A Metagenomic Analysis of Bacterial Microbiota in the Digestive Tract of Triatomines

Nicolas Carels1, Marcial Gumiel2, Fabio Faria da Mota3, Carlos José de Carvalho Moreira4 and Patrícia Azambuja2,5

1Laboratório de Modelagem de Sistemas Biológicos, National Institute for Science and Technology on Innovation in Neglected Diseases (INCT-IDN), Centro de Desenvolvimento Tecnológico em Saúde (CDTS), Fundação Oswaldo Cruz (FIOCRUZ), Rio de Janeiro, Brazil.
2Laboratório de Bioquímica e Fisiologia de Insetos, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz (IOC/FIOCRUZ), Rio de Janeiro, Brazil.
3Laboratório de Biologia Computacional e Sistemas, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz (IOC/FIOCRUZ), Rio de Janeiro, Brazil.
4Laboratório de Doenças Parasitárias, Fundação Oswaldo Cruz (IOC, FIOCRUZ), Rio de Janeiro, Brazil.
5Departamento de Entomologia Molecular, Instituto Nacional de Entomologia Molecular (INCT-EM), Rio de Janeiro, Brazil.

ABSTRACT: The digestive tract of triatomines (DTT) is an ecological niche favored by microbiota whose enzymatic profile is adapted to the specific substrate availability in this medium. This report describes the molecular enzymatic properties that promote bacterial prominence in the DTT. The microbiota composition was assessed previously based on 16S ribosomal DNA, and whole sequenced genomes of bacteria from the same genera were used to calculate the GC level of rare and prominent bacterial species in the DTT. The enzymatic reactions encoded by coding sequences of both rare and common bacterial species were then compared and revealed key functions explaining why some genera outcompete others in the DTT. Representativeness of DTT microbiota was investigated by shotgun sequencing of DNA extracted from bacteria grown in liquid Luria-Bertani broth (LB) medium. Results showed that GC-rich bacteria outcompete GC-poor bacteria and are the dominant components of the DTT microbiota. In addition, oxidoreductases are the main enzymatic components of these bacteria. In particular, nitrate reductases (anaerobic respiration), oxygenases (catabolism of complex substrates), acetate-CoA ligase (tricarboxylic acid cycle and energy metabolism), and kinase (signaling pathway) were the major enzymatic determinants present together with a large group of minor enzymes including hydrogenases involved in energy and amino acid metabolism. In conclusion, despite their slower growth in liquid LB medium, bacteria from GC-rich genera outcompete the GC-poor bacteria because their specific enzymatic abilities impart a selective advantage in the DTT.

KEYWORDS: GC content, genome size, gene number, EC number, ecological niche, midgut

Introduction

Chagas disease remains a serious health concern in South American countries, with approximately 8 million people in the chronic phase of this parasitosis. Trypanosoma cruzi, the causative agent, is mainly transmitted to humans by insects from the Triatominae subfamily distributed throughout the American continent.1

The host-parasite relationship between T. cruzi and vertebrate hosts has been extensively studied and studies continue to develop new drugs and vaccines.2,3 In contrast, there are still few studies on the host-parasite relationship involving T. cruzi and its interaction with the microbiota in the triatomine vector gut. Pioneer work by Azambuja et al4 showed that Serratia marcescens, belonging to the family Enterobacteriaceae, is a major component of the bacterial microbiota in the digestive tract of triatomines (DTT) that may kill T. cruzi through mannose-sensitive fimbriae5,6 and could thus affect the epidemiology of Chagas disease. An investigation of the bacterial composition in the DTT was only recently undertaken at a molecular level through 16S ribosomal RNA (rDNA) characterization by da Mota et al7 and Gumiel et al.8 These 2 investigations found that the bacterial microbiota diversity is low (less than 10 major species) and varies in composition depending on the species of host triatomine. Apart from the intracellular endosymbiont genera, Arsenophonus, Wolbachia, and Candidatus Robertsioria, the major bacterial species found in the DTT were from Serratia genera and from the suborder Corynebacterineae (Mycobacterium, Rhodococcus, Gordonia, Corynebacterium, and Dietzia). Another microbiota with a low level of diversity has been described in female mosquitoes (Anopheles gambiæ and Aedes aegypti), which are also hematophagous insects.9

This low number of major bacterial species found in the DTT provides an opportunity to investigate their molecular determinants. Aside from the major bacterial species mentioned above, da Mota et al7 and Gumiel et al8 also found several bacterial species that previously have not been reported to
reproduce significantly in DTT. These species belong to the following genera: Acinetobacter, Actinomyces, Alhbaeribacter, Bradyrhizobium, Chryseobacterium, Comamonas, Diaphorobacter, Enterococcus, Eruinia, Geobacillus, Haemophilus, Hydrogenophilus, Janthinobacterium, Marinomonas, Microvirga, Pectobacterium, Protoniphacterium, Providencia, Pseudomonas, Shinella, Sphingomonas, Staphylococcus, Stenotrophomonas, Streptococcus, Streptophyta, Williamsia, and Xanthobacter. If some bacterial species reproduce optimally in the DTT, there must be a biochemical basis and it needs elucidation due to the possibility of controlling Chagas disease through paratransgenesis to reduce vector competence with genetically modified symbionts.10,11

The aim of this investigation was to identify enzymatic determinants resulting in the success of bacterial genera in the DTT, such as Serratia and the members of Corynebacterineae, in contrast to minor genera, including Staphylococcus, Streptococcus, Haemophilus, or Enterococcus. Thus, the biochemical differences were analyzed between the GC-rich (rich in guanine + cytosine) and GC-poor bacterial species found in the DTT, emphasizing specific enzymatic functions that could explain the success of the former compared with the latter bacteria.12–16

Materials and Methods
The successive methodological steps followed to analyze the genome properties and enzymatic determinants in the particular niche of DTT were as follows:

1. To set up a metagenomic model of wild microbiota in DTT from species of bacteria having their genome completely sequenced and being as close as possible (from the same genera) to those identified by 16S rDNA sequencing;
2. To characterize the gross genomic features (genome size, gene number, GC level, enzymatic annotation) from bacteria of the metagenomic model;
3. To set up a working hypothesis based on the metagenomic model to justify the enzymatic determinants that make certain bacteria outcompete others in the DTT niche;
4. To validate the inference drawn from the metagenomic model through a bench experiment using a quantitative marker. The bench experiment was a shotgun sequencing model through a bench experiment using a quantitative marker. The bench experiment was a shotgun sequencing model through a bench experiment using a quantitative marker. The bench experiment was a shotgun sequencing model through a bench experiment using a quantitative marker. The bench experiment was a shotgun sequencing model through a bench experiment using a quantitative marker.

DNA extraction from bacterial cultures
After incubation in LB medium for 48 hours at 30°C without agitation for circa 48 hours until turbidity, due to bacterial growth, became evident.

Ethics statement
The animals used for blood feeding the triatomines at FIOCRUZ were treated according to the Ethical Principles in Animal Experimentation approved by the Ethics Committee in Animal Experimentation (CEUA/FIOCRUZ) under the license numbers LW-24/2013 and following the protocol from Conselho Nacional de Experimentação Animal/Ministério de Ciência e Tecnologia. Triatomines were captured under the license L14323-7 given by the Sistema de Autorização e Informação em Biodiversidade (SISBIO) of the Instituto Chico Mendes de Conservação da Biodiversidade/Ministério do Meio Ambiente (MMA).

Triatomine colonies, gut dissection, and bacterial cultures
Triatoma infestans, Triatoma vitticeps, Panstrongylus megistus and Rhodnius neglectus are of epidemiologic importance.17 Dipetalogaster maximus has no epidemiologic importance, but presents good susceptibility to _T. cruzi_, is used in xenodiagnosis, and is limited to Southern California and Mexico where it lives in a rocky habitat in association with lizards.18

The male and female triatomines used were in the fifth instar and maintained on chicken blood over approximately 20 generations in the Laboratório de Doenças Parasitárias (Instituto Oswaldo Cruz—IOC, Fundação Oswaldo Cruz—FIOCRUZ). Triatomines were dissected 7 to 10 days after feeding by opening the dorsal side from the posterior end of the abdomen to the last thoracic segment. Meticulous dissection of the midgut (stomach and intestine) and hindgut (rectum) was performed using a sterile ultratine insulin syringe needle. Feces were obtained by abdominal compression or spontaneous ejections immediately after feeding. Guts and feces were collected together in sterile Eppendorf tubes and maintained at −20°C until use. All steps were performed under aseptic conditions.

Three guts and their feces of each _T. infestans, T. vitticeps, D. maximus, P. megistus, and R. neglectus_ were then incubated in 2-mL Eppendorf tubes filled with liquid LB (Sigma-Aldrich Brasil Ltda., Sao Paulo, Brazil) at 30°C without agitation for circa 48 hours until turbidity, due to bacterial growth, became evident.

Microbial composition of triatomine digestive tract, sequence databases, and GC content
The predominant bacterial genera identified by denaturing gradient gel electrophoresis (DGGE) in the digestive tract of _T. infestans, T. vitticeps, D. maximus, P. megistus, and R. neglectus_...
were Serratia, Erwinia, Candidatus Rebrkolberia, Providence, Pectobacterium, and Arsenophonus. Of these 5 triatomine species, the genus Triatoma had a more diverse microbiota.

A previous more detailed study of the microbiota composition from digestive tracts of Triatoma brasiliensis collected in the field showed that the most abundant bacterial genera found by 454 sequencing of 16S rDNA were Gordonia sp. (36%), Serratia sp. (18%), Myxobacterium sp. (18%), Corynebacterium (6%), and Rhodococcus sp. (6%). Serratia was the most widely distributed genus among the triatomine species investigated here (Table 1). Complete genomes were sequenced for at least one species in most of the bacterial genera identified in this work (Table 1) and can be downloaded from ftp://ftp.ncbi.nih.gov/genomes/genbank/bacteria/. The coding sequences (CDS) were retrieved from the sequences of these genomes, available in *.fna files (see Table 1), by homologous comparison (tBLASTn) with their protein sequences, available in *.faa files (ent—stands for the name of the bacterial species under consideration). The average GC content was then calculated for (1) whole CDS or for (2) the first (GC1), second (GC2), and third (GC3) codon positions of each genome set, using a Perl script. When a complete genome sequence was not available for a given bacterial genus, as in the case of Arsenophonus sp. and Dietzia sp., a CDS sample was retrieved from GenBank (release 208—June 15, 2015) using the Infobiongen server (see http://www.infobiongen.fr) and the ACNUC/QUERY retrieval system19 with the options * = cds. The CDS samples used here for the GCx (x = 1, 2, 3, or the average of them) calculations are from bacterial species that, in most cases, were not the same as those diagnosed by da Mota et al2 and Gumiel et al8 although the GC content obtained from these species was considered representative of the genus to which they belonged. Reference was made to Takahashi et al13 who stated that the construction of phylogenetic trees based on oligonucleotide frequency of bacterial species with similar GC contents led to topologies that were congruent at genus and family levels with those constructed from homologous genes. The bacterial genomes in GC-poor and GC-rich species were divided according to whether their GC3 was lower or higher than 50%.

