

Horizontal transfer of the n-TASE gene cluster in *Burkholderia seminalis* TC3.4.2R3 inferred from phylogenetic and molecular methods

Sarina Tsui

USP: Universidade de Sao Paulo

Wellington Luiz Araújo (✉ wlaraujo@usp.br)

USP: Universidade de Sao Paulo <https://orcid.org/0000-0002-2120-8172>

Research Article

Keywords: n-TASE cluster, HGT event, Aquamicrobium sp. SK-2, B. contaminans LMG23361, phylogenetic analysis

Posted Date: August 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-749681/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

This study describes the n-TASE cluster in *Burkholderia seminalis* TC3.4.2R3, which was present in *B. contaminans* (CP046609.1), but absent in other related *Burkholderia* species. Phylogeny, comparative genomics and molecular analysis indicated it is not common to *B. seminalis* species, presenting similarity with homologous genes presents *Aquamicrobium* sp. SK-2 and *B. contaminans* LMG23361, probably acquired by an HGT (Horizontal Gene Transfer) event. It was not possible to determine which was the most likely donor strain of the n-TASE cluster. The HGT event did not occur in all strains of the Bcc group, nor in the *B. seminalis*, but it did occur punctually in the strain *B. seminalis* TC34.2R3. It has a correlation in biotechnological applications related processes. Aiming at understanding the involvement of the n-TASE cluster in the interaction of this bacterium in the environment, genes in this cluster will be inactivated, next.

Introduction

The genus *Burkholderia* (Yabuuchi et al. 1992), formerly classified as *Pseudomonas*, was described in 1992 and is composed of closely related bacteria that occur naturally in various environments. *Burkholderia* species are located into two groups, being the first composed by beneficial and environmental plant-associated species and the second by pathogenic species (Suárez-Moreno et al. 2012; Eberl and Vandamme 2016). The first group have been reclassified into a new genus named *Paraburkholderia* and, although members of this group fail to have pathogenic biomolecular markers (Sawana et al. 2014) the inclusion of these species in this new genus is still controversial since isolates pathogenic to animals can be beneficial to plants (Eberl and Vandamme 2016). The second group include species pathogenic to plant and animals (including human being), and human pathogens, including members of the “pseudomallei” clade and the *Burkholderia cepacia* complex (Bcc) clades (Suárez-Moreno et al. 2012). Species belonging to Bcc group has been isolated from different environments, such as water, soil, rhizosphere and plant tissues (Eberl and Vandamme, 2016; Loveridge et al. 2017). Industrial and agricultural sectors are interested in Bcc species because of their potential use in bioremediation and plant growth promotion and their ability to produce an array of secondary metabolites (Vial et al. 2007). However, the biotechnological applications of Bcc species should to be considered with caution, as they are associated to pneumonia and inflammatory process induction in chronic infections in immunosuppressed and cystic fibrosis (CF) patients. This fact has halted studies aimed at their biotechnological use, especially in agriculture (Vial et al. 2011).

Burkholderia species produce an array of antimicrobial compounds (Parke and Gurian-Sherman 2001), as cepacins, pyrrolnitrins, cepaciamides, cepacidines, quinolones, phenazines, siderophores, and lipopeptides (Mao et al. 2006), making them potential biocontrol agents. The *B. seminalis* strain TC3.4.2R3 belongs to the Bcc group; however, it was isolated from the sugarcane inner root tissues (Mendes et al. 2007; Luvizotto et al. 2010). It is an agent that controls symptoms of orchid necrosis caused by *Burkholderia gladioli* (Araújo et al. 2016). The virulence against *Galleria melonella* and the antifungal production are regulated by temperature; the virulence increased, and antifungal production

decreased at 37°C (Gonçalves et al. 2018). *B. seminalis* can be found in water (Fang et al. 2011), soil (Panhwar et al. 2014), plants (Araújo et al. 2016), and in the sputum of CF patients (Vanlaere et al. 2008). The strain TC3.4.2R3 produces pyochelin and other metabolites associated with the inhibition of *Fusarium oxysporum* (Araújo 2017b), while rhamnolipid synergistically combined with other unidentified diffusible metabolites is involved in the inhibition of cacao pathogens *Moniliophthora perniciosa*, *Phytophthora citrophthora*, *P. capsici*, and *P. palmivora* (Araújo et al. 2017a).

The evolution of prokaryote genomes involves a number of mechanisms associated to rearrangements such as duplication (paralogy), gene loss and gene gain through horizontal gene transfer (HGT, xenology) (Rocha 2008; Kuo and Ochman 2009). HGT it is a major source of phenotypic innovation and a mechanism of niche adaptation, since can introduce genes from distant lineages/species into genomes resulting in new functions (Hiramatsu et al. 2001), which allow the bacterium to acquire antibiotic resistance and virulence determinants gene (Vos et al. 2015) as well as explore new environments and substrates. However, in order to understand the role of HGT in evolution of prokaryote genomes, accurate and precise methods are crucial to quickly and precisely identify genes horizontally acquired (Flutre et al. 2011).

