Vol.58, n.3: pp. 421-430, May-June 2015
http://dx.doi.org/10.1590/S1516-8913201500474
ISSN 1516-8913 Printed in Brazil

BRAZILIAN ARCHIVES OF
BIOLOGY AND TECHNOLOGY
AN INTERNATIONAL JOURNAL

Multiple Gene Sequence Analysis Using Genes of The Bacterial DNA Repair Pathway

Miguel Rotelok Neto[1], Carolina Weigert Galvão[1], Leonardo Magalhães Cruz[2], Dieval Guizelini[2], Leilane Caline Silva[2], Jarem Raul Garcia[1] and Rafael Mazer Etto[1*]
[1]Departamento de Biologia Estrutural, Molecular e Genética; Universidade Estadual de Ponta Grossa; Ponta Grossa - PR - Brasil. [2]Departamento de Bioquímica e Biologia Molecular; Universidade Federal do Paraná; Curititba - PR - Brasil

ABSTRACT

The ability to recognize and repair abnormal DNA structures is common to all forms of life. Physiological studies and genomic sequencing of a variety of bacterial species have identified an incredible diversity of DNA repair pathways. Despite the amount of available genes in public database, the usual method to place genomes in a taxonomic context is based mainly on the 16S rRNA or housekeeping genes. Thus, the relationships among genomes remain poorly understood. In this work, an approach of multiple gene sequence analysis based on genes of DNA repair pathway was used to compare bacterial genomes. Housekeeping and DNA repair genes were searched in 872 completely sequenced bacterial genomes. Seven DNA repair and housekeeping genes from distinct metabolic pathways were selected, aligned, edited and concatenated head-to-tail to form a super-gene. Results showed that the multiple gene sequence analysis using DNA repair genes had better resolution at class level than the housekeeping genes. As housekeeping genes, the DNA repair genes were advantageous to separate bacterial groups at low taxonomic levels and also sensitive to genes derived from horizontal transfer.

Key words: Bacterial genome, multiple gene sequence analysis, taxonomic assignment, DNA repair genes, HGT

INTRODUCTION

DNA repair processes are indispensable for maintaining the integrity of genetic information in all organisms. Environmental agents such as chemicals, UV light, and ionizing radiation, as well as errors in DNA metabolism challenge the chemical structure and stability of the genome. These threats lead to a variety of alterations in the normal DNA structure such as single- and double-strand breaks, chemically modified bases, abasic sites, inter- and intra-strand cross-links, and base-pairing mismatches. Given this diversity of threats and their effects, it is not surprising that there is a corresponding diversity of DNA repair pathways (Eisen et al. 1999). The diversity of specificity, functions, and complexity of repair pathways is best understood by comparing the mechanisms of action among pathways. Such comparisons are simplified by the division of repair processes into three major classes based on general mechanism of action: (1) direct repair, in which abnormalities are chemically reversed; (2) recombinational repair, in which homologous recombination is used to repair abnormalities; and (3) excision repair, in which a section of the DNA strand containing an abnormality is removed and a repair patch is synthesized using the intact strand as a template. Within each of these classes, there are multiple types and sometimes, even subtypes of repair (Eisen et al. 1999). Despite the high diversity of DNA repair genes and its importance for

[*]Author for correspondence: mazeretto@hotmail.com

maintaining the integrity of genetic information, most studies use 16S rRNA or others housekeeping genes to establish relationships among the bacterial genomes.

The multiple gene sequence analysis has to be used to understand the relationships among genomes. There are many ways of multiple gene sequence analysis (de Queiroz et al. 1995; Huelsenbeck et al. 1996; Yang 1996; Nei et al. 2001; Suchard et al. 2003), but two fundamentally different ways are more often considered. In one, the analyses are done after the gene sequences are concatenated head-to-tail to form a super-gene alignment. In the other, the analyses are done separately for each gene and the resulting dendrograms are used to generate a consensus dendrogram. The dendrogram based on the concatenation approach has been more used because of its presumed statistical advantages and accurate generation of dendrograms, even when the concatenated sequences have evolved with very different substitution patterns (Gadagkar et al. 2005; Lang et al. 2013).

The evolution of species has been largely assumed to be strictly bifurcating, a tree-like process (Felsenstein 2004). However, many argue that such a tree-like depiction of the history of prokaryote species is not valid because of the high incidences of horizontal gene transfer (HGT), which obscures a vertical line of descent (Hilario and Gogarten 1993; Bapteste et al. 2005; Dagan et al. 2008; Koonin and Wolf 2008). In general, HGT tends to occur between the closely related organisms, more extensively between the genera and species than at the level of phylum or family (Staley 2009). In this work, bacterial genomes were compared using multiple gene sequence analysis of genes involved on DNA repair pathway. To validate this approach, the same methodology of analysis was applied also on the classical housekeeping genes. Both analyses were compared with the dendrogram based on 16S rRNA gene. The influence of possible HGT on the super-gene alignments of ubiquitous housekeeping or DNA repair genes from many completely sequenced genomes was evaluated.

MATERIAL AND METHODS

Selection of bacterial genes

The 52 housekeeping genes commonly used in the multilocus sequence typing approach and the 60 genes that encode proteins involved in DNA repair were searched in 872 bacterial genomes completely sequenced from National Center for Biotechnology Information (NCBI) database in June 2010 using the GenBank Database Explorer program (GDE) (Guizelini 2010). The ubiquity and presence in distinct pathways were the criteria used to choose the genes. Seven genes of the DNA repair and seven housekeeping genes were selected (Table 1). The genes that had more than one copy per genome were compared to the NCBI database using the BLAST tool and the copy that had the highest identity score was chosen.

Alignment and edition

Individual gene alignments were carried out with Clustal W (Larkin 2007), using the translated sequences and applying a gap opening penalty of 3.0 and a gap extension penalty of 1.8 as suggested by Hall (2011). After alignment, the amino acid sequences were converted again to nucleotides and edited with GBlocks (Catresana 2000) to remove highly divergent regions and to reduce the influence of the length of gene in the dendrogram. Fragments of approximately 500 bp were generated. The parameters used in the program GBlocks are shown in Table 1.

Table 1 - Parameter of the GBlocks program used for each gene.

Gene	Minimum block length (bp)	Permission of Gap
adk	2	none
fusA	100	none
guaA	38	none
gyrB	70	none
leuS	140	none
rplB	10	none
rpoB	40	none
dnaE	220	*50%
dut	2	none
mutS	2	none
nth	5	none
recA	144	none
ruvB	85	none
uvrC	50	none
16S rRNA	2	50%

*The gene showed low homology, so in this case was permitted the presence of 50% gaps in the sequences. The chosen parameters generated sequences of approximately 500 bp, except for 16S rDNA which was full used. Thus, it was possible to minimize the effect of the length of genes in analysis.

Concatenation of genes to form super-gene alignments

The gene sequences edited were concatenated head-to-tail to form a super-gene alignment. The order of super-DNA-repair-gene alignment was: *dnaE*, *dut*, *mutS*, *nth*, *recA*, *ruvB* and *uvrC*. The order of super-housekeeping-gene alignment was: *adk*, *fusA*, *guaA*, *gyrB*, *leuS*, *rplB* and *rpoB.*

Dendrogram of super-gene alignments

The aligned concatenated sequence of DNA repair and housekeeping genes were analyzed by MEGA 5.05 program (Tamura et al. 2011) to choose the best nucleotide substitution model. The dendrograms were constructed by Maximum Likelihood (ML). The statistical test used 500 bootstrap replications and the cutoff value was set at 65 Bootstrap values.