Enzymatic profiling of bacterial genomes used as references

For this study, the classification Enzyme Commission (EC) numbers from the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) was used. If different enzymes catalyze the same reaction, then they receive the same EC number. Approximately 5500 enzyme reactions have already been classified according to a 4-digit hierarchy that is used to progressively refine classification descriptions. Briefly, the first digit reports on the type of reaction considered, which is divided into 6 main categories: Oxidoreductases, transferases, hydrolases, lyases, isomerases, and ligases. The second digit of the EC number describes the type of chemical object the reaction is acting on, whereas the third digit often describes the type of donor or acceptor group. Finally, the fourth digit associates the enzyme with its reaction name (see a complete description at http://www.chem.qmul.ac.uk/submb/enzyme/).

Protein sequences of files (1) NC_018221.faa (Enterococcus fecalis), NC_010519.faa (Haemophilus somnus), NC_007350.faa (Staphylococcus saprophyticus), NC_015291.faa (Streptococcus oralis) and (2) NC_010612.faa (Mycobacterium marinum), NC_012522.faa (Rhodococcus opacus), NC_016906.faa (Gordonia polyisoprenivorans), NC_020064.faa (Serratia marcescens), and NC_021663.faa (Corynebacterium terpenotabidum) were considered representative of GC-poor and GC-rich bacterial species found in the DTT, respectively.8 By taking only the best hits (E value < 0.0001) into account, the protein sequences (BLASTp) of each of the files outlined above were compared with the enzyme sequences from the database of the Kyoto Encyclopedia of Genes and Genomes (KEGG version from June 2015; http://www.genome.jp/kegg/) where the EC numbers are available. A homologous hit was considered significant when its identity rate was at least 60% for at least 33 amino acids. For each file of the homology comparison, the EC numbers (http://www.enzyme-database.org/class.php) were grouped according to their first (6 classes), second (67 subclasses), third (264 subclasses), and fourth digits (the whole EC number set of 5549 approved enzymes as available from BREnda—online release as of June, 30, 2015; http://www.brenda-enzymes.org/all_enzymes.php; see the “Results” section). Finally, the relative frequency of EC numbers per functional category were compared between GC-poor and GC-rich genomes of bacterial species considered representative of the bacterial genera diagnosed in DTT.

Shotgun sequencing and analysis of DNA from bacterial cultures of triatomine guts

The shotgun sequencing of bacteria from DTT incubated in LB medium was performed to determine the validity of the model proposed above. At the time of the experiment, a representative amplification by polymerase chain reaction would have needed an amount of DNA that was not compatible with that obtained from direct extraction of triatomine feces. Thus, a culture step was introduced prior to shotgun library construction knowing that it would introduce a bias due to the different growth conditions in LB and DTT.

The 723 543 reads obtained by 454 sequencing were mounted into 16 435 contigs using Velvet to http://ged.msu.edu/angus/tutorials-2011/short-read-assembly-velvet.html (k = 31, ie, 31mers were looking for overlaps between reads) and further assembled with CAP3 to finally obtain 14 269 nonredundant sequences (Supplementary file S1). We then extracted the 738 297 open reading frames (ORF) from both positive and negative strands of these 14 269 sequences and filtered them out for CDS ORFs (cORFs) larger than 99 bp (base pairs) using the universal feature method (UFM),22 ending up with 35 105 cORFs compatible with the purine bias found in the CDSs.
### Table 1. Sequence materials for GC and enzymatic characterization.

| GENERA           | FREQ | WHOLE GENOME SEQUENCES                                      | GC         |
|------------------|------|-------------------------------------------------------------|------------|
| Acinetobacter²   | 2    | NA                                                          |            |
| Actinomyces²     | 1    | NA                                                          |            |
| Adhaeribacter²   | 1    | NA                                                          |            |
| Arsenophonus³    |      | Endo GenBank                                                 | Poor       |
| Bradyrhizobium²  | 1    | Bradyrhizobium_japonicum_USDA_6_u1d158851/NC_017249.fna    | Rich       |
| Chryseobacterium²| 1    | NA                                                          |            |
| Comamonas³       | 1    | Comamonas_testosteroni_CNB_2_u1d62961/NC_013446.fna        | Rich       |
| Corynebacterium² | 2015 | Corynebacterium_terpenotabidum_Y_11_u1d210639/NC_021663.fna| Rich       |
| Diaphorobacter²  | 2    | NA                                                          |            |
| Dietzia²         | 5008 | GenBank                                                     | Rich       |
| Enterococcus²    | 6    | Enterococcus_faecalalis_D32_u1d171261/NC_018221.fna         | Poor       |
| Erwinia³         | Low  | Erwinia_amylovora_ATCC_49946_u1d46943/NC_013971.fna         | Rich       |
| Geobacillus²     | 2    | NA                                                          |            |
| Gordonia²        | 11825| Gordonia_polyisopenivorans_VH2_u1d86651/NC_016906.fna       | Rich       |
| Haemophilus²     | 1    | Haemophilus_sommus_2336_u1d57979/NC_010519.fna              | Poor       |
| Hydrogenophilus² | 15   | NA                                                          |            |
| Janthinobacterium²| 1    | NA                                                          |            |
| Marinomonas²     | 1    | Marinomonas_posidonica_IVIA_Po_181_u1d67323/NC_015559.fna  | Poor       |
| Microvirga²      | 2    | NA                                                          |            |
| Mycobacterium²   | 5737 | Mycobacterium_marinum_M_u1d59423/NC_010612.fna              | Rich       |
| Pectobacterium³  | Low  | Pectobacterium_carotovorum_PC1_u1d59295/NC_012917.fna       | Rich       |
| Propionibacterium²| 5    | Propionibacterium_propionicum_F0230a_u1d170533/NC_018142.fna| Rich       |
| Pseudomonas²     | 4    | Pseudomonas_aeruginosa_RP73_u1d209328/NC_021577.fna         | Rich       |
| Rhodococcus²     | 1855 | Rhodococcus_opacus_B4_u1d13791/NC_012522.fna                | Rich       |
| Serratia²        | 4917 | Serratia_marcescens_FGI94_u1d185180/NC_020064.fna           | Rich       |
| Shinella²        | 2    | NA                                                          |            |
| Sphingomonas²    | 1    | Sphingomonas_wittichii_RW1_u1d58691/NC_009511.fna           | Rich       |
| Staphylococcus²  | 3    | Staphylococcus_saprophyticus_ATCC_15305_u1d58411/NC_007350.fna| Poor       |
| Stenotrophomonas²| 1    | Stenotrophomonas_maltophilia_D457_u1d162199/NC_017671.fna  | Rich       |
| Streptococcus²   | 1    | Streptococcus_oralis_Uo5_u1d65449/NC_015291.fna             | Poor       |
| Streptophyta²    | 2    | NA                                                          |            |
| Williamsia²      | 15   | NA                                                          |            |
| Wolbachia³       | Endo | Wolbachia_endosymbiont_of_Culex_quinquefasciatus_Pel_u1d61645/NC_010981.fna | Poor       |
| Xanthobacter²    | 1    | Xanthobacter_autotrophicus_Py2_u1d58453/NC_009720.fna       | Rich       |

Abbreviation: NA, not available.

*Bacterial species detected by Gumiel et al.*³

*Bacterial species detected by da Mota et al.*⁷

Freq. is for the numbers in Table 2 of Gumiel et al.⁸ indicating the maximum absolute number of times a genus was detected over each of 4 samples. Low is for the low but uncharacterized level of detection of a genus by DGGE in da Mota et al.⁷ Endo is for the high but uncharacterized level of detection of an endosymbiont by DGGE in da Mota et al.⁷
(supplementary file S2). These sequences were then compared (BLASTx) with a data set composed by the protein sequences of GC-rich bacteria from the whole genomes of _C. terrenstabanum_ (NC_021663.fna), _G. polysporaevorans_ (NC_016606.fna), _M. marinum_ (NC_010612.fna), _R. opacus_ (NC_012522.fna), and _S. marcescens_ (NC_002064.fna), downloaded from ftp://ftp.ncbi.nih.gov/genomes/Bacteria/ and filtered out the homologies for identity rates ≥ 60% over ≥ 33 amino acids. We retrieved the sequence subset corresponding to these homologies with a Perl script and compared (BLASTx) them with the subset of protein sequences from KEGG corresponding to the list of EC numbers that are more frequent in GC-rich bacteria compared with GC-poor forms according to their first 3 digits. Finally, we did the same exercise with the protein sequences from the whole genomes of GC-poor bacteria, ie, _S. apophyticus_ (NC_007350.fna), _H. somnus_ (NC_010519.fna), _H. sylvarum_ (NC_018221.fna), _E. facialis_ (NC_018221.fna), and _S. oralis_ (NC_015291.fna), and compared the results.

**Statistics**

Due to the different environmental conditions, differential growth of the GC-poor and GC-rich bacteria in DTT and LB were to be expected. Thus, a marker is needed to verify that the model matches the bench experiment (shotgun sequences). Therefore, as a maker, the ratio (proportion) of (1) enzymatic functions that are overrepresented in GC-rich bacteria compared with GC-poor ones relative to (2) the whole DNA-encoded protein sample for the type of bacteria considered (GC-poor or GC-rich) were chosen. Because the relative ratio of these enzymatic functions is different in GC-rich and GC-poor bacteria, then rejection of the null hypothesis of equality of both proportions (1 for GC-rich and 1 for GC-poor bacteria) in the bench experiment is to be expected if it mirrors the model (where both proportions are different).

There are at least 4 different methods to test the equality of 2 proportions, but 1 based on the Z score (https://online-courses.science.psu.edu/stat414/node/268) is presented here. According to this method, the hypothesis of equality of 2 proportions \( \hat{p}_1 = \hat{p}_2 \) can be rejected if a quantity, \( Z \), is larger than a theoretical value (1.96) of reference for a probability risk \( \alpha = 0.05 \). The quantity \( Z \) is calculated using formula (1):

\[
Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p}) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \tag{1}
\]

where,

\[
\hat{p} = \frac{Y_1 + Y_2}{n_1 + n_2} \tag{2}
\]

is the proportion of successes in the 2 samples combined (\( Y_1 \) and \( Y_2 \) are the absolute frequency of success in samples 1 and 2, respectively. The sample sizes of \( Y_1 \) and \( Y_2 \) are referred to as \( n_1 \) and \( n_2 \), respectively).

**Results**

In genera of bacteria isolated from _Triatoma_ spp., those with GC-rich genomes surpass in relative number (75%) the genera of bacteria with GC-poor genomes (25%). In addition, among the genera described by Gumiel et al, 8 the most widely represented include species with GC-rich genomes, whereas the genera only marginally represented include bacterial species with GC-poor genomes (Table 2). Obviously, being GC-rich is not sufficient for a bacterium to outperform others present in the gut of triatomines because 45% of the other minor bacterial species were also GC-rich. However, because all outperforming bacteria (6 belonging to 6 genera in 6 different families with 5 from Actinomycetales and 1 from Enterobacteriales) were GC-rich, the possibility of the GC level being a key factor for these bacteria in the DTT cannot be ignored (Table 2).

