the 5S rRNA intergenic spacer regions (Rombout *et al*., 2001), migratory DNA from *Trichinella* larvae (Li *et al*., 2010) and aminopeptidase (TsAP) gene (Zhang *et al*., 2013).

One effective strategy for specific detection of *Trichinella* infection is identification of a genomic or proteomic determinant, which is absent in host genomes or proteomes (Duplessis and Moineau, 2001). Beta carbonic anhydrase (β-CA) could be one of such host-specific markers (Fasseas *et al*., 2010; Syrjanen *et al*., 2010; Zolfaghari Emameh *et al*., 2014; 2015). Six evolutionary divergent gene families encode for CAs (α, β, γ, δ, ζ and η) (Lane *et al*., 2005; Xu *et al*., 2008; Ferry, 2010; Del Prete *et al*., 2014; Zolfaghari Emameh *et al*., 2014). Different CA isozymes are involved in many biological processes, such as respiration involving transport of CO_2 and bicarbonate between metabolizing tissues, pH homeostasis, electrolyte transfer, bone resorption, and calcification. Moreover, they participate in some biosynthetic reactions, such as gluconeogenesis, lipogenesis and ureagenesis in animals (Vullo *et al*., 2004; 2005; Alterio *et al*., 2006; Nishimori *et al*., 2007; Supuran, 2008), growth of some bacterial species (Merlin *et al*., 2003; Mitsuhashi *et al*., 2004) and photosynthesis in plants and algae (Fawcett *et al*., 1990; Funke *et al*., 1997). The previous investigations have confirmed that *β-CA* gene sequences are completely absent in vertebrate genomes, whereas they can be found in plants, algae, yeasts, bacteria, archaea, protozoa and invertebrate metazoans (Fawcett *et al*., 1990; Funke *et al*., 1997; Smith *et al*., 2000; Merlin *et al*., 2003; Mitsuhashi *et al*., 2004; Aguilera *et al*., 2005; Fasseas *et al*., 2010; Syrjanen *et al*., 2010; Zolfaghari Emameh *et al*., 2014). By this background, we hypothesized that genomic *β-CA* sequences from prokaryotic pathogens and eukaryotic parasites would be an ideal host-specific determinant for molecular diagnostics, with minimal interfering effects from mammalian genomes.

In this study, we designed and piloted a PCR method for rapid detection of three species of *Trichinella* larvae, including *T. spiralis*, *T. pseudospiralis* and *T. nativa* in meat samples.

Results

Multiple sequence alignment (MSA)

MSA of the β-CA protein sequence from *T. spiralis* (UniProt ID: E5SH53) against the sequences from *Drosophila melanogaster* (UniProt ID: Q9VHJ5) and *Caenorhabditis elegans* (UniProt ID: Q22460) revealed that they all contain the first (CxDxR; C: cystein, D: aspartic acid, R: arginine, and x: any residue) and second (HxxC; H: histidine, C: cystein, and x: any residue) highly conserved residues (Fig. 1). Hence, the β-CA protein sequence from *T. spiralis* (UniProt ID: E5SH53) extracted correctly through the Basic Local Alignment Search Tool (BLAST) homology analysis and contained identical highly conserved residues to the predefined β-CA protein sequences.

Primer design and PCR

Among the four different pairs of primers (Nos. 1–4) only the primer set No. 2 (PCR product size 191 bp) showed a clear positive DNA band on agarose gel electrophoresis for detection of *T. spiralis* β-CA genomic sequence in larvae and meat samples (data not shown). It also produced positive bands on larvae samples from *T. pseudospiralis* and *T. nativa* (Fig. 2). The next rounds of PCR analyses were performed to find out whether it was possible to design primers specific for *T. pseudospiralis* and *T. nativa.* We used the same reverse primer (primer set No. 2) which worked well for the first PCR round. The PCR products representing partial sequences of *T. pseudospiralis* and *T. nativa β-CA* gene were sequenced, and the forward primers were designed based on the obtained sequences. These primers were included in the sets Nos. 5–7 for *T. pseudospiralis* and 8–11 for *T. nativa*, and the PCR

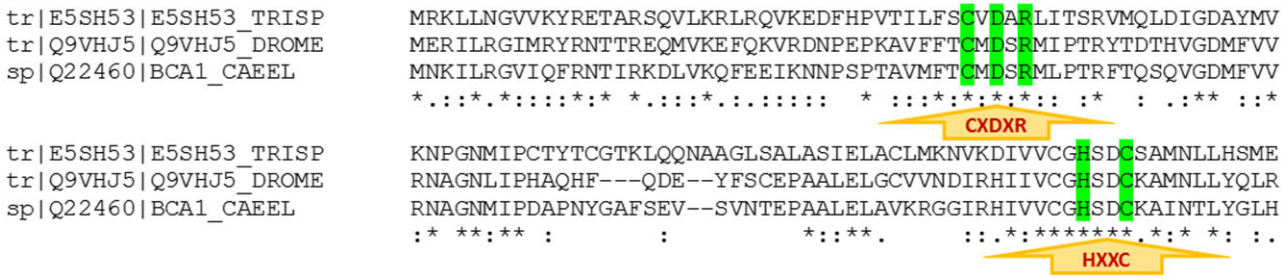


Fig. 1. Multiple sequence alignment (MSA) of β-CA protein sequences from *D. melanogaster* (UniProt ID: Q9VHJ5), *C. elegans* (UniProt ID: Q22460) and *T. spiralis* (UniProt ID: E5SH53). The first (CxDxR) and second (HxxC) highly conserved residues are shown by arrows and marked with green colour.

results are shown in Fig. 3. Most primers produced multiple bands. When the most promising primers (No. 6 for *T. pseudospiralis* and No. 10 for *T. nativa*) were tested against all three species of *Trichinella*, they completely cross-amplified because of the high sequence similarity (data not shown). Due to the interspecies cross-amplification we further designed four new reverse primers which were used together with the previous forward primers resulting in primer set Nos. 12–16.

The results revealed that the primer set No. 13 from *T. nativa* and primer set No. 12 from *T. pseudospiralis* produced the strongest DNA bands on the agarose gel electrophoresis (Fig. 4). For evaluation of cross-amplification these primers were tested on all three *Trichinella* spp. Figure 5 shows that there was again a complete cross-amplification between different *Trichinella* spp. by these primers. This finding led us to conclude that finding of species-specific primers for *Trichinella β-CA* genes is probably impossible due to the high sequence similarity. Based on our results the primer set No. 2 was considered the most potential tool for diagnostic purposes, even though it was not able to discriminate between different *Trichinella* spp. Therefore, its specificity was further tested on prepared samples from *Toxoplasma gondii*, *Toxocara cati*, and *Parascaris equorum*. No false-positive reactions were detectable (Fig. 6).

Discussion

Many protozoan, arthropod, and nematode species (invertebrate metazoans) contain *β-CA* gene and protein sequences in their genomes and proteomes respectively. BLAST homology analysis of predefined β-CA protein sequences from *D. melanogaster* (Uniprot ID: Q9VHJ5) and *C. elegans* (Uniprot ID: Q22460) resulted in identification of more than 75 β-CA protein sequences from protozoans and invertebrate metazoans (Zolfaghari Emameh *et al.*, 2014). One of them was β-CA protein sequence from *T. spiralis* (UniProt ID: E5SH53). Our previous results also revealed that β-CA proteins show high

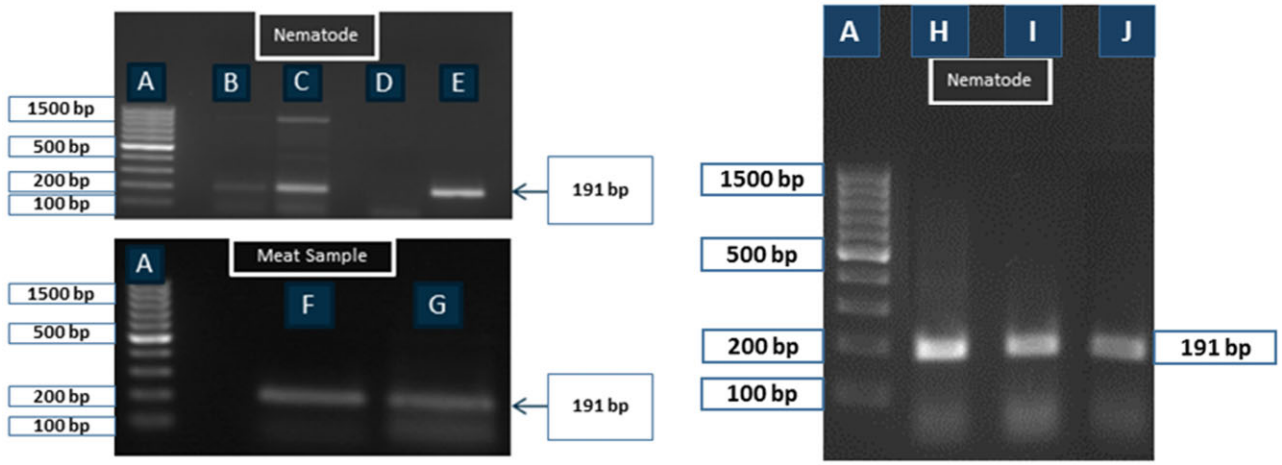


Fig. 2. PCR-based detection of genomic β-CA sequence from *Trichinella* larvae and infected meat samples. The primer set No. 2 showed a clear PCR product (size = 191 kb) in all *T. spiralis*-positive samples. (A) 100 bp ladder; (B) one larva; (C) five larvae; (D) negative control (PCR reaction mixture without digested sample); (E) more than 10 larvae; (F and G) 5 mg of *T. spiralis* infected mouse meat samples; (H) *T. nativa* (10 larvae); (I) *T. pseudospiralis* (10 larvae); (J) *T. spiralis* (10 larvae).

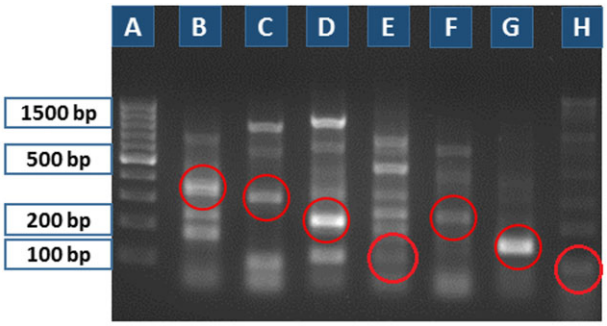


Fig. 3. Second round of PCR-based diagnosis tested on *T. nativa* and *T. pseudospiralis* samples (10 larvae/each lane). Primer set Nos. 5–7 for *T. pseudospiralis* and primer set Nos. 8–11 for *T. nativa*. (A) 100 bp ladder; (B) primer set No. 8 (372 bp), (C) primer set No. 9 (317 bp), (D) primer set No. 10 (226 bp), (E) primer set No. 11 (116 bp) for *T. nativa*; and (F) primer set No. 5 (251 bp), (G) primer set No. 6 (158 bp), (H) primer set No. 7 (103 bp) for *T. pseudospiralis*. The relevant PCR products are shown in red circles.

similarity among various *Drosophila* species and other species of dipteran flies (Zolfaghari Emameh *et al*., 2014). This led us to hypothesize that the high similarity is extensible to other species of one defined genus, such as *Trichinella* spp. including the clinically most relevant species: *T. spiralis*, *T. pseudospiralis*, and *T. nativa*. In principle, the high sequence homology at gene level could provide a straightforward avenue to develop a genus-specific PCR assay for diagnostics of *Trichinella* infection (trichinellosis). On the other hand, the high sequence similarity could be a difficult challenge for development of a species-specific assay based on the *β-CA* gene.

A number of PCR-based diagnostic methods have been developed for detection of infectious diseases, and some of them have been applied for parasitic diseases, such as trichinellosis (Li *et al*., 2010). In our study, we designed

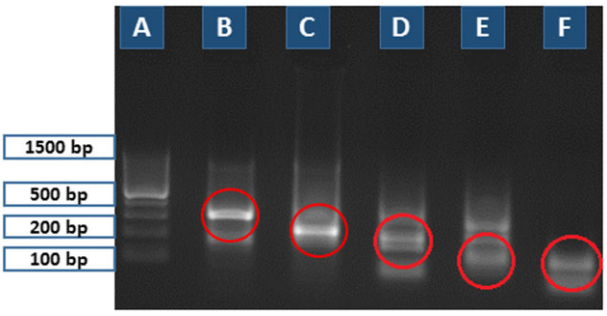


Fig. 4. Third round of PCR tested on *T. nativa* and *T. pseudospiralis* samples (10 larvae/each lane). Primer set No. 12 for *T. pseudospiralis* and primer set Nos. 13–16 for *T. nativa*. (A) 100 bp ladder; (B) primer set No. 13 (319 bp), (C) primer set No. 14 (171 bp), (D) primer set No. 15 (168 bp), (E) primer set No. 16 (129 bp) for *T. nativa*, and (F) primer set No. 12 (116 bp) for *T. pseudospiralis*. The relevant PCR products are shown in red circles.

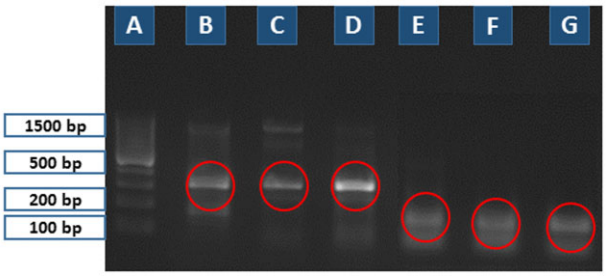


Fig. 5. Evaluation of possible cross-amplification between the primers designed for *T. nativa* and *T. pseudospiralis* β-CA genes and templates from *T. spiralis*, *T. nativa*, and *T. pseudospiralis* (10 larvae/each lane). (A) 100 bp ladder; tested primer set No. 13 (product size 319 bp) on *T. nativa* (B), *T. pseudospiralis* (C), and *T. spiralis* (D); and primer set No. 12 (product size 116 bp) on *T. nativa* (E), *T. pseudospiralis* (F), and *T. spiralis* (G).

and tested 16 different pairs of primers for detection of β-CA genomic sequences of *Trichinella* spp. At the beginning of assay development, it is necessary to test the efficiency of designed primers. Our method using the primer set No. 2 seemed to detect minute amounts of DNA (a single larva/reaction). Thereafter, the set No. 2 was considered the most promising primer pair for the development of genus-specific tool for *Trichinella* diagnostics. The results also revealed that interspecies cross-amplification is a recurrent challenge in this assay system when the DNA sequences are highly similar. None of the tested primers were able to discriminate between different *Trichinella* spp. As already pointed out, the described method was no able to discriminate between various species of *Trichinella*. This same problem hampers most detection methods reported to date. All species and genotypes of the genus *Trichinella* are morphologically indistinguishable at all developmental stages. Consequently, only biochemical or molecular methods can be used to identify the genotype of the parasite (Gajadhar *et al*., 2009).

Our results showed that the developed assay is robust enough for screening of *Trichinella* infection in routine meat samples. The positive signals were genus-specific since no false-positive bands were detectable when samples from other parasites, a protozoan *T. gondii* and two nematodes *T. cati* and *P. equorum*, were used as templates. Importantly, no *β-CA* genes are present in mammalian genomes (Syrjanen *et al*., 2010; Zolfaghari Emameh *et al*., 2014). Therefore, host-derived false-positive reactions would be unlikely. Other advantages of the present method include both ease of use and rapidity of detection. The meat samples are lysed overnight, and thereafter, the sample is ready for PCR amplification. No separate DNA isolation procedure is needed for the preparation of the template. The PCR reaction and gel electrophoresis take about 3 h after which the positive bands are clearly visible for documentation.

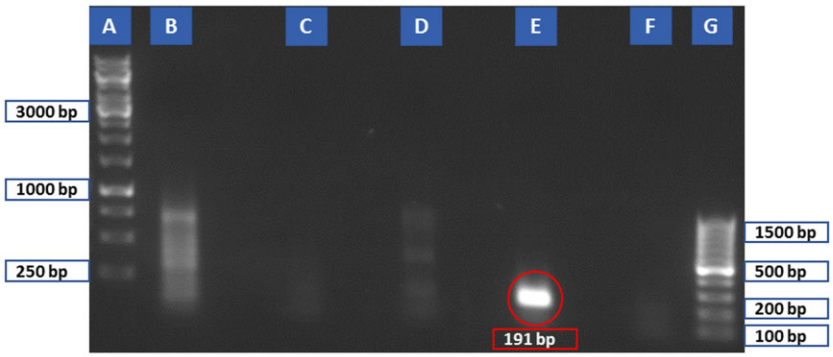


Fig. 6. Evaluation of primer set No. 2 on different lysed parasites. (A) 1 kb ladder, (B) *T. gondii*, (C) *T. cati*, (D) *P. equorum*, (E) *T. spiralis* as the positive control, (F) negative control, and (G) 100 bp ladder. Red colour circle shows 191 bp PCR product of primer set No. 2 on *T. spiralis*.

As an overall conclusion, even though there are many diagnostic methods available for trichinellosis, most of them are either expensive, time-consuming or require special expertise. Therefore, new simple tools are urgently needed for screening of biological samples in livestock slaughterhouses and quality controlling laboratories of meat processing companies. Our developed method is based on detection of β-CA genomic sequences from *Trichinella* spp. The greatest advantage of this method is that mammalian hosts do not contain any *β-CA* gene sequences in their genomes. The results indicated that the PCR method is *Trichinella* genus-specific and robust enough for routine screening purposes.

Experimental procedures

Identification of β-CA protein sequence from T. spiralis

β-CA protein sequence identification number (ID: E5SH53) of *Trichinella spiralis* in UniProt database (http://www.uniprot.org/) was retrieved by BLAST homology analysis using the previously defined β-CA protein sequences from *D. melanogaster* (Uniprot ID: Q9VHJ5) and *C. elegans* (Uniprot ID: Q22460) (Pearson, 2014; Zolfaghari Emameh *et al*., 2014). The corresponding genomic sequence ID of *T. spiralis* (EFV55868) was obtained from Ensembl Metazoa (http://metazoa.ensembl.org/index.html). Moreover, MSA by Clustal Omega algorithm (Sievers *et al*., 2011) within EMBL-EBI database (http://www.ebi.ac.uk/Tools/msa/clustalo/) was applied to approve the blast homology result. MSA revealed highly conserved sequences within β-CA protein sequences from *T. spiralis*, *D. melanogaster*, and *C. elegans*.

Preparation of parasites and infected meat samples

Parasite samples and *Trichinella* spp.-infected mouse muscle specimens were prepared at the Department of Veterinary Biosciences, University of Helsinki, Finland. The Animal Experiment Board in Finland had approved the study protocol. The tested *Trichinella* spp. involved *T. spiralis*, *T. pseudospiralis* and *T. nativa*. The blinded samples, including both infected and uninfected meat, were packed and shipped to Tissue Biology group, School of Medicine, University of Tampere, Finland, for PCR analysis. After arriving in the laboratory, the samples were transferred to −20°C freezer. To test the specificity of different primers, three more parasite samples were analysed including *T. gondii* (in culture medium), *T. cati* and *P. equorum* (in 70% ethanol).