Dendrogram of *16S rRNA* gene

A dendrogram based only on the *16S rRNA* was constructed and compared to the dendrogram of the super-gene alignment. The *16S rRNA* gene sequences were searched in the NCBI database also with the help of the GDE program (Guizelini 2010). The *16S rRNA* sequences were aligned with Clustal W using default parameters (Larkin 2007). Alignments were edited with GBlocks.

Prediction of putative Horizontal Gene Transfer

The Alien Hunter program was used to check the possibility of horizontal gene transfer (HGT). The program analyzed the genome of the organism, predicted putative HGT events with the implementation of Interpolated Variable Order Motifs (Vernikos 2006) and marked them with a color code compatible with the Artemis genomic analysis program (Rutherford 2000).

RESULTS AND DISCUSSION

The results of DNA repair genes in the 872 completely sequenced bacterial genomes from NCBI, confirmed that the *recA* was the main DNA repair gene, because it was found in every bacterial genomes. The universality of RecA indicated both that it was an ancient gene and that its activity was irreplaceable. Despite the existence of more than 60 genes involved in DNA repair pathway, only seven genes were ubiquitous in 182 bacterial genomes. Although the DNA repair genes were indispensable for maintaining the genetic information, it was surprising that only seven genes were omnipresent in 20% bacterial genomes analyzed. Since the analyzed bacteria occupy different habitat, this result suggested that the environmental threats were responsible for maintaining the DNA repair genes diversity in bacterial genomes.

Based on an approach of gene sequences concatenation, a comparison was made between the bacterial genomes present in the NCBI database using super-gene alignments. Seven housekeeping and seven DNA repair genes present in 182 bacterial genomes were utilized, corresponding to seven phyla, fourteen classes and thirty one orders. Since *Gammaproteobacteria* is the class, which has more completely sequenced genomes deposited in databases (Lagensen 2010), it was the most representative in this analysis.

The dendrogram based on the concatenated sequences of housekeeping or repair genes was more efficient to distinct the bacterial groups at low taxonomic levels. On the other hand, the *16S rRNA* gene dendrogram showed better resolution at high taxonomic levels. The *Escherichia* and *Shigella* genera were grouped into the same branch in the *16S rRNA* dendrogram (Fig. 1) but they were separated in the housekeeping (Fig. 2) and in the DNA repair (Fig. 3) genes super-alignment dendrogram. In contrast, the housekeeping genes dendrogram showed incongruities at the class level. Some alpha proteobacteria clustered with epsilonproteobacteria and two species of *Bacteroidetes* grouped with *Chlamydiae* (Fig. 2). These clustering incongruities were not showed on the DNA-repair genes dendrogram (Fig. 3).

The housekeeping or the DNA repair genes dendrograms grouped taxonomically distant species. The *Brachyspira hyodysenteriae* from *Espirochaetes* phylum was grouped with *Clostridium botulinum* from *Firmicutes* phylum. *B. hyodysenteriae* genome analysis using the Alien Hunter program suggested that *rpoB*, *fusA*, *dut* and *dnaE* genes could have been acquired via horizontal transfer (Fig. 4). Both *B. hyodysenteriae* and *C. botulinum* inhabited the large intestine of pigs (Bellgard et al. 2009; Hafström et al. 2011). It justified the dendrogram results of housekeeping or repair genes concatenated sequences (Figs. 2 and 3). The relationship between the environmental and genetic similarity was observed also in the *Bacteroidetes* group. The *Salinibacter ruber* bacterium was distant from other bacteria of the same phylum *Bacteroidetes* (*Cytophaga hutchinsonni* and *Gramella forsetii*). However, none of the genes used in the construction of the super alignment was identified as coming from the horizontal transfer by Alien Hunter program (data not shown). The result observed by super-gene alignment could be attributed to evolutionary convergence with species of *Halobacteria* class of the *Euryarchaeota* phylum that were common in the habitat of *S. ruber* (Mongodin 2005). In the dendrogram based on *Bacteria* and *Archaea* 16S rRNA gene, it was possible to see that *S. ruber* grouped with the archaea *Haloarcula marismortui* with a strong bootstrap support and did not group with other *Bacteroidetes* species (Fig. 5).