On comparing GC3 with genome size and gene number for the species of the genera identified by Gumiel et al for which a complete genome sequence was available, a significant positive correlation was found for GC3 vs genome size (\( r = .66, P < .01 \)), GC3 vs gene number (\( r = .61, P < .01 \)), and genome size vs gene number (\( r = .99, P < .01 \)) (Table 3).

If the last correlation may seem trivial in bacteria, the first one is not (the second is a consequence of the first given the third). As a consequence of the positive correlation between GC3 and genome size, GC-rich genomes of DTT bacterial microbiota have a potentially more complex metabolism than that of GC-poor genomes, which seems to be an advantage in this system. A more careful analysis of Table 3 shows that _Corynebacterium_ has a small genome (at least in the species considered here), but this fact is not necessarily a contradiction because several other Corynebacterineae in DTT have large genomes. It simply suggests that _Corynebacterium_ (at least the species considered here) may be in a process of genome reduction on the basis of the enzymatic apparatus of the family.

When comparing enzymatic activities in GC-poor and GC-rich bacteria through the evaluation of their relative frequency according to the first digit of the EC numbers (Table 4), oxidoreductases might explain the success of GC-rich bacteria because they were 2 times more frequent, on average, than in GC-poor ones.

According to this observation, the enzymatic comparison of the second digit (Table 5) revealed a larger number of subcategories (6, ie, acting on the CH-CH group of donors—EC:1.3.-.-, acting on the CH-NH2 group of donors—EC:1.4.-.-, acting on single donors with incorporation of molecular oxygen—EC:1.13.-.-, acting on paired donors with incorporation or reduction of molecular oxygen—EC:1.14.-.-, acting on iron-sulfur proteins as donors—EC:1.18.-.-) with a larger EC number frequency (≥2 times more frequent) in GC-rich compared with GC-poor bacteria in oxidoreductases than in the other 4 categories of the first digit level, ie, 2 (acting on ether bonds—EC:3.3.-.-, acting on carbon–carbon bonds—EC:3.7.-.-) in hydrolases, 1 (intramolecular lyses—EC:5.5.-.-) in isomerases, and 1 (forming carbon–sulfur bonds—EC:6.2.-.-) in ligases. The largest differences of EC relative
Table 2. GC content of coding sequences associated with bacterial genera in digestive tract of triatomines.

| Phylum                | Class       | Order             | Suborder   | Family                | Species                          | N   | GC, %   | σGC  | GC1, % | σGC1 | GC2, % | σGC2 | GC3, % | σGC3 | Fr* |
|-----------------------|-------------|-------------------|------------|-----------------------|----------------------------------|-----|---------|------|---------|------|---------|------|---------|------|-----|
| Firmicutes            | Bacilli     | Bacillales        | NA         | Staphylococcaceae     | Staphylococcus saprophyticus      | 1844| Poor    | 33.8 | 4.6     | 46.8 | 5.4     | 31.7 | 4.6     | 22.9 | 3.8 |
| Proteobacteria        | α-Proteobacteria | Rickettsiales     | NA         | Anaplasmataceae       | Wolbachia sp.                     | 1037| Poor    | 34.4 | 4.5     | 44.3 | 4.8     | 32.9 | 4.3     | 26.1 | 4.3 |
| Proteobacteria        | γ-Proteobacteria | Pasteurellales    | NA         | Pasteurellaceae       | Haemophilus somnus                | 1601| Poor    | 37.5 | 5.1     | 49.5 | 5.1     | 34.9 | 4.5     | 28.3 | 5.7 |
| Firmicutes            | Bacilli     | Lactobacillales   | NA         | Enterococcaceae       | Enterococcus faecalis             | 2249| Poor    | 37.8 | 4.7     | 49.4 | 5.0     | 34.3 | 4.7     | 29.8 | 4.5 |
| Firmicutes            | Bacilli     | Lactobacillales   | NA         | Streptococcaceae      | Streptococcus oralis              | 1451| Poor    | 41.3 | 5.9     | 52.2 | 5.8     | 34.0 | 4.6     | 37.7 | 7.2 |
| Proteobacteria        | γ-Proteobacteria | Oceanospirillales | NA         | Oceanospirillaceae    | Marinomonas polishonica           | 2795| Poor    | 44.6 | 4.6     | 54.0 | 4.4     | 37.9 | 3.9     | 41.7 | 5.3 |
| Proteobacteria        | γ-Proteobacteria | Enterobacterales  | NA         | Enterobacteriaceae    | Arsenophonus nasoiae              | 306 | Poor    | 42.3 | 6.0     | 45.2 | 5.9     | 39.6 | 6.0     | 42.2 | 6.1 |
| Proteobacteria        | γ-Proteobacteria | Enterobacterales  | NA         | Enterobacteriaceae    | Porteobacterium carotovorum       | 3322| Rich    | 52.2 | 6.4     | 59.0 | 6.0     | 40.9 | 4.7     | 56.7 | 8.6 |
| Proteobacteria        | γ-Proteobacteria | Enterobacterales  | NA         | Enterobacteriaceae    | Erwinia amylovora                 | 2478| Rich    | 53.9 | 6.7     | 60.2 | 6.4     | 41.7 | 4.9     | 59.6 | 9.0 |
| Actinobacteria        | Actinomycetales | Actinomycesales   | Corynebacterineae | Dietziaceae            | Dietzia sp.                       | 157 | Rich    | 68.0 | 4.7     | 69.1 | 5.3     | 66.3 | 4.3     | 68.6 | 4.4 |
| Proteobacteria        | γ-Proteobacteria | Enterobacterales  | NA         | Enterobacteriaceae    | Senatia marcescens                | 3620| Rich    | 59.4 | 7.5     | 63.2 | 6.6     | 42.7 | 5.1     | 72.5 | 10.8 |
| Proteobacteria        | β-Proteobacteria | Burkholterales    | NA         | Comamonadaceae        | Comamonas testosteron              | 3941| Rich    | 62.0 | 6.2     | 65.2 | 5.3     | 45.9 | 4.9     | 74.8 | 8.3 |
| Actinobacteria        | Actinomycetales | Actinomycesales   | Corynebacterineae | Mycobacteriaceae        | Mycobacterium marinum             | 4464| Rich    | 65.6 | 5.4     | 68.0 | 4.9     | 50.6 | 5.5     | 78.0 | 5.8 |
| Proteobacteria        | α-Proteobacteria | Rhizobiales       | NA         | Bradyrhizobiales      | Bradyrhizobium japonicum          | 7029| Rich    | 64.0 | 6.1     | 65.1 | 4.9     | 47.9 | 4.9     | 78.9 | 8.6 |
| Actinobacteria        | Actinomycetales | Actinomycesales   | Corynebacterineae | Gordonaceae             | Gordonia polyisoprenovans         | 3870| Rich    | 67.1 | 5.4     | 69.0 | 5.0     | 51.1 | 5.2     | 81.3 | 6.0 |
| Actinobacteria        | Actinomycetales | Actinomycesales   | NA         | Propionibacteriaceae  | Propionibacterium propionionum    | 2292| Rich    | 66.4 | 6.3     | 68.4 | 5.5     | 49.2 | 5.4     | 81.6 | 8.0 |
| Proteobacteria        | γ-Proteobacteria | Xanthomonadales   | NA         | Xanthomonadaceae      | Stenotrophomonas maltophilia      | 3275| Rich    | 67.2 | 5.7     | 69.8 | 5.6     | 48.8 | 5.4     | 83.0 | 6.2 |
| Actinobacteria        | Actinomycetales | Actinomycesales   | Corynebacterineae | Corynebacteriaceae      | Corynebacterium terpenediabium     | 1861| Rich    | 67.3 | 5.4     | 68.8 | 5.1     | 49.7 | 5.3     | 83.3 | 5.7 |
| Actinobacteria        | Actinomycetales | Actinomycesales   | Corynebacterineae | Nocardiaceae            | Rhodococcus opacus                | 5008| Rich    | 67.9 | 5.7     | 69.7 | 4.9     | 50.6 | 5.3     | 83.5 | 6.7 |
| Proteobacteria        | α-Proteobacteria | Rhizobiales       | NA         | Xanthobacteriaceae    | Xanthobacter autotrophicus        | 3783| Rich    | 67.6 | 6.0     | 69.1 | 5.6     | 49.7 | 5.3     | 84.0 | 7.3 |
| Proteobacteria        | γ-Proteobacteria | Pseudomonadales   | NA         | Pseudomonadaceae      | Pseudomonas aeruginosa             | 4927| Rich    | 66.6 | 6.4     | 68.7 | 5.8     | 46.8 | 5.8     | 84.3 | 7.5 |
| Proteobacteria        | α-Proteobacteria | Pseudomonadales   | NA         | Pseudomonadaceae      | Pseudomonas stutzeri              | 4069| Rich    | 68.7 | 5.9     | 69.3 | 5.6     | 50.0 | 5.2     | 86.7 | 6.9 |

Abbreviation: NA, not available.
| Fr is for the frequency of bacterial species reported in Gumiel et al.9 |
| L is for low frequency. |
| Endo is for endosymbiont. |
| H is for high frequency and is highlighted in gray background. |
| The shading regions in Table 2 is to improve the contrast between GC-rich (gray) and GC-poor (white) genomes. |
frequency, according to the second digit within those of the first digit category, between GC-poor and GC-rich bacteria were due to acting on single donors with incorporation of molecular oxygen—EC:1.13.-.- (difference of ~24 times) and acting on paired donors with incorporation or reduction of molecular oxygen—EC:1.14.-.- (difference of ~8 times). The differences due to acting on iron-sulfur proteins as donors—EC:1.18.-.- (difference of ~5 times), acting on ether bonds—EC:3.3.-.- (difference of ~6 times), acting on carbon-carbon bonds—EC:3.7.-.- (difference of ~4 times), and forming carbon-sulfur bonds—EC:6.2.-.- (difference of ~4 times) were also relatively large (the other differences being around 2 times). Thus, the functional variability of the enzymatic apparatus seems to be important for a bacterium to be able to outperform the others in the intestinal environment of triatomines. In addition, we found that even if the function acting on diphenols and related substances as donors (EC:1.10.-.-) exists only at a low rate in all GC-rich bacteria of Table 5, it is simply absent from the GC-poor bacteria found in DTT. This kind of function is an example of a larger metabolic complexity in GC-rich bacteria in the DTT environment.