Previous reports highlighted the essential role played by HGT in *Burkholderia* evolution. For example, *Burkholderia xenovorans* LB400, a degrader of polychlorinated biphenyl (PCB), apparently acquired several aromatic degradation capabilities by gene transfer, enabling adaptation to an environment with a recalcitrant carbon source (Chain et al. 2006). The *Burkholderia cepacia* FX5 strain, isolated from *Zea mays* root tissues, can grow on phenol and reduce its concentration. Phenol and other monocyclic aromatic compounds are highly water-soluble and volatile pollutants that plants cannot degrade completely. A plasmid isolated from this FX5 strain possesses a gene encoding the key enzyme in phenol degradation processes, the catechol 2, 3-dioxygenase (C230). The C230 gene is horizontally transferred between endophytic and rhizosphere bacteria, assisting the bioremediation process once these microorganisms can use pollutants as a nutrient source (Wang et al. 2007). *B. cenocepacia* J2315, a human pathogen, contains 14 genomic islands (GIs) that promote survival and pathogenesis in CF lungs (Holden et al. 2009).

Burkholderia gladioli Lv-StB a non-culturable strain and protective symbiont of *Lagria* beetles that produce the polyketide lagriamide, a compound structurally similar to bistramides produced by marine tunicates, is responsible for protecting *Lagria villosa* offspring from fungal pathogens (Flórez and Kaltenpoth 2017; Flórez et al. 2017). This polyketide is biosynthesised by a gene cluster that was probably acquired by HGT, highlighting the potential of microbial symbionts and HGT as critical sources of ecological innovation (Flórez et al. 2018).

The analysis of the evolutionary history behind the diversity and pathogenicity of 60 complete *Burkholderia* genomes revealed that HGT events occurred extensively in the adaptive evolution of this genus. It was also observed that HGT plays an essential role in adaptation and pathogenicity. Most of these acquired genes encode hypothetical proteins or *Burkholderia*-specific proteins of unknown function.

For these reasons, a better understanding of their role in adaptation and strains divergence should be achieved using coexpression analysis of these gene products. The localization of these transferred genes is frequently the small chromosomes, and they were probably acquired millenia ago, contributing to essential differences among species (Zhu et al. 2011).

A range of indirect strategies are necessary to demonstrate a convincing case of an HGT event; once, they are difficult to prove (Buades and Moya 1996). Genomic data are updated and increased constantly, facilitating the determination of how organisms increase their genetic diversity through horizontally acquired genes. The identification of HGT events involves different strategies based on the analysis of DNA composition, including GC content and codon usage in comparison to the genome of the species, presence of mobile elements and discrepancy in phyletic patterns and phylogenetic tree topology (Ragan 2001).

The genome of *B. seminalis* TC3.4.2R3 was sequenced and compared with other *Burkholderia* strains (Araújo et al. 2016). The authors observed a 4,378 bp gene cluster in chromosome 1 that is also present in *B. contaminans* (CP046609.1) but absent in other related *Burkholderia* species (including *B. cenocepacia* and *B. ambifaria*). This cluster was composed of four genes that encode a glycosyltransferase (Bsem_02857), a methyltransferase (Bsem_02858), and two hypothetical proteins (Bsem_02859 and Bsem_02860). In most *Burkholderia* strains, glycosyltransferases (GTFs) are involved in synthesizing compounds essential for the adaptation and competition of the bacteria in their environment (Videira et al. 2005; Liang and Qiao 2007; Hanuszkiewicz et al. 2014). In *B. seminalis* TC3.4.2R3, a glycosyltransferase gene was associated with pyochelin and rhamnolipid production (Araújo et al. 2017b). Generally, O-methyltransferases have been frequently characterized in the biosynthetic pathways of secondary metabolites and have been utilized for biotechnological modifications in several compounds, including flavonoids, alkaloids, and antibiotics (Darsandhari et al. 2018). These findings suggest that glycosyltransferase and methyltransferases could be critical enzymes associated with synthesizing secondary metabolites in *Burkholderia* spp.

In the present study, we characterized a 4,378 bp gene cluster and determined its evolutionary history. Due to the lack of knowledge about the four genes belonging to this cluster and their potential function in TC3.4.2R3, we named it the n-TASE cluster (**n**- for **not** knowing and **TASE** – for **transferases**). The n-TASE cluster refers to the four-gene cluster composed of Bsem_02857, Bsem_02858, Bsem_02859, and Bsem_02860. We focused on the origin of the n-TASE cluster in *B. seminalis* TC3.4.2R3, an agricultural and biotechnological relevant strain, to characterize the evolutionary mechanisms of the horizontal acquisition of genes related to bacterial interactions in the environment. The HGT phenomenon in endophytes highlights a biological mechanism that is important for their evolutionary adaptation within the host plant and simultaneously confers “novel traits” (Wang et al. 2010).