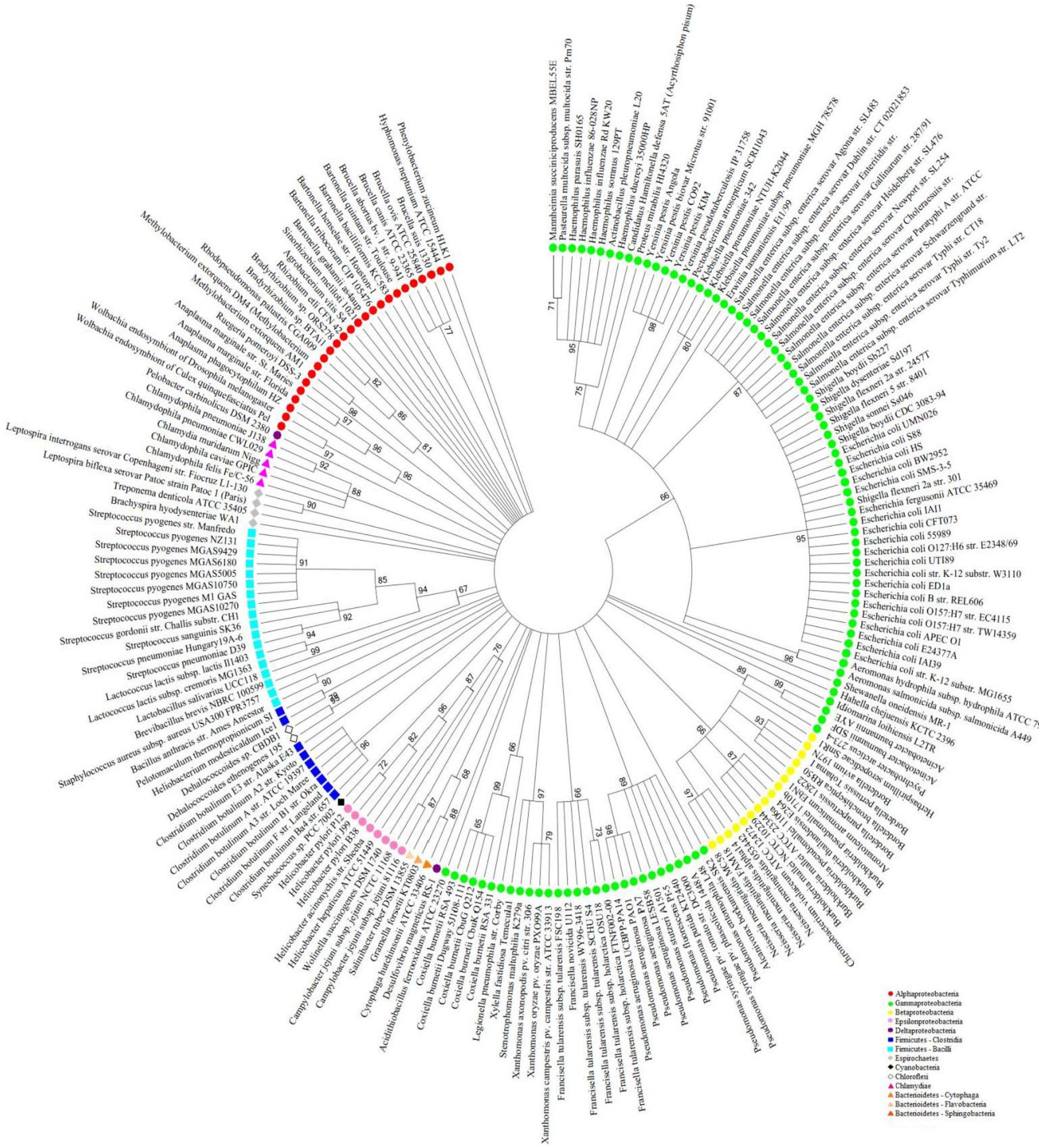
Mannheimia succiniciproducens MBEL55E
Pasteurella multocida subsp. multocida str. Pm70
Haemophilus parasuis SH0165
Haemophilus influenzae 86-028NP
Haemophilus influenzae Rd KW20
Haemophilus somnus 129PT
Actinobacillus pleuropneumoniae L20
Haemophilus ducreyi 35000HP
Candidatus Hamiltonella defensa 5AT (Acyrthosiphon pisum)
Proteus mirabilis HI4320
Yersinia pestis Angola
Yersinia pestis biovar Microtus str. 91001
Yersinia pestis CO92
Yersinia pestis KIM
Yersinia pseudotuberculosis IP 31758
Pectobacterium atrosepticum SCRI1043
Klebsiella pneumoniae 342
Klebsiella pneumoniae NTUH-K2044
Klebsiella pneumoniae subsp. pneumoniae MGH 78578
Erwinia tasmaniensis Et1/99
Salmonella enterica subsp. enterica serovar Agona str. SL483
Salmonella enterica subsp. enterica serovar Dublin str. CT 02021853
Salmonella enterica subsp. enterica serovar Enteritidis str.
Salmonella enterica subsp. enterica serovar Gallinarum str. 287/91
Salmonella enterica subsp. enterica serovar Heidelberg str. SL476
Salmonella enterica subsp. enterica serovar Newport str. SL254
Salmonella enterica subsp. enterica serovar Choleraesuis str.
Salmonella enterica subsp. enterica serovar Paratyphi A str. ATCC
Salmonella enterica subsp. enterica serovar Schwarzengrund str.
Salmonella enterica subsp. enterica serovar Typhi str. CT18
Salmonella enterica subsp. enterica serovar Typhi str. Ty2
Salmonella enterica subsp. enterica serovar Typhimurium str. LT2
Shigella boydii Sb227
Shigella dysenteriae Sd197
Shigella flexneri 2a str. 2457T
Shigella flexneri 5 str. 8401
Shigella sonnei Ss046
Shigella boydii CDC 3083-94
Escherichia coli UMN026
Escherichia coli S88
Escherichia coli HS
Escherichia coli BW2952
Escherichia coli SMS-3-5
Shigella flexneri 2a str. 301
Escherichia fergusonii ATCC 35469
Escherichia coli IAI1
Escherichia coli CFT073
Escherichia coli 55989
Escherichia coli O127:H6 str. E2348/69
Escherichia coli UTI89
Escherichia coli str. K-12 substr. W3110
Escherichia coli ED1a
Escherichia coli B str. REL606
Escherichia coli O157:H7 str. EC4115
Escherichia coli O157:H7 str. TW14359
Escherichia coli APEC O1
Escherichia coli E24377A
Escherichia coli IAI39
Escherichia coli str. K-12 substr. MG1655
Aeromonas hydrophila subsp. hydrophila ATCC 7966
Aeromonas salmonicida subsp. salmonicida A449
Shewanella oneidensis MR-1
Hahella chejuensis KCTC 2396
Idiomarina loihiensis L2TR
Acinetobacter baumannii AYE
Acinetobacter baumannii SDF
Psychrobacter arcticus 273-4
Herbaspirillum seropedicae SmR1
Bordetella avium 197N
Bordetella pertussis Tohama I
Bordetella bronchiseptica RB50
Bordetella parapertussis 12822
Aromatoleum aromaticum EbN1
Burkholderia thailandensis E264
Burkholderia pseudomallei 1710b
Burkholderia pseudomallei 1106a
Burkholderia mallei NCTC 10229
Burkholderia mallei ATCC 23344
Chromobacterium violaceum ATCC 12472
Neisseria meningitidis 053442
Neisseria meningitidis alpha14
Neisseria meningitidis FAM18
Neisseria meningitidis MC58
Alcanivorax borkumensis SK2
Pseudomonas entomophila L48
Pseudomonas syringae pv. phaseolicola 1448A
Pseudomonas syringae pv. tomato str. DC3000
Pseudomonas putida KT2440
Pseudomonas fluorescens Pf-5
Pseudomonas stutzeri A1501
Pseudomonas aeruginosa LESB58
Pseudomonas aeruginosa PA7
Pseudomonas aeruginosa PAO1
Pseudomonas aeruginosa UCBPP-PA14
Francisella tularensis subsp. holarctica
Francisella tularensis subsp. holarctica FTNF002-00
Francisella tularensis subsp. holarctica OSU18
Francisella tularensis subsp. tularensis SCHU S4
Francisella tularensis subsp. tularensis WY96-3418
Francisella novicida U112
Francisella tularensis subsp. tularensis FSC198
Xanthomonas campestris pv. campestris str. ATCC 33913
Xanthomonas oryzae pv. oryzae PXO99A
Xanthomonas axonopodis pv. citri str. 306
Stenotrophomonas maltophilia K279a
Xylella fastidiosa Temecula1
Legionella pneumophila str. Corby
Coxiella burnetii RSA 331
Coxiella burnetii CbuK Q154
Coxiella burnetii CbuG Q212
Coxiella burnetii Dugway 5J108-111
Coxiella burnetii RSA 493
Acidithiobacillus ferrooxidans ATCC 23270
Desulfovibrio magneticus RS-1
Cytophaga hutchinsonii ATCC 33406
Gramella forsetii KT0803
Salinibacter ruber DSM 13855
Campylobacter jejuni subsp. jejuni 81116
Campylobacter jejuni subsp. jejuni NCTC 11168
Wolinella succinogenes DSM 1740
Helicobacter hepaticus ATCC 51449
Helicobacter acinonychis str. Sheeba
Helicobacter pylori B38
Helicobacter pylori J99
Helicobacter pylori P12
Synechococcus sp. PCC 7002
Clostridium botulinum Ba4 str. 657
Clostridium botulinum F str. Langeland
Clostridium botulinum B1 str. Okra
Clostridium botulinum A3 str. Loch Maree
Clostridium botulinum A str. ATCC 19397
Clostridium botulinum A2 str. Kyoto
Clostridium botulinum E3 str. Alaska E43
Dehalococcoides ethenogenes 195
Dehalococcoides sp. CBDB1
Heliobacterium modesticaldum Ice1
Pelotomaculum thermopropionicum SI
Bacillus anthracis str. Ames Ancestor
Staphylococcus aureus subsp. aureus USA300 FPR3757
Brevibacillus brevis NBRC 100599
Lactobacillus salivarius UCC118
Lactococcus lactis subsp. cremoris MG1363
Lactococcus lactis subsp. lactis Il1403
Streptococcus pneumoniae D39
Streptococcus pneumoniae Hungary19A-6
Streptococcus sanguinis SK36
Streptococcus gordonii str. Challis substr. CH1
Streptococcus pyogenes MGAS10270
Streptococcus pyogenes M1 GAS
Streptococcus pyogenes MGAS10750
Streptococcus pyogenes MGAS5005
Streptococcus pyogenes MGAS6180
Streptococcus pyogenes MGAS9429
Streptococcus pyogenes NZ131
Streptococcus pyogenes str. Manfredo
Brachyspira hyodysenteriae WA1
Treponema denticola ATCC 35405
Leptospira biflexa serovar Patoc strain Patoc 1 (Paris)
Leptospira interrogans serovar Copenhageni str. Fiocruz L1-130
Chlamydophila felis Fe/C-56
Chlamydophila caviae GPIC
Chlamydia muridarum Nigg
Chlamydophila pneumoniae CWL029
Chlamydophila pneumoniae J138
Pelobacter carbinolicus DSM 2380
Wolbachia endosymbiont of Culex quinquefasciatus Pel
Wolbachia endosymbiont of Drosophila melanogaster
Anaplasma phagocytophilum HZ
Anaplasma marginale str. Florida
Anaplasma marginale str. St. Maries
Ruegeria pomeroyi DSS-3
Methylobacterium extorquens AM1
Methylobacterium extorquens DM4 (Methylobacterium
Rhodopseudomonas palustris CGA009
Bradyrhizobium sp. BTAi1
Bradyrhizobium sp. ORS278
Rhizobium etli CFN 42
Agrobacterium vitis S4
Sinorhizobium meliloti 1021
Bartonella grahamii as4aup
Bartonella tribocorum CIP 105476
Bartonella henselae str. Houston-1
Bartonella bacilliformis KC583
Bartonella quintana str. Toulouse
Brucella abortus bv. 1 str. 9-941
Brucella canis ATCC 23365
Brucella ovis ATCC 25840
Brucella suis 1330
Hyphomonas neptunium ATCC 15444
Phenylobacterium zucineum HLK1
● Alphaproteobacteria
● Gammaproteobacteria
● Betaproteobacteria
● Epsilonproteobacteria
● Deltaproteobacteria
■ Firmicutes - Clostridia
■ Firmicutes - Bacilli
◆ Espirochaetes
◆ Cyanobacteria
◇ Chloroflexi
▲ Chlamydiae
▲ Bacterioidetes - Cytophaga
▲ Bacterioidetes - Flavobacteria
▲ Bacterioidetes - Sphingobacteria