The comparison of the third digit place of EC numbers (Table 6) for difference of enzymatic activity between GC-poor and GC-rich bacteria of the DTT showed a consistently larger number of enzymes involved in oxygen and nitrogen processing suggesting a much larger ability to cope with the degradation of complex substrates of higher chemical stability such as those containing aromatic rings (aryl). These enzymes can be mainly grouped under EC numbers 1.13.11.-, 1.4.3.-, 1.3.99.-, and 1.14.99.- but also to a lesser extent in 1.13.12.-, 1.14.11.-, 1.14.12.-, 1.14.13.-, 1.1.99.-, and 1.7.99.-.

Table 3. Relationships between GC3, genome size, and gene number in the representative bacterial species with complete genome sequence of bacterial genera found in the intestinal tract of triatomines.

| SPECIES                          | GC3 | GENOME, BP | GENE, NB |
|----------------------------------|-----|------------|----------|
| Staphylococcus saprophyticus     | 22.9| 2516573    | 2445     |
| Haemophilus somnus               | 28.3| 2263855    | 1980     |
| Enterococcus faecalis            | 29.8| 2987449    | 2876     |
| Streptococcus oralis             | 37.7| 1958688    | 1905     |
| Marinomonas posidonica           | 41.7| 3899938    | 3491     |
| Pectobacterium carotovorum       | 56.7| 4862911    | 4246     |
| Erwinia amylovora                | 59.6| 3805872    | 3437     |
| Serratia marcescens              | 72.5| 4858215    | 4361     |
| Comamonas testosteroni           | 74.8| 5373642    | 4802     |
| Mycobacterium marinum            | 78.0| 6636826    | 5423     |
| Bradyrhizobium japonicum         | 78.9| 9207382    | 8826     |
| Gordonia polysoprenivorans       | 81.3| 5669804    | 4945     |
| Propionibacterium propionicum    | 81.6| 3449358    | 2938     |
| Stenotrophomonas maltophilia     | 83.0| 4769154    | 4101     |
| Corynebacterium terpenobium      | 83.3| 2751232    | 2369     |
| Rhodococcus opacus               | 83.5| 7913449    | 7246     |
| Xanthobacter autotrophicus       | 84.0| 5309392    | 4746     |
| Pseudomonas aeruginosa           | 84.3| 6342033    | 5762     |
| Sphingomonas wittichii           | 86.7| 5382259    | 4850     |

Correlations

- $r_{GC3 \times \text{GenomSz}} = 0.66$
- $r_{GC3 \times \text{GeneNb}} = 0.61$
- $r_{\text{GenomSz} \times \text{GeneNb}} = 0.99$

*Correlation for GC3 vs genome size.
*Correlation for GC3 vs gene number.
*Correlation for genome size vs gene number.

The shading regions in Table 3 is to improve the contrast between GC-rich (gray) and GC-poor (white) genomes.
Table 4. Relative frequency (%) of enzymatic functions in GC-poor and GC-rich bacterial species found in triatomine digestive tract according to the first EC number digit.

| ECNO. | SS  | EF  | SO  | HS  | FS  | EF  | SO  | FS  | HS  | AVERAGE | SD  | SM  | MM  | GP  | CT  | RO  | AVERAGE | SD  | AVGCR/AVGCPa |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|---------|-----|-----|-----|-----|-----|-----|---------|-----|-------------|
| GC3,% | 22.9| 28.3| 29.8| 37  | 7.7 | 72.5| 78.0| 81.3| 83.3| 83.5    |     | 6.4 | 0.9 | 2.6 | 3.8 | 2.6 | 3.8     | 0.7 | 1.1         |
|       | 1. -.-.- | 17.9| 14.4| 12.7| 12.8| 14.4| 2.4 | 21.2| 31.6| 34.2    |     | 2.7 | 1.2 | 2.6 | 3.8 | 2.6 | 3.8     | 0.7 | 1.1         |
|       | 2. -.-.- | 31.9| 33.8| 31.6| 33.7| 32.8| 1.2 | 29.5| 31.6| 34.2    |     | 2.7 | 1.2 | 2.6 | 3.8 | 2.6 | 3.8     | 0.7 | 1.1         |
|       | 3. -.-.- | 29.3| 27.2| 35.5| 33.9| 31.5| 3.9 | 28.3| 21.2| 25.3    |     | 2.2 | 1.2 | 2.6 | 3.8 | 2.6 | 3.8     | 0.7 | 1.1         |
|       | 4. -.-.- | 8.2 | 10.0| 6.9 | 7.2 | 8.1 | 1.4 | 10.1| 8.5 | 9.6     |     | 3.3 | 1.1 | 2.6 | 3.8 | 2.6 | 3.8     | 0.7 | 1.1         |
|       | 5. -.-.- | 5.3 | 7.5 | 7.1 | 7.1 | 7.1 | 1.1 | 5.3 | 7.5 | 7.1     |     | 7.1 | 1.1 | 2.6 | 3.8 | 2.6 | 3.8     | 0.7 | 1.1         |
|       | 6. -.-.- | 7.5 | 7.1 | 6.6 | 7.3 | 7.1 | 1.1 | 5.3 | 7.5 | 7.1     |     | 7.1 | 1.1 | 2.6 | 3.8 | 2.6 | 3.8     | 0.7 | 1.1         |

Abbreviations: Ct, Corynebacterium terpentinabidum; EC no., Enzyme Commission number; Ef, Enterococcus faecalis; Gp, Gordonia polyisoprenivorans; Hs, Haemophilus somnus; Mm, Mycobacterium marinum; Ro, Rhodococcus opacus; Sm, Stenotrophomonas maltophilia; So, Staphylococcus saprophyticus; Ss, Staphylococcus oralis.

The shading regions in Table 4 is to improve the contrast between GC-rich (gray) and GC-poor (white) genomes.

With the list of EC numbers (n = 30) that are overrepresented in GC-rich bacteria according to the first 3 digits (Table 6), the protein sequences were retrieved corresponding to the EC numbers fully described on the 4 digits (n = 778) in (1) S saprophyticus (n = 39), H somnus (n = 25), E faecalis (n = 29), and S eralis (n = 11), ie, 26 EC numbers on average (σ = 11.6) for GC-poor bacteria and (2) C terpentinabidum (n = 55), G polyisoprenivorans (n = 96), M marinum (n = 73), R opacus (n = 116), and S marcescens (n = 95), ie, 87 EC numbers on average (σ = 23.0) for GC-rich bacteria. This statistic means that GC-rich bacteria have enzymes with 3.3 times more enzymatic functionalities than GC-poor ones, on average, according to the list of Table 6, which sustains the hypothesis that GC-rich bacteria outperform GC-poor bacteria because of their more complex metabolism, which seems to be an advantage in the DTT. Most enzymatic activities were found in dehydrogenases (aldehyde and amino acid), oxygenases (mono and di), and ligase (acetate-CoA), which are enzymatic activities involved in the very first steps of molecular degradation and synthesis.

Table 7 shows that among 35 enzymatic reactions that are overrepresented in GC-rich bacteria, the large majority are from oxidoreductases (74%) followed by transferases and ligases (9% each) with hydrolases and lyases in last position accounting for only 6% and 3%, respectively. A closer look at Table 7 enables understanding that the enzyme groups, which most explain the differences between GC-rich and GC-poor bacteria, are ranked by decreasing level of factor difference (AvGCr/AvGCp) and that the data of Table 7 can be reorganized as shown in Table 8.

In conclusion, we can say from the divisions in Table 8 that oxygenases (incorporation of oxygen in organic substrates) and CaA ligases (a central function in energy storage) make up the main significant differences, in comparison with the more basic metabolic functions, between the GC-poor and GC-rich bacteria. In addition, this difference emphasizes the existence of a larger metabolic variety of enzymatic systems in GC-rich bacteria than in GC-poor ones in the DTT. Also, the relative frequency of the enzymes ranked AvGCr/AvGCp < 2 is low in GC-rich bacteria, but because they are absent in GC-poor bacteria, they are probably of significance too.

The shotgun sequencing of bacteria from DTT grown in LB medium produced a total number of 723,543 readings whose size followed a bimodal distribution with the significant mode at 350 bp (Figure 1A). The average size significantly increased after contig assembling as can be seen in Figure 1B; however, most of the ORFs remained below 100 bp (Figure 1C). Filtering of cORFs with UFM resulted in a final sample of 35,105 cORFs mostly in the range of 100 to 300 bp (Figure 1D), which is in the acceptable limit to perform homology comparison.

From the shotgun sequence samples, 2233 significant homologies (best hit) were found among the putative 35,105 cORFs with the representative species of the genera of GC-rich bacteria reported by Gumiel et al. Among the 2233 sequences, 425 (19.0%) had significant homologies...
Table 5. Relative frequency (%) of enzymatic functions in GC-poor and GC-rich bacterial species found in triatomine digestive tract according to the 2 first EC number digits.

| ECNO.     | SS   | HS   | EF   | SO   | AVERAGE | SD | SM   | MM   | GP   | CT   | RO   | AVERAGE | SD | AVGCp/AVGCr^a | CLASS OF ENZYME FUNCTION                                                                 |
|-----------|------|------|------|------|----------|----|------|------|------|------|------|----------|----|----------------|----------------------------------------------------------------------------------------|
| 1.3.--    | 1.3  | 1.4  | 1.2  | 1.3  | 1.3      | 0.1| 1.6  | 5.1  | 4.6  | 2.8  | 5.2  | 3.8      | 1.6| 2.9            | Acting on the CH-CH group of donors                                                                 |
| 1.4.--    | 0.9  | 0.2  | 0.3  | 0.3  | 0.4      | 0.3| 0.9  | 0.9  | 0.9  | 1.3  | 1.6  | 1.1      | 0.3| 2.7            | Acting on the CH-NH2 group of donors (amino acid oxidoreductase)                             |
| 1.10.--   | 0.0  | 0.0  | 0.0  | 0.0  | 0.0      | 0.0| 0.4  | 0.1  | 0.2  | 0.3  | 0.1  | 0.2      | 0.1| ∞              | Acting on diphenols and related substances as donors (diphenol oxidoreductases)           |
| 1.13.--   | 0.0  | 0.1  | 0.0  | 0.0  | 0.0      | 0.1| 0.6  | 0.6  | 0.8  | 0.1  | 1.5  | 0.7      | 0.5| 24.5           | Acting on single donors with incorporation of molecular oxygen (monooxygenases)           |
| 1.14.--   | 0.8  | 0.6  | 0.1  | 0.1  | 0.4      | 0.4| 1.4  | 4.3  | 4.9  | 2.1  | 3.8  | 3.3      | 1.5| 8.2            | Acting on paired donors, with incorporation or reduction of molecular oxygen (dioxygenases) |
| 1.18.--   | 0.1  | 0.0  | 0.2  | 0.0  | 0.1      | 0.1| 0.3  | 0.2  | 0.4  | 0.4  | 0.5  | 0.4      | 0.1| 5.0            | Acting on iron-sulfur proteins as donors                                                                 |
| 3.3.--    | 0.1  | 0.0  | 0.1  | 0.0  | 0.0      | 0.0| 0.1  | 0.4  | 0.2  | 0.3  | 0.3  | 0.3      | 0.1| 6.1            | Acting on ether bonds                                                                                                           |
| 3.7.--    | 0.1  | 0.0  | 0.0  | 0.0  | 0.0      | 0.0| 0.1  | 0.1  | 0.2  | 0.0  | 0.2  | 0.1      | 0.1| 4.4            | Acting on carbon-carbon bonds                                                                                           |
| 5.5.--    | 0.0  | 0.1  | 0.0  | 0.1  | 0.1      | 0.1| 0.2  | 0.1  | 0.2  | 0.1  | 0.1  | 0.1      | 0.0| 2.2            | Intramolecular lyases                                                                                                                                |
| 6.2.--    | 0.6  | 0.5  | 0.5  | 0.1  | 0.4      | 0.2| 0.5  | 2.3  | 1.7  | 0.6  | 3.3  | 1.7      | 1.2| 4.0            | Forming carbon-sulfur bonds                                                                                                                   |

Abbreviations: Ct, Corynebacterium terpenotabidum; EC no., Enzyme Commission number; Ef, Enterococcus faecalis; Gp, Gordonia polysoprenivorans; Hs, Haemophilus somnus; Mm, Mycobacterium marinum; Ro, Rhodococcus opacus; Sm, Stenotrophomonas maltophilia; So, Streptococcus oralis; Ss, Staphylococcus saprophyticus.