Lysis of the parasites and meat samples

The meat and larvae samples were handled under the laminar hood. *Toxoplasma gondii* parasites were separated from the culture medium by centrifugation (Heraeus Biofuge Fresco, Thermo Scientific, Waltham, MA) at 13 000 RPM for 10 min (Leroux *et al*., 2015). Then the following lysis protocol was performed on all the samples: 300 µl of lysis buffer (DirectPCR Tail Lysis reagent, Viagen Biotech, LA) was added to 5 mg of samples. Then 6 µl of proteinase K (Thermo Scientific) was added to the mixture. The lysis tubes were incubated in rotating incubator (HYBAID, Thermo Scientific) at 55°C overnight.

Primer design

Three sets of primers for genomic sequence of β-CA (exonic and intronic) from *T. spiralis* (Ensembl Metazoa ID: EFV55868) were designed automatically by NCBI Primer-BLAST tool (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) (Ye *et al*., 2012), and one additional primer set was designed manually (Table 1). The expected sizes of the PCR products were calculated by Oligo Calc (http://www.basic.northwestern.edu/biotools/oligocalc.html) (Kibbe, 2007). To obtain genomic sequences for β-CA of *T. pseudospiralis* and *T. nativa* we first performed PCR amplification using the designed *T. spiralis* primers, and *T. pseudospiralis* and *T. nativa* larvae were used as templates. The corresponding bands for *T. pseudospiralis* and *T. nativa* on the agarose gel were cut and sequenced as described in the sequencing section. The obtained β-CA genomic sequences from *T. pseudospiralis* and *T. nativa* were aligned by Clustal Omega algorithm (http://www.ebi.ac.uk/Tools/msa/clustalo/) (Sievers *et al*., 2011) within EMBL-EBI database (http://www.ebi.ac.uk/Tools/msa/clustalo/) against the genomic sequence of β-CA from *T. spiralis*, and the most non-identical regions were selected for the second round of primer design. We designed 3 and 4 different forward primers manually for *T. pseudospiralis* (primer set Nos. 5–7) and *T. nativa* (primer set Nos. 8–11), respectively, and used with the same reverse primer No. 2

Table 1. Designed primers for *β-CA* genomic sequences from *T. spiralis*, *T. pseudospiralis* and *T. nativa*.

No.	Designed primers	Species name	Product length (bp)
1	**Forward:** 5′-AGAGACTGCCCGTTCACAAG-3′ **Reverse:** 3′-CTGGGAGAGTTTGTCAGCGT-5′	*T. spiralis*	882
2	**Forward:** 5′-TTTGAGCGCACTAGCATCCA-3′ **Reverse:** 3′-TCCATTCTGCATCACGCTGT-5′	*T. spiralis*	191
3	**Forward:** 5′-AGACTGCCCGTTCACAAGTT-3′ **Reverse:** 3′-TGGATGCTAGTGCGCTCAAA-5′	*T. spiralis*	530
4	**Forward:** 5′-ATTAATAAGTGAAAAGCACA-3′ **Reverse:** 3′-GTTTTAGAACTGGACACTGT-5′	*T. spiralis*	83
5	**Forward:** 5′-CTCACCCATCACCCCGGCTT-3′ **Reverse:** 3′-TCCATTCTGCATCACGCTGT-5′	*T. pseudospiralis*	251
6	**Forward:** 5′-GGCCAGGCCGTTCATCTGGT-3′ **Reverse:** 3′-TCCATTCTGCATCACGCTGT-5′	*T. pseudospiralis*	158
7	**Forward:** 5′-ATCGCGCTCTCGCGATTGGG-3′ **Reverse:** 3′-TCCATTCTGCATCACGCTGT-5′	*T. pseudospiralis*	103
8	**Forward:** 5′-TCCCAGACCAGCGGNAGCAC-3′ **Reverse:** 3′-TCCATTCTGCATCACGCTGT-5′	*T. nativa*	372
9	**Forward:** 5′-GACCCAGCGCGCTTTCGTTG-3′ **Reverse:** 3′-TCCATTCTGCATCACGCTGT-5′	*T. nativa*	317
10	**Forward:** 5′-CGGATACCACGGGCCGATGT-3′ **Reverse:** 3′-TCCATTCTGCATCACGCTGT-5′	*T. nativa*	226
11	**Forward:** 5′-AGTCGCCCAGCTTGATCGCG-3′ **Reverse:** 3′-TCCATTCTGCATCACGCTGT-5′	*T. nativa*	116
12	**Forward:** 5′-ATCGCGCTCTCGCGATTGGG-3′ **Reverse:** 3′-CAACCGATACCGAACGGACC-5′	*T. pseudospiralis*	116
13	**Forward:** 5′-TCCCAGACCAGCGGNAGCAC-3′ **Reverse:** 3′-CGGACCGAACTCTGGTACAG-5′	*T. nativa*	319
14	**Forward:** 5′-CGGATACCACGGGCCGATGT-3′ **Reverse:** 3′-CGGACCGAACTCTGGTACAG-5′	*T. nativa*	171
15	**Forward:** 5′-TCCCAGACCAGCGGNAGCAC-3′ **Reverse:** 3′-ACATCGGCCCGTGGTATCCG-5′	*T. nativa*	168
16	**Forward:** 5′-TCCCAGACCAGCGGNAGCAC-3′ **Reverse:** 3′-GTGAGTCCAGCAGCAACCCG-5′	*T. nativa*	129

which was described for the first round. In the third round of primer design, one new reverse primer was designed for *T. pseudospiralis* (primer set No. 12) and used with its own forward primer No. 7. Also, four new reverse primers were designed for *T. nativa* and tested with its own forward primers Nos. 8 and 10 (primer set Nos. 13–16). Three other species, *T. gondii*, *T. cati*, and *P. equorum* were tested for any false-positive cross-amplifications with other parasites.

PCR

PCR was performed based on the following reaction mixture: 12.5 μl of 2X KAPA ReadyMix (KAPA 2G Robust HotStart ReadyMix PCR Kit, Kapa Biosystems, Wilmington, MA), 1.25 μl of forward primer (Oligomer, Helsinki, Finland), 1.25 μl of reverse primer (Oligomer), 9 μl of dH_2O and 1 μl of lysed samples. Lysed *T. spiralis* larvae served as a positive control and the negative control was the PCR master mix without any lysed sample. The PCR assay was run on the thermocycler (PTC100, MJ Research Inc, Waltham, MA) according to the following details: 95°C (3 min), [95°C (10 s), 53°C (10 s), 72°C (10 s)] × 36 cycles, 72°C (5 min).

Sequencing of β-CA genomic sequences from T. pseudospiralis *and* T. nativa

We cut the PCR products of *T. pseudospiralis* and *T. nativa* templates from an agarose gel. Illustra-GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, Buckinghamshire, UK) was applied for extraction and purification of the corresponding DNA bands. At the final step, the overnight-dried samples were suspended in HiDi (Life Technologies Europe, Applied Biosystems, Finland), which was followed by vortexing and spinning. The samples were heated at 95°C for 2 min and transferred to ice for 2 min. The samples were sent to Core Facilities and Research Services, BioMediTech, Tampere, Finland for DNA sequencing. The sequencing was carried out with HITACHI 3130x/Genetic Analyzer (Life Technologies Europe, Applied Biosystems).

Authors' contributions

RZE carried out blast homology analysis and MSA on *T. spiralis*, *T. pseudospiralis*, *T. nativa*, *C. elegans*, and *D. melanogaster* β-CA sequences. RZE and MK participated in the primers design and setting up the PCR method. AN and AS prepared the parasite and infected meat samples. All authors participated in the design of the study. RZE and MK drafted the first version of the paper. All authors read and approved the final manuscript.

Conflict of Interest

None declared.

References

Aguilera, J., Van Dijken, J.P., De Winde, J.H., and Pronk, J.T. (2005) Carbonic anhydrase (Nce103p): an essential biosynthetic enzyme for growth of *Saccharomyces cerevisiae* at atmospheric carbon dioxide pressure. *Biochem J* **391:** 311–316.

Airas, N., Saari, S., Mikkonen, T., Virtala, A.M., Pellikka, J., Oksanen, A., *et al*. (2010) Sylvatic *Trichinella* spp. infection in Finland. *J Parasitol* **96:** 67–76.

Alban, L., Pozio, E., Boes, J., Boireau, P., Boue, F., Claes, M., *et al*. (2011) Towards a standardised surveillance for *Trichinella* in the European Union. *Prev Vet Med* **99:** 148–160.

Alterio, V., Vitale, R.M., Monti, S.M., Pedone, C., Scozzafava, A., Cecchi, A., *et al*. (2006) 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* **128:** 8329–8335.

Ancelle, T. (1998) History of trichinellosis outbreaks linked to horse meat consumption 1975–1998. *Euro Surveill* **3:** 86–89.

Blaga, R., Fu, B., Le Rhun, D., Le Naour, E., Heckman, A., Zocevic, A., *et al*. (2009) Use of mitochondrial RNA genes for the differentiation of four *Trichinella* species by multiplex PCR amplification. *J Helminthol* **83:** 121–128.

Borsuk, P., Moskwa, B., Pastusiak, K., and Cabaj, W. (2003) Molecular identification of *Trichinella spiralis* and *Trichinella britovi* by diagnostic multiprimer large mitochondrial rRNA amplification. *Parasitol Res* **91:** 374–377.

Clausen, M.R., Meyer, C.N., Krantz, T., Moser, C., Gomme, G., Kayser, L., *et al*. (1996) *Trichinella* infection and clinical disease. *QJM* **89:** 631–636.

Del Prete, S., Vullo, D., Fisher, G.M., Andrews, K.T., Poulsen, S.A., Capasso, C., and Supuran, C.T. (2014) Discovery of a new family of carbonic anhydrases in the malaria pathogen *Plasmodium falciparum* – the eta-carbonic anhydrases. *Bioorg Med Chem Lett* **24:** 4389–4396.

Duplessis, M., and Moineau, S. (2001) Identification of a genetic determinant responsible for host specificity in *Streptococcus thermophilus* bacteriophages. *Mol Microbiol* **41:** 325–336.

Fasseas, M.K., Tsikou, D., Flemetakis, E., and Katinakis, P. (2010) Molecular and biochemical analysis of the beta class carbonic anhydrases in *Caenorhabditis elegans*. *Mol Biol Rep* **37:** 2941–2950.

Fawcett, T.W., Browse, J.A., Volokita, M., and Bartlett, S.G. (1990) Spinach carbonic anhydrase primary structure deduced from the sequence of a cDNA clone. *J Biol Chem* **265:** 5414–5417.

Ferry, J.G. (2010) The gamma class of carbonic anhydrases. *Biochim Biophys Acta* **1804:** 374–381.

Froscher, W., Gullotta, F., Saathoff, M., and Tackmann, W. (1988) Chronic trichinosis. Clinical, bioptic, serological and electromyographic observations. *Eur Neurol* **28:** 221–226.

Funke, R.P., Kovar, J.L., and Weeks, D.P. (1997) Intracellular carbonic anhydrase is essential to photosynthesis in *Chlamydomonas reinhardtii* at atmospheric levels of CO2. Demonstration via genomic complementation of the high-CO2-requiring mutant ca-1. *Plant Physiol* **114:** 237–244.

Gajadhar, A.A., Pozio, E., Gamble, H.R., Nockler, K., Maddox-Hyttel, C., Forbes, L.B., *et al*. (2009) Trichinella diagnostics and control: mandatory and best practices for ensuring food safety. *Vet Parasitol* **159:** 197–205.

Gamble, H.R., Bessonov, A.S., Cuperlovic, K., Gajadhar, A.A., van Knapen, F., Noeckler, K., *et al*. (2000) International Commission on Trichinellosis: recommendations on methods for the control of *Trichinella* in domestic and wild animals intended for human consumption. *Vet Parasitol* **93:** 393–408.

Gottstein, B., Pozio, E., and Nockler, K. (2009) Epidemiology, diagnosis, treatment, and control of trichinellosis. *Clin Microbiol Rev* **22:** 127–145. Table of Contents.

Guenther, S., Nockler, K., von Nickisch-Rosenegk, M., Landgraf, M., Ewers, C., Wieler, L.H., and Schierack, P. (2008) Detection of *Trichinella spiralis*, *T. britovi* and *T. pseudospiralis* in muscle tissue with real-time PCR. *J Microbiol Methods* **75:** 287–292.

Kibbe, W.A. (2007) OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res* **35:** W43–W46.

Korinkova, K., Kovarcik, K., Pavlickova, Z., Svoboda, M., and Koudela, B. (2008) Serological detection of *Trichinella spiralis* in swine by ELISA (enzyme-linked immunosorbent assay) using an excretory-secretory (E/S) antigen. *Parasitol Res* **102:** 1317–1320.

Krivokapich, S.J., Pozio, E., Gatti, G.M., Prous, C.L., Ribicich, M., Marucci, G., *et al*. (2012) *Trichinella patagoniensis* n. sp. (Nematoda), a new encapsulated species infecting carnivorous mammals in South America. *Int J Parasitol* **42:** 903–910.

Kumar, V., Pozio, E., de Borchgrave, J., Mortelmans, J., and De Meurichy, W. (1990) Characterization of a *Trichinella* isolate from polar bear. *Ann Soc Belg Med Trop* **70:** 131–135.

La Rosa, G., Marucci, G., Zarlenga, D.S., and Pozio, E. (2001) *Trichinella pseudospiralis* populations of the Palearctic region and their relationship with populations of the Nearctic and Australian regions. *Int J Parasitol* **31:** 297–305.

Lane, T.W., Saito, M.A., George, G.N., Pickering, I.J., Prince, R.C., and Morel, F.M. (2005) Biochemistry: a cadmium enzyme from a marine diatom. *Nature* **435:** 42.

Leroux, L.P., Dasanayake, D., Rommereim, L.M., Fox, B.A., Bzik, D.J., Jardim, A., and Dzierszinski, F.S. (2015) Secreted *Toxoplasma gondii* molecules interfere with expression of MHC-II in interferon gamma-activated macrophages. *Int J Parasitol* **45:** 319–332.

Li, F., Wang, Z.Q., and Cui, J. (2010) Early detection by polymerase chain reaction of migratory *Trichinella spiralis* larvae in blood of experimentally infected mice. *Foodborne Pathog Dis* **7:** 887–892.

Li, L.Z., Wang, Z.Q., Jiang, P., Zhang, X., Ren, H.J., and Cui, J. (2011) Molecular identification of a *Trichinella* isolate from a naturally infected pig in Tibet, China. *Korean J Parasitol* **49:** 381–384.

Lin, Z., Cao, J., Zhang, H., Zhou, Y., Deng, M., Li, G., and Zhou, J. (2013) Comparison of three molecular detection methods for detection of *Trichinella* in infected pigs. *Parasitol Res* **112:** 2087–2093.

Marucci, G., La Grange, L.J., La Rosa, G., and Pozio, E. (2009) *Trichinella nelsoni* and *Trichinella* T8 mixed infection

in a lion (*Panthera leo*) of the Kruger National Park (South Africa). *Vet Parasitol* **159:** 225–228.

Merlin, C., Masters, M., McAteer, S., and Coulson, A. (2003) Why is carbonic anhydrase essential to *Escherichia coli*? *J Bacteriol* **185:** 6415–6424.

Mitsuhashi, S., Ohnishi, J., Hayashi, M., and Ikeda, M. (2004) A gene homologous to beta-type carbonic anhydrase is essential for the growth of *Corynebacterium glutamicum* under atmospheric conditions. *Appl Microbiol Biotechnol* **63:** 592–601.

Nelson, M., Wright, T.L., Pierce, A., and Krogwold, R.A. (2003) A common-source outbreak of trichinosis from consumption of bear meat. *J Environ Health* **65:** 16–19, 24.

Nishimori, I., Minakuchi, T., Onishi, S., Vullo, D., Scozzafava, A., and Supuran, C.T. (2007) Carbonic anhydrase inhibitors. DNA cloning, characterization, and inhibition studies of the human secretory isoform VI, a new target for sulfonamide and sulfamate inhibitors. *J Med Chem* **50:** 381–388.

Nockler, K., Pozio, E., Voigt, W.P., and Heidrich, J. (2000) Detection of *Trichinella* infection in food animals. *Vet Parasitol* **93:** 335–350.

Pearson, W.R. (2014) BLAST and FASTA similarity searching for multiple sequence alignment. *Methods Mol Biol* **1079:** 75–101.

Pozio, E. (2007) World distribution of *Trichinella* spp. infections in animals and humans. *Vet Parasitol* **149:** 3–21.

Pozio, E., and Darwin Murrell, K. (2006) Systematics and epidemiology of *Trichinella*. *Adv Parasitol* **63:** 367–439.

Pozio, E., Foggin, C.M., Marucci, G., La Rosa, G., Sacchi, L., Corona, S., *et al*. (2002) *Trichinella zimbabwensis* n.sp. (Nematoda), a new non-encapsulated species from crocodiles (*Crocodylus niloticus*) in Zimbabwe also infecting mammals. *Int J Parasitol* **32:** 1787–1799.

Pozio, E., Owen, I.L., Marucci, G., and La Rosa, G. (2005) Inappropriate feeding practice favors the transmission of *Trichinella papuae* from wild pigs to saltwater crocodiles in Papua New Guinea. *Vet Parasitol* **127:** 245–251.

Rombout, Y.B., Bosch, S., and Van Der Giessen, J.W. (2001) Detection and identification of eight *Trichinella* genotypes by reverse line blot hybridization. *J Clin Microbiol* **39:** 642–646.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., *et al*. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7:** 539.

Smith, K.S., Cosper, N.J., Stalhandske, C., Scott, R.A., and Ferry, J.G. (2000) Structural and kinetic characterization of an archaeal beta-class carbonic anhydrase. *J Bacteriol* **182:** 6605–6613.

Supuran, C.T. (2008) Carbonic anhydrases: novel therapeutic applications for inhibitors and activators. *Nat Rev Drug Discov* **7:** 168–181.

Syrjanen, L., Tolvanen, M., Hilvo, M., Olatubosun, A., Innocenti, A., Scozzafava, A., *et al*. (2010) 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* **11:** 28.

Tantrawatpan, C., Intapan, P.M., Thanchomnang, T., Lulitanond, V., Boonmars, T., Wu, Z., *et al*. (2012) Differential detection of *Trichinella papuae*, *T. spiralis* and *T. pseudospiralis* by real-time fluorescence resonance energy transfer PCR and melting curve analysis. *Vet Parasitol* **185:** 210–215.