Figure 1 - Maximum Likelihood tree of bacterial 16S rRNA gene. The dendrogram was estimated by neighbor joining using the k2+G model, with MEGA 5.05 program. Support values are calculated from 500 bootstrap replicates, with a bootstrap cutoff of 65. The colored symbols represent different taxonomic groups. This representation is a radial cladogram, in which branch length is not proportional to time, and some branches may be elongated so that the names of the taxa appear on the circumference of the circle.

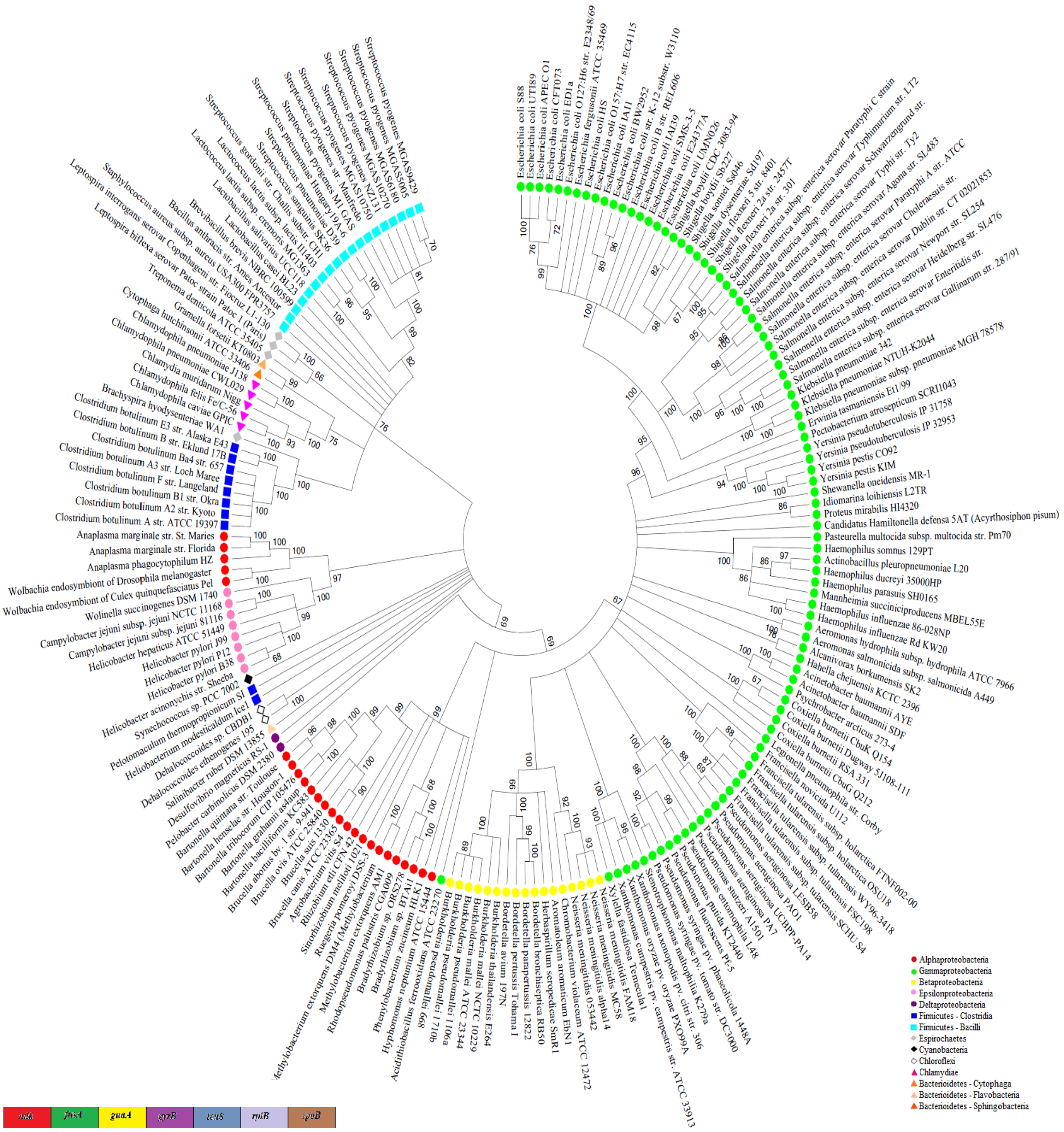
Escherichia coli S88
Escherichia coli UTI89
Escherichia coli APEC O1
Escherichia coli CFT073
Escherichia coli ED1a
Escherichia coli O127:H6 str. E2348/69
Escherichia fergusonii ATCC 35469
Escherichia coli HS
Escherichia coli O157:H7 str. EC4115
Escherichia coli IAI1
Escherichia coli BW2952
Escherichia coli str. K-12 substr. W3110
Escherichia coli B str. REL606
Escherichia coli IAI39
Escherichia coli SMS-3-5
Escherichia coli E24377A
Escherichia coli UMN026
Shigella boydii CDC 3083-94
Shigella boydii Sb227
Shigella sonnei Ss046
Shigella dysenteriae Sd197
Shigella flexneri 5 str. 8401
Shigella flexneri 2a str. 2457T
Shigella flexneri 2a str. 301
Salmonella enterica subsp. enterica serovar Paratyphi C strain
Salmonella enterica subsp. enterica serovar Typhimurium str. LT2
Salmonella enterica subsp. enterica serovar Schwarzengrund str.
Salmonella enterica subsp. enterica serovar Typhi str. Ty2
Salmonella enterica subsp. enterica serovar Agona str. SL483
Salmonella enterica subsp. enterica serovar Paratyphi A str. ATCC
Salmonella enterica subsp. enterica serovar Choleraesuis str.
Salmonella enterica subsp. enterica serovar Dublin str. CT 02021853
Salmonella enterica subsp. enterica serovar Newport str. SL254
Salmonella enterica subsp. enterica serovar Heidelberg str. SL476
Salmonella enterica subsp. enterica serovar Enteritidis str.
Salmonella enterica subsp. enterica serovar Gallinarum str. 287/91
Klebsiella pneumoniae 342
Klebsiella pneumoniae NTUH-K2044
Klebsiella pneumoniae subsp. pneumoniae MGH 78578
Erwinia tasmaniensis Et1/99
Pectobacterium atrosepticum SCRI1043
Yersinia pseudotuberculosis IP 31758
Yersinia pseudotuberculosis IP 32953
Yersinia pestis CO92
Yersinia pestis KIM
Shewanella oneidensis MR-1
Idiomarina loihiensis L2TR
Proteus mirabilis HI4320
Candidatus Hamiltonella defensa 5AT (Acyrthosiphon pisum)
Pasteurella multocida subsp. multocida str. Pm70
Haemophilus somnus 129PT
Actinobacillus pleuropneumoniae L20
Haemophilus ducreyi 35000HP
Haemophilus parasuis SH0165
Mannheimia succiniciproducens MBEL55E
Haemophilus influenzae 86-028NP
Haemophilus influenzae Rd KW20
Aeromonas hydrophila subsp. hydrophila ATCC 7966
Aeromonas salmonicida subsp. salmonicida A449
Alcanivorax borkumensis SK2
Hahella chejuensis KCTC 2396
Acinetobacter baumannii AYE
Acinetobacter baumannii SDF
Psychrobacter arcticus 273-4
Coxiella burnetii CbuK Q154
Coxiella burnetii Dugway 5J108-111
Coxiella burnetii RSA 331
Coxiella burnetii CbuG Q212
Legionella pneumophila str. Corby
Francisella novicida U112
Francisella tularensis subsp. holarctica FTNF002-00
Francisella tularensis subsp. holarctica OSU18
Francisella tularensis subsp. tularensis WY96-3418
Francisella tularensis subsp. tularensis FSC198
Francisella tularensis subsp. tularensis SCHU S4
Pseudomonas aeruginosa LESB58
Pseudomonas aeruginosa PAO1
Pseudomonas aeruginosa UCBPP-PA14
Pseudomonas aeruginosa PA7
Pseudomonas stutzeri A1501
Pseudomonas entomophila L48
Pseudomonas putida KT2440
Pseudomonas fluorescens Pf-5
Pseudomonas syringae pv. phaseolicola 1448A
Pseudomonas syringae pv. tomato str. DC3000
Stenotrophomonas maltophilia K279a
Xanthomonas axonopodis pv. citri str. 306
Xanthomonas oryzae pv. oryzae PXO99A
Xanthomonas campestris pv. campestris str. ATCC 33913
Xylella fastidiosa Temecula1
Neisseria meningitidis FAM18
Neisseria meningitidis MC58
Neisseria meningitidis alpha14
Neisseria meningitidis 053442
Chromobacterium violaceum ATCC 12472
Aromatoleum aromaticum EbN1