^aFactor difference where AVGCp is for average of GC-rich and AVGCr is for average of GC-poor.

The shading regions in Table 5 is to improve the contrast between GC-rich (gray) and GC-poor (white) genomes and is expected to improve table readability.
Table 6. Relative frequency (%) of enzymatic functions in GC-poor and GC-rich bacterial species found in the triatomine digestive tract according to the 3 first EC number digits.

| ECNO   | SS   | HS   | EF   | SO   | AVERAGE | SD | SM   | MM   | GP   | CT   | RO   | AVERAGE | SD | AVGCR/AVGCP | CLASS OF ENZYME FUNCTION |
|--------|------|------|------|------|----------|----|------|------|------|------|------|----------|----|---------------|----------------------------|
| GC3,%  | 22.9 | 28.3 | 29.8 | 37.7 | 72.5     | 78.0| 81.3 | 83.3 | 83.5 |       |      |          |      |               | With a cytochrome as acceptor |
| 1.1.2.-| 0.0  | 0.0  | 0.0  | 0.0  | 0.0      | 0  | 0.1  | 0.2  | 0.1  | 0.1  | 0.0  | 0.1      | 0  | 0           | With oxygen as acceptor |
| 1.1.3.-| 0.0  | 0.0  | 0.1  | 0.1  | 0.1      | 0  | 0.0  | 0.2  | 0.3  | 0.0  | 0.4  | 0.2      | 0  | 2.8        | With unknown physiological acceptors |
| 1.1.99-| 0.1  | 0.0  | 0.1  | 0.0  | 0.1      | 0  | 0.5  | 0.3  | 0.3  | 0.1  | 0.2  | 0.2      | 0  | 4.8        | With NAD+ or NADP+ as acceptor |
| 1.2.1.-| 1.5  | 0.9  | 0.9  | 0.3  | 0.9      | 0.5| 1.5  | 2.1  | 2.6  | 1.8  | 3.5  | 2.3      | 0.8| 2.6        | With unknown physiological acceptors |
| 1.3.99-| 0.3  | 0.6  | 0.2  | 0.1  | 0.3      | 0.2| 0.6  | 4.5  | 3.4  | 1.8  | 4.7  | 3.0      | 1.8| 9.6        | With unknown physiological acceptors |
| 1.4.1.-| 0.6  | 0.1  | 0.3  | 0.1  | 0.3      | 0.2| 0.5  | 0.3  | 0.5  | 0.7  | 0.9  | 0.6      | 0.2| 2.0        | With NAD+ or NADP+ as acceptor |
| 1.4.3.-| 0.0  | 0.1  | 0.0  | 0.0  | 0.0      | 0  | 0.3  | 0.5  | 0.4  | 0.3  | 0.6  | 0.4      | 0.1| 14.2       | With oxygen as acceptor |
| 1.4.4.-| 0.2  | 0.0  | 0.0  | 0.0  | 0.0      | 0  | 0.1  | 0.2  | 0.0  | 0.3  | 0.1  | 0.1      | 0  | 2.8        | With disulfide as acceptor |
| 1.5.3.-| 0.0  | 0.0  | 0.0  | 0.0  | 0.0      | 0  | 0.2  | 0.1  | 0.5  | 0.1  | 0.2  | 0.2      | 0  | 2.8        | With oxygen as acceptor |
| 1.6.1- | 0.0  | 0.2  | 0.0  | 0.0  | 0.1      | 0  | 0.2  | 0.3  | 0.3  | 0.1  | 0.1  | 0.2      | 0  | 3.5        | With NAD+ or NADP+ as acceptor |
| 1.7.99-| 0.0  | 0.1  | 0.0  | 0.0  | 0.0      | 0  | 0.7  | 0.1  | 0.0  | 0.0  | 0.0  | 0.2      | 0.3| 6.7        | With unknown physiological acceptors |
| 1.9.3- | 0.0  | 0.0  | 0.1  | 0.0  | 0.0      | 0  | 0.0  | 0.1  | 0.2  | 0.3  | 0.1  | 0.1      | 0.1| 5.3        | With oxygen as acceptor |
| 1.13.11-| 0.0 | 0.1  | 0.0  | 0.0  | 0.0      | 0  | 0.5  | 0.4  | 0.6  | 0.0  | 1.2  | 0.5      | 0.4| 17.5       | With incorporation of 2 atoms of oxygen |
| 1.13.12-| 0.0 | 0.0  | 0.0  | 0.0  | 0.0      | 0  | 0.3  | 0.3  | 0.3  | 0.1  | 0.5  | 0.2      | 0.2| 18.2       | With incorporation of 1 atom of oxygen (internal monoxygenases or internal mixed function oxidases) |
| 1.14.11-| 0.0 | 0.0  | 0.0  | 0.0  | 0.0      | 0  | 0.2  | 0.1  | 0.3  | 0.0  | 0.2  | 0.1      | 0  | 4.3        | With 2-oxoglutarate as 1 donor, and incorporation of 1 atom each of oxygen into both donors |
| 1.14.12-| 0.2 | 0.0  | 0.0  | 0.0  | 0.1      | 0  | 0.2  | 0.0  | 0.2  | 0.4  | 0.4  | 0.2      | 0  | 4.5        | With NADH or NADPH as 1 donor, and incorporation of 2 atoms of oxygen into 1 donor |
| 1.14.13-| 0.5 | 0.5  | 0.1  | 0.0  | 0.3      | 0.3| 0.6  | 1.0  | 3.0  | 0.6  | 1.7  | 1.4      | 1.0| 4.9        | With NAD or NADH as 1 donor, and incorporation of 1 atom of oxygen |
| 1.14.99-| 0.1 | 0.0  | 0.0  | 0.0  | 0.0      | 0  | 0.0  | 0.1  | 0.3  | 0.3  | 0.4  | 0.2      | 0  | 9.2        | Miscellaneous |
| 1.17.7- | 0.0 | 0.1  | 0.0  | 0.0  | 0.0      | 0  | 0.1  | 0.1  | 0.1  | 0.1  | 0.0  | 0.1      | 0  | 2.1        | With an iron-sulfur protein as acceptor |
| 1.18.1- | 0.1 | 0.0  | 0.2  | 0.0  | 0.1      | 0  | 0.2  | 0.2  | 0.4  | 0.4  | 0.6  | 0.4      | 0.2| 5.1        | With NAD+ or NADP+ as acceptor |
| 2.7.11- | 0.4 | 0.0  | 0.2  | 0.3  | 0.2      | 0  | 0.2  | 0.8  | 1.1  | 0.6  | 1.3  | 0.8      | 0.4| 3.6        | Protein-serine/threonine kinases |
| 2.8.3- | 0.1 | 0.2  | 0.0  | 0.0  | 0.1      | 0  | 0.2  | 0.7  | 0.5  | 0.6  | 1.4  | 0.7      | 0  | 7.7        | CoA-transferases |
| ECNO. | SS | HS | EF | SO | AVERAGE | SD | MM | GP | CT | RO | AVERAGE | SD | AvGCr/ AvGCp | CLASS OF ENZYME FUNCTION |
|-------|----|----|----|----|----------|----|----|----|----|----|----------|----|-------------|----------------------------|
| GC3. %|    | 22.9 | 29.3 | 29.8 | 37.7 | 72.5 | 78.0 | 81.3 | 83.3 | 83.5 | 72.5 | 78.0 | 81.3 | 83.3 | 83.5 | 2.1 | Molybdenumtransferases or tungstentransferases with sulfide groups as acceptors |
|       | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.1 | 0.0 | 0.0 | 0.1 | 0.0 | 2.1 | Molybdenumtransferases or tungstentransferases with sulfide groups as acceptors |
|       | 0.0 | 0.2 | 0.0 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.0 | 0.1 | 0.1 | 0.1 | 2.4 | Sulfuric ester hydrolases |
|       | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 0.1 | 0.1 | 0.1 | 0.0 | 0.1 | 0.1 | 0.1 | 5.1 | Ether hydrolases |
|       | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.1 | 0.0 | 0.1 | 4.1 | Amine-lyases |
|       | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.1 | 0.0 | 0.1 | 0.1 | 0.0 | 0.1 | 5.1 | Ether hydrolases |
|       | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.1 | 0.0 | 0.1 | 0.1 | 0.0 | 0.1 | 5.1 | Ether hydrolases |
|       | 0.1 | 0.0 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.0 | 0.1 | 0.1 | 0.0 | 1.3 | Transposing C=C bonds |
|       | 0.0 | 0.1 | 0.0 | 0.0 | 0.1 | 0.1 | 0.1 | 0.1 | 0.0 | 0.1 | 0.1 | 0.0 | 0.1 | 5.1 | Ether hydrolases |
|       | 0.6 | 0.5 | 0.5 | 0.1 | 0.4 | 0.2 | 0.6 | 2.6 | 1.9 | 0.6 | 3.6 | 1.9 | 1.3 | 4.3 | Miscellaneous |

Abbreviations: Ct, Corynebacterium terpenotabidum; EC no., Enzyme Commission number; Ef, Enterococcus faecalis; Gp, Gordonia polyisoprenivorans; Hs, Haemophilus somnus; Mm, Mycobacterium marinum; Ro, Rhodococcus opacus; Sm, Stenotrophomonas maltophilia; So, Streptococcus oralis; Ss, Staphylococcus saprophyticus.