We combined phylogenetics and comparative genomics, including similarity searches, GC content analysis, statistical comparisons of codon adaptation index (CAI), comparative genomics, and molecular analysis to determine if the evolutionary origin of the n-TASE cluster involves an HGT event or if it was a

result of other gene adaptation phenomenon. Our findings suggested that the n-TASE cluster arose from an HGT. To the best of our knowledge, this the first report exploring the horizontal acquisition of a gene cluster in the *B. seminalis* species. A better understanding of the function of this cluster in TC3.4.2R3 will provide insights into the evolutionary mechanisms of the transition from an endophytic bacterium to an opportunistic pathogen.

Materials And Methods

Bacterial strains

The *B. seminalis* strain TC3.4.2R3 belongs to the microbial collection of the Laboratory of Molecular Biology and Microbial Ecology (Department of Microbiology, ICB / USP) and was isolated from sugar cane inner root tissues (Mendes et al. 2007; Luvizotto et al. 2010). The *B. seminalis* strains LMG19587, LMG24067, LMG24271, LMG24272, and LMG24273 were obtained from BCCM/LMG, Belgian Coordinated Collections of Micro-Organisms. *B. cenocepacia* H111 was kindly provided by Professor Leo Eberl from the University of Zurich. All strains were stored at -80°C in Luria-Bertani broth (LB) and 20% glycerol. *Burkholderia* spp. strains were grown in LB medium for 24 h at 28°C .

Integrated analysis of genomic islands

The software IslandViewer (available at: <http://www.pathogenomics.sfu.ca/islandviewer/>) was used to predict and interactively visualize genomic GIs that were considered regions, probably originating from HGT (Bertelli et al. 2017) in bacterial and Archaea genomes. Chromosome 1 from the TC3.4.2R3 genome was analyzed using the software's default settings for integrated prediction of GI regions.

PCR detection of the n-TASE cluster in *Burkholderia* spp.

A pair of primers (forward primer, HGT_P2_RIGHT, GAGTACATCCTGTCGTCGGAGA and reverse primer, HGT_P1_LEFT, ATGGGTATGGACATGTAATCGAC) were designed for the detection of n-TASE cluster sequence in the TC3.4.2R3 strain in other five *B. seminalis* strains (LMG19587, LMG24067, LMG24271, LMG24272, and LMG24273), and the *B. cenocepacia* H111 strain. The genomic DNA of each strain was isolated using the Wizard® Genomic DNA Purification Kit (Promega), and PCR was performed using Q5® High-Fidelity DNA Polymerase (New England Biolabs) in 25- μl PCR mix aliquots in a thermocycler programmed to 98°C for 30 s followed by 35 cycles of 98°C for 10 s, 66°C for 30 s, and 72°C for 2 min. The final extension was at 72°C for 2 min. PCR products were all evaluated on 1% electrophoresis gels.

Phylogenetic analysis

All gene sequences (16S rRNA, pyrimidine kinase, chaperonin GROEL, glycosyltransferase, methyltransferase, and hypothetical proteins) used for *Burkholderia* strains for phylogenetic analysis were retrieved from the IMG website database (www.img.jgi.doe.gov) by identifying homologous genes from the pyrimidine kinase and chaperonin GROEL genes locus tags of the *B. ambifaria* AMMD genome. These were as follows: [*B. ambifaria* AMMD (NCBI Taxon ID 339670); *B. gladioli* ATCC 10248 (NCBI

Taxon ID 28095); *B. cepacia* AMMD (NCBI Taxon ID 339670); *B. cenocepacia* K56-2 (NCBI Taxon ID 985076); *B. ambifaria* RZ2MS16 (NCBI Taxon ID 152480); *Burkholderia* sp. AU31652 (NCBI Taxon ID 2015354); *B. unamae* TAtl-371 (NCBI Taxon ID 95486); *B. anthina* MSMB1496 (NCBI Taxon ID 179879); *B. cenocepacia* J2315 (NCBI Taxon ID 216591); *B. cenocepacia* H111 (NCBI Taxon ID 1055524) *B. seminalis* FL-5-4-10-S1-D7 (NCBI Taxon ID 488731); *B. seminalis* FL-5-5-10-S1-D0 (NCBI Taxon ID 488731); *B. ubonensis* MSMB761WGS (NCBI Taxon ID 101571); *B. territorii* MSMB2156WGS (NCBI Taxon ID 1503055); *Paraburkholderia sacchari* LMG 19450 (NCBI Taxon ID 159450); *Aquamicrobium* sp. SK-2 (NCBI Taxon ID 1560338); *Burkholderia* sp. AU6039 (NCBI Taxon ID 2015344); *B. cenocepacia* DWS 37E-2 (NCBI Taxon ID 95486); *B. contaminans* LMG23361 (NCBI Taxon ID 1334628); *B. latens* AU17928 (NCBI Taxon ID 488446); *Burkholderia* sp. AU31652 (NCBI Taxon ID 2015354); *Burkholderia* sp. HI2500 (NCBI Taxon ID 2015358); *Mumia flava* MUSC201 (NCBI Taxon ID 1348852), and *B. territorii* MSMB1919WGS (NCBI Taxon ID 1503055)].