Herbaspirillum seropedicae SmR1
Bordetella bronchiseptica RB50
Bordetella parapertussis 12822
Bordetella pertussis Tohama I
Bordetella avium 197N
Burkholderia thailandensis E264
Burkholderia mallei NCTC 10229
Burkholderia mallei ATCC 23344
Burkholderia pseudomallei 1106a
Burkholderia pseudomallei 1710b
Burkholderia pseudomallei 668
Acidithiobacillus ferrooxidans ATCC 23270
Hyphomonas neptunium ATCC 15444
Phenylobacterium zucineum HLK1
Bradyrhizobium sp. BTAi1
Bradyrhizobium sp. ORS278
Rhodopseudomonas palustris CGA009
Methylobacterium extorquens AM1
Methylobacterium extorquens DM4 (Methylobacterium
Ruegeria pomeroyi DSS-3
Sinorhizobium meliloti 1021
Rhizobium etli CFN 42
Agrobacterium vitis S4
Brucella canis ATCC 23365
Brucella suis 1330
Brucella ovis ATCC 25840
Brucella abortus bv. 1 str. 9-941
Bartonella bacilliformis KC583
Bartonella grahamii as4aup
Bartonella tribocorum CIP 105476
Bartonella henselae str. Houston-1
Bartonella quintana str. Toulouse
Pelobacter carbinolicus DSM 2380
Desulfovibrio magneticus RS-1
Salinibacter ruber DSM 13855
Dehalococcoides ethenogenes 195
Dehalococcoides sp. CBDB1
Heliobacterium modesticaldum Ice1
Pelotomaculum thermopropionicum SI
Synechococcus sp. PCC 7002
Helicobacter acinonychis str. Sheeba
Helicobacter pylori B38
Helicobacter pylori P12
Helicobacter pylori J99
Helicobacter hepaticus ATCC 51449
Campylobacter jejuni subsp. jejuni 81116
Campylobacter jejuni subsp. jejuni NCTC 11168
Wolinella succinogenes DSM 1740
Wolbachia endosymbiont of Culex quinquefasciatus Pel
Wolbachia endosymbiont of Drosophila melanogaster
Anaplasma phagocytophilum HZ
Anaplasma marginale str. Florida
Anaplasma marginale str. St. Maries
Clostridium botulinum A str. ATCC 19397
Clostridium botulinum A2 str. Kyoto
Clostridium botulinum B1 str. Okra
Clostridium botulinum F str. Langeland
Clostridium botulinum A3 str. Loch Maree
Clostridium botulinum Ba4 str. 657
Clostridium botulinum B str. Eklund 17B
Clostridium botulinum E3 str. Alaska E43
Brachyspira hyodysenteriae WA1
Chlamydophila caviae GPIC
Chlamydophila felis Fe/C-56
Chlamydia muridarum Nigg
Chlamydophila pneumoniae CWL029
Chlamydophila pneumoniae J138
Cytophaga hutchinsonii ATCC 33406
Gramella forsetii KT0803
Treponema denticola ATCC 35405
Leptospira biflexa serovar Patoc strain Patoc 1 (Paris)
Leptospira interrogans serovar Copenhageni str. Fiocruz L1-130
Staphylococcus aureus subsp. aureus USA300 FPR3757
Bacillus anthracis str. Ames Ancestor
Brevibacillus brevis NBRC 100599
Lactobacillus casei BL23
Lactobacillus salivarius UCC118
Lactococcus lactis subsp. cremoris MG1363
Lactococcus lactis subsp. lactis Il1403
Streptococcus gordonii str. Challis substr. CH1
Streptococcus sanguinis SK36
Streptococcus pneumoniae D39
Streptococcus pneumoniae Hungary19A-6
Streptococcus pyogenes M1 GAS
Streptococcus pyogenes str. Manfredo
Streptococcus pyogenes MGAS10750
Streptococcus pyogenes NZ131
Streptococcus pyogenes MGAS10270
Streptococcus pyogenes MGAS6180
Streptococcus pyogenes MGAS5005
Streptococcus pyogenes MGAS9429
● Alphaproteobacteria
● Gammaproteobacteria
● Betaproteobacteria
● Epsilonproteobacteria
● Deltaproteobacteria
■ Firmicutes - Clostridia
■ Firmicutes - Bacilli
◆ Espirochaetes
◆ Cyanobacteria
◇ Chloroflexi
▲ Chlamydiae
▲ Bacterioidetes - Cytophaga
▲ Bacterioidetes - Flavobacteria
▲ Bacterioidetes - Sphingobacteria
adk
fusA
guaA
gyrB
leuS
rpiB
rpoB

Figure 2 - Maximum Likelihood tree of concatenated housekeeping gene. Dendrogram inferred from a concatenated, partitioned alignment of 7 housekeeping genes. The dendrograms was estimated by neighbor joining using the k2+G model, with MEGA 5.05 program. Support values were calculated from 500 rapid bootstrap replicates, with a bootstrap cutoff of 65. The colored symbols represent different taxonomic groups. This representation is a radial cladogram, in which branch length is not proportional to time, and some branches may be elongated so that the names of the taxa appear on the circumference of the circle.