*Factor difference where AvGCr is for average of GC-rich and AvGCp is for average of GC-poor. The shading regions in Table 6 is to improve the contrast between GC-rich (gray) and GC-poor (white) genomes.
| ECNO | CLASS OF ENZYME FUNCTION | SS  | HS  | EF  | SD  | AVERAGE | SD | AVGCR/AVGCP | GC3,% |
|------|--------------------------|-----|-----|-----|-----|----------|---|-------------|-------|
| 1.1.2 | Choline dehydrogenase     | 1.0 | 0.0 | 1.0 | 1.0 | 2.0      | 1.0 | 0.0         | 0.0   |
| 1.2.1 | Formate dehydrogenase     | 3.0 | 0.0 | 1.0 | 0.0 | 1.4      | 0.0 | 0.0         | 0.0   |
| 1.2.2 | Aldehyde dehydrogenase (NAD+) | 6.0 | 0.0 | 1.0 | 0.0 | 1.0      | 0.0 | 0.0         | 0.0   |
| 1.2.3 | Biotin-adenylyl-carboxylase | 3.0 | 0.0 | 1.0 | 0.0 | 1.0      | 0.0 | 0.0         | 0.0   |
| 1.2.4 | Glutathione-C-tRNA reductase (phosphorylating) | 1.0 | 1.0 | 1.0 | 1.0 | 1.0      | 1.0 | 0.0         | 0.0   |
| 1.3.1 | Succinic dehydrogenase     | 2.0 | 3.0 | 1.0 | 0.0 | 1.5      | 1.3 | 8.0         | 6.0   |
| 1.3.2 | Alanine dehydrogenase      | 2.0 | 0.0 | 0.0 | 0.0 | 0.5      | 1.0 | 2.0         | 0.0   |
| 1.4.1 | Glutamate dehydrogenase    | 2.0 | 0.0 | 0.0 | 0.0 | 0.3      | 1.0 | 1.0         | 0.0   |
| 1.4.2 | Glutamate synthase (NADP+) | 1.0 | 0.0 | 0.0 | 0.0 | 0.3      | 1.0 | 2.0         | 1.0   |
| 1.5.1 | Dimethylamine monoxygenase (N-oxide-forming) | 1.0 | 0.0 | 0.0 | 0.0 | 1.0      | 1.0 | 0.0         | 0.0   |
| 1.6.1 | Heme oxygenase (decyclizing) | 1.0 | 0.0 | 0.0 | 0.0 | 1.0      | 1.0 | 0.0         | 0.0   |
Table 7. (Continued)

| EC NO. | SS | HS | EF | SO | AVERAGE | SD | SM | MM | GP | CT | RO | AVERAGE | SD | AvGCor/AvgCp | CLASS OF ENZYME FUNCTION |
|--------|----|----|----|----|---------|----|----|----|----|----|----|---------|----|--------------|--------------------------|
| GC3, % | 22.9 | 28.3 | 29.8 | 377 | 0.3      | 0.5 | 1.0 | 2.0 | 1.0 | 0.0 | 1.0 | 1.0      | 0.7 | 4.0          | 4-Hydroxy-3-methylbut-2-en-1-yldiphosphate synthase |
| 1.17.7.1 | 1.0 | 0.0 | 2.0 | 0.0 | 0.8      | 1.0 | 2.0 | 1.0 | 3.0 | 2.0 | 5.0 | 2.6      | 1.5 | 3.5          | Ferredoxin—NADP+ reductase |
| 1.18.1.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0      | 0.0 | 1.0 | 1.0 | 1.0 | 1.0 | 6.0 | 2.0      | 2.2 | ∞           | Ferredoxin—NAD+ reductase |
| 1.18.1.3 | 1.0 | 0.0 | 1.0 | 2.0 | 1.0      | 0.8 | 2.0 | 12.0| 12.0| 4.0 | 28.0 | 11.8     | 10.6 | 11.8        | Nonspecific serine/threonine protein kinase |
| 2.7.11.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0      | 0.0 | 1.0 | 9.0 | 4.0 | 1.0 | 17.0 | 6.4      | 6.8 | ∞           | Propionate CoA-transferase |
| 2.8.3.1  | 0.0 | 1.0 | 0.0 | 0.0 | 0.3      | 0.5 | 1.0 | 2.0 | 1.0 | 0.0 | 1.0 | 1.0      | 0.7 | 4.0          | Molybdopterin molybdotransferase |
| 3.1.6.1  | 1.0 | 0.0 | 1.0 | 0.0 | 0.5      | 0.6 | 2.0 | 5.0 | 0.0 | 1.0 | 7.0 | 3.0      | 2.9 | 6.0          | Arylsulfatase |
| 3.3.2.1  | 1.0 | 0.0 | 1.0 | 1.0 | 0.5      | 0.6 | 1.0 | 4.0 | 1.0 | 0.0 | 1.0 | 1.4      | 1.5 | 2.8          | Isochorismatase |
| 4.3.3.7  | 0.0 | 0.0 | 1.0 | 0.0 | 0.3      | 0.5 | 5.0 | 0.0 | 0.0 | 1.0 | 0.0 | 1.2      | 2.2 | 4.8          | 4-Hydroxy-tetrahydrodipicolinate synthase |
| 6.2.1.1  | 2.0 | 0.0 | 0.0 | 0.0 | 0.5      | 1.0 | 2.0 | 2.0 | 6.0 | 1.0 | 15.0 | 5.2      | 5.8 | 10.4         | Acetate-CoA ligase |
| 6.2.1.3  | 1.0 | 1.0 | 1.0 | 0.0 | 0.8      | 0.5 | 4.0 | 5.0 | 6.0 | 2.0 | 9.0 | 5.2      | 2.6 | 6.9          | Long-chain-fatty-acid-CoA ligase |
| 6.2.1.26 | 0.0 | 1.0 | 1.0 | 0.0 | 0.5      | 0.6 | 1.0 | 1.0 | 1.0 | 0.0 | 3.0 | 1.2      | 1.1 | 2.4          | α-Succinylbenzoate-CoA ligase |

Abbreviations: Ct, Corynebacterium terpenotabidum; EC no., Enzyme Commission number; Ef, Enterococcus faecalis; Gp, Gordonia polyisoprenivorans; Hs, Haemophilus somnus; Mm, Mycobacterium marinum; Ro, Rhodococcus opacus; Sm, Stenotrophomonas maltophilia; So, Staphylococcus oreaf; St, Staphylococcus saprophyticus.

*Factor difference where AVGCor is for average of GC-rich and AVGCP is for average of GC-poor.

The shading regions in Table 7 is to improve the contrast between GC-rich (gray) and GC-poor (white) genomes.
| $\text{AV}_{\text{GC}}/\text{AV}_{\text{GCP}}^{a}$ | S. NO. | EC NO. | ENZYMATIC FUNCTION |
|---------------------------------|--------|--------|---------------------|
| 10 to 16 | 1 | 1.13.11.2 | Catechol 2,3-dioxygenase |
| 2 | 1.7.99.4 | Nitrate reductase |
| 3 | 1.14.13.1 | Salicylate 1-monooxygenase |
| 4 | 2.7.11.1 | Nonspecific serine/threonine protein kinase |
| 5 | 6.2.1.1 | Acetate—CoA ligase |
| 5 to <10 | 1 | 1.2.1.27 | Methylmalonate-semialdehyde dehydrogenase |
| 2 | 1.2.1.8 | Betaine-aldehyde dehydrogenase |
| 3 | 6.2.1.3 | Long-chain-fatty-acid-CoA ligase |
| 4 | 3.1.6.1 | Arylsulfatase |
| 5 | 1.1.99.1 | Choline dehydrogenase |
| 6 | 1.2.1.3 | Aldehyde dehydrogenases |
| 7 | 1.13.11.24 | Quercetin 2,3-dioxygenase |
| 8 | 1.14.99.3 | Heme oxygenase—biliverdin-producing |
| 9 | 1.4.1.13 | Glutamate synthase—NADPH |
| 10 | 1.4.1.1 | Alanine dehydrogenase |
| 4 to <5 | 1 | 4.3.3.7 | 4-Hydroxy-tetrahydrodipicolinate synthase |
| 2 | 1.2.1.2 | Formate dehydrogenase |
| 3 | 1.6.1.2 | NAD(P)+ transhydrogenase—Re/Si-specific |
| 4 | 1.3.99.1 | Succinate dehydrogenase |
| 5 | 1.4.4.2 | Glycine dehydrogenase—aminomethyl-transferring |
| 6 | 1.17.7.1 | Cytidine diphosphate-4-dehydro-6-deoxyglucose reductase |
| 2 to <4 | 1 | 2.10.1.1 | Molybdopterin molybdotransferase |
| 2 | 1.2.1.7 | Benzaldehyde dehydrogenase |
| 3 | 1.4.1.2 | Glutamate dehydrogenase |
| 4 | 1.18.1.2 | Ferredoxin—NADP+ reductase |
| 5 | 1.2.1.38 | N-acetyl-g-glutamyl-phosphate reductase |
| 6 | 3.3.2.1 | Isochorismatase |
| 7 | 6.2.1.26 | o-Succinylbenzoate—CoA ligase |
| 8 | 1.2.1.10 | Acetaldehyde dehydrogenase |
| 9 | 1.2.1.70 | Glutamyl-tRNA reductase |
| <2 | 1 | 1.4.3.1 | o-Aspartate oxidase |
| 2 | 1.5.3.1 | Sarcosine oxidase |
| 3 | 1.14.13.8 | Flavin-containing monooxygenase |
| 4 | 1.18.1.3 | Ferredoxin—NAD+ reductase |
| 5 | 2.8.3.1 | Propionate CoA-transferase |

Abbreviation: EC no., Enzyme Commission number.

$^{a}$Factor difference where $\text{AV}_{\text{GC}}$ is for average of GC-rich and $\text{AV}_{\text{GCP}}$ is for average of GC-poor.
with KEGG for the EC number list of Table 6. In contrast, 19,715 significant homologies (best hit) were found among the putative 35,105 cORFs with the representative species of the genera of GC-poor bacteria reported by Gumiel et al. Among the 19,715 sequences, 1,424 (7.2%) had significant homologies with KEGG for the EC number list of Table 6. Because the sample sizes are very different, one must be concerned with a statistical consistency of the factor ~2.7 (close to the theoretical value of 3.3 found with the model) of overrepresented enzymes (Tables 6 and 7) observed in GC-rich compared with GC-poor bacteria. The \( Z \) test applied to the comparison of 2 proportions allows the formal conclusion that the null hypothesis of proportion equality must be rejected because \( Z_{\text{obs}} (19) > Z_{\text{th}} (1.96) \). Thus, despite a bias introduced by LB fermentation is favorable for the growth of GC-poor bacteria, the conclusion from the shotgun DNA sequencing is that overrepresented enzymatic activities are more frequent, in relative terms, in a medium fermented by GC-rich than in a medium fermented by GC-poor bacteria, as suggested by the model analysis based on complete genome sequences available from National Center for Biotechnology Information (NCBI).