Before the phylogenetic analysis, sequences of pyrimidine kinase and chaperonin GROEL from all *Burkholderia* strains were concatenated into a ~2.5 kb sequence; the four genes composing the n-TASE cluster in TC3.4.2R3 homologous sequences from *Burkholderia* strains presenting the n-TASE cluster homologous regions [*Aquamicrobium* sp. SK-2 (NCBI Taxon ID 1560338); *Burkholderia* sp. AU6039 (NCBI Taxon ID 2015344); *B. cenocepacia* DWS 37E-2 (NCBI Taxon ID 95486); *B. contaminans* LMG23361 (NCBI Taxon ID 1334628); *B. latens* AU17928 (NCBI Taxon ID 488446); *Burkholderia* sp. AU31652 (NCBI Taxon ID 2015354); *Burkholderia* sp. HI2500 (NCBI Taxon ID 2015358); *Mumia flava* MUSC201 (NCBI Taxon ID 1348852) and *B. territorii* MSMB1919WGS (NCBI Taxon ID 1503055)] were also concatenated (glycosyltransferase, methyltransferase, hypothetical protein, hypothetical protein) in a ~4.5 kb sequence. This genetic composition from TC3.4.2R3 was also made for comparisons.

Each gene composition set was aligned using ClustalW (Batada and Hurst 2007) using MEGA-X software, which was also to phylogenetic analysis based on neighbor-joining method. The phylogenetic analysis was performed for glycosyltransferase and methyltransferase genes from the genomes of TC3.4.2R3, and the IMG website retrieved *Burkholderia* strains (*B. seminalis* FL-5-4-10-S1-D7, *B. cepacia* AMMD, *B. cenocepacia* J2315, *B. cenocepacia* DWS 37E-2, *B. cenocepacia* H111, *B. contaminans* LMG 23361, and *Aquamicrobium* sp. SK-2). Transferases annotated with different functions or classified as different groups like those from TC3.4.2R3 were downloaded from the IMG website. Sequences for each transferase were aligned using ClustalW, and a phylogenetic tree was constructed using the maximum likelihood method using Kimura's 2-parameter distance correction.

GC content and codon bias

GC content and CAI of the n-TASE cluster and other compared genes were calculated using CAIcal server software (<http://genomes.urv.es/CAIcal/>).

Similarity search between the n-TASE and flanking genes in *Burkholderia* strains

For the similarity search, nucleotide sequences retrieved from the IMG website for all 24 bacteria were compared to the gene sequence (nucleotide) of TC3.4.2R3 strains. These genes were kinase (Bsem_02856), glycosyltransferase (Bsem_02857), methyltransferase (Bsem_02858), two hypothetical proteins (Bsem_02859 and Bsem_02860), and chaperonin GROEL (Bsem_02861). The “two-sequence alignment” tool in the nucleotide BLAST – Local Basic Alignment Search Tool (BlastN) was performed using the sequences from TC3.4.2R3 as the query.

The protein BLAST – Local Basic Alignment Search Tool (BlastP) using amino acid sequences was also used with the “two-sequence alignment” tool, using TC3.4.2R3 as the query for strains, for which possible nucleotide similarity was not found: *Burkholderia* sp. AU6039; *B. cenocepacia* DW 37E-2; *B. latens* AU17928; *Burkholderia* sp. AU31652; *Burkholderia* sp. HI2500; and *B. territori* MSMB1919WGS.

Each gene composing the n-TASE cluster in TC3.4.2R3 strain was used as a query in a nucleotide BLAST (BlastN) search against the genomes of *Burkholderia* (*B. ambifaria* AMMD; *B. cenocepacia* J2315; *B. seminalis* FL-5-4-10-S1-D7; *Aquamicrobium* sp. SK-2; *Burkholderia* sp. AU6039; *B. cenocepacia* DWS 37E-2; *B. contaminans* LMG23361; *B. latens* AU17928; *Burkholderia* sp. AU31652 *Burkholderia* sp. HI2500; *Mumia flava* MUSC201, and *B. territori* MSMB1919WGS) using a two-sequence alignment.

Results

Integrated analysis of genomic islands

The n-TASE cluster sequence from *B. seminalis* TC3.4.2R3 was not identified by IslandViewer as a potential genomic island (Fig. 1). The prediction based on IslandPath-DIMOB uses six ORFs as a cluster because single-ORF dinucleotide bias is highly variable. Previous codon-based analysis showed that a minimum cluster of genes of approximately 4.5 kb (corresponding to approximately 6–8 ORFs) is required for reliable estimation of nucleotide composition. Moreover, IslandPath cannot detect HGT of individual genes or islands obtained from organisms with similar DNA signals (Hsiao et al. 2003). In the SIGI-HMM approach, GI prediction is performed on the gene level (Waack et al. 2006). GIs usually range in size from 5–500 kb in the IslandPick prediction. GIs smaller than 8 kb were not targeted, and only those 8–31 kb in size can be predicted as GIs using this method (Langille et al. 2008).