Van De, N., Trung, N.V., Ha, N.H., Nga, V.T., Ha, N.M., Thuy, P.T., *et al*. (2012) An outbreak of trichinosis with molecular identification of *Trichinella sp*. in Vietnam. *Korean J Parasitol* **50:** 339–343.

Vullo, D., Franchi, M., Gallori, E., Antel, J., Scozzafava, A., and Supuran, C.T. (2004) Carbonic anhydrase inhibitors. Inhibition of mitochondrial isozyme V with aromatic and heterocyclic sulfonamides. *J Med Chem* **47:** 1272–1279.

Vullo, D., Innocenti, A., Nishimori, I., Pastorek, J., Scozzafava, A., Pastorekova, S., and Supuran, C.T. (2005) Carbonic anhydrase inhibitors. Inhibition of the transmembrane isozyme XII with sulfonamides – a new target for the design of antitumor and antiglaucoma drugs? *Bioorg Med Chem Lett* **15:** 963–969.

Wang, L., Cui, J., Hu, D.D., Liu, R.D., and Wang, Z.Q. (2014) Identification of early diagnostic antigens from major excretory-secretory proteins of *Trichinella spiralis* muscle larvae using immunoproteomics. *Parasit Vectors* **7:** 40.

Xu, Y., Feng, L., Jeffrey, P.D., Shi, Y., and Morel, F.M. (2008) Structure and metal exchange in the cadmium carbonic anhydrase of marine diatoms. *Nature* **452:** 56–61.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13:** 134.

Zarlenga, D.S., Chute, M.B., Martin, A., and Kapel, C.M. (1999) A multiplex PCR for unequivocal differentiation of all encapsulated and non-encapsulated genotypes of *Trichinella*. *Int J Parasitol* **29:** 1859–1867.

Zarlenga, D.S., Rosenthal, B.M., La Rosa, G., Pozio, E., and Hoberg, E.P. (2006) Post-Miocene expansion, colonization, and host switching drove speciation among extant nematodes of the archaic genus *Trichinella*. *Proc Natl Acad Sci USA* **103:** 7354–7359.

Zhang, Y., Wang, Z., Li, L., and Cui, J. (2013) Molecular characterization of *Trichinella spiralis* aminopeptidase and its potential as a novel vaccine candidate antigen against trichinellosis in BALB/c mice. *Parasit Vectors* **6:** 246.

Zolfaghari Emameh, R., Barker, H., Tolvanen, M.E., Ortutay, C., and Parkkila, S. (2014) Bioinformatic analysis of beta carbonic anhydrase sequences from protozoans and metazoans. *Parasit Vectors* **7:** 38.

Zolfaghari Emameh, R., Syrjanen, L., Barker, H., Supuran, C.T., and Parkkila, S. (2015) *Drosophila melanogaster*: a model organism for controlling Dipteran vectors and pests. *J Enzyme Inhib Med Chem* **30:** 505–513.

Zolfaghari Emameh et al. Parasites & Vectors (2016) 9:152
DOI 10.1186/s13071-016-1415-7

Parasites & Vectors

RESEARCH Open Access

Horizontal transfer of β-carbonic anhydrase genes from prokaryotes to protozoans, insects, and nematodes

Reza Zolfaghari Emameh[1,2,3*], Harlan R. Barker[1], Martti E. E. Tolvanen[4], Seppo Parkkila[1,3] and Vesa P. Hytönen[2,3]

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 (β-CA) plays a crucial role in metabolic functions of many living organisms, the origin of *β-CA* genes in eukaryotic species remains unclear.

Methods: This study was conducted using phylogenetics, prediction of subcellular localization, and identification of β-CA, transposase, integrase, and resolvase genes on the MGEs of bacteria. We also structurally analyzed β-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 *β-CA* gene sequences on mobile genetic elements (MGEs) which have facilitated the mobility of *β-CA* genes from bacteria to protozoan, insect, and nematode species. The prokaryotic origin of protozoan, insect, and nematode β-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 *β-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 λ 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
[1]School of Medicine, University of Tampere, Medisiinarinkatu 3, FI-33520 Tampere, Finland
[2]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 *α-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 α, β, γ, δ, ζ, and η [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 β-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 β-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 *β-CA* gene sequences in protozoans, insects, and nematodes by HGT from ancestral prokaryotes using phylogenetics, prediction of subcellular localization, and identification of β-CA, transposase, integrase, and resolvase genes on the MGEs of bacteria. We also structurally analyzed β-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 *β-CA* genes across multiple species living together in distinct environments.

Methods

Identification of *β-CA* gene and protein sequences

We collected all β-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 β-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 β-CA protein sequences from protozoans, insects, and nematodes in the EMBL-EBI BLAST database (http://www.ebi.ac.uk/Tools/sss/fasta/) to define bacterial β-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 transfered genes
Wolbachia	*Aedes aegypti* (yellow fever mosquito), *Anopheles gambiae* (malaria mosquito), and *Drosophila melanogaster*	Many prokaryotic genes, such as gag-pol, D34 immunodominant antigen, actin and aminotransferase genes [65, 66]
Escherichia coli	*Caenorhabditis elegans*	Antibiotic-resistance genes [67]
Prokaryotes	Anaerobic protozoans: *Trichomonas vaginalis*, *Entamoeba histolytica,* and *Naegleria gruberi*	Alcohol dehydrogenase (*adh* gene) and *Pyruvate:ferredoxin oxidoreductase* genes [1, 68]
Prokaryotes	*Dictyostelium discoideum* (soil-living amoeba)	18 prokaryotic genes [48]
Prokaryotes	Trypanosomatids: *Leishmania* spp., *Angomonas deanei* and *Strigomonas culicis*	Bacterial amino acid pathways [58]
α-proteobacteria	*Leishmania* spp.	Mitochondria (initiation point of apoptosis) [69]
β-proteobacteria and *γ-proteobacteria*	Trypanosomatids: *Leishmania* spp., *Angomonas deanei* and *Strigomonas culicis*	*Heme synthesis* gene [50]
Peptostreptococcus harei	*Trichomonas vaginalis*	Lateral gene transfer fragment (TvLF) [51]

protozoan, insect, and nematode β-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 β-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 β-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 β-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 β-CA protein sequences was performed using a subcellular signal prediction tool. Mitochondrial and secretory targeting peptides in β-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 β-CA protein sequences, which had a predicted common ancestor with protozoan, insect, or nematode β-CA protein sequences. Specifically, this included *Afipia felis* (K8NQ88), *Bradyrhizobium japonicum* (G7D846), *Cesiribacter andamanensis* (M7MX87), *Colwellia psychrerythraea* (Q47YG3), *Corallococcus coralloides* (H8MJ17), *Leptospira kirschneri* (M6X652), *Magnetospirillum magneticum* (Q2VZD0),

Table 2 Predicted sources of the β-CA genes. The tentative prokaryotic endosymbionts and their hosts are listed

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

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

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

Identification of β-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 *β-CA* gene sequences on protozoan, insect, and nematode genomic DNA

Analyses regarding determination of precise locations of protozoan, insect and nematode *β-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 *β-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 β-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* β-CA PDB 1I6P; Clade B: *Pisum sativum* β-CA PDB 1EKJ; Clade C: *Mycobacterium tuberculosis* β-CA PDB 1YM3; and Clade D: *Methanobacterium thermoautotrophicum* β-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 β-CA protein sequences from defined bacterial, protozoan, insect, and nematode species

Multiple sequence alignment (MSA) of β-CA protein sequences from protozoan, insect, nematode species with bacterial β-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 β-CA protein sequences revealed the common ancestor of protozoan, insect, and nematode β-CAs within bacterial β-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) β-CA proteins probably contain mitochondrial targeting peptides. Even three bacterial β-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 β-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 β-CA proteins (Additional file 2).

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

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

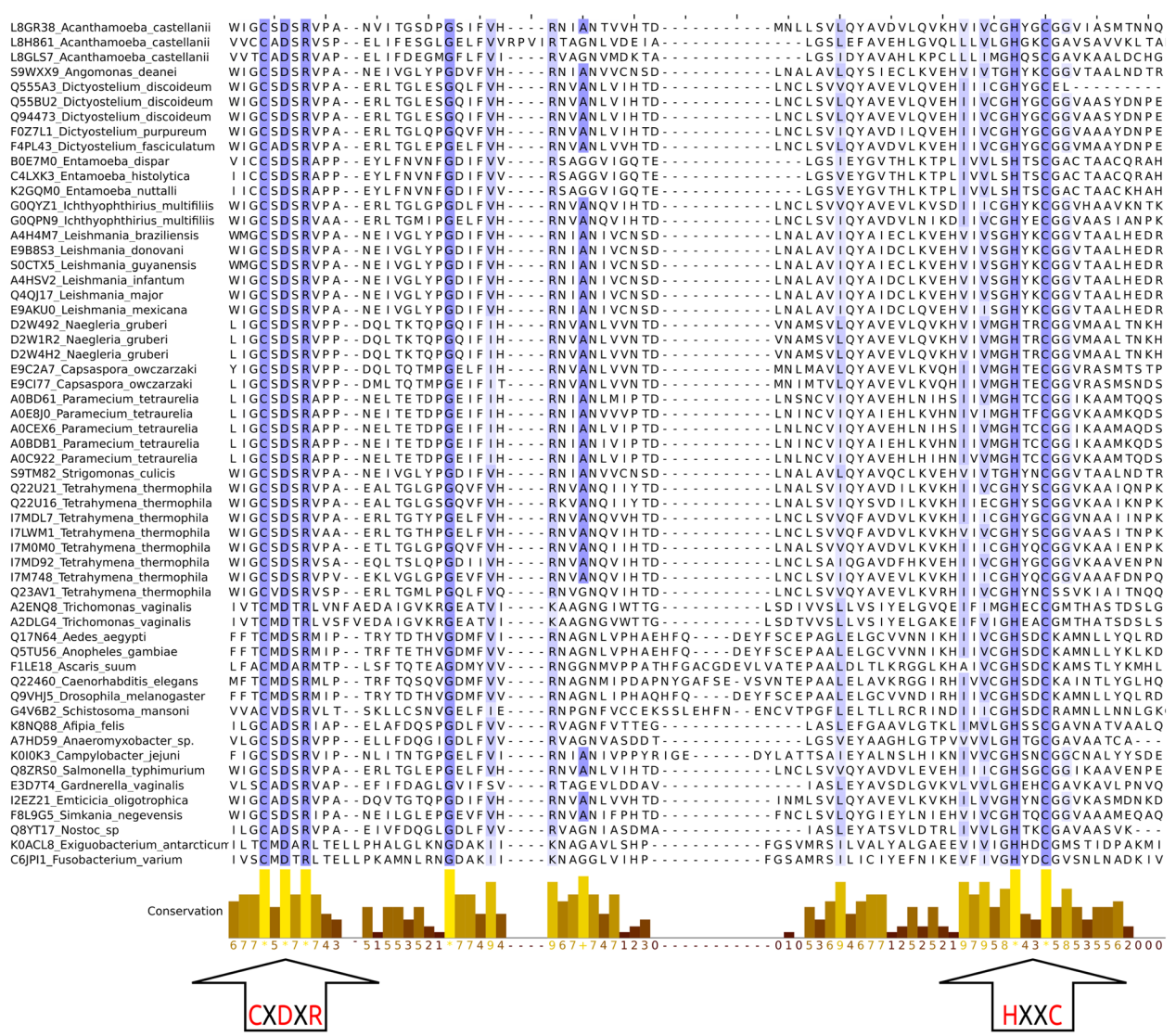


Fig. 1 Multiple sequence alignment (MSA) of 57 β-CA protein sequences. They include sequences (102 amino acid residues starting three amino acid residues prior to the first highly conserved sequence; CXDXR) from defined protozoan, insect, and nematode species, as well as ten β-CA protein sequences from bacterial endosymbionts of protozoans, insects, and nematodes, and *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). First (CXDXR) and second (HXXC) highly conserved motifs of β-CAs are shown with two black arrows at the bottom of the figure

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 *β-CA* genes located within MGEs in different prokaryotic species. Each bacterial MGE contained only one *β-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 β-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 β-CA is found in evolutionary conserved modules of MGEs, even at the most stringent significance thresholds. The locations of β-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 *β-CA* gene in this plasmid

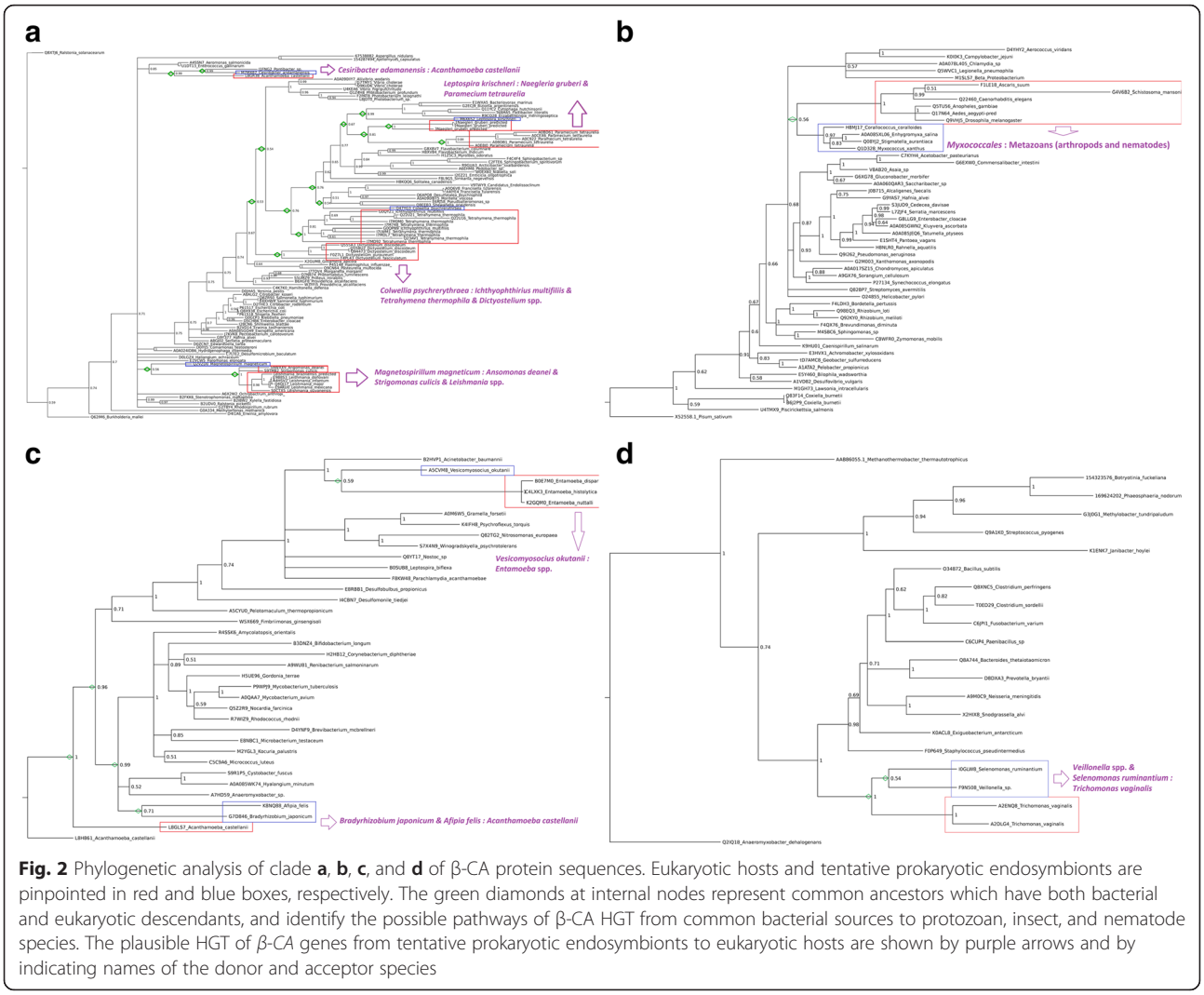


Fig. 2 Phylogenetic analysis of clade **a**, **b**, **c**, and **d** of β-CA protein sequences. Eukaryotic hosts and tentative prokaryotic endosymbionts are pinpointed in red and blue boxes, respectively. The green diamonds at internal nodes represent common ancestors which have both bacterial and eukaryotic descendants, and identify the possible pathways of β-CA HGT from common bacterial sources to protozoan, insect, and nematode species. The plausible HGT of *β-CA* genes from tentative prokaryotic endosymbionts to eukaryotic hosts are shown by purple arrows and by indicating names of the donor and acceptor species

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

Identification of *β-CA* gene sequences on protozoan, insect, and nematode genomic DNA

Analysis of the precise location of *β-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 *β-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 *β-CA* gene sequences included only one exon. The definitive locations of *β-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 Phylogentic Analysis

β-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

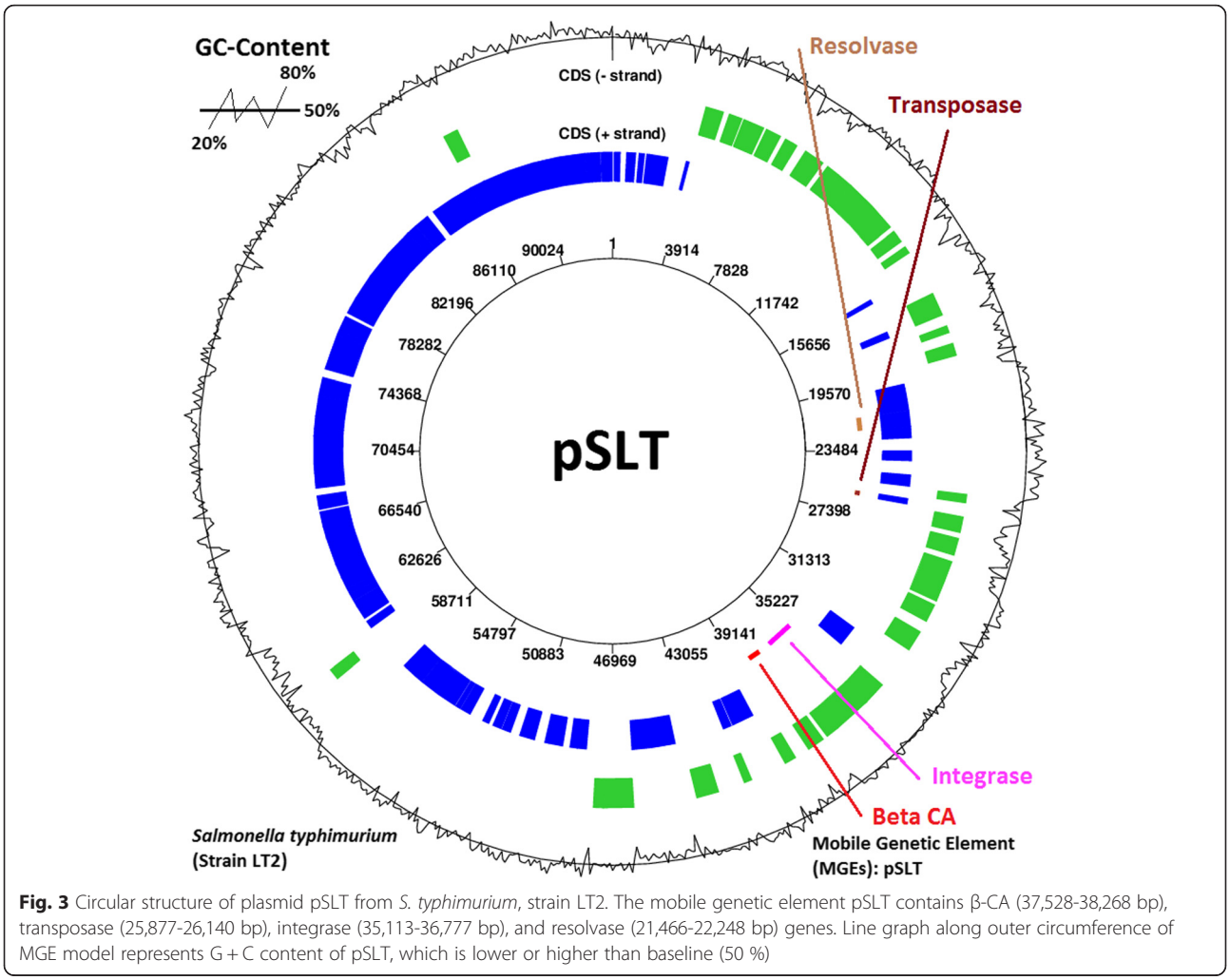


Fig. 3 Circular structure of plasmid pSLT from *S. typhimurium*, strain LT2. The mobile genetic element pSLT contains β-CA (37,528-38,268 bp), transposase (25,877-26,140 bp), integrase (35,113-36,777 bp), and resolvase (21,466-22,248 bp) genes. Line graph along outer circumference of MGE model represents G + C content of pSLT, which is lower or higher than baseline (50 %)