**Discussion**

A recent investigation compared the microbiota of DTT in the presence or absence of \( T. cruzi \). Globally, it showed a predominance of GC-rich bacterial species (without considering the intracellular endosymbiont \( \text{Arsenophonus} \)) as previously described except for \( \text{Staphylococcus} \) which predominated in some individuals of \( P. megistus \) and \( T. infestans \). However, the results of Díaz et al. may be equivocal because the V3-V4 hypervariable region of 16S rDNA produces a less accurate quantitative description than with the V1-V3 region used by Gumiel et al. particularly for \( \text{Staphylococcus} \). Also, with the V3-V4 region, differences between expected and observed frequencies of 10 to 300 times were reported by Zheng et al. for this latter genus, whereas the measure obtained using V1-V3 region was close to the expected value.

The GC level of a genome is an interesting variable to consider because it is robust in the sense that it is expected to be globally conserved at the level of the family rank. Thus, if one GC-poor bacterial species is present in one family, there is a major likelihood that another species of that family will also be GC-poor. The same reasoning also applies to GC-rich organisms. However, in special situations, such as in endosymbiosis where the selective constraints are not those normally encountered by the members of the family, the above tendency is violated because endosymbionts are generally GC-poor, independent of the family they belong to. The GC-poor trait of endosymbionts may be due to an evolutionary convergence induced by the peculiar constraints imposed by the intracellular environment although this is debatable. The fact that the luminal environment of DTT in which \( T. cruzi \) thrives is very

![Figure 1](image-url)

**Figure 1.** Relative frequency of shotgun sequences associated with bacteria found in the digestive tract of triatomines. (A) Size of reads obtained by 454 Titanium technology, (B) contig size after successive read assembling with Velvet and CAP3, (C) size of ORFs extracted from read contigs, and (D) size of coding ORFs after UFM filtering. cORFs indicate coding open reading frames; ORFs, open reading frames.
different compared with that of intestinal epithelial cells was suf-\ficient to eliminate the endosymbionts, Arsenophonus, Wobachia, Candidatus, and Rrobkrolbertia from the present analysis.

Predominant bacterial species are GC-rich, which raises the question of whether a cause and effect relationship exists between a bacterial species being GC-rich and its growth suc-

cess in the DTT environment. In fact, the positive corre-
lation between GC3 and genome size suggests that, in the DTT, bacterial species with GC-rich genomes have a potentially more complex metabolism than those with GC-poor genomes, which would be an advantage for the bacteria in this niche.

The most significant differences that we found between the bacterial groups were due to oxidoreductase enzymes that are much more numerous in GC-rich than GC-poor bacteria and seem to confer a metabolic advantage to GC-rich bacteria in an environment such as blood, in particular, nitrate reductases and oxygenases that are common in GC-rich bacteria.

In the enzymatic reaction involving nitrate reductases (EC 1.7.99.4, KEGG map 910), the electron transport system is similar to that of aerobic respiration. It can be complemented by vitamin K to generate the energy required to survive in anaerobic conditions. Oxygenases are enzymes that oxidize a substrate by the transference of gaseous oxygen. Dioxygenases transfer both oxygen atoms of O2 into the substrate, whereas monoxygenases, such as phenolases (cytobrome P450 oxidases), incorporate only one atom of molecular oxygen into a substrate, such as phenols, and the other atom is reduced to H2O. Oxygenases are usual in soil bacteria because oxygen reactivity plays important roles in the degradation of complex substrates. In particular, ring-cleaving dioxygenases catalyze key reactions in the aerobic microbial degradation of aromatic compounds. Many pathways converge to catecholic intermediates. An example of the degradation of a complex substrate is catechol 2,3-dioxygenases (EC 1.13.12.1) that catalyzes the opening of the benzene ring (KEGG maps 361, 362, 622, 643) and converts catechol into semialdehyde (OHC-C=COOH). Ring-cleaving dioxygenases that are active toward ring compounds belong to the cupin superfamily. Cupin-type dioxygenases also involve quercet-

inases (flavonol 2,4-dioxygenases), which open up 2 C-C bonds of the heterocyclic ring of quercetin, a widespread plant fla-

vonol. In GC-rich bacteria, several other enzymes involved in ring modification or heteroatom oxidation are also available such as (1) arylsulfatases (EC 3.1.6.1), (2) benzaldehyde dehydro-
genes—NADP (EC 1.2.1.7), and (3) flavin-containing monoxygenases (EC 1.14.13.8), which can oxidize a wide array of heteroatoms, particularly soft nucleophiles, such as amines, sulfides, and phosphites from xeno-substrates, with no common structural features, to facilitate their excretion.

In the DTT, oxygenases have been shown to allow the access to iron of bacteria that encode that enzymatic system via hemoglobin degradation with heme oxygenases—biliverdin—pro-

ducing (EC 1.14.99.3). Heme oxygenase is an enzyme that catalyzes the degradation of heme and produces biliverdin, iron, and carbon monoxide. Biliverdin is subsequently con-

verted to bilirubin by biliverdin reductases. Iron is an essential nutrient required for the survival of most bacteria. Bioavailability of iron in many environments such as soil or sea is limited by the very low solubility of the Fe3+ ion. Microbes release siderophores to scavenge iron from these mineral phases by formation of soluble Fe3+ complexes that can be taken up by active transport mechanisms. Many siderophores are nonribo-

somal peptides, although several are biosynthesized inde-

pendently. Some pathogenic bacteria, such as S marcescens, can use heme and hemoproteins as iron sources, independently of siderophore production, by mechanisms involving outer mem-

brane heme-binding proteins and heme transport systems. The iron-binding protein, transferrin, produces a marked increase in S marcescens hemolytic activity.

The levels of extracellular iron available within a host are limited, with most of the free iron being complexed to high-affinity binding proteins such as transferrin. To circumvent this low iron availability, pathogens have developed sophisticated mechanisms to use the host’s iron-containing and heme-containing proteins. The mechanism by which gram-positive bacteria, such as Corynebacterium diphtheriae, acquire heme is similar to the heme transport with siderophores and involves iron-chelating mole-

cules excreted in the bacterial environment. Once the heme has been transported across the outer membrane and is localized within the cytoplasm, it is degraded by heme oxygenase.

In contrast to oxygenases, oxidases (that reduce molecular oxygen to hydrogen peroxide or to water) and dehydrogenases (by transferring hydrogen from one substance to another) are mainly, if not exclusively, involved in energy metabolism. Many of the hydrogenases predominating in GC-rich bacteria are involved in many different central pathways such as glycolysis, the triacylglycerol acid cycle and oxidative phosphorylation that are essential to cell success in their environment.

As shown by Unrean and Srienc, a “cell system has a natu-

ral tendency to evolve with time towards an asymptotic state with maximum rate of entropy production.” In addition, De Martino et al showed that “growth rate can be explained in terms of a trade-off between the higher fitness of fast-growing phenotypes and the higher entropy of slow-growing ones.” From the results of this article and those of Unrean and Srienc and De Martino et al., it can be deduced that the success of GC-rich compared with GC-poor bacteria in the DTT is due to their enhanced ability to metabolize chemically complex substrates. The higher entropy of their metabolic networks may at least result from the predominance of hydrogenase functions in central metabolic pathways such as those for amino acid and nucleotide metabolism. In parallel with these increases in enzymatic functions, a conserved set of CoA enzymes was also found to be predominant in GC-rich bacteria and involved in different pathways such as the synthesis of chemical bond between large molecules (EC 6.2.1.1), toxic compound
degradation (EC 2.8.3.1, EC 1.2.1.10), fatty acid (EC 6.2.1.3) and amino acid (EC 4.3.3.7) metabolism.

The higher metabolic activity found in GC-rich bacteria suggests that signaling proteins should also be significantly increased. Indeed, a large difference was found for the nonspecific serine/threonine protein kinases (EC 2.7.11.1), which belong to the family of transferases, specifically protein-serine/threonine kinases. These enzymes transfer phosphates to the oxygen atom of a serine or threonine side chain in proteins. This process is called phosphorylation and is known to regulate most of the cellular pathways, especially those involved in signal transduction.60

In agreement with De Martino et al.56 the size inversion of GC-rich and GC-poor bacterial population found in the shotgun sequencing analysis is not surprising because the rich LB medium is more favorable for fast-growing bacteria with small genomes and less enzymatic abilities (lower metabolic network entropy). During experiments in this study, to have sufficient DNA, it was necessary to amplify it for sequencing and bacterial culturing were necessary steps, but, of course, at a cost of a bias. The population bias favored GC-poor bacteria and demonstrates the importance of using culture-independent techniques for in situ microbiota investigation. In this respect, the strategy of describing the microbiota composition by 16S rDNA sequencing prior to any further metagenomic description is surely the best, provided that the complete genome sequences for the metagenomes investigated are available. Thus, complete genome sequences allow the construction of a model suitable to determine what can be reasonably expected from the present experiments. Despite its bias, the shotgun analysis undertaken shows that the inferences proposed through the present model are still relevant. Therefore, the species believed to be representative of their respective genera are indeed representative in the context of this work because the proportion of predominant enzymes in the experiments is similar to that of the model.

Shotgun sequencing of microbiota is expensive, and the large amount of data provided can be difficult to analyze, especially when a eukaryote vector and blood meal source are involved as most of the sequences come from the host gut and not from the rare microbiota it contains.61 For instance, the genome of *Rhodnius prolixus* RproC1 was predicted to be about 733 Mb, whereas the average size of each bacterial genome sequence of its digestive tract is only about 4 to 5 Mb.62 Another limitation of shotgun sequencing is that the information it provides on the composition of a microbiota depends on a reference set of microbial genomes which is still only small, typically in the range of few thousand genomes.63 In contrast, large numbers of 16S rDNA gene sequences are available for comparative analyses. For example, the RDP Release 11.5 of September 30, 2016 consisted of 3356 809 aligned and annotated 16S rDNA sequences (http://rdp.cme.msu.edu/).

The 1550 bp of the 16S gene consist of 8 highly conserved regions (U1-U8) and 9 variable regions across the bacterial domain.64 Identifying the organisms populating a microbial community and their relative abundances is the typical primary objective of investigations based on 16S rDNA amplicon characterization. A similarity comparison of 16S gene sequences is usually used as the gold standard for taxonomic identification at least at the genus level.65 The characterization of 16S amplicons by DGGE has been a useful technique for rapid assessment of the composition of DTT microbiota and is particularly suitable for a first-pass comparison of multiple samples.66 However, sequencing the 16S gene is currently the most common approach used in microbial classification.67 The application of next-generation sequencing to microbial ecology has shown that the diversity in microbial populations is significantly higher than previously estimated by traditional culture-based and conventional molecular methods.67 New technologies of DNA microarray (PhyloChip, Second Genome Inc, South San Francisco, USA) are now supporting microbiota investigations by 16S rDNA classification and offer the benefit of simultaneous detection of thousands of genes in a single shot.68,69 Core genes enriched for housekeeping functions are also used to enrich classification based on 16S rDNA and to improve the resolution of microbial community structure.70

**Conclusions**

The qualitative and quantitative description of a microbiota, as adapted from Gumiel et al.,8 is more precise than a blind metagenomic analysis by DNA shotgun as long as complete genome sequences exist for the bacterial genera diagnosed by 16S rDNA. This is precisely the case in the present investigation as most of the genomes in the list of bacterial species identified by 16S rDNA sequencing by Gumiel et al8 had companion species effectively sequenced that can be downloaded from the NCBI server. The most striking differences in overrepresented enzymatic functions that are found in GC-rich bacteria (prominent in the colonization of the triatomine digestive tract compared with GC-poor ones) are for the most part due to *oxidoreductases*. We conclude that this group of enzymatic functions allows GC-rich bacteria to outcompete GC-poor ones in an environment where the fermentation of a medium such as fresh blood may need some specific metabolic activities such as iron recycling and oxygen management. In such a context, GC-rich bacteria would have a comparative advantage in the colonization of their environment, thanks to their more complex enzymatic apparatus, however, at the cost of a larger genome that is slower to replicate. In consequence, invertebrate vectors are valuable systems in which to study the properties that may favor one particular microbial community as opposed to another.71

**Acknowledgements**

The authors thank Norman A Ratcliffe for reading and editing the manuscript.