Comparative analysis of the genetic structure and PCR detection of n-TASE cluster

The common 23,720 bp encoding regions of TC3.4.2R3 (comprising genes from locus tag Bsem_02847 to Bsem_02869) were used as a reference for the development of a comparative syntenic gene map (Fig. 2 and Supplementary material Table S1) of *Burkholderia* strains (with or without the n-TASE cluster sequence). The upstream (Bsem_02847 to Bsem_02856) and downstream (Bsem_02861 to Bsem_02869) homologous genes flanking the n-TASE cluster were considered for comparative analysis. The genetic organization was developed based on the gene map available at the IMG website (details such as gene and intergenic spaces were represented based on the IMG gene map).

The structures of the n-TASE cluster and flanking genes were conserved between *B. seminalis* TC3.4.2R3 and *B. cenocepacia* DWS37E-2. Except for the absence of the n-TASE cluster, the structure of this region was similar to *B. cenocepacia* J2315, H111, and *B. ambifaria* AMMD. *Aquamicrobium* sp. SK-2 and *B. contaminans* LMG23361 presented a similar structure, including the n-TASE cluster. However, there is an extra insertion with six genes between Bsem_02862 and Bsem_02863 (co-chaperonin GroES gene and carbohydrate-selective porin OprB gene, respectively in TC3.4.2R3) in *Aquamicrobium* sp. SK-2 and *B. contaminans* LMG23361 strains (Fig. 2), responsible for encoding proteins homologous to the ABC or the iron transport system. The strain *B. seminalis* FL-5-4-10-S1-D7 also presents this extra six-gene insertion, which is also related to transport; however, the n-TASE was absent. The structure of this 23,720 bp common region with n-TASE was more similar between *B. seminalis* TC3.4.2R3 and *B. cenocepacia* DWS37E-2 than between *B. seminalis* TC3.4.2R3 and *B. seminalis* FL-5-4-10-S1-D7. PCR analysis confirmed a ~ 4,378 bp cluster in the TC3.4.2R3 strain absent in *B. cenocepacia* H111 and other *B. seminalis* strains (from the BCCM/LMG Bacteria Collection), which generated a 500-bp amplified DNA fragment (Fig. 3).

Sequence similarity searches and genome composition

The TC3.4.2R3 strain n-TASE cluster presented $\geq 80\%$ identity and e -value of $9e-178$ with four *B. contaminans* strains (X173, SK875, ZCC, and CH-1) (Supplementary material Table S2). No homologous sequences of other *Burkholderia* species or other bacteria were returned in this sequence similarity search.

We conducted further homolog searches using the Gene Ortholog Neighborhoods tool in the IMG Gene ID website. The search started with the Bamb_0734 (locus tag for putative pyrimidine kinase gene in the *B. ambifaria* AMMD genome), homologs of pyrimidine kinase, chaperonin GROEL, and 16S rRNA gene sequences from representative *Burkholderia* genomes that were downloaded from the IMG website. For the genomes containing the n-TASE cluster region, the same methodology was applied; however, in the course of a further search among the genomes returned as orthologs for the Bamb_0734 locus tag, some showed the four-gene cluster between the same pyrimidine kinase and chaperonin GROEL genes. The strains also had their gene sequences of interest (16S rRNA, pyrimidine kinase, glycosyltransferase, methyltransferase, hypothetical protein, hypothetical protein, and chaperonin GROEL) downloaded from the IMG website for further analyses.

We selected nine bacteria and 14 *Burkholderia* strains with or without the n-TASE cluster (bacterial names, strain identification, and NCBI taxon ID are listed in the materials and methods section) to perform the 16S rRNA gene phylogeny, GC content, and nucleotide identity with *B. seminalis* TC3.4.2R3 (Fig. 4). The phylogenetic analysis based on the 16s rRNA gene revealed that species that presents the n-TASE cluster homologous sequence are not evolutionary related. This result suggest that the origin of the n-TASE cluster can not be explained by this phylogeny, which shows that *B. seminalis* is closely related to *B. cenocepacia*. *B. cenocepacia* DWS 37E-2 strain presented higher synteny in this n-TASE region with *B. seminalis* TC3.4.2R3 grouped in a distinct clade of *B. cenocepacia* species. These results suggested

that the n-TASE cluster has an extensive history of independent evolution, although it undoubtedly shares a common ancestor with the *Aquamicrobium* sp. SK-2 and *B. contaminans* LMG23361 strains.