Analysis of the genes on circular mitochondrial DNA from *A. castellanii* revealed that none of the protozoan *β-CA*s 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 β-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 β-CA protein sequences with previously defined bacterial β-CA proteins, revealed that all of the evaluated sequences contained the first (CXDXR) and second (HXXC) highly conserved motifs characteristic of β-CA proteins. Phylogenetic analysis revealed that protozoan, insect, and nematode β-CA protein sequences are

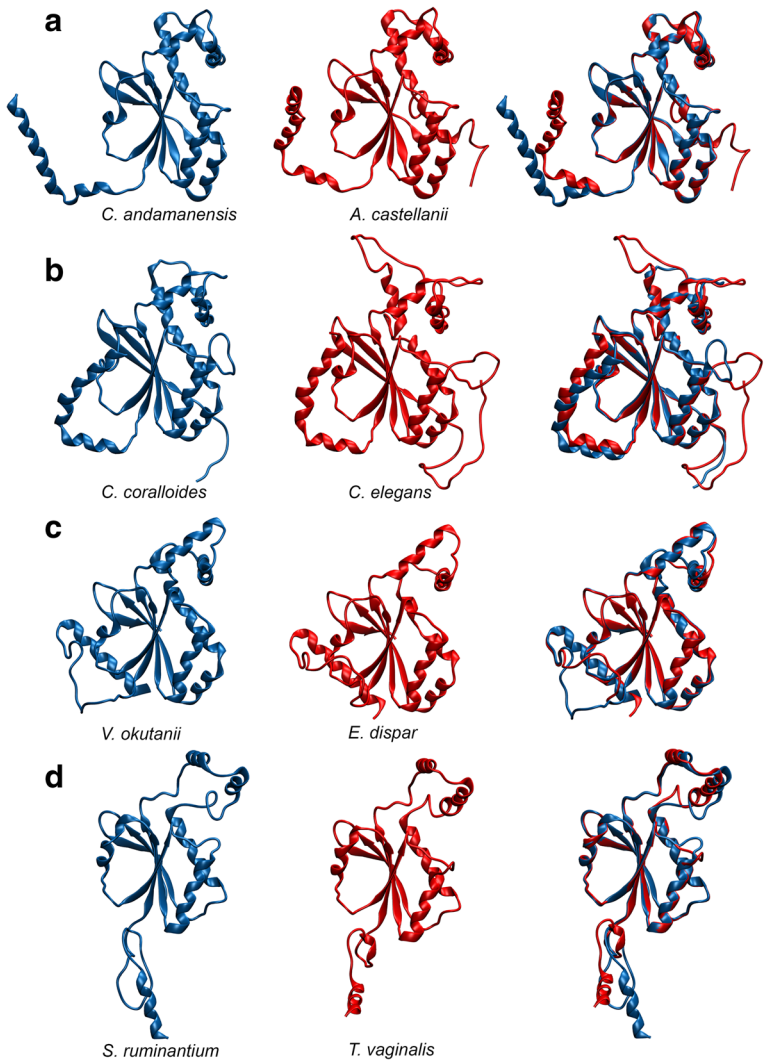


Fig. 4 Homology models of representative pairs of β-CAs from clades **a**, **b**, **c**, and **d**. The blue protein models correspond to prokaryotic proteins and the red models to eukaryotic proteins. The superimposed models were shown in the third column at the right side

mostly categorized as clade A or B β-CA protein structures, respectively.

Based on our phylogenetic analysis, *A. castellanii* possesses two *β-CA* genes, one from clade A and one from clade C. Our results in Fig. 2c, suggest that the *β-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 β-CAs (Fig. 2a), showed that all β-CAs in *N. gruberi* and *P. tetraurelia* protozoa have a common source with the single β-CA from spirochaetes bacteria, *L. kirschneri* (M6X652). Potentially, after HGT of a *β-CA* gene from the common source to these two protozoan hosts, the gene duplicated and created three different *β-CA* genes for *N. gruberi* (Predicted 1, 2, 3) and five for *P. tetraurelia* (A0BD61, A0CEX6, A0C922, A0BDB1, A0E8I0).

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 *β-CA* in *I. multifiliis* (G0QYZ1, G0QPN9), *T. thermophila* (Q22U21, Q22U16, I7M0M0, I7M748, I7LWM1, I7MDL7, Q23AV1, I7MD92), and *Dictyostelium* spp. (Q555A3, Q55BU2, 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 *β-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 β-CAs belong to clade B and suggests that they may have a common source with myxobacterial β-CAs. The various myxobacteria *Corallococcus, Enhygromyxa, Stigmatella,* and *Myxococcus*, are part of the same subtree that contains insect and nematode β-CAs. However, a larger analysis with more insect, nematode, and plant β-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 *β-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 phyologenetic trees do not provide conclusive evidence for this, and it is thus possible to speculate that the β-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 β-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 β-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 *β-CA* genes from *Entamoeba* spp. (B0E7M0, 1C4LXK3, K2GQM0) have a common source with the *β-CA* gene of gammaproteobacterium *V. okutanii* (A5CVM8). From this result, we propose that *β-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 *β-CA* genes in *T. vaginalis* (A2ENQ8, A2DLG4) have a common source with *β-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 *β-CA* gene sequence into the *T. vaginalis* genome is possible.

Prediction of subcellular signals of β-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 β-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, mitosoms, 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 α-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 β-CA proteins in prokaryotes, leading to their mitochondrial localization after HGT into protozoans and possibly insects. Supporting this idea, the β-CA of *D. melanogaster* has been experimentally shown to be localized in mitochondria [28, 29].

Identification of β-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 *β-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, β-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 *β-CA* gene sequences were located on a single chromosome, although the precise chromosomal location for some protozoan *β-CA* genes is still pending (Additional file 4).

In order to evaluate the structural features of the identified β-CA proteins, we first analyzed the functional roles of the conserved residues. β-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 β-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 β-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 β-CAs, but not in the β-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 *β-CA* gene sequences are completely absent in the genomes of vertebrates. In protozoan and invertebrate metazoans, including insect and nematode species, *β-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 *β-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 *β-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 *β-CA* gene sequence in the germ line, it may have then been removed by genetic assortment of the vertebrate hosts. Therefore, the lack of *β-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 α-CAs in all vertebrates.

Conclusions

Many prokaryotic MGEs contain necessary enzyme gene sequences, such as transposase, integrase, and resolvase, together with β-CA. These enzymes can facilitate HGT of *β-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 *β-CA* gene sequences from endosymbiont bacteria to protozoan, insect, and nematode hosts by MGEs. The phylogenetic analysis suggests that different protozoan *β-CA* genes have various common ancestors among prokaryotes, divided between clades A, C and D of β-CAs. In contrast, the case of insect and nematode *β-CA* genes is more complex. We propose that they may have had a single common ancestor from a bacterial *β-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 *β*-CA gene in eukaryotic species.

Additional files

Additional file 1: β-CA expressing prokaryotes and their endosymbiotic protozoan, insect, and nematodes hosts. (PDF 301 kb)

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

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

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

Additional file 5: Location of *β-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 *β-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 β-CA, transposase, integrase, resolvase, and conjugation complex protein genes from bacterial mobile genetic elements and genomic location of protozoan β-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 β-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
[1]School of Medicine, University of Tampere, Medisiinarinkatu 3, FI-33520 Tampere, Finland. [2]BioMediTech, University of Tampere, FI-33520 Tampere, Finland. [3]Fimlab Laboratories Ltd and Tampere University Hospital, FI-33520 Tampere, Finland. [4]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

1. Andersson JO. Lateral gene transfer in eukaryotes. Cell Mol Life Sci. 2005; 62(11):1182–97.
2. Keeling PJ, Palmer JD. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet. 2008;9(8):605–18.
3. 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.
4. 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.
5. 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.
6. Toussaint A, Merlin C. Mobile elements as a combination of functional modules. Plasmid. 2002;47(1):26–35.
7. Hou Q, He J, Yu J, Ye Y, Zhou D, Sun Y, et al. A case of horizontal gene transfer from to C6/36 cell line. Mob Genet Elements. 2014;4(1):e28914.
8. Wernegreen JJ. Endosymbiosis: lessons in conflict resolution. PLoS Biol. 2004;2(3):E68.
9. 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.
10. Greub G, Raoult D. Microorganisms resistant to free-living amoebae. Clin Microbiol Rev. 2004;17(2):413–33.
11. Skriwan 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.
12. Bertelli C, Greub G. Lateral gene exchanges shape the genomes of amoeba-resisting microorganisms. Front Cell Infect Microbiol. 2012;2:110.
13. 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.
14. 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.
15. Fujishima M, Kodama Y. Endosymbionts in Paramecium. Eur J Protistol. 2012;48(2):124–37.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.
21. 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.
22. Elleuche S, Poggeler S. Carbonic anhydrases in fungi. Microbiology. 2010; 156(Pt 1):23–9.
23. 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.
24. 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.
25. 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 isoform VI, a new target for sulfonamide and sulfamate inhibitors. J Med Chem. 2007;50(2):381–8.
26. 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.
27. 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 antiglaucoma drugs? Bioorg Med Chem Lett. 2005;15(4):963–9.
28. 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.
29. 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.
30. 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.
31. 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.
32. 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.
33. Smith KS, Jakubzick 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.
34. Sievers F, Higgins DG. Clustal omega. Curr Protoc Bioinformatics. 2014;48:3.13.1–16.
35. 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.
36. 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.
37. Slater GS, Birney E. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. 2005;6:31.
38. 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.
39. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna 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.
40. Guindon S, Delsuc F, Dufayard JF, Gascuel O. Estimating maximum likelihood phylogenies with PhyML. Methods Mol Biol. 2009;537:113–37.
41. 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.
42. 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. Aurrecoechea 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, Sucgang R, Berriman M, et al. The genome of the social amoeba *Dictyostelium discoideum*. Nature. 2005;435(7038):43–57.
49. Huang J, Mullapudi 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. *Afipia 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, Nusbaum 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, Brieba 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.

Mobilome genomics: new evidences in horizontal transfer of **β**-carbonic anhydrase genes from prokaryotes to parasitic and non-parasitic eukaryotes.

Reza Zolfaghari Emameh[1,2,3], Harlan R. Barker[1], Seppo Parkkila[1,3]

[1]School of Medicine, University of Tampere, FI-33520 Tampere, Finland; [2]BioMediTech, University of Tampere, FI-33520 Tampere, Finland; [3]Fimlab Laboratories Ltd and Tampere University Hospital, FI-33520 Tampere, Finland.

E-mail addresses:

Reza Zolfaghari Emameh: zolfaghari.emameh.reza.x@student.uta.fi

Harlan R. Barker: barker.harlan.r@student.uta.fi

Seppo Parkkila: seppo.parkkila@staff.uta.fi

Correspondence to:

Reza Zolfaghari Emameh, MSc

School of Medicine, Medisiinarinkatu 3,

FI-33520 University of Tampere, Finland.

Tel: +358 46 882 8000

Fax: +358 3 213 4473

E-mail: zolfaghari.emameh.reza.x@student.uta.fi

Abstract

Mobilomes, including Genomic Islands (GIs) and plasmids, refer to both mobile linear and circular DNA sequences which contain genes coding for their own accessory enzymatic proteins. These proteins, such as transposase and integrase are encoded by Insertion Sequence (IS) elements, enable transportation of vital coding sequences between prokaryotes or from prokaryotes to eukaryotes. This transfer of genetic sequences, which occurs outside of reproduction, is called Horizontal Gene Transfer (HGT). Beta carbonic anhydrase (β-CA) plays a critical role in biochemical pathways of many protozoans and metazoans, including parasitic and non-parasitic species. Importantly, we believe that these enzymes represent potential therapeutic targets in many parasitic species. Identification of horizontal transfer events where *β-CA* genes were transferred from archaeal and bacterial mobilomes to parasitic and non-parasitic eukaryotic hosts, can reveal the common ancestor of *β-CA* genes in these lineages.

In this work, we investigated archaeal and bacterial mobilomes such as GIs, and plasmids containing *β-CA,* transposase, and integrase genes. In order to reveal the distinct movements of these genes among a wide variety of organisms we utilized BLAST homology search, multiple sequence alignment, and phylogenetic analysis methods as well as IslandViewer, ACLAME, plasmid browser of EMBL-EBI, Jena Prokaryote Genome Viewer, and IS Finder databases.

We identified horizontal transfer of *β-CA* genes from mobilomes, including chromosomal GIs and plasmids of ancestral archaea and bacteria to parasitic and non-parasitic eukaryotes by phylogenetic analysis and identification of *β-CA* genes, and IS-encoded transposase, and integrase on the bacterial main chromosomes and plasmids. We also defined that β-CA plays as a virulence factor in *Mycobacterium tuberculosis* (strain H37Rv).

Our results revealed that mobilomes include a complete set of enzymatic tools, which are involved in the horizontal transfer of *β-CA* genes from archaeal GIs to archaeal and bacterial plasmids and then from prokaryotic plasmids to parasitic and non-parasitic eukaryotic hosts.

Keywords: Mobilome; Plasmid; Genomic islands; IS elements; β-carbonic anhydrase; Prokaryotes; Parasites; Transposase; Integrase; Evolution

Introduction

Horizontal Gene Transfer (HGT) is an evolutionary phenomenon by which a gene, or set of genes, can be exchanged between different species. This evolutionary process is mediated by mobile DNA, or Mobile Genetic Elements (MGEs), including: Genomic Islands (GIs), plasmids (circular and linear), transposons, retrotransposons, and prophages (Koonin, Makarova et al. 2001, Thomas and Nielsen 2005, Aminov 2011) which are known as a mobilome. Also, Polintons (the most complex DNA transposons) are widespread in the genomes of eukaryotes and diverged from linear plasmids about one billion years ago to play a critical role in evolution of protists, fungi, and animals, such as: *Entamoeba histolytica, Trichomonas vaginalis, Tetrahymena thermophila*, *Bombyx mori* (Silkworm), *Caenorhabditis* spp., *Ciona intestinalis* (Sea squirt), *Crassostrea gigas* (Pacific oyster), *Glyptapanteles flavicoxis* (Endoparasitoid wasp), *Hydra magnipapillata, Drosophila* spp., *Nasonia vitripennis* (Parasitoid wasp), *Nematostella vectensis* (Starlet sea anemone), *Strongylocentrotus purpuratus* (Purple sea urchin), *Tribolium castaneum* (Flour beetle), *Anolis carolinensis* (Carolina lizard) , *Gallus gallus* (Chicken), *Xenopus tropicalis* (Western clawed frog), *Danio rerio* (Zebrafish), and *Phakopsora* spp. (Soybean rust) (Kapitonov and Jurka 2006, Novick, Smith et al. 2011, Loehrer, Vogel et al. 2014, Krupovic and Koonin 2015).

Transformation, conjugation, and transduction are each distinct methods of HGT in bacteria. Varieties of crucial genes are transferred between prokaryotes (Pok-Pok) or from prokaryotes to eukaryotes (Pok-Euk) (Zolfaghari Emameh, Barker et al. 2016) through HGT, including virulence factors, antibiotic resistance, and toxin genes (Koonin, Makarova et al. 2001, Thomas and Nielsen 2005, Aminov 2011). In 1990, some clusters of virulence genes were identified as pathogenicity islands (PAIs) in *Escherichia coli*, which transferred through HGT (Hacker, Bender et al. 1990). Later GIs were defined as any cluster of genes (10–200 kb) that has been acquired by HGT (Hacker,

Blum-Oehler et al. 1997). GIs represent a part of a cell's chromosome, which are recognized as discrete DNA segments, which differ between closely related strains. Different GIs families have been recognized on the basis of sequence and functional homologies by GI prediction tools (Juhas, van der Meer et al. 2009).

There are two computational methods for prediction of GIs, including: (1) evaluation of sequence compositions, such as SIGI-HMM (Waack, Keller et al. 2006), IslandPath-DIMOB (Hsiao, Wan et al. 2003), PAI-IDA (Tu and Ding 2003), and Centroid (Rajan, Aravamuthan et al. 2007), and (2) application of comparative genomics, such as BLAST homology search and whole-genome sequence alignment. Among the sequence composition analysis methods, SIGI-HMM (Waack, Keller et al. 2006) and IslandPath-DIMOB (Hsiao, Wan et al. 2003) have shown the highest overall accuracy. Prediction based on comparative genomics relies on comparing of multiple genome sequences to detect GIs. There are two computational methods for prediction of GIs based on comparative genomics, including IslandPick (Langille, Hsiao et al. 2008) and MobilomeFINDER (Ou, He et al. 2007). Prediction based on IslandPick is supported by Islandviewer database (http://www.pathogenomics.sfu.ca/islandviewer/) (Dhillon, Laird et al. 2015). MobilomeFINDER focuses on identification of those islands that are associated with tRNA genes; not all GIs use tRNA genes as insertion sites, which limits the usage of MobilomeFINDER compared to the IslandPick method (Langille, Hsiao et al. 2010).