**Author Contributions**

NC, MG, FFDm, CJdCM, and PA conceived and designed the experiments. NC analyzed the data and wrote the first draft of
the manuscript. PA contributed to the writing of the manu-
script. All authors reviewed and approved the final manuscript.

Disclosures and Ethics
As a requirement of publication, author(s) have provided to the
publisher signed confirmation of compliance with legal and
ethical obligations including but not limited to the following:
authorship and contributorship, conflicts of interest, privacy
and confidentiality, and (where applicable) protection of human
and animal research subjects. The authors have read and con-
firmed their agreement with the ICMJE authorship and con-
flict of interest criteria. The authors have also confirmed that
this article is unique and not under consideration or published
in any other publication, and that they have permission from
rights holders to reproduce any copyrighted material. Any dis-
closures are made in this section. The external blind peer
reviewers report no conflicts of interest.

REFERENCES
1. World Health Organization. WHO media centre: Chagas disease (American
trypanosomiasis). http://www.who.int/mediacentre/factsheets/fs634/en/. Pub-
lished 2014. Accessed January 9, 2015.
2. Moraes CB, Giardini MA, Kim H, et al. Nitroheterocyclic compounds are more
efficacious than CYP51 inhibitors against Trypanosoma cruzi: implications for
Chagas disease drug discovery and development. Sci Rep. 2014;4:4703.
3. Quijano-Hernandez I, Dumontel E. Advances and challenges towards a vaccine
against Chagas disease. Hum Vaccin. 2011;7:1184–1191.
4. Azambuja P, Feder D, Garcia E. Isolation of Serratia marcescens in the midgut of
Rhodius prolixus: impact on the establishment of the parasite Trypanosoma cruzi in the vector. Exp Parasitol. 2004;107:89–96.
5. Castro D, Morales C, Garcia E, Azambuja P. Inhibitory effects of d-mannose on
trypanosomatid lysis induced by Serratia marcescens. Exp Parasitol. 2007;115:200–204.
6. Garcia E, Genta F, Azambuja P, Schaub GA. Interactions between intestinal com-
ponents of triatomines and Trypanosoma cruzi. Trends Parasitol. 2010;26:499–505.
7. da Mota FF, Marinho LP, Moreira CJC, et al. Cultivation-independent methods
for the genomic and metagenomic analysis of Trypanosoma cruzi. PLoS Negl Trop Dis.
2012;6:e6331.
8. Gumiel M, da Mota FF, Rizzo VS, et al. Characterization of the microbiota in
the guts of Triatoma infestans and Triatoma pseudomaculata infected by Try-
panosoma cruzi in natural conditions using culture independent methods. Parasite
Vector. 2015;8:245.
9. Dennison NJ, Jupatanakul N, Dimopoulos G. The mosquito microbiota influences
vector competence for human pathogens. Curr Opin Insect Sci. 2014;3:6–13.
10. Beard CB, Cordon-Rosales C, Durvasula RV. Bacterial symbionts of the triato-
mines and their potential use in control of Chagas disease. Trends Parasitol.
2002;18:727–734.
11. Del Rio A, Menaker J, Tisserat M. Application of a molecular biological meth-
odology to the identification and quantification of Trypanosoma cruzi in the plasma
of patients infected with Chagas disease. J Clin Microbiol. 1989;27:1767–1774.
12. Ponce de Leon M, de Miranda A, Alvarez-Valin F, Carsels N. The purine bias of
coding frame in transcriptome sequences. PLoS One. 2014;9:e86977.
13. Zerbino DR, Birney E. Velvet: algorithms for de novo short reads assembly using
de Bruijn graphs. Genome Res. 2008;18:821–829.
14. Segers FH, Kešnerová L, Kryukov K, Saitou N. Estimation of bacterial species phylogeny
and homogeneity within prokaryotes. Bioinformatics. 2009;25:141–147.
15. Miethke M, Marahiel M. Siderophore-based iron acquisition and pathogen con-
trol. Nat Rev Microbiol. 2007;5:733–744.
16. Ziegler D. An overview of the mechanism, substrate specificities, and structure
of FMOs. Drug Metab Rev. 2002;34:503–511.
17. Chiang Y, Hider RC, Kong X. Chemistry and biology of siderophores. Curr Opin
Pharm. 2011;15:682–688.
18. Acerbi A, Magrini V, Sighinolfi C, et al. The HMMER web server: an exten-
sion to relational database search. Nucleic Acids Res. 2010;38:2682–2688.
19. Durvasula RV, Sundaram RK, Kirsch P, et al. Genetic transformation of a cory-
nebacterial symbiont from the Chagas disease vector Triatoma infestans. Exp Par-
asitol. 2008;115:99–94.
20. Bentley SD, Parkhill J. Comparative genomic structure of prokaryotes. Curr
Opin Microbiol. 2010;13:382–389.
21. Zerbino DR, Birney E. Velvet: algorithms for de novo short reads assembly using
de Bruijn graphs. Genome Res. 2008;18:821–829.
22. Ponce de Leon M, de Miranda A, Alvarez-Valin F, Carsels N. The purine bias of
coding frame in transcriptome sequences. Bioinform Biol Insights. 2009;3:141–154.
23. Gouy M, Delmotte S. Remote access to ACNUC nucleotide and protein
sequence databases at PBIL. Genome Res. 2004;14:733–734.
24. Blevins CR, Bridges SC, Deatherage DE, et al. The Caenorhabditis elegans genome
project: current status and future directions. Nucleic Acids Res. 2006;34:367–380.
25. Létoffé S, Ghigo JM, Létoffé S, Ghigo JM. Heme oxygenase/monooxygenase
microsomal enzyme involved in heme metabolism. Biochem J. 2001;354:1–11.
26. Hidber RC, Kong X. Chemistry and biology of siderophores. Curr Opin
Pharm. 2011;15:682–688.
51. Poole K, Braun V. Iron regulation of Serratia marcescens hemolysin gene expres-
sion. Infect Immun. 1988;56:2967–2971.
52. Fukushima T, Allred BE, Sia AK, Nichiporuk R, Andersen UN, Raymond KN.
Gram-positive siderophore-shuttle with iron-exchange from Fe-siderophore to
aposiderophore by Bacillus cereus YxeB. Proc Natl Acad Sci U S A.
2013;110:13821–13826.
53. Matsui T, Furukawa M, Unno M, Tomita T, Ikeda-Saito M. Roles of distal Asp
in heme oxygenase from Corynebacterium diphtheriae, HmuO: a water-driven
oxygen activation mechanism. J Biol Chem. 2005;280:2981–2989.
54. Lai W, Chen H, Matsui T, et al. Enzymatic ring-opening mechanism of verdo-
heme by the heme oxygenase: a combined X-ray crystallography and QM/MM
study. J Am Chem Soc. 2010;132:12960–12970.
55. Unrean P, Srienc F. Metabolic networks evolve towards states of maximum
entropy production. Metab Eng. 2011;13:666–673.
56. De Martino D, Capuani F, De Martino A. Growth against entropy in bacterial
metabolism: the phenotypic trade-off behind empirical growth rate distributions
in E. coli. Phys Biol. 2016;13:036005. doi:10.1088/1478-3975/13/3/036005.
57. Gardner JC, Grundy FJ, Henkin TM, Escalante-Semerena JC. Control of ace-
tyl-coenzyme A synthetase (AcS) activity by acetylation/deacetylation without
NAD(+) involvement in Bacillus subtilis. J Bacteriol. 2006;188:5460–5468.
58. Manjasetty BA, Powolowsk i J, Vrielink A. Crystal structure of a bifunctional
aldolase-dehydrogenase: sequestering a reactive and volatile intermediate. Proc
Natl Acad Sci U S A. 2003;100:6992–6997.
59. Mirwaldt C, Korndörfer I, Huber R. The crystal structure of dihydrodipicolinate
 synthase from Escherichia coli at 2.5 A resolution. J Mol Biol. 1995;246:227–239.
60. Wolanin PW, Thomason PA, Stock JB. Histidine protein kinases: key signal
transducers outside the animal kingdom. Genome Biol. 2002;3:reviews3013.
doi:10.1186/gb-2002-3-10-reviews3013.
61. Fraker MH, O’Toole PW, Quigley EM. Techniques used to characterize the gut
microbiota: a guide for the clinician. Nat Rev Gastroenterol Hepatol. 2012;9:312–322.
62. Mesquita RD, Vionette-Amaral RJ, Lowenberger C, et al. Genome of Rhod-
nius prolixus, an insect vector of Chagas disease, reveals unique adaptations to
hematophagy and parasite infection. Proc Natl Acad Sci U S A. 2015;112:14936–14941.
63. Markowitz VM, Chen IM, Palaniappan K, et al. IMG: the Integrated Microbial
Genomes database and comparative analysis system. Nucleic Acids Res.
2012;40:D115–D122.
64. Jonasson J, Olofsson M, Monstein HJ. Classification, identification and subtyp-
ing of bacteria based on pyrosequencing and signature matching of 16S rDNA
fragments. APPLIS. 2002;110:263–272.
65. Clarridge JE III. Impact of 16S rRNA gene sequence analysis for identification
of bacteria on clinical microbiology and infectious diseases. Clin Microbial Rev.
2004;17:840–862.
66. Gentry TJ, Wickham GS, Schadt CW, He Z, Zhou J. Microarray applications in
microbial ecology research. Microb Ecol. 2006;52:159–175.
67. DeSantis TZ, Brodie EL, Moberg JP, Zubieta IX, Piceno YM, Andersen GL.
High-density universal 16S rRNA microarray analysis reveals broader diversity
than typical clone library when sampling the environment. Microb Ecol.
2007;53:371–383.
68. Segata N, Huttenhower C. Toward an efficient method of identifying core genes
for evolutionary and functional microbial phylogenies. PLoS ONE.
2011;6:e24704.
69. Newton IL, Sheehan KB, Lee FJ, Horton MA, Hicks RD. Invertebrate sys-
tems for hypothesis-driven microbiome research. Microbiome Sci. 2013;1:1–
9. doi:10.2478/mics-2013-0001.