Almost all compared sequences were from the Bcc group, belonging to opportunistic pathogenic strains, except for *Paraburkholderia sacchari*, *B. gladioli*, and the non-*Burkholderia* strains. The differences in the GC content of n-TASE cluster genes and their flanking genes, pyrimidine kinase, and chaperonin GROEL genes were slight for all compared sequences between the *Burkholderia* strains (Fig. 4, numbers showed in red). The GC content of the pyrimidine kinase gene for most of the *Burkholderia* ranged from 69.4% to 70% for *B. cenocepacia* DWS 37E-2 and *B. latens* AU17928 was 67.9% (slightly lower). For the chaperonin GROEL gene, the GC content ranged from 60.7% to 62.7% for all compared sequences. The GC content for the four genes of the n-TASE cluster showed a pattern value around 60% in the TC3.4.2R3 strain. This value was also observed in strains containing this sequence.

Aquamicrobium sp. SK-2 showed the highest nucleotide identity in comparison with the n-TASE cluster of TC3.4.2R3. There was 88% of identity for the glycosyltransferase gene, 86% for the methyltransferase gene, and 80% and 81% identity for the hypothetical proteins (Bsem_02859 and Bsem_02860, respectively). Some sequences showed “no possible alignment” with the TC3.4.2R3 nucleotide sequences, as in *Burkholderia* sp. HI2500, AU6039 and AU31652 strains, where the methyltransferase gene showed no correspondence alignment (*ns* in Fig. 4) with the methyltransferase (Bsem_02858) of TC3.4.2R3. For the *B. cenocepacia* DWS 37E-2, the alignment was possible only for the latter hypothetical protein (61%); for *B. latens* AU17928, none of the four genes showed possible alignment with the TC3.4.2R3 n-TASE cluster. Most of the compared strains showed 89%–99% nucleotide identity with the pyrimidine kinase gene; the same range was observed for the chaperonin GROEL gene, except for *B. contaminans* LMG233361 (63% identity) and *B. cenocepacia* DWS 37E-2 (63% identity) (Fig. 4).

Comparing the amino acid sequences of the strains that presented a result of “no significant similarity found” in at least one of the genes composing the n-TASE cluster in the search for nucleotide identity (*Burkholderia* sp. AU6039; *B. cenocepacia* DW 37E-2; *B. latens* AU17928; *Burkholderia* sp. AU31652; *Burkholderia* sp. HI2500 and *B. territori* MSMB1919WGS), we found that the four genes that compose the n-TASE cluster had a range of 76–87% amino acid identity with the n-TASE cluster of TC3.4.2R3 (Fig. 5); however, all chaperonin GROEL gene sequences from these strains resulted in “no significant similarity” for amino acid comparison (Fig. 5).

CAI analysis was performed with genes located at the upstream position of the n-TASE cluster (locus tag ranging from Bsem_02850) and genes located downstream of the n-TASE cluster (locus tag ranging to Bsem_02870) to determine if the host genome composition of TC3.4.2R3 differs from the n-TASE cluster sequence. The CAI failed to show a significant difference between genes around the cluster and the n-TASE cluster. The CAI value for genes Bsem_02850 to Bsem_02856 and Bsem_02861 to Bsem_02870 was 0.714 ± 0.028 ; for n-TASE genes, the CAI value was 0.7 ± 0.009 (Supplementary material Table S3).

The same calculation was performed for GC content. For the n-TASE upstream and downstream genes, the surrounding genes' GC content was $69.5\% \pm 4.59$, and for the n-TASE genes, GC content was $60.65\% \pm$

2.24. According to the t-test, these values (GC content and CAI) did not differ statistically.

Phylogenetic analysis

Genes from several strains of *Burkholderia* and non-*Burkholderia* species (i.e., *Aquamicrobium* sp. and *Mumia flava*) were used to perform gene-by-gene phylogenetic analyses. The topology of the phylogenetic trees was used to understand the evolutionary history of the n-TASE cluster (Fig. 6). Genes from the n-TASE cluster region in other strains are distant from those of the TC3.4.2R3 strain. Interestingly, the n-TASE cluster genes from *Aquamicrobium* sp. SK-2 are more closely related to genes in the *B. contaminans* LMG23361 than other *Burkholderia* strains.

The pyrimidine kinase and chaperonin GROEL genes (Bsem_02856 and Bsem_02861 in TC3.4.2R3) were concatenated to build a phylogenetic tree (Fig. 6a). Sequences from strains without n-TASE cluster clustered closely; however, strains containing the n-TASE cluster region clustered in two major groups, in which *B. cenocepacia* DWS 37-2, *B. latens* AU17928, and *B. territori* MSMB1919WGS formed a cluster, and all other strains formed another clade (Fig. 6a).

The phylogenetic tree made only with the pyrimidine kinase gene concatenated with the chaperonin GROEL from strains containing the n-TASE region (Fig. 6b, also showing the kinase/GROEL sequence from TC3.4.2R3 as a distinct group. The flanking genes of the n-TASE region in TC3.2.4R3, although coding the same gene product and showing the same genetic organization, probably have a completely different evolutionary history.