Insertion sequences (IS), like transposase and integrase genes, are regions of the genome that serve an accessory role in HGT (Mahillon and Chandler 1998, Berger and Haas 2001). IS elements are defined as small segments of DNA (2.5 kb) which have a simple genetic organization and can be inserted at multiple sites of a DNA molecule. IS elements move from one position to other positions in the same chromosome or other DNA structures. One class of IS is the Mariner family,

which are transposable elements that have been isolated and characterized in a broad spectrum of organisms, including fungi, ciliates, rotifers, insects, nematodes, plants, fish, and mammals (Robertson 1993, Mahillon and Chandler 1998, Plasterk, Izsvak et al. 1999). IS elements encode for enzymes, including transposases and integrase, which mediate transposition reactions. The majority of IS elements exhibit short terminal inverted-repeat (IR) sequences (10-40 bp), which can be divided into two functional domains, including IRR and IRL (Mahillon and Chandler 1998).

Carbonic anhydrases (CAs) are ubiquitous metalloenzymes, which are categorized into six evolutionary disparate genes, including α, β, γ, δ, ζ, and η (Elleuche and Poggeler 2010, Del Prete, Vullo et al. 2014). CAs are involved in many important biochemical pathways including pH homeostasis, electrolyte transfer, transport of CO_2 and bicarbonate between metabolizing tissues, calcification, bone resorption, some biosynthetic processes, and tumor progression (Vullo, Franchi et al. 2004, Vullo, Innocenti et al. 2005, Alterio, Vitale et al. 2006, Nishimori, Minakuchi et al. 2007). A variety of putative β-CAs have been discovered in pathogenic and non-pathogenic protozoans, rotifers, sea louses, molluscs, starlet sea anemones, purple sea urchins, arthropods, nematodes, and trematodes (Fasseas, Tsikou et al. 2010, Syrjanen, Tolvanen et al. 2010, Zolfaghari Emameh, Barker et al. 2014), as well as in archaea, bacteria and other eukaryotes such as fungi, algae, and plants (Smith, Jakubzick et al. 1999). *β-CA* gene sequences are present in the genomes of most living organisms, except for vertebrates (Syrjanen, Tolvanen et al. 2010, Zolfaghari Emameh, Barker et al. 2014).

The present study shows the role of mobilomes (GIs and plasmids) in horizontal transfer of *β-CA* gene sequences from archaea to bacteria and from bacteria to parasitic and non-parasitic protozoans and metazoans, including arthropods and nematodes. In addition, this study

demonstrates the importance of IS elements in HGT and *β-CA* gene exchange between GIs and plasmids.

Methods

BLAST homology search

The presence of *β-CA* genes on archaeal chromosomal GIs was first predicted in *Sulfolobus islandicus* (UniProt ID: F0NNH7) using the IslandViewer (http://www.pathogenomics.sfu.ca/islandviewer/) (Dhillon, Laird et al. 2015) database. Then the archaeal β-CAs most similar to the *S. islandicus* (UniProt ID: F0NNH7) β-CA protein sequence, predicted by BLASTp search (default parameters) of the NCBI BLAST homology database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (Coordinators 2015), were retrieved for multiple sequence alignment (MSA) and phylogenetic analysis.

Identification of **β**-CA protein sequences from defined archaeal, bacterial, protozoan, and nematode species

We collected β-CA protein sequences from bacterial mobilomes using the Jena Prokaryotic Genome Viewer (plasmids) (Romualdi, Felder et al. 2007) and IslandViewer 3 databases (GIs) (Dhillon, Laird et al. 2015). We then added β-CA protein sequences from bacteria *Acaryochloris marina* (Uniprot ID: A8ZKS7) and archaea *S. islandicus* (Uniprot ID: F0NNH7) as representative samples from β-CA expressing plasmids and GIs, respectively. The β-CA protein sequence from *Cesiribacter andamanensis* (Uniprot ID: M7MX87) was then also added as a representative sample that shared a common ancestor with protozoans (Zolfaghari Emameh, Barker et al. 2016). Some

protozoan species contain more than one β-CA protein sequence, in which case we used only one of them as a representative sequence from each protozoan species, including: *Acanthamoeba castellanii* (Uniprot ID: L8GR38), *Dictyostelium* spp. (Uniprot ID: Q555A3), *E. histolytica* (Uniprot ID: C4LXK3), *Ichthyophthirius multifiliis* (Uniprot ID: G0QYZ1), *Leishmania braziliensis* (Uniprot ID: A4H4M7), *Naegleria gruberi* (Uniprot ID: D2W492), *Paramecium tetraurelia* (Uniprot ID: A0BD61), *Tetrahymena thermophila* (Uniprot ID: Q22U21), and *Trichomonas vaginalis* (Uniprot ID: A2ENQ8). The β-CA protein sequence from hookworm *Necator americanus* (Unirprot ID: W2SJ13) was included as a representative sample from nematode species. Finally, a MSA was computed from these 22 β-CA protein sequences from archaeal, bacterial, protozoan, and nematode species (85 amino acid residues, starting three amino acid residues prior to first highly conserved motif; CXDXR). The MSA was computed using the Clustal Omega algorithm and visualized in JalView (http://www.jalview.org/) (Waterhouse, Procter et al. 2009).

Identification of ***β****-CA* gene sequences on the archaeal and bacterial mobilomes

Identification of β-CA gene sequences from archaeal and bacterial mobilomes was performed using ACLAME version 0.4 (http://aclame.ulb.ac.be/) (Leplae, Lima-Mendez et al. 2010), the EMBL-EBI plasmid browser (http://www.ebi.ac.uk/genomes/plasmid.html), the Jena Prokaryote Genome Viewer (JPGV) (http://jpgv.fli-leibniz.de/cgi/index.pl) (Romualdi, Felder et al. 2007) databases for bacterial plasmids, and IslandViewer version 3 (http://www.pathogenomics.sfu.ca/islandviewer/) (Dhillon, Laird et al. 2015) database for both archaeal and bacterial chromosomal GIs. The JPGV database was used to generate a circular plot of the plasmid with annotations of: negative and positive strands of DNA, G+C percentage, and exact location of *β-CA* genes. Similarly, the IslandViewer 3 webtool provides the ability to draw main circular genomes of defined prokaryotes

containing GIs as well as search for *β-CA* genes throughout whole prokaryotic main chromosomal genomes.

Identification of IS elements from archaeal and bacterial plasmids and chromosomal genomes

The studies revealed that IS elements encode enzymes such as transposases and integrases, which are necessary for horizontal transfer of genes from prokaryotes to prokaryotes or eukaryotes (Mahillon and Chandler 1998). Identification of archaeal and bacterial IS elements was performed using the IS Finder (https://www-is.biotoul.fr/) (Siguier, Perochon et al. 2006) database. IS Finder has collected an updated list of nominated archaeal and bacterial IS elements, which are located on prokaryotic plasmids and chromosomal genomes.

Phylogenetic analysis

A total of 107 β-CA sequences with chromosomal, plasmid, or GI origins were retrieved from the UniProt (UniProt 2015) and NCBI (Pirovano, Boetzer et al. 2015) databases using a BLAST search as the query. Of these, nine protein sequences from six species displayed a significant lack of homology with any others, owing to missing or spurious exons or poor quality of gene prediction. Therefore, the genomes of these species were retrieved from the RefSeq database (Pruitt, Brown et al. 2014) and, using all other sequences as template queries, analyzed using the exonerate program (Slater and Birney 2005). Subsequently, four of these sequences were replaced with corrected versions, two were fragments of the same protein and replaced with a single prediction, and no corresponding prediction could be made for the remaining three. Additionally, the exonerate analyses produced four proteins which were not revealed in our initial BLAST search. Individual BLAST searches revealed that three appear to be novel. See Table S3 for a summary of these results.

The final set of protein sequences was aligned using Clustal Omega (Figure S1) (Sievers and Higgins 2014), and a corresponding alignment of coding sequences (CDS) was created by Pal2Nal (Suyama, Torrents et al. 2006). The sequences were analyzed within the MrBayes v3.2.3 program (Ronquist, Teslenko et al. 2012) using the General Time Reversible (GTR) nucleotide model for 9,000,000 generations until the standard deviation of split frequencies was 0.008032. A consensus tree was generated from the 6,749 trees. The phylogenetic analyses were performed using supercomputer resources provided by the Finnish IT Center for Science (https://www.csc.fi/).

Results

Identification of **β**-CA protein sequences from defined archaeal, bacterial, protozoan, and nematode species

BLAST homology analysis of a chromosomal GI-encoded β-CA protein sequence from *S. islandicus* (UniProt ID: F0NNH7) revealed six β-CA protein sequences with the greatest identity, including: *Acidianus* spp. (Uniprot ID: G0WXL9), *S. solfataricus* (Uniprot ID: Q97YU3), *Cyanothece* spp. (Uniprot ID: E0UFL0), *Hassallia byssoidea* (Uniprot ID: A0A0C1XKI0), *Calothrix* spp. (Uniprot ID: K9PEP8), and *Acidithiobacillus thiooxidans* (Uniprot ID: S5FU55).

Also, multiple sequence alignment (MSA) of 85 amino acid residues of 22 β-CA protein sequences from defined archaeal, bacterial, protozoan, and nematode species revealed that all aligned β-CA protein sequences contain the first (CXDXR; C: Cysteine, D: Aspartic acid, R: Arginine, and X: any amino acids) and second (HXXC; H: Histidine, C. Cysteine, and X: any residues) highly conserved motifs which are characteristic of a β-CA protein (Figure 1).

Identification of **β**-CA gene sequences on the archaeal and bacterial mobilomes

Even though the plasmids are mostly small circular DNA molecules, such as pOU1113 from *Salmonella enterica* (Figure 2), previous studies have defined that there are also some linear plasmids in *Borrelia burgdorferi* (Ferdows and Barbour 1989), *Streptomyces coelicolor* (Kinashi and Shimaji-Murayama 1991), *S. lividans* (Chen, Yu et al. 1993), *Rhodococcus fascians* (Crespi, Messens et al. 1992), and *Agrobacterium tumefaciens* (Allardet-Servent, Michaux-Charachon et al. 1993). Through identification of *β-CA* genes from archaeal and bacterial mobilomes, 18 plasmids (Table S1) and 49 chromosomal GIs (Table S2) were detected in archaeal and bacterial species which contain *β-CA* genes (e.g. chromosomal GI-encoded β-CA of *Mycobacterium tuberculosis* (strain H37Rv) (Figure 3)). Three out of 18 plasmids-encoded β-CAs were linear plasmids, including pAH1134_566, pRHL1, and pFRL6 from *Bacillus cereus* (strain AH1134) (strain F12), *Rhodococcus jostii* (strain RHA1), and *Streptomyces* spp., respectively. One GI-encoded β-CA was determined in archaea *S. islandicus* (Strain HVE10/4).

Identification of IS elements from archaeal and bacterial plasmids and chromosomal genomes

Our analysis revealed that both plasmids and bacterial chromosomal genomes contain IS elements (Table S2). IS elements are able to express either transposase or integrase. Some IS elements contained two separate starting codons and overlapped open reading frames (ORFs) coding for both transposase and integrase, including IS families: IS3 (ISAzs24, ISMra1, ISMra4, ISRel21, ISRhosp5, ISSen3, ISSen4), IS21 (ISAcma26, ISAzs2, ISRel5, ISRel16), and IS630 (ISRel6) from plasmids and IS3 (ISYps8), IS6 (ISMtsp1, ISMtsp2, ISMtsp4), and IS21 (ISFK1) from bacterial main

chromosomes. The only predicted IS element from a linear plasmid was IS3 (ISRhosp5) from pRHL1 of *R. jostii* (strain RHA1).

Phylogenetic analysis

The result of the phylogenetic analysis is shown in ten different panels A-J (Figure 4 & Table 1). Our analysis identified that both GIs and plasmids functioned as the shuttle for horizontal transfer of *β-CA* genes between archaea, bacteria, protozoans, arthropods, and nematodes. We identify a "Golden Box" within the phylogenetic tree (Figure 4) where both GIs and plasmids have played a critical role in horizontal transfer of *β-CA* genes within archaea and bacteria (both cyanobacteria and proteobacteria). The "Golden Box" clearly shows that *β-CA* genes from proteobacteria *A. thiooxidans* (S5FU55) and cyanobacteria *H. byssoidea* (A0A0C1XKI0), *Calothrix* spp. (K9PEP8), and *Cyanothece* spp. (E0UFL0) have a common ancestor with archaea *Halopiger xanaduensis* (F8DDX3), which is located on circular plasmid pHALXA01. In addition, we could define that pHALXA01-encoded β-CA from archaea *H. xanaduensis* (F8DDX3) has a common ancestor with a *β-CA* gene from chromosomal GI of archaea *S. islandicus* (F0NNH7). *β-CA* genes which we have identified as being horizontally transferred, through the action of prokaryotic mobilomes, to protozoans, arthropods, and nematodes are presented in Table S3.

Discussion

Bacterial mobilomes and IS elements represent important modes of adaptation to environmental stimuli. They provide a platform for recombination and HGT events that spread genetic material between genera (Figure 5) (Sorensen, Bailey et al. 2005), for example acquisition of the genes

responsible for degradation of pesticides and other xenobiotic compounds (Izmalkova, Mavrodi et al. 2006, Dealtry, Holmsgaard et al. 2014, Jorgensen, Kiil et al. 2014).

BLAST homology analysis of GI-encoded β-CA protein sequence from archaea *S. islandicus* (UniProt ID: F0NNH7) revealed that this enzyme and the other six archaeal β-CA protein sequences from *Acidianus* spp. (Uniprot ID: G0WXL9), *S. solfataricus* (Uniprot ID: Q97YU3), *Cyanothece* spp. (Uniprot ID: E0UFL0), *H. byssoidea* (Uniprot ID: A0A0C1XKI0), *Calothrix* spp. (Uniprot ID: K9PEP8), and *A. thiooxidans* (Uniprot ID: S5FU55) are more identical to each other than bacterial β-CA protein sequences. An MSA of 22 β-CA protein sequences from archaeal, bacterial, protozoan, arthropod, and nematode species defined that all aligned β-CA protein sequences contain the first (CXDXR) and second (HXXC) highly conserved motifs in catalytic active site of β-CA. Identification of *β-CA* genes both within archaeal and bacterial mobilomes and in close proximity with IS elements on plasmids and chromosomes, strongly suggests that *β-CA* genes were transferred horizontally from prokaryotes to prokaryotes (Pok-Pok) or eukaryotes (Pok-Euk) (Zolfaghari Emameh, Barker et al. 2016). Our studies identified that horizontal transfer of *β-CA* genes has been occurred by mobilome system, plasmid or chromosomal GI, giving a functional and metabolic adaptability strength to archaea or bacteria. Surprisingly, we could discover a potent HGT strategy in *Rhizobium etli* (strain CFN 42) containing different *β-CA* genes on both plasmid (p42d and pB) and chromosomal GI (Gene name: RHE_CH01064). This property provides two powerful arms to *R. etli* (strain CFN 42) for Pok-Pok or Pok-Euk horizontal transfer of *β-CA* genes. Even though no difference in HGT ability has been yet reported between linear and circular plasmids, our results showed that circular plasmids have higher diversity in horizontal transfer of *β-CA* genes than linear plasmids. In archaeal species, we only observed HGT of *β-CA* genes from chromosomal GIs. This prediction seems logical, because after the creation of archaeal species around 3.5 billion years

ago (Doolittle 1997), they probably applied chromosomal GIs more than plasmids for HGT. After the transformation of the Earth's environment to a more oxygen-rich atmosphere, bacterial species flourished; through their utilization of the plasmids and IS elements the speed of their evolution quickly outpaced that of archaea leading to a wider distribution (Makiuchi and Nozaki 2014).

Our studies have revealed that β-CA is a potential virulence factor for some pathogenic bacteria, such as *M. tuberculosis* (strain H37Rv). These enzymatic virulence factors can be attenuated by attenuators or RNA-based regulatory strategies which lead to premature termination of transcription of protein (Naville and Gautheret 2009). Identification of highly mobile transposase and integrase genes flanking *β-CA* genes in archaeal and bacterial plasmids and chromosomes signifies that these enzymatic recombination tools are expressed by IS elements and have played a major role in horizontal transfer of *β-CA* genes.

During HGT, *β-CA* genes from prokaryotic plasmids are transferred to the genome of another prokaryote (Pok-Pok) or eukaryote (Pok-Euk) (Figure 6). In another HGT possibility, *β-CA* genes are transferred from prokaryotic chromosomal GIs to the different genomic target sites of another prokaryotic or eukaryotic species (Figure 7).

Because of the complexity of eukaryotic genomes, it is poorly understood which DNA transposons and retrotransposons are responsible for integration of *β-CA* genes therein. Even so, recent studies showed that most Polintons encode two viral capsid proteins, suggesting that Polintons can produce virions that could infect new hosts. Therefore, it seems that Polintons combine the features of viruses and transposons (Polintoviruses) (Krupovic and Koonin 2015). Discovery of an evolutionary relationship between Polintons and prokaryotic mobilomes could unravel the mystery of horizontal transfer of *β-CA* genes from prokaryotes to eukaryotes. Based on the

bacteriophage origin of Polintons (Krupovic and Koonin 2015), it is proposed that IS element-encoded transposase and integrase were derived from Polintons. However, to date we still have no evidence of *β-CA* genes in bacteriophages or viral particles. It is tempting to speculate that Polintons might have facilitated the HGT of structural and metabolic genes from prokaryotes to eukaryotes. Previous studies determined that transposons tend to be create exonic (Exonization) and intronic (Intronization) sequences in the genome of vertebrates (*Gallus gallus* and *Danio rerio*) and invertebrates (*D. melanogaster* and *C. elegans*) (Sela, Kim et al. 2010). Hence, the main reason for the absence of β-CAs in vertebrates might be related to the inability of β-CA-coding mobilomes to detect an integration site in the vertebrate egg or sperm genomes to insert *β-CA* genes as exonic sequences.

As the overall conclusions, several archaeal and bacterial mobilomes, including plasmids and GIs, have been equipped with IS elements, as well as *β-CA* genes. IS elements are able to encode for transposase and integrase, which are known enzymatic tools that facilitate horizontal transfer of *β-CA* genes from archaeal and bacterial mobilomes to other prokaryotic (Pro-Pro), parasitic, or non-parasitic eukaryotic (Pro-Euk) species. The phylogenetic analysis suggests that *β-CA* genes are horizontally transferred from archaeal GIs to other archaeal or bacterial plasmids, while prokaryotic species horizontally transfer *β-CA* genes from either plasmids or GIs to another prokaryote, parasitic, or non-parasitic eukaryote. The present results identified that *β-CA* genes primarily originated from archaeal GIs and then distributed to other archaeal and bacterial plasmids by HGT mechanism and other facilitators such as IS elements-encoded transposase and integrase. Further studies need to be performed for identification of the origin of *β-CA* gene in archaeal species.