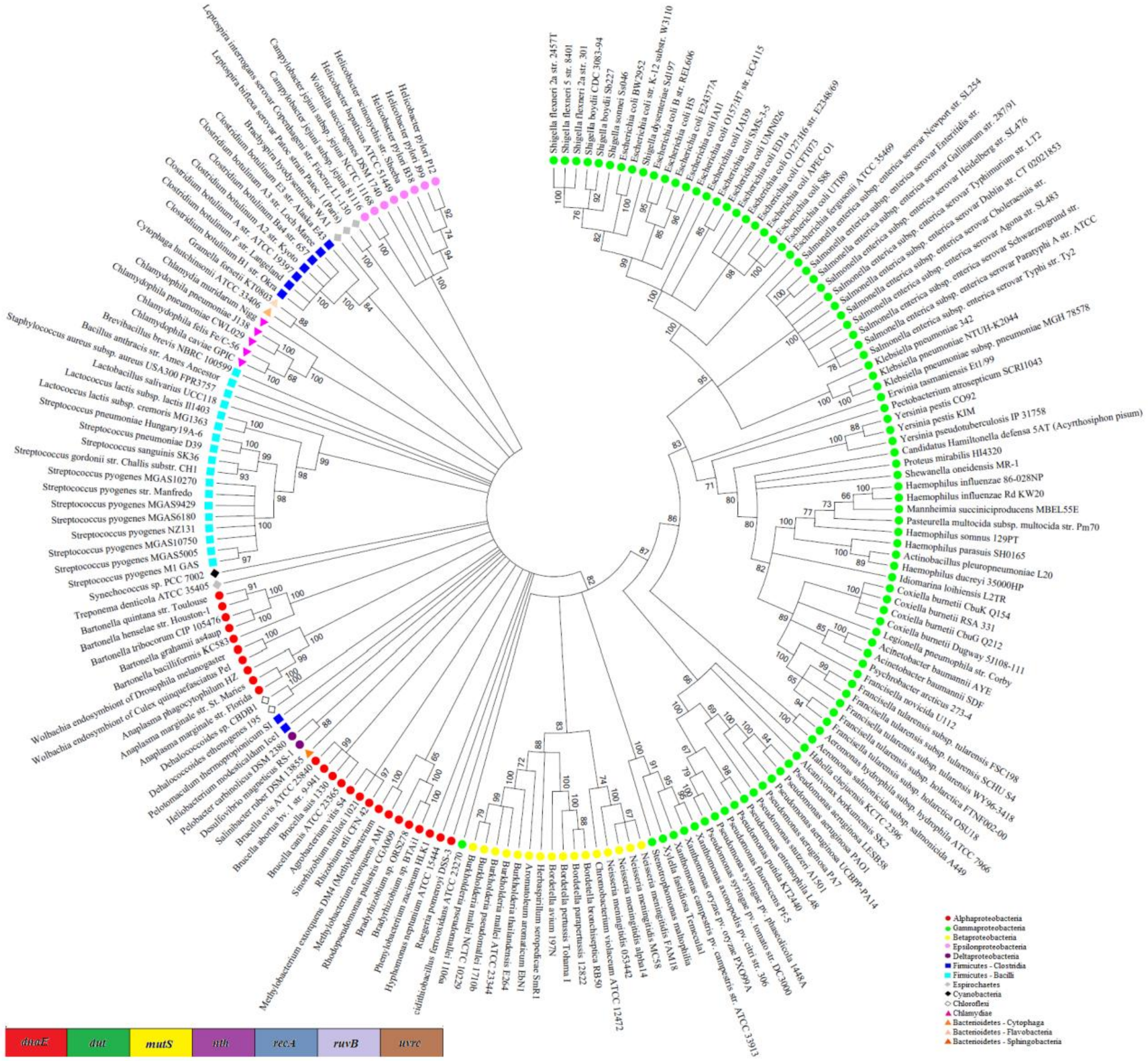

Figure 3 - Maximum Likelihood tree of concatenated DNA repair gene. Dendrogram inferred from a concatenated, partitioned alignment of 7 DNA repair gene. The dendrograms was estimated by neighbor joining using the k2+G model, with MEGA 5.05 program. Support values were calculated from 500 rapid bootstrap replicates, with a bootstrap cutoff of 65. The colored symbols represent different taxonomic groups. This representation is a radial cladogram, in which branch length is not proportional to time, and some branches may be elongated so that the names of the taxa appear on the circumference of the circle.

In all dendrograms, the thermophilic bacterium *Pelotomaculum thermopropionicum*, present in anaerobic reactors (Kosaka et al. 2008), and the photosynthetic bacteria *Heliobacterium modesticaldum* (Sattley et al. 2008) were separated from the *Firmicutes* phylum cluster. Alien Hunter genomic analyses of these two species suggested that *P. thermopropionicum rpoB*, *fusA*, *rplB* and *mutS* genes (Fig. 6) and *H. modesticaldum dut* and *dnaE* genes (Fig. 7), used in the analyzes, could have been acquired via horizontal transfer.

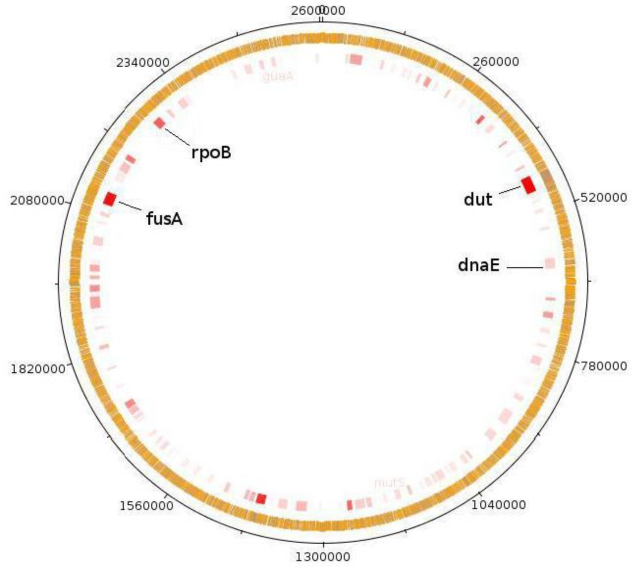

Figure 4 - Chromosomal islands in *Brachyspira hyodysenteriae*. Putative horizontal transferred genes are mapped in the genome in different color tones according to the Alien Hunter program scores. Dark Red = highest HGT scores and Light red = lowest HGT scores. The target genes used in DNA repair and housekeeping super-gene construction are indicated.