The phylogenetic tree for all four genes from the n-TASE cluster presented a five-clade tree (Fig. 6c), in which the n-TASE cluster from TC3.4.2R3 once more did not cluster with any other compared sequence; however, it formed a sister group to the clade composed by *Aquamicrobium* sp. SK-2; *B. contaminans* LMG23361 and *Burkholderia* sp. strains (AU3165, HI2500, and AU6039). Comparing the 16S rRNA gene phylogeny (Fig. 4) and the genetic composition for the n-TASE cluster and flanking genes (Fig. 6c) showed no agreement between these trees.

The analysis of GC content failed to support the notion of HGT of the n-TASE cluster because there is little variation between the compared values. Furthermore, genome composition analysis based on the CAI comparison did not support the HGT event. On the other hand, the 16S rRNA revealed an incongruency in phylogeny in which the groups evolutionarily closest to TC3.4.2R3 did not show the n-TASE cluster region in their genomes, leading us to hypothesize that the n-TASE cluster has a long horizontally acquisition history.

To perform a glycosyltransferase comparative analysis, 48 putative glycosyltransferase gene sequences were selected and downloaded from the IMG website. There were five glycosyltransferase sequences from *B. cepacia* AMMD and *B. cenocepacia* DWS37E-2, seven sequences from *B. seminalis* FL-5-4-10-S1-D7, and *B. cenocepacia* H1111, two sequences from *B. cenocepacia* J2315, four sequences from *B. contaminans*, and three sequences from *Aquamicrobium* sp. SK-2 (Supplementary material Table S4). In

the TC3.4.2R3 genome, we selected 15 glycosyltransferases with diverse family annotation, including the Bsem_02857 glycosyltransferase. For the strains containing the n-TASE cluster, the Bsem_02857 homologous gene was also selected for this analysis [*Aquamicrobium* sp. SK-2 (Ga0098313_1081193); *B. contaminans* LMG23361 (WR31_RS18360) and *B. cenocepacia* DWS37Ee-2 (DM40_RS21930)].

The Bsem_02857 glycosyltransferase gene (TC3.4.2R3_GTF11) clustered (node G1) with gene sequences from *Aquamicrobium* sp. (SK2_GTF48) and *B. contaminans* LMG23361 (LMG 23361_GTF42), forming a sister group (node G2) with the sequence DWS37E-2_GTF32 from *B. cenocepacia* DWS37E-2. Sequences from this node form a sister group (node G3) with another clade (node G4 – Fig. 7) formed by three *B. seminalis* sequences, FL-5-4-10-S1-D7 (FL-5-4-10-S1-D7_GTF19, FL-5-4-10-S1-D7_21), FL-5-4-10-S1-D7_22, and a sequence from *B. cenocepacia* H111 (H111_GTF41).

Despite the distant phylogeny according to the 16S rRNA gene tree, the homologous genes of Bsem_02857 showed the maximum evolutionary history with glycosyltransferase in question; 42% of selected glycosyltransferases from *B. seminalis* FL-5-4-10-S1-D7 clustered with a sister group in the clade containing Bsem_02857. None of the 14 selected glycosyltransferase sequences selected from TC3.4.2R3 strain grouped with Bsem_02857. In this comparison, there were no homologous Bsem_02857 glycosyltransferase genes in the genomes (Fig. 7). The Bsem_02857 probably shares a common ancestor with the homologous glycosyltransferase from *Aquamicrobium* sp. SK-2 and *B. contaminans* LMG23361.

We selected 70 putative methyltransferase gene sequences to perform a methyltransferase gene comparison to identify possible homologous sequences in the TC3.4.2R3 genome or the other seven bacterial genomes. There was a diverse function annotation of these selected genes. Of these, there were 30 methyltransferase gene sequences (including the Bsem_02858) from TC3.4.2R3 strain, 11 sequences from *B. seminalis* FL-5-4-10-S1-D7, five sequences from *B. cepacia* AMMD; *B. cenocepacia* H111 and *B. cenocepacia* J2315, 11 sequences from *B. cenocepacia* DWS37Ee-2, four sequences from *B. contaminans* LMG23361, and nine sequences from *Aquamicrobium* sp. SK-2 (Supplementary material Table S4).

The frequency of gene sequences annotated as methyltransferases varied from strain to strain, *B. cepacia* AMMD, *B. cenocepacia* H111, *B. cenocepacia* J2315, and *B. contaminans* LMG23361 were strains with fewer available methyltransferase sequences. The Bsem_02858 homologous genes from the selected n-TASE containing strains were included in this comparison [*Aquamicrobium* sp. SK-2 (Ga0098313_1081194); *B. contaminans* LMG23361 (WR31_RS18355) and *B. cenocepacia* DWS37Ee-2 (DM40_RS21925)]. The 70 collected putative methyltransferase gene sequences were encoded in chromosomes. Genes encoded in plasmids were not taken into account for this comparison.