Authors' contributions

All authors participated in the design of the study. RZE carried out the bioinformatics searches on bacterial, parasitic and non-parasitic eukaryotic species, as well as identification of β-CAs, IS elements, and mobilomes containing *β-CA*, transposase, and integrase genes. RZE designed the schematic figures of mobilomes and HGT. HRB made protein sequence corrections and predictions. RZE and HRB performed the MSA and phylogenetic analysis. RZE and HRB 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. This work was also supported by research grants from the Academy of Finland to SP, Finnish Cultural Foundation (Pirkanmaa Regional Fund to RZE and Maili Autio Fund to HRB), Sigrid Juselius Foundation to SP, Jane and Aatos Erkko Foundation to SP, Tampere Tuberculosis Foundation to SP, and Competitive Research Funding of the Tampere University Hospital.

References

Allardet-Servent, A., S. Michaux-Charachon, E. Jumas-Bilak, L. Karayan and M. Ramuz (1993). "Presence of one linear and one circular chromosome in the Agrobacterium tumefaciens C58 genome." J Bacteriol 175(24): 7869-7874.

Alterio, V., R. M. Vitale, S. M. Monti, C. Pedone, A. Scozzafava, A. Cecchi, G. De Simone and C. T. Supuran (2006). "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 128(25): 8329-8335.

Aminov, R. I. (2011). "Horizontal gene exchange in environmental microbiota." Front Microbiol 2: 158.

Berger, B. and D. Haas (2001). "Transposase and cointegrase: specialized transposition proteins of the bacterial insertion sequence IS21 and related elements." Cell Mol Life Sci 58(3): 403-419.

Chen, C. W., T. W. Yu, Y. S. Lin, H. M. Kieser and D. A. Hopwood (1993). "The conjugative plasmid SLP2 of Streptomyces lividans is a 50 kb linear molecule." Mol Microbiol 7(6): 925-932.

Chen, L., Z. Xiong, L. Sun, J. Yang and Q. Jin (2012). "VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors." Nucleic Acids Res 40(Database issue): D641-645.

Coordinators, N. R. (2015). "Database resources of the National Center for Biotechnology Information." Nucleic Acids Res 43(Database issue): D6-17.

Crespi, M., E. Messens, A. B. Caplan, M. van Montagu and J. Desomer (1992). "Fasciation induction by the phytopathogen Rhodococcus fascians depends upon a linear plasmid encoding a cytokinin synthase gene." EMBO J 11(3): 795-804.

Dealtry, S., P. N. Holmsgaard, V. Dunon, S. Jechalke, G. C. Ding, E. Krogerrecklenfort, H. Heuer, L. H. Hansen, D. Springael, S. Zuhlke, S. J. Sorensen and K. Smalla (2014). "Shifts in abundance and diversity of mobile genetic elements after the introduction of diverse pesticides into an on-farm biopurification system over the course of a year." Appl Environ Microbiol 80(13): 4012-4020.

Del Prete, S., D. Vullo, G. M. Fisher, K. T. Andrews, S. A. Poulsen, C. Capasso and C. T. Supuran (2014). "Discovery of a new family of carbonic anhydrases in the malaria pathogen Plasmodium falciparum--the eta-carbonic anhydrases." Bioorg Med Chem Lett 24(18): 4389-4396.

Dhillon, B. K., M. R. Laird, J. A. Shay, G. L. Winsor, R. Lo, F. Nizam, S. K. Pereira, N. Waglechner, A. G. McArthur, M. G. Langille and F. S. Brinkman (2015). "IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis." Nucleic Acids Res.

Doolittle, W. F. (1997). "Fun with genealogy." Proc Natl Acad Sci U S A 94(24): 12751-12753.

Elleuche, S. and S. Poggeler (2010). "Carbonic anhydrases in fungi." Microbiology 156(Pt 1): 23-29.

Fasseas, M. K., D. Tsikou, E. Flemetakis and P. Katinakis (2010). "Molecular and biochemical analysis of the beta class carbonic anhydrases in Caenorhabditis elegans." Mol Biol Rep 37(6): 2941-2950.

Ferdows, M. S. and A. G. Barbour (1989). "Megabase-sized linear DNA in the bacterium Borrelia burgdorferi, the Lyme disease agent." Proc Natl Acad Sci U S A 86(15): 5969-5973.

Hacker, J., L. Bender, M. Ott, J. Wingender, B. Lund, R. Marre and W. Goebel (1990). "Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extraintestinal Escherichia coli isolates." Microb Pathog 8(3): 213-225.

Hacker, J., G. Blum-Oehler, I. Muhldorfer and H. Tschape (1997). "Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution." Mol Microbiol 23(6): 1089-1097.

Hsiao, W., I. Wan, S. J. Jones and F. S. Brinkman (2003). "IslandPath: aiding detection of genomic islands in prokaryotes." Bioinformatics 19(3): 418-420.

Izmalkova, T. Y., D. V. Mavrodi, S. L. Sokolov, I. A. Kosheleva, K. Smalla, C. M. Thomas and A. M. Boronin (2006). "Molecular classification of IncP-9 naphthalene degradation plasmids." Plasmid 56(1): 1-10.

Jorgensen, T. S., A. S. Kiil, M. A. Hansen, S. J. Sorensen and L. H. Hansen (2014). "Current strategies for mobilome research." Front Microbiol 5: 750.

Juhas, M., J. R. van der Meer, M. Gaillard, R. M. Harding, D. W. Hood and D. W. Crook (2009). "Genomic islands: tools of bacterial horizontal gene transfer and evolution." FEMS Microbiol Rev 33(2): 376-393.

Kapitonov, V. V. and J. Jurka (2006). "Self-synthesizing DNA transposons in eukaryotes." Proc Natl Acad Sci U S A 103(12): 4540-4545.

Kinashi, H. and M. Shimaji-Murayama (1991). "Physical characterization of SCP1, a giant linear plasmid from Streptomyces coelicolor." J Bacteriol 173(4): 1523-1529.

Koonin, E. V., K. S. Makarova and L. Aravind (2001). "Horizontal gene transfer in prokaryotes: quantification and classification." Annu Rev Microbiol 55: 709-742.

Krupovic, M. and E. V. Koonin (2015). "Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution." Nat Rev Microbiol 13(2): 105-115.

Langille, M. G., W. W. Hsiao and F. S. Brinkman (2008). "Evaluation of genomic island predictors using a comparative genomics approach." BMC Bioinformatics 9: 329.

Langille, M. G., W. W. Hsiao and F. S. Brinkman (2010). "Detecting genomic islands using bioinformatics approaches." Nat Rev Microbiol 8(5): 373-382.

Leplae, R., G. Lima-Mendez and A. Toussaint (2010). "ACLAME: a CLAssification of Mobile genetic Elements, update 2010." Nucleic Acids Res 38(Database issue): D57-61.

Loehrer, M., A. Vogel, B. Huettel, R. Reinhardt, V. Benes, S. Duplessis, B. Usadel and U. Schaffrath (2014). "On the current status of Phakopsora pachyrhizi genome sequencing." Front Plant Sci 5: 377.

Mahillon, J. and M. Chandler (1998). "Insertion sequences." Microbiol Mol Biol Rev 62(3): 725-774.

Makiuchi, T. and T. Nozaki (2014). "Highly divergent mitochondrion-related organelles in anaerobic parasitic protozoa." Biochimie 100: 3-17.

Naville, M. and D. Gautheret (2009). "Transcription attenuation in bacteria: theme and variations." Brief Funct Genomic Proteomic 8(6): 482-492.

Nishimori, I., T. Minakuchi, S. Onishi, D. Vullo, A. Scozzafava and C. T. Supuran (2007). "Carbonic anhydrase inhibitors. DNA cloning, characterization, and inhibition studies of the human secretory isoform VI, a new target for sulfonamide and sulfamate inhibitors." J Med Chem 50(2): 381-388.

Novick, P. A., J. D. Smith, M. Floumanhaft, D. A. Ray and S. Boissinot (2011). "The evolution and diversity of DNA transposons in the genome of the Lizard Anolis carolinensis." Genome Biol Evol 3: 1-14.

Ou, H. Y., X. He, E. M. Harrison, B. R. Kulasekara, A. B. Thani, A. Kadioglu, S. Lory, J. C. Hinton, M. R. Barer, Z. Deng and K. Rajakumar (2007). "MobilomeFINDER: web-based tools for in silico and experimental discovery of bacterial genomic islands." Nucleic Acids Res 35(Web Server issue): W97-W104.

Pirovano, W., M. Boetzer, M. F. Derks and S. Smit (2015). "NCBI-compliant genome submissions: tips and tricks to save time and money." Brief Bioinform.

Plasterk, R. H., Z. Izsvak and Z. Ivics (1999). "Resident aliens: the Tc1/mariner superfamily of transposable elements." Trends Genet 15(8): 326-332.

Pruitt, K. D., G. R. Brown, S. M. Hiatt, F. Thibaud-Nissen, A. Astashyn, O. Ermolaeva, C. M. Farrell, J. Hart, M. J. Landrum, K. M. McGarvey, M. R. Murphy, N. A. O'Leary, S. Pujar, B. Rajput, S. H. Rangwala, L. D. Riddick, A. Shkeda, H. Sun, P. Tamez, R. E. Tully, C. Wallin, D. Webb, J. Weber, W. Wu, M. DiCuccio, P. Kitts, D. R. Maglott, T. D. Murphy and J. M. Ostell (2014). "RefSeq: an update on mammalian reference sequences." Nucleic Acids Res 42(Database issue): D756-763.

Rajan, I., S. Aravamuthan and S. S. Mande (2007). "Identification of compositionally distinct regions in genomes using the centroid method." Bioinformatics 23(20): 2672-2677.

Robertson, H. M. (1993). "The mariner transposable element is widespread in insects." Nature 362(6417): 241-245.

Romualdi, A., M. Felder, D. Rose, U. Gausmann, M. Schilhabel, G. Glockner, M. Platzer and J. Suhnel (2007). "GenColors: annotation and comparative genomics of prokaryotes made easy." Methods Mol Biol 395: 75-96.

Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Hohna, B. Larget, L. Liu, M. A. Suchard and J. P. Huelsenbeck (2012). "MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space." Syst Biol 61(3): 539-542.

Sela, N., E. Kim and G. Ast (2010). "The role of transposable elements in the evolution of non-mammalian vertebrates and invertebrates." Genome Biol 11(6): R59.

Sievers, F. and D. G. Higgins (2014). "Clustal omega." Curr Protoc Bioinformatics 48: 3 13 11-13 13 16.

Siguier, P., J. Perochon, L. Lestrade, J. Mahillon and M. Chandler (2006). "ISfinder: the reference centre for bacterial insertion sequences." Nucleic Acids Res 34(Database issue): D32-36.

Slater, G. S. and E. Birney (2005). "Automated generation of heuristics for biological sequence comparison." BMC Bioinformatics 6: 31.

Smith, K. S., C. Jakubzick, T. S. Whittam and J. G. Ferry (1999). "Carbonic anhydrase is an ancient enzyme widespread in prokaryotes." Proc Natl Acad Sci U S A 96(26): 15184-15189.

Sorensen, S. J., M. Bailey, L. H. Hansen, N. Kroer and S. Wuertz (2005). "Studying plasmid horizontal transfer in situ: a critical review." Nat Rev Microbiol 3(9): 700-710.

Suyama, M., D. Torrents and P. Bork (2006). "PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments." Nucleic Acids Res 34(Web Server issue): W609-612.

Syrjanen, L., M. Tolvanen, M. Hilvo, A. Olatubosun, A. Innocenti, A. Scozzafava, J. Leppiniemi, B. Niederhauser, V. P. Hytonen, T. A. Gorr, S. Parkkila and C. T. Supuran (2010). "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 11: 28.

Thomas, C. M. and K. M. Nielsen (2005). "Mechanisms of, and barriers to, horizontal gene transfer between bacteria." Nat Rev Microbiol 3(9): 711-721.

Tu, Q. and D. Ding (2003). "Detecting pathogenicity islands and anomalous gene clusters by iterative discriminant analysis." FEMS Microbiol Lett 221(2): 269-275.

UniProt, C. (2015). "UniProt: a hub for protein information." Nucleic Acids Res 43(Database issue): D204-212.

Waack, S., O. Keller, R. Asper, T. Brodag, C. Damm, W. F. Fricke, K. Surovcik, P. Meinicke and R. Merkl (2006). "Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models." BMC Bioinformatics 7: 142.

Waterhouse, A. M., J. B. Procter, D. M. Martin, M. Clamp and G. J. Barton (2009). "Jalview Version 2--a multiple sequence alignment editor and analysis workbench." Bioinformatics 25(9): 1189-1191.

Vullo, D., M. Franchi, E. Gallori, J. Antel, A. Scozzafava and C. T. Supuran (2004). "Carbonic anhydrase inhibitors. Inhibition of mitochondrial isozyme V with aromatic and heterocyclic sulfonamides." J Med Chem 47(5): 1272-1279.

Vullo, D., A. Innocenti, I. Nishimori, J. Pastorek, A. Scozzafava, S. Pastorekova and C. T. Supuran (2005). "Carbonic anhydrase inhibitors. Inhibition of the transmembrane isozyme XII with sulfonamides-a new target for the design of antitumor and antiglaucoma drugs?" Bioorg Med Chem Lett 15(4): 963-969.

Zolfaghari Emameh, R., H. Barker, M. E. Tolvanen, C. Ortutay and S. Parkkila (2014). "Bioinformatic analysis of beta carbonic anhydrase sequences from protozoans and metazoans." Parasit Vectors 7: 38.

Zolfaghari Emameh, R., H. R. Barker, M. E. Tolvanen, S. Parkkila and V. P. Hytonen (2016). "Horizontal transfer of beta-carbonic anhydrase genes from prokaryotes to protozoans, insects, and nematodes." Parasit Vectors 9(1): 152.

Table 1 Several layers of common ancestors of ***β****-CA* genes from archaea, bacteria, protozoans, and metazoans.

Box	Suggested first layer common ancestor	Suggested second layer common ancestor	Suggested third layer common ancestor	Relative protozoans, arthropods, and nematodes
A	Common ancestor with "Aa"	Follow through Figure 4*		
B	Common ancestor with "Bb"	Follow through Figure 4*		
C	*Cesiribacter adamanensis* (M7MX87), *Leptospira kirschneri* (M6X652)	*Colwellia psychrerythraea* (Q47YG3)	*Proteus mirabilis*_GI (YP_008396903.1)	*Dictyostelium* spp.**1
			Vibrio spp_GI (A7K5V9)	*Ichthyophthirius multifiliis* (G0QPN9 & G0QYZ1), *Tetrahymena thermophila***2
D		*Corallococcus coralloides* (H8MJ17)	*Acaryochloris marina*_PLASMID (A8ZKS7)	*Haemonchus contortus* (U6PDI1), *Necator americanus* (W2SJ13), *Musca domestica* (T1PI17)
E		*Afipia felis* (K8NQ88)	*Deinococcus maricopensis*_GI (E8UBN1)	*Acanthamoeba castellanii* (L8H861 & L8GLS7)
F		*Vesicomyosocius okutanii* (A5CVM8)	*Haliscomenobacter hydrossis*_GI (F4KZC7)	*Entamoeba* spp.**3
G		*Selenomonas ruminantium* (I0GLW8)	*Bacillus bombysepticus*_PLASMID (A0A023PDY3)	*Trichomonas vaginalis* (A2ENQ8 & A2DLG4)

		Veillonella spp. (F9N508)		
H		*Leptospira kirschneri* (M6X652)	*Saccharopolyspora erythraea*_GI (A4F6V7)	*Naegleria gruberi* (Predicted 1 & 2) *Paramecium tetraurelia*[**4]
I	*Cesiribacter adamanensis* (M7MX87)	*Magnetospirillum magneticum* (Q2VZD0)	*Pseudomonas stutzeri*_GI (A4VPG8)	*Leishmania* spp.[**5]
J	*Cesiribacter adamanensis* (M7MX87)	*Cesiribacter adamanensis* (M7MX87)	*Burkholderia pseudomallei*_Predicted 2[***]	*Acanthamoeba castellanii* (L8GR38)

*: As the suggested first layer common ancestor, "Aa" covers all the phylogenetic tree, starting from *Leptospira kirschneri* (M6X652) at the top of Figure 4. Also, "Bb" covers all the phylogenetic tree, starting from *Colwellia psychrerythraea* (Q47YG3) at the top of Figure 4.

**: IDs: (1) Q555A3, Q55BU2, Q94473, F0Z7L1, F4PL43; (2) I7LWM1, I7MDL7, Q23AV1, Q22U21, I7M0M0, I7M748, I7MD92; (3) B0E7M0, C4LXK3, K2GQM0; (4) A0BD61, A0CEX6, A0C922, A0E8J0, A0BDB1; (5) A4H4M7, E9B8S3, A4HSV2, Q4QJ17, E9AKU0, S0CTX5.

***: Predicted from the *Burkholderia pseudomallei* genome using all other sequences as query templates (Figure S1).

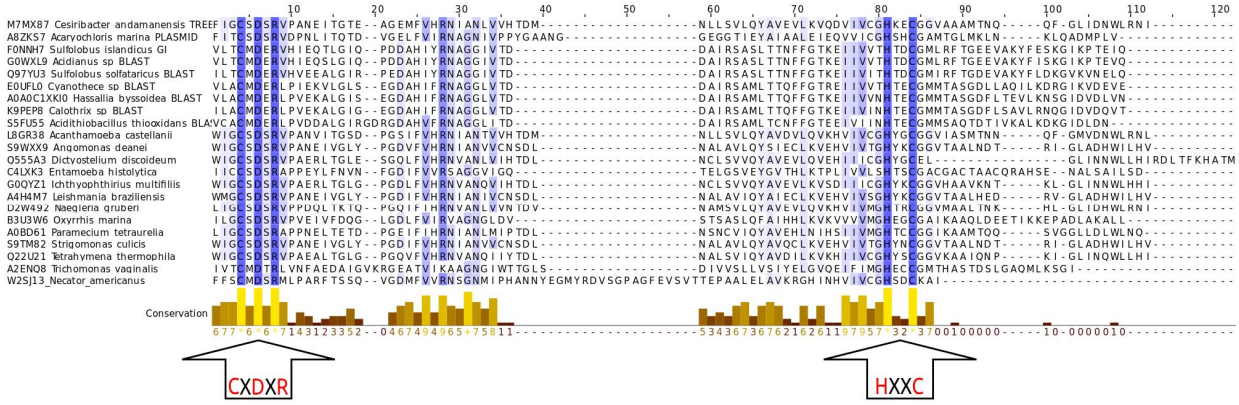


Figure 1 Multiple sequence alignment (MSA) of **β**-CA protein sequences from defined archaeal, bacterial, protozoan, and nematode species. The aligned 22 β-CA protein sequences (starting 3 amino acid residues prior to the first highly conserved residues (CXDXR) show that they all contain the first (CXDXR; C: Cysteine, D: Aspartic acid, R: Arginine, and X: any residues) and second (HXXC; H: Histidine, C. Cysteine, and X: any residues) highly conserved motifs of β-CAs.