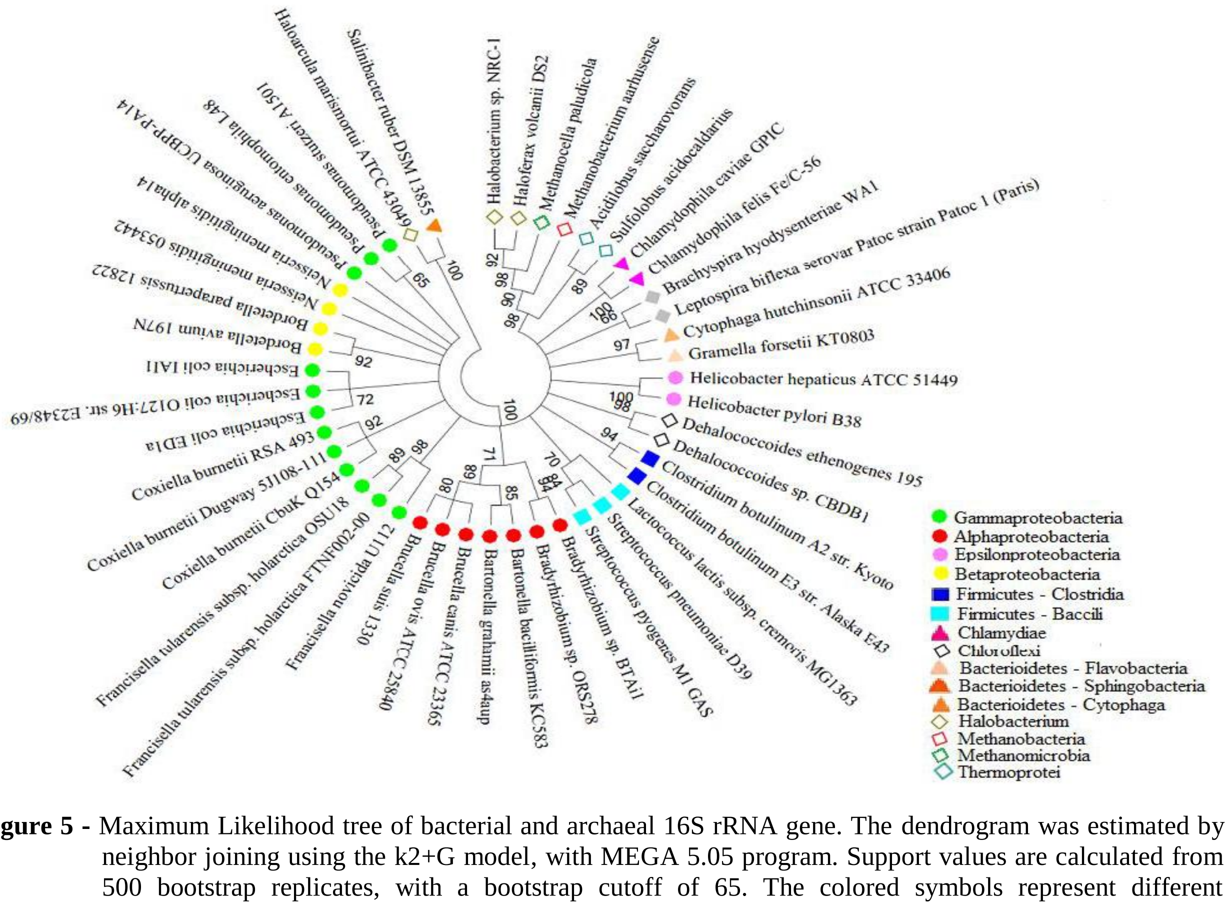
Halobacterium sp. NRC-1
Haloferax volcanii DS2
Methanocella paludicola
Methanobacterium aarhusense
Acidilobus saccharovorans
Sulfolobus acidocaldarius
Chlamydophila caviae GPIC
Chlamydophila felis Fe/C-56
Brachyspira hyodysenteriae WA1
Leptospira biflexa serovar Patoc strain Patoc 1 (Paris)
Cytophaga hutchinsonii ATCC 33406
Gramella forsetii KT0803
Helicobacter hepaticus ATCC 51449
Helicobacter pylori B38
Dehalococcoides ethenogenes 195
Dehalococcoides sp. CBDB1
Clostridium botulinum A2 str. Kyoto
Clostridium botulinum E3 str. Alaska E43
Lactococcus lactis subsp. cremoris MG1363
Streptococcus pneumoniae D39
Streptococcus pyogenes M1 GAS
Bradyrhizobium sp. BTAi1
Bradyrhizobium sp. ORS278
Bartonella bacilliformis KC583
Bartonella grahamii as4aup
Brucella canis ATCC 23365
Brucella ovis ATCC 25840
Brucella suis 1330
Francisella novicida U112
Francisella tularensis subsp. holarctica FTNF002-00
Francisella tularensis subsp. holarctica OSU18
Coxiella burnetii CbuK Q154
Coxiella burnetii Dugway 5J108-111
Coxiella burnetii RSA 493
Escherichia coli ED1a
Escherichia coli O127:H6 str. E2348/69
Escherichia coli IAI1
Bordetella avium 197N
Bordetella parapertussis 12822
Neisseria meningitidis 053442
Neisseria meningitidis alpha14
Pseudomonas aeruginosa UCBPP-PA14
Pseudomonas entomophila L48
Pseudomonas stutzeri A1501
Haloarcula marismortui ATCC 43049
Salinibacter ruber DSM 13855
100
65
92
72
92
89
98
80
98
71
85
70
94
84
100
94
100
98
97
100
99
89
98
90
98
92
● Gammaproteobacteria
● Alphaproteobacteria
● Epsilonproteobacteria
● Betaproteobacteria
■ Firmicutes - Clostridia
■ Firmicutes - Baccili
▲ Chlamydiae
◇ Chloroflexi
▲ Bacterioidetes - Flavobacteria
▲ Bacterioidetes - Sphingobacteria
▲ Bacterioidetes - Cytophaga
◇ Halobacterium
□ Methanobacteria
◇ Methanomicrobia
◇ Thermoprotei

Figure 5 - Maximum Likelihood tree of bacterial and archaeal 16S rRNA gene. The dendrogram was estimated by neighbor joining using the k2+G model, with MEGA 5.05 program. Support values are calculated from 500 bootstrap replicates, with a bootstrap cutoff of 65. The colored symbols represent different taxonomic groups. This representation is a radial cladogram, in which branch length is not proportional to time, and some branches may be elongated so that the names of the taxa appear on the circumference of the circle.

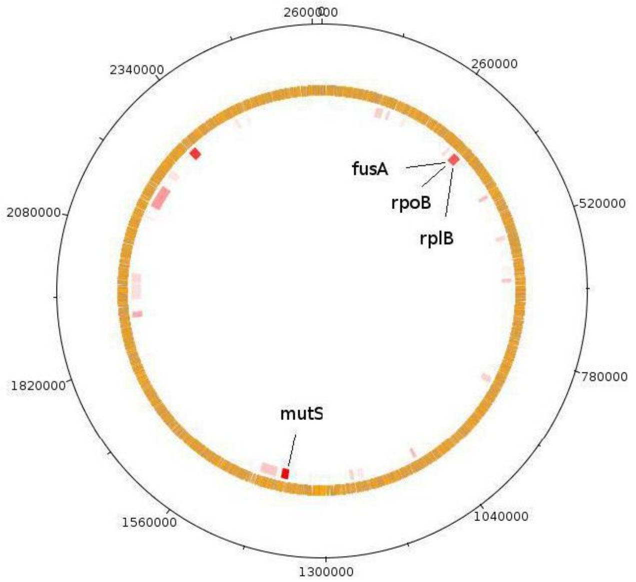

Figure 6 - Chromosomal islands in *Pelotomaculum thermopropionicus*. Putative horizontal transferred genes are mapped in the genome in different color tones according to the Alien Hunter program scores. Dark Red = highest scores and Light red = lowest scores. The target genes used in DNA repair and housekeeping super-gene construction are indicated.