A methyltransferase maximum likelihood tree was constructed, and Bsem_02858 (TC3.4.2R3_MTF20) grouped with the methyltransferase sequence from *B. cenocepacia* DWS 37E-2 (DWS 37E-2_MTF49) (node M1 - Fig. 8). This consists of a sister group of node M3 composed of methyltransferases from *Aquamicrobium* sp. SK-2 (SK2_MTF69) and *B. contaminans* LMG23361 (LMG23361_58), which form the

clade M2. Sequences were saved from the IMG website, and the locus tag used in this website were not always the same as those used in the “Burkholderia Genome DB” (<https://www.burkholderia.com/>). The methyltransferase genes that clustered (node M3) with the Bsem_02858 sequence (TC3.4.2R3_MTF20) were homologous in the respective genomes of *Aquamicrobium* sp. SK-2 (Ga0098313_1081194), *B. contaminans* LMG23361 (WR31_RS18355), and *B. cenocepacia* DWS37Ee-2 (DM40_RS21925). None of the 19 compared methyltransferases selected from TC3.4.2R3 grouped with the Bsem_02858 sequence. It is clear that Bsem_02858 shares no common ancestor with any other compared methyltransferase genes, even those that were classified as its homologous gene. There are no methyltransferases in the TC3.4.2R3 genome homologous to the Bsem_02858 methyltransferase gene (Fig. 8). All other methyltransferase sequences from TC3.4.2R3 clustered in two groups (the respective nodes are marked with red circles in figure 8). Neither transferase enzyme from the n-TASE cluster showed any similarity nor shared evolutionary background with any other enzymes in the genome of TC3.4.2R3.

Discussion

Organisms acquire foreign genes or DNA fragments across species boundaries via HGT, which accelerates the evolution and innovation in genomes because it introduces newly evolved donor genes into the host genomes by avoiding the slow steps of *ab initio* gene creation (Jain et al. 1999; Jain et al. 2003). Some special function genes, including those for antibiotics resistance and extreme environment adaptation, would be spread among organisms by HGT (Hanage et al. 2009; Gootz 2010), allowing species to occupy different niches or habitats.

Although IslandViewer indicated that several regions of chromosome 1 resulted from HGT, the n-TASE cluster may not be detectable by the program because the *ab initio* patterns or parametric parameters that might consider the sequence composition data and genomic signatures (nucleotide composition, oligonucleotide spectrum, DNA structural characteristics, and genomic context) are insufficiently sensitive to detect small anomalous patterns or determine the analyzed region as a genomic island; therefore, it could not be determined if it were an HGT (Ravenhall et al. 2015). Furthermore, the n-TASE cluster has 4,378 bp, and it might be difficult for automatized GIs prediction programs to infer this region as an HGT acquired sequence. Once, one of the GI prediction tools (IslandPath-DIMOB) required a minimum cluster size of 4.5 kb for reliable estimation of nucleotide composition (Hsiao et al. 2003).

Araújo et al. (2016) analyzed the *B. seminalis* TC3.4.2R3 genome and found a 4,378-bp four-gene cluster comprising genes corresponding to locus tags Bsem_02857 to Bsem_02860. Curiously, this four-gene cluster that we named n-TASE was not found in the closest Bcc group strain of TC3.4.2R3, *B. cenocepacia* J2315. We aimed to determine whether the n-TASE cluster was horizontally acquired from a non-*Burkholderia* donor to an ancestor of Bcc group strains or if the HGT event was exclusive to the *B. seminalis* TC3.4.2R3 strain.

The detection of HGT depends on search strategies for differences in nucleotide composition patterns in the genome (Putonti et al. 2006), since, depending on the evolutionary distance of the donor organism,

there may be differences in the content of GC and preferential codons (Rocha et al. 2006). al. 2006). To evaluate these patterns, the use of the CAI value, was proposed as an estimative of the use of certain codons in relation to a set of reference genes. The CAI value ranges from 0 to 1.0 and genes that present values closest to 1.0 present codon usage pattern similar to the reference gene (Sharp et al. 1987). Substitute methods, which do not employ phylogenetic tree construction or other direct phylogenetic analyses, can be used to identify regions acquired by horizontal transfer (Ragan, 2001). However, these surrogate methods can result in a high rate of false positives, as the intragenomic variation of certain codons may be large enough to be identified as different and therefore attributed to horizontal transfer events (Guindon and Perrière 2001). Although, codon bias and the ratio of base composition could generate inadequate signs for the detection of genes acquired by HGT (Koski et al. 2001) the use of combined strategies could improve the quality of this identification.

Based on codon bias (represented by the CAI analysis in the present study) and base composition (GC content and nucleotide identity among compared genes from selected strains), there was no evidence for an HGT event for the n-TASE cluster origin because there were no significant differences among these values between strains. Furthermore, a comparison between n-TASE cluster genes and their adjacent gene base composition features showed no significant differences. This result could be related to the time since the occurrence of the horizontal transfer, whereas after acquisition, the genes may undergo a phenomenon called amelioration which result in gaining of the host genome codon usage and compositional values similar to the host (Marri and Golding 2008), reducing the differences and the sensitivity of surrogate methods (Becq et al. 2010). In another hand, it has been suggested that for successful acquisition of