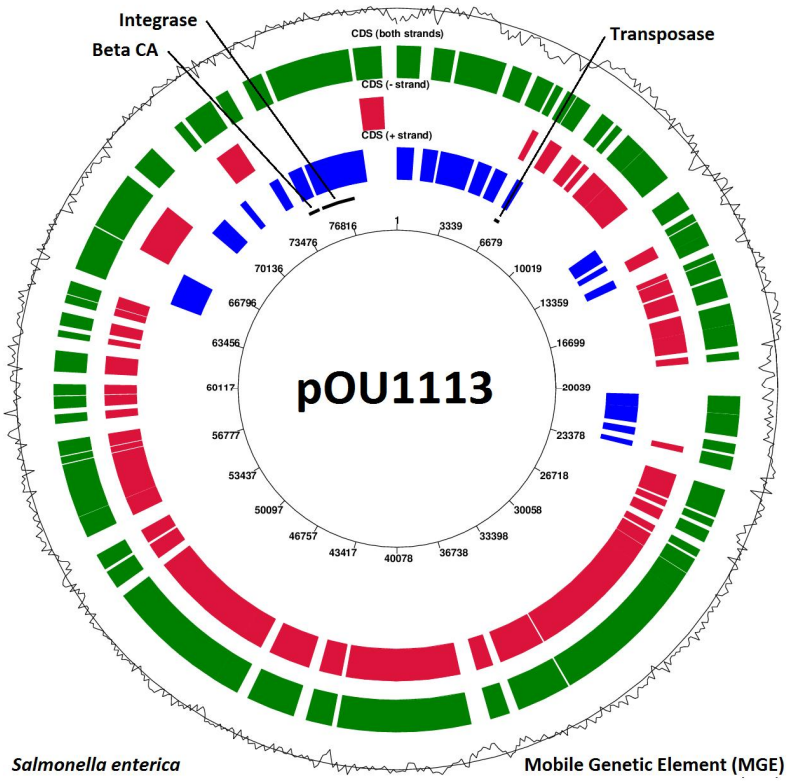


Figure 2 Circular plasmid pOU1133 from *Salmonella enterica*. pOU1133 was visualized within the Jena Prokaryote Genome Viewer. The length of pOU1113 is 80,156 bp and it contains *β-CA* (74,210-74,950 bp), transposase (6,653-7,003 bp), and integrase (75,157-77,365 bp) genes. The outermost ring indicates G+C content of the plasmid.

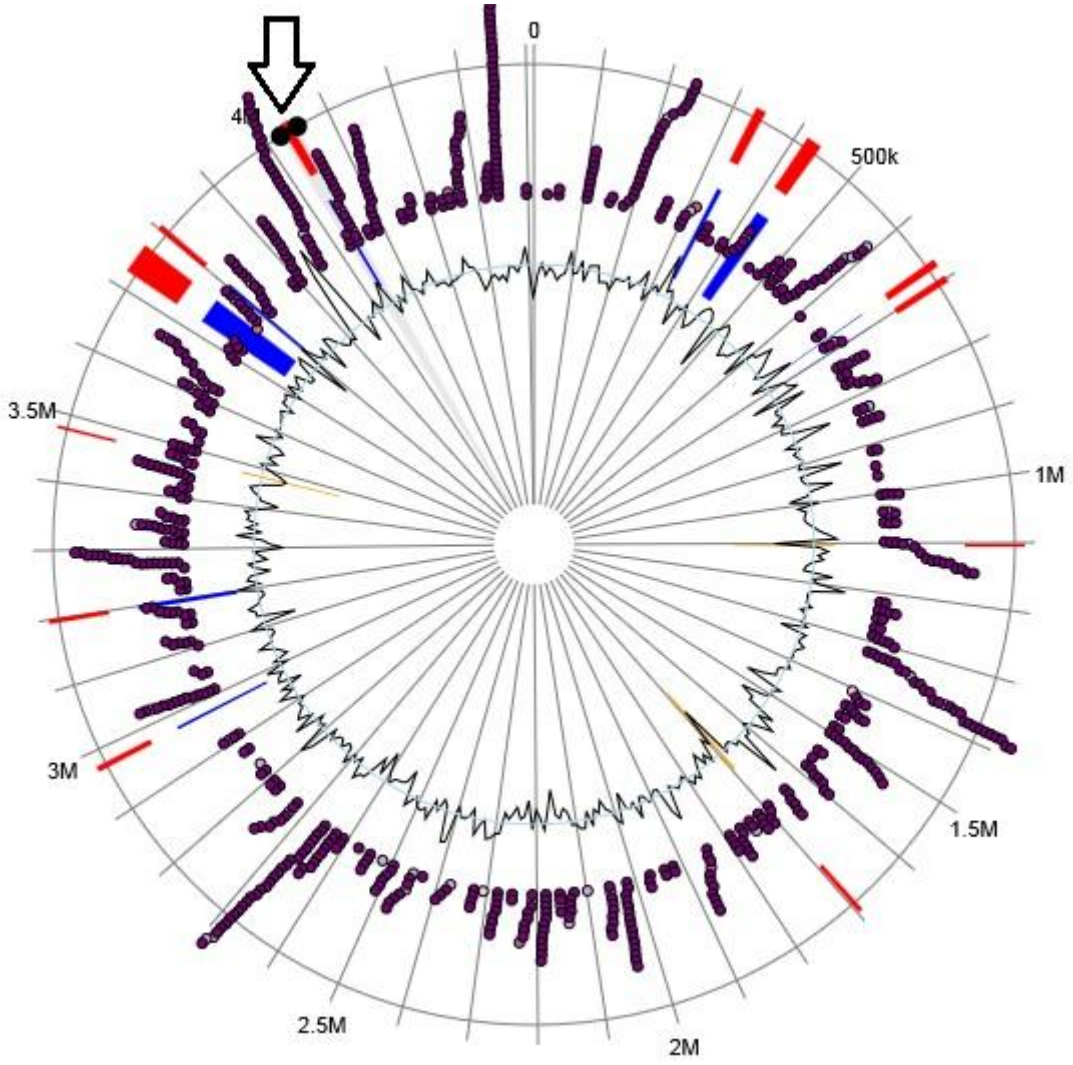


Figure 3 The main chromosome of *Mycobacterium tuberculosis* (strain H37Rv). The *β-CA* gene (Gene name: *canB*, NCBI ID: NP_218105.1, and sequence coordinates: 4,029,871-4,030,494 bp) from *M. tuberculosis* (strain H37Rv) was located on a chromosomal GI, and is shown by an arrow. The GIs were predicted by IslandViewer version 3. IslandPath-DIMOB (Hsiao, Wan et al. 2003) and SIGI-HMM (Waack, Keller et al. 2006) methods (Blue and yellow color bands, respectively) were used to predict GIs. The ultimate prediction result is represented as red colored bands. β-CA was predicted as a virulence factor in *M. tuberculosis* virulence factor database (Chen, Xiong et al. 2012). The height of purple columns indicate the strength of prediction for the presence of virulence factors.

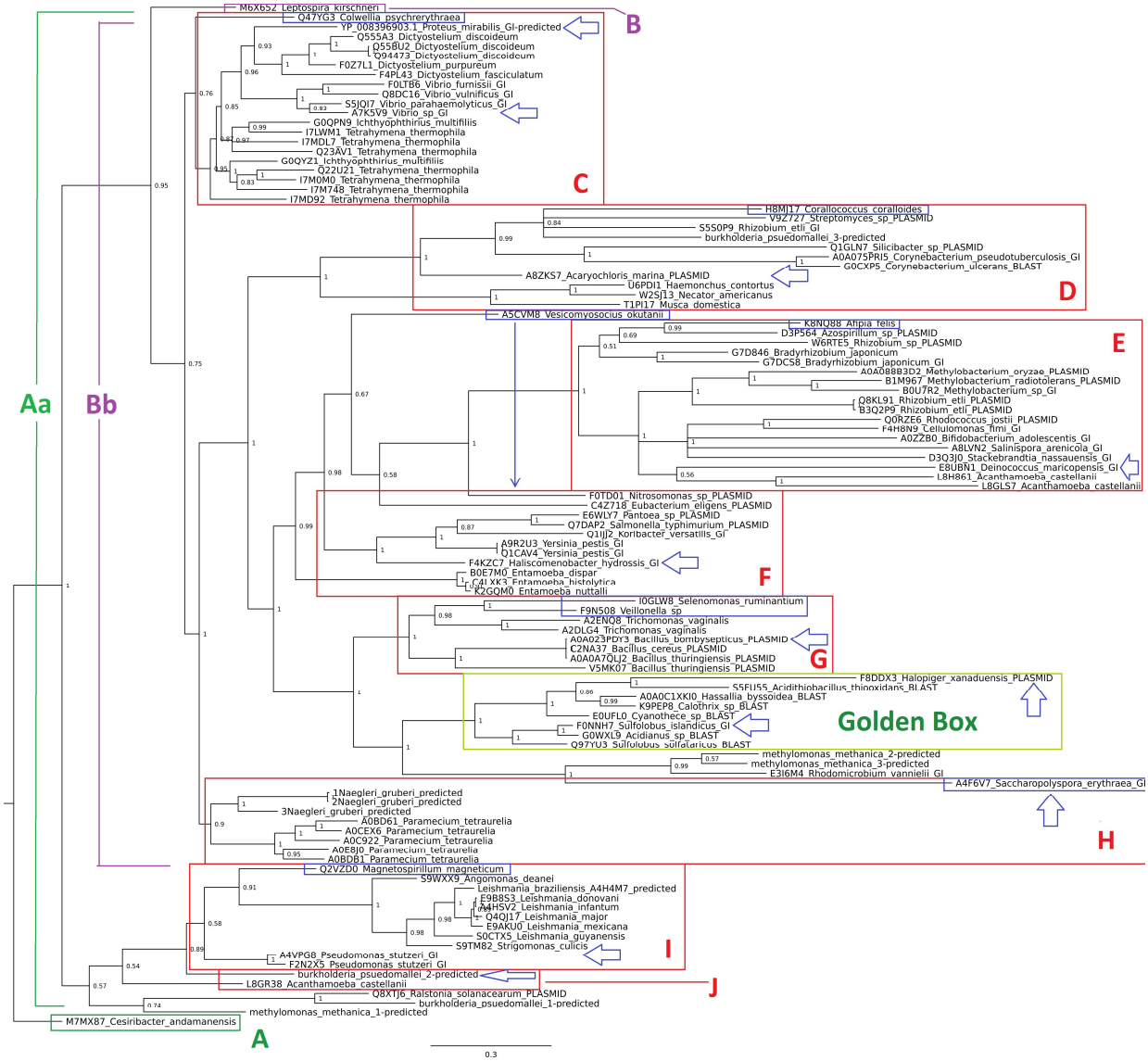


Figure 4 Phylogenetic analysis of **β**-CA proteins from prokaryotic mobilomes and parasitic and non-parasitic eukaryotic chromosomal genomes. Species in boxes (A) and (B) (Zolfaghari Emameh, Barker et al. 2016) had common ancestors with species “Aa” and “Bb” respectively. In boxes (C) and (E-J) (protozoans) and box (D) (arthropods and nematodes) at least one prokaryote is found grouped with the other members of the clade. The “Golden Box” shows that these β-CAs from archaeal, cyanobacterial, and proteobacterial species had a common ancestor. Also, this box displayed that GI and plasmid *β-CA* genes, from two archaeal species, had a common ancestor with each other. The relationships of the phylogenetic tree are explained in Table 1.

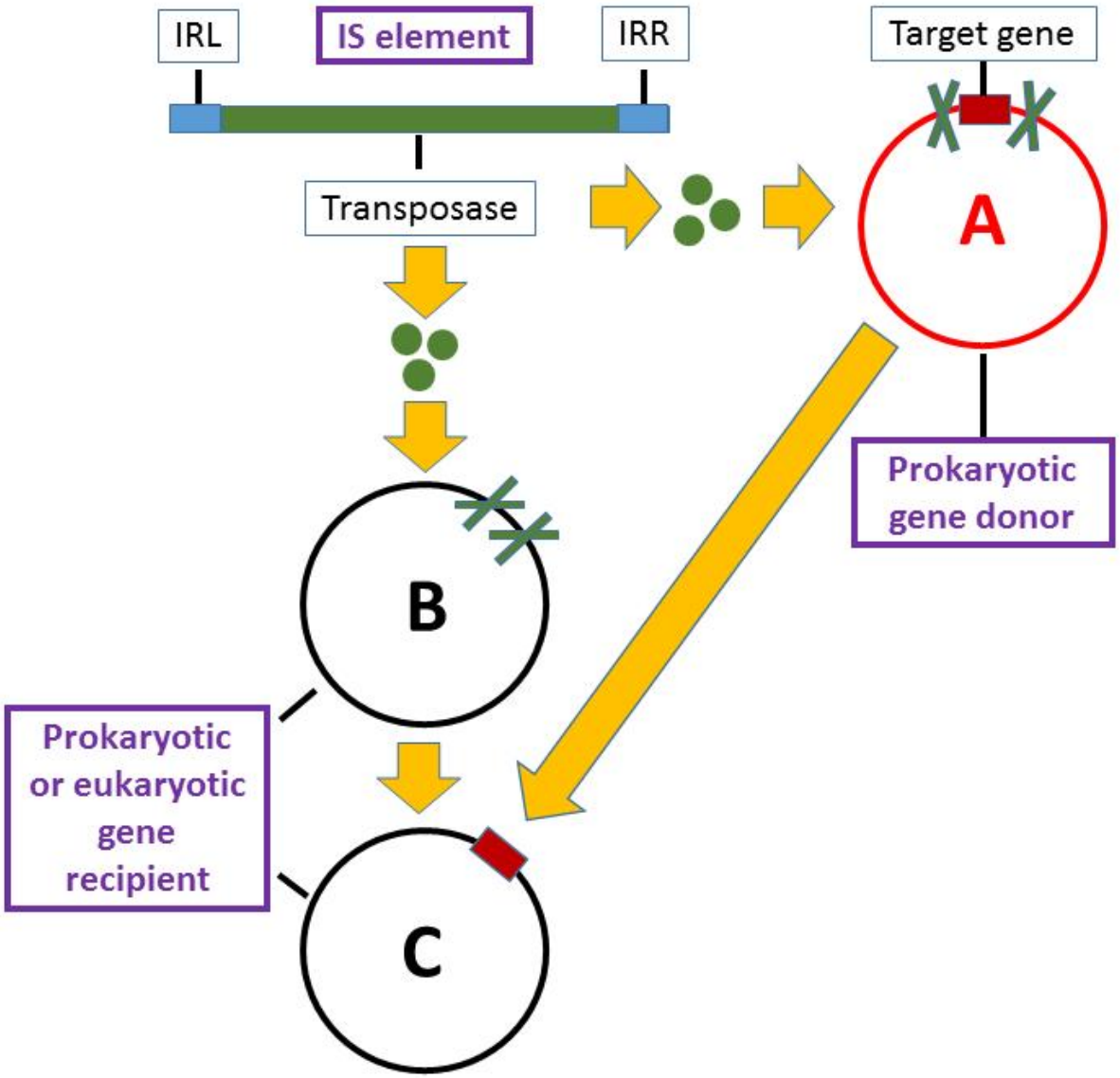


Figure 5 Schematic molecular mechanism of HGT of a target gene. An IS element includes: IRL and IRR, a promoter for transposase or integrase in IRL, and transposase or integrase open reading frame (ORF). (A) The encoded transposase cut or copy a target gene from a prokaryotic DNA; (B) The encoded transposase cut a specific sequence on another prokaryotic or eukaryotic DNA; and (C) The target gene is horizontally transferred from a prokaryotic DNA to another prokaryotic or eukaryotic DNA.

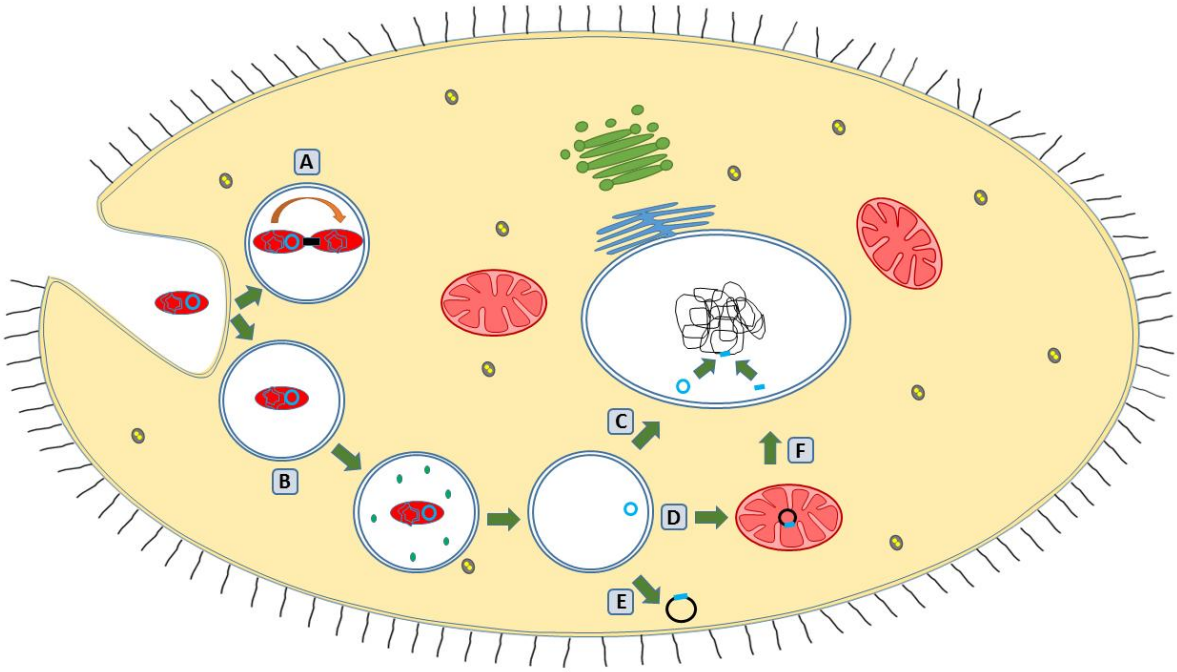


Figure 6 Horizontal transfer of prokaryotic ***β****-CA* genes from plasmids to other prokaryotic or protozoan species. (A) After endocytosis of bacteria by a protozoan ciliate, a bilayer vesicle or vacuole is formed which surrounds the bacteria. Bacterial species are able to horizontally transfer *β-CA* genes from plasmids to another recipient bacterial species by conjugation pilus. (B) Bacterial species, which entered the bilayer vacuole through endocytosis survived as intracellular endosymbionts. Therefore, horizontal transfer of prokaryotic *β-CA* genes to the protozoan genome can be followed through four possibilities: (C) Prokaryotic *β-CA* gene from plasmid is integrated into protozoan main chromosome, (D) mitochondrial DNA, or (E) mobilome. (F) In addition, integrated *β-CA* gene into circular mitochondrial DNA might have migrated from mitochondria to the main chromosome.