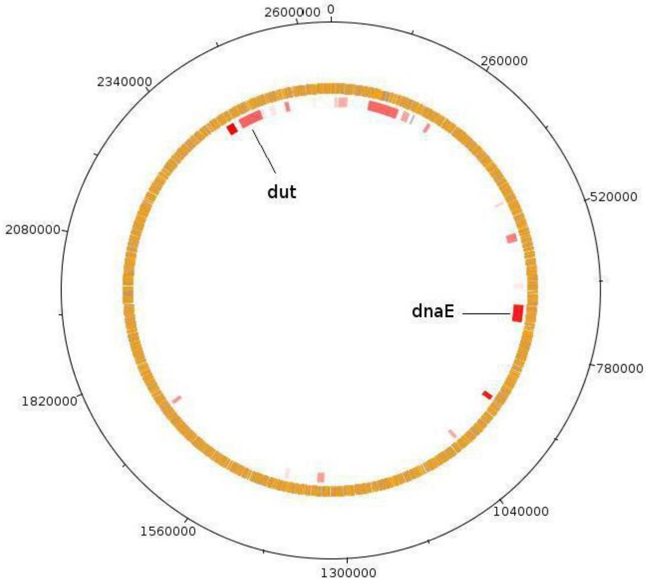

Figure 7 - Chromosomal islands in *Heliobacterium modesticaldum*. Putative horizontal transferred genes are mapped in the genome in different color tones according to the Alien Hunter program scores. Dark Red = highest HGT scores and Light red = lowest HGT scores. The target genes used in DNA repair and housekeeping super-gene construction are indicated.

In conclusion, the results of DNA repair genes in 872 bacterial genomes showed that few genes were present in most species analyzed. The diversity of DNA repair genes seemed to be a result of different environmental threats to which bacterial genomes were exposed. Nevertheless, the multiple gene sequence analysis using DNA repair genes showed better resolution at class level than the multiple gene sequence analysis of housekepping genes. As housekeeping genes, the dendrogram based on DNA repair genes showed to be an accurate approach to distinct bacterial groups at low taxonomic levels and also sensitive to genes derived from horizontal transfer. These results also suggested that the genes of DNA repair pathway could substitute the classical housekepping genes on the multiple gene sequence analysis.

ACKNOWLEGMENTS

We would like to thank the Brazilian Program of National Institutes of Science and Technology-INCT/Brazilian Research Council-CNPq/MCT, CAPES and Fundação Araucária of Paraná State for the financial support.

REFERENCES

Bapteste E, Susko E, Leigh J, MacLeod D, Charlebois RL, Doolittle WF. Do orthologous gene phylogenies really support tree-thinking? *BMC Evol Biol.* 2005; 5: 33.

Bellgard MI, Wanchanthuek P, La T, Ryan K, Moolhuijzen P, Albertyn Z, et al. Genome sequence of the pathogenic intestinal spirochete *Brachyspira hyodysenteriae* reveals adaptations to its lifestyle in the porcine large intestine. *PLoS ONE.* 2009; 4: e4641.

Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* 2000; 17: 540–552.

Dagan T, Artzy-Randrup Y, Martin W. Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc Natl Acad Sci. U.S.A.* 2008; 105: 10039-10044.

de Queiroz A, Donoghue MJ, Kim J. Separate versus combined analysis of phylogenetic evidence. *Annu Rev Ecol Syst.* 1995; 26: 657-681.

Eisen JA, Hanawalt PC. A phylogenomic study of DNA repair genes, proteins and processes. *Mutat Res.* 1999; 435(3): 171-213.

Felsenstein J. Inferring phylogenies. Sunderland: Sinauer Associates, 2004, 580p.

Gadagkar SR, Rosenberg MS, Kumar S. Inferring species phylogenies from multiple genes: Concatenated sequence tree versus consensus gene tree. *J Exp Zool B Mol Dev Evol.* 2005; 304(1): 64-74.

Guizelini D, Pedrosa FO, Raittz RT. Banco de Dados Biológico no Modelo Relacional para Mineração de Dados em Genomas Completos De Procariotos Disponibilizados pelo NCBI GenBank. 2010; 148 p.

Hafström T, Jansson DS, Segerman B. Complete genome sequence of *Brachyspira intermedia* reveals unique genomic features in *Brachyspira* species and phage-mediated horizontal gene transfer. *BMC Genomics.* 2011; 12: 395.

Hall BG. Phylogenetic Trees Made Easy: A How-To Manual. 4 ed. Sunderland: Sinauer Associates, 2011, 282 p.

Hilario E, Gogarten JP. Horizontal transfer of ATPase genes–the tree of life becomes a net of life. *Biosystems.* 1993; 31: 111-119.

Huelsenbeck JP, Bull JJ, Cunningham CW. Combining data in phylogenetic analysis. *Trends Ecol Evol.* 1996; 11: 152-158.

Koonin EV, Wolf YI. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* 2008; 36: 6688-6719.

Kosaka T, Kato S, Shimoyama T, Ishii S, Abe T, Watanabe K. The genome of *Pelotomaculum thermopropionicum* reveals niche-associated evolution in anaerobic microbiota. *Genome Res.* 2008; 18: 442-448.

Lagesen K, Ussery DW, Wassenaar TM. Genome update: the 1000th genome - a cautionary tale. *Microbiology.* 2010; 156: 603-608.

Lang JM, Darling AE, Eisen JA. Phylogeny of Bacterial and Archaeal Genomes Using Conserved Genes: Supertrees and Supermatrices. *PLoS ONE.* 2013; 8: e62510.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007; 23: 2947-2948.

Mongodin EF, Nelson KE, Daugherty S, Deboy RT, Wister J, Khouri H, et al. The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and archaea. *Proc Natl Acad Sci. U.S.A.* 2005; 102: 18147-18152.

Nei M, Xu P, Glazko G. xEstimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. *Proc Natl Acad Sci. U.S.A.* 2005; 98: 2497-2502.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. *Bioinformatics.* 2000; 16: 944-945.

Sattley WM, Madigan MT, Swingley WD, Cheung PC, Clocksin KM, Conrad AL, et al. The genome of *Heliobacterium modesticaldum*, a phototrophic representative of the Firmicutes containing the simplest photosynthetic apparatus. *J Bacteriol.* 2008; 190: 4687-4696.

Suchard MA, Kitchen CMR, Sinsheimer JS, Weiss RE. Hierarchical phylogenetic models for analyzing multipartite sequence data. *Syst Biol.* 2003; 52: 649-664.

Staley JT. The phylogenomic species concept for Bacteria and Archaea. *Microbe.* 2009; 4: 361-365.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 2011; 28: 2731-2739.

Vernikos GS, Parkhill J. Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics.* 2006; 22: 2196-2203.

Yang ZH. Maximum-likelihood models for combined analyses of multiple sequence data. *J Mol Evol.* 1996; 42: 587-596.

Received: December 12, 2014;
Accepted: March 01, 2015.