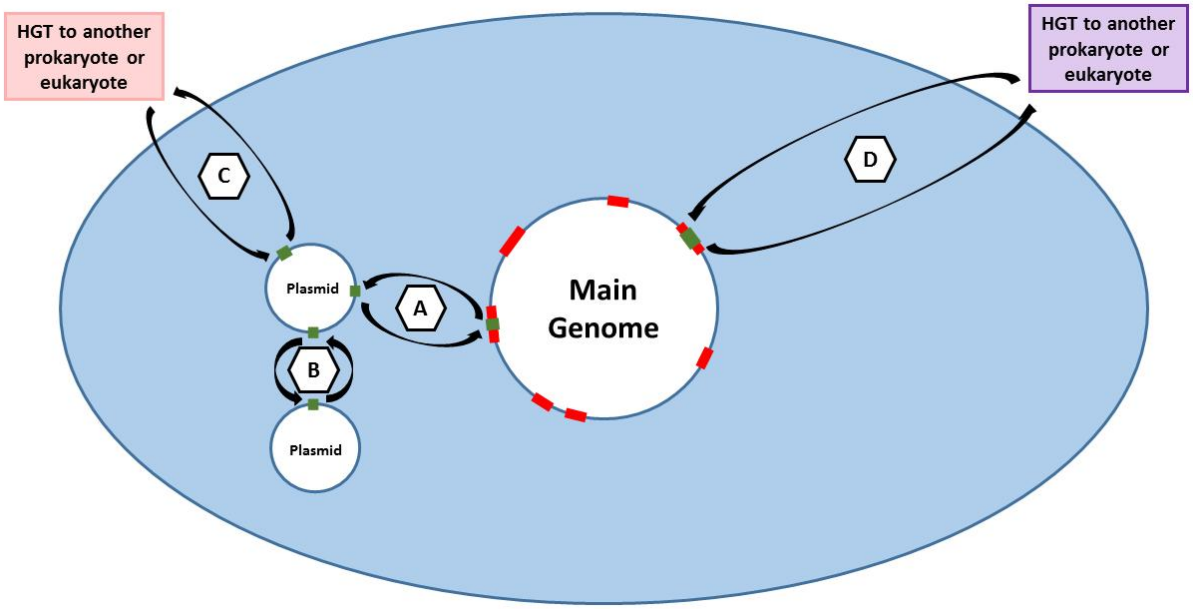


Figure 7 Schematic representation of horizontal transfer of prokaryotic ***β****-CA* genes from chromosomal GIs to other hosts. (A) Horizontal transfer of *β-CA* gene from a plasmid to the GI region of the main chromosome of the same prokaryote. (B) Horizontal transfer of *β-CA* gene between two different plasmids from the same prokaryote. (C) Horizontal transfer of *β-CA* gene from a prokaryotic plasmid to another prokaryote or eukaryote, and (D) Horizontal transfer of *β-CA* gene from the GI region of the main genome to another prokaryote, parasitic, or non-parasitic eukaryote.

Supporting information

Table S1 Archaeal or bacterial plasmids containing ***β****-CA* gene sequences and IS elements.

Archaeal or bacterial species	Plasmid	Plasmid structure	IS family	IS group	Enzymatic activity
Acaryochloris marina (strain MBIC11017)	pREB1	Circular	IS1 (ISAcma3)	ISMhu11	Transposase
			IS4 (ISAcma47)	IS10	Transposase
			IS21 (ISAcma26)	-	Transposase & Integrase
			IS256 (ISAcma7)	-	Transposase
			ISL3 (ISAcma10)	-	Transposase
Azospirillum spp. (strain B510)	pAB510c	Circular	IS21 (ISAzs2)	-	Transposase & Integrase
			IS1380 (ISAzs3)	-	Transposase
			ISAs1 (ISAzs5, ISAzs11, ISAzs12, ISAzs15)	-	Transposase
			IS701 (ISAzs6)	-	Transposase
			IS5 (ISAzs9)	IS427	Transposase
			IS66 (ISAzs20)	-	Transposase
			IS3 (ISAzs24)	IS51	Transposase & Integrase
			IS30 (ISAzs27)	-	Integrase
Bacillus cereus (strain AH1134)	pAH1134_566	Linear	ND	-	-
Bacillus thuringiensis	pBMB0231	Circular	ND	-	-
Burkholderia spp.	p1	Circular	ND	-	-
Burkholderia phymatum	pBPHY01	Circular	ND	-	-

Halopiger xanaduensis (strain SH-6)	pHALXA01	Circular	IS200/IS605 (ISHxa1) IS5 (ISHxa2)	IS1341 ISH1	Transposase Transposase
Methylobacterium radiotolerans (strain JCM 2831)	pMRAD01	Circular	IS3 (ISMra1, ISMra4)	-	Transposase & Integrase
			IS5 (ISMra2, ISMra5)	IS427	Integrase
Nitrosomonas spp.	pNAL21201	Circular	ND	-	-
Ralstonia solanacearum (strain GMI1000)	GMI1000M megaplasmid	Circular	ND	-	-
Rhizobium etli (strain CFN 42)	p42d	Circular	IS21 (ISRel5, ISRel16)	-	Transposase & Integrase
			IS630 (ISRel6)	-	Transposase & Integrase
			IS66 (ISRel8, ISRel15, ISRel19, ISRel24)	-	Transposase
			IS110 (ISRel9, ISRel25)	IS1111	Transposase
			IS5 (ISRel11, ISRel13, ISRel20)	-	Transposase
			IS3 (ISRel21)	IS407	Transposase & Integrase
	pB	Circular	ND	-	-
Rhizobium spp.	pLPU83d	Circular	ND	-	-
Rhodococcus jostii (strain RHA1)	pRHL1	Linear	IS3 (ISRhosp5)	IS51	Transposase & Integrase
Salmonella enterica, serovar *Choleraesuis* (strain OU7519)	pOU7519	Circular	IS21 (ISSen3)	-	Transposase & Integrase
			IS3 (ISSen4)	IS407	Transposase

					& Integrase
Salmonella typhimurium (strain LT2)	pSLT	Circular*	ND	-	-
Silicibacter spp.	TM1040 mega plasmid	Circular	ND	-	-
Streptomyces spp. (strain F12)	pFRL6	Linear	ND	-	-

*: Linear from of pSLT is present in *S. typhimurium* (strain STm12).

Table S2 Archaeal and bacterial chromosomal GIs containing ***β****-CA* genes and IS elements.

Bacterial species	Genomic Islands (GIs) containing **β**-CA genes	IS family	IS group	Enzymatic activity
Bifidobacterium adolescentis (ATCC 15703)	icfA NCBI ID: YP_908875.1 Sequence: 15,555-16,250	IS21	-	Integrase
		IS3	IS150	Integrase
Bradyrhizobium japonicum (Strain USDA 6)	BJ6T_77910 NCBI ID: YP_005612624.1 Sequence: 8,050,287-8,051,030	IS21 (ISFK1)	-	Transposase & Integrase
		IS630 (ISRj1)	-	Integrase
Burkholderia spp. (Strain YI23)	BYI23_C006340 NCBI ID: YP_004979218.1 Sequence: 740,540-741,148	ND	-	-
Candidatus Koribacter versatilis (Strain Ellin345)	Acid345_3958 NCBI ID: YP_593032.1 Sequence: 4,682,343-4,683,095	-	-	-
Cellulomonas fimi (ATCC 484)	Celf_2925 NCBI ID: YP_004454434.1 Sequence: 3,266,000-3,266,632	-	-	-
Corynebacterium pseudotuberculosis (Strain 316)	cynT NCBI ID: YP_005303201.1 Sequence: 601,932-602,654	-	-	-
Deinococcus maricopensis (Strain DSM 21211)	Deima_2841 NCBI ID: YP_004172135.1 Sequence: 3,019,134-3,019,817	-	-	-
Haliscomenobacter hydrossis (Strain DSM 1100)	Halhy_0512 NCBI ID: YP_004445295.1 Sequence: 657,785-658,525	-	-	-
Haliscomenobacter hydrossis (Strain DSM 1100)	Halhy_0815 NCBI ID: YP_004445295.1 Sequence: 1,020,594-1,021,334	-	-	-
Methylobacterium spp. (Strain 4-46)	M446_3323 NCBI ID: YP_001770160.1 Sequence: 3,693,418-3,694,149	IS6 (ISMtsp1, ISMtsp2, ISMtsp4)	-	Transposase & Integrase

		IS3 (ISMtsp5, ISMtsp18)	IS2	Integrase
		IS110 (ISMtsp6, ISMtsp7, ISMtsp17)	IS1111	Transposase
		IS21 (ISMtsp8)	-	Integrase
		IS256 (ISMtsp13)	-	Transposase
		IS5 (ISMtsp14, ISMtsp20)	IS1031	Transposase
		IS481 (ISMtsp15, ISMtsp16)	-	Integrase
		IS701 (ISMtsp19)	-	Transposase
Methylomonas methanica (Strain MC09)	Metme_4307 NCBI ID: YP_004515155.1 Sequence: 4,747,353-4,748,012	-	-	-
Mycobacterium africanum (Strain GM041182)	MAF_36010 NCBI ID: YP_004725229.1 Sequence: 4,011,663-4,012,286	-	-	-
Mycobacterium bovis (Strain AF2122/97)	Mb3619c NCBI ID: NP_857258.1 Sequence: 3,973,086-3,973,709	-	-	-
Mycobacterium bovis BCG (Strain Korea 1168P)	K60_037270 NCBI ID: YP_007432785.1 Sequence: 4,011,005-4,011,628	-	-	-
Mycobacterium bovis BCG (strain Mexico)	canB NCBI ID: YP_005173102.1 Sequence: 3,988,124-3,988,747	-	-	-
Mycobacterium bovis BCG (Strain Pasteur 1173P2)	canB	-	-	-

	NCBI ID: YP_979732.1 Sequence: 4,008,759-4,009,382			
Mycobacterium bovis BCG (strain Tokyo 172)	canB NCBI ID: YP_002646694.1 Sequence: 4,006,121-4,006,744	IS21 (ISMbo1)	-	Integrase
Mycobacterium tuberculosis (Strain 7199-99)	MT7199_3651 NCBI ID: YP_007353388.1 Sequence: 4,039,132-4,039,755	-	-	-
Mycobacterium tuberculosis (Strain CCDC5079)	CCDC5079_3329 NCBI ID: YP_005914751.1 Sequence: 4,016,572-4,017,195	-	-	-
Mycobacterium tuberculosis (Strain CTRI-2)	MTCTRI2_3653 NCBI ID:YP_005918673.1 Sequence: 4,015,336-4,015,959	-	-	-
Mycobacterium tuberculosis (Strain EAI5)	M943_18450 NCBI ID: YP_008227678.1 Sequence: 4,009,506-4,010,129	-	-	-
Mycobacterium tuberculosis (Strain F11)	TBFG_13621 NCBI ID: YP_001289548.1 Sequence: 4,042,652-4,043,275	-	-	-
Mycobacterium tuberculosis (Strain H37Rv)	canB NCBI ID: NP_218105.1 Sequence: 4,029,871-4,030,494	IS5 (ISMt1) IS21 (ISMt2, ISMt3)	IS427 -	Transposase Integrase
Mycobacterium tuberculosis (Strain KZN 1435)	TBMG_03627 NCBI ID: YP_003033630.1 Sequence: 4,015,341-4,015,964	-	-	-
Mycobacterium tuberculosis (Strain RGTB423)	MRGA423_22670 NCBI ID: YP_005924688.1 Sequence: 4,024,611-4,025,234	-	-	-
Mycobacterium tuberculosis (Strain Erdman)	ERDMAN_3936 NCBI ID: YP_007612385.1 Sequence: 4,010,917-4,011,540	-	-	-
Mycobacterium tuberculosis (strain Haarlem)	TBHG_03529 NCBI ID: YP_002158994.1 Sequence: 4,026,696-4,027,319	-	-	-

Proteus mirabilis (Strain BB2000)	can NCBI ID: YP_008396903.1 Sequence: 383,873-384,526	IS3 (ISPmi1)	IS51	Integrase
Pseudomonas stutzeri (Strain A1501)	PST_3234 NCBI ID: YP_001173711.1 Sequence: 3,494,946-3,495,590	IS5 (ISPst12)	-	Transposase
Pseudomonas stutzeri (Strain CCUG 29243)	A458_04495 NCBI ID: YP_006456574.1 Sequence: 1,010,741-1,011,385	ND	-	-
Pseudomonas stutzeri (Strain DSM 4166)	PSTAA_3393 NCBI ID: YP_005940001.1 Sequence: 3,589,959-3,590,603	ND	-	-
Rhizobium etli (Biovar *mimosae*, strain Mim1)	cynT NCBI ID: YP_008363860.1 Sequence: 1,118,237-1,118,941	ND	-	-
Rhizobium etli (Strain CFN 42)	RHE_CH01064 NCBI ID: YP_468599.1 Sequence: 1,114,742-1,115,446	IS4 (ISRel1)	IS4Sa	Transposase
		IS481 (ISRel2)	-	Integrase
		IS630 (ISRel6)	-	Integrase
		IS66 (ISRel15, ISRel19)	-	Transposase
Rhodomicrobium vannielii (ATCC 17100)	Rvan_1414 NCBI ID: YP_004011770.1 Sequence: 1,560,926-1,561,507	-	-	-
Saccharopolyspora erythraea (Strain NRRL2338)	SACE_0433 NCBI ID: YP_001102707.1 Sequence: 487,286-487,819	IS110 (ISSer1)	-	Transposase
		IS256 (ISSer2)	-	Transposase
Salinispora arenicola (Strain CNS-205)	Sare_1607 NCBI ID: YP_001536491.1 Sequence: 1,836,918-1,837,652	IS5 (ISSar1)	ISL2	Transposase
Stackebrandtia nassauensis (Strain DSM 44728)	Snas_2345 NCBI ID: YP_003511124.1 Sequence: 2,485,950-2,486,648	-	-	-
Sulfolobus islandicus (Strain HVE10/4)	SiH_0575	ND	-	-

	NCBI ID: YP_005645149.1 Sequence: 558,791-559,405			
Vibrio furnissii (Strain NCTC 11218)	vfu_A00886 NCBI ID: YP_004992094.1 Sequence: 907,010-907,678	-	-	-
Vibrio parahaemolyticus (Biovar O1:K33, strain CDC_K4557)	M636_09385 NCBI ID: YP_008315248.1 Sequence: 236,646-237,314	IS5 (ISVpa3)	IS903	Transposase
Vibrio spp. (Strain Ex25)	VEA_002555 NCBI ID: YP_003285182.1 Sequence: 1,045,195-1,045,863	ND	-	-
Vibrio vulnificus (Strain CMCP6)	VV1_1637 NCBI ID: NP_760528.1 Sequence: 1,605,648-1,606,316	IS1 (ISVvu1)	-	Transposase
		IS4 (ISVvu2)	-	Transposase
Vibrio vulnificus (Strain MO6-24/O)	VVMO6_00537 NCBI ID: YP_004187762.1 Sequence: 597,960-598,628	IS1 (ISVvu1)	-	Transposase
Yersinia pestis (Strain Angola)	YpAngola_A3272 NCBI ID: YP_001607619.1 Sequence: 3,471,564-3,472,241	IS1 (ISYps7)	-	Transposase
Yersinia pestis (Strain Antiqua)	YPA_0450 NCBI ID: YP_650363.1 Sequence: 522,628-523,383	IS1 (ISYps7)	-	Transposase
Yersinia pestis (Biovar *Medievalis*, strain Harbin 35)	YPC_1002 NCBI ID: YP_005622966.1 Sequence: 1,009,190-1,009,945	-	-	-
Yersinia pestis (Biovar *Microtus,* strain 91001)	cynT2 NCBI ID: NP_994143.1 Sequence: 3,151,234-3,151,989	IS1 (ISYps7)	-	Transposase
pestis Pestoides (Strain F)	YPDSF_0692 NCBI ID: YP_001162071.1 Sequence: 783,077-783,832	-	-	-
Yersinia pseudotuberculosis (Strain IP32953)	YPTB3068 NCBI ID: YP_071573.1 Sequence: 3,618,183-3,618,938	IS3 (ISYps8)	IS407	Transposase & Integrase

Table S3 Prediction of correct ***β****-CA* gene sequences from poor quality gene sequences.

Species	Original sequence IDs	Predictions	Prediction BLAST results
Naegleria gruberi	D2W492	Nagleria_gruberi_2-predicted	-
	D2W1R2	Nagleria_gruberi_1-predicted	-
	D2W4H2	Nagleria_gruberi_1-predicted	-
	-	Nagleria_gruberi_3-predicted	Novel
Burkholderia spp.	G8MGR3	-	-
	A0A060PJQ8	burkholderia_pseudomallei_1-predicted	
	-	burkholderia_pseudomallei_2-predicted	Novel
Burkholderia phymatum	B2JW96	burkholderia_pseudomallei_3-predicted	-
Methylomonas methanica	G0A2Z7	methylomonas_methanica_2-predicted	-
	-	methylomonas_methanica_1-predicted	Novel
	-	methylomonas_methanica_3-predicted	Previously identified
Oxyrrhis marina	B3U3W6	-	-
Tetrahymena thermophila	Q22U16	-	-

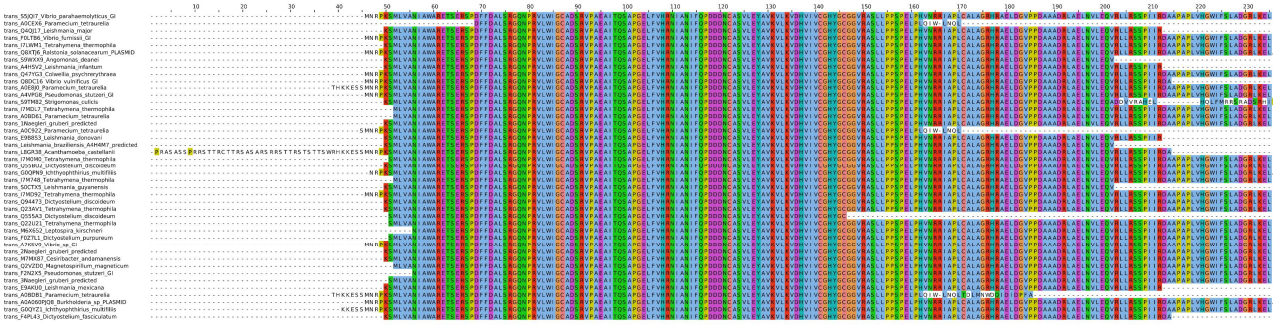


Figure S1 Prediction of **β**-CA from the *Burkholderia pseudomallei* genome using all other sequences as query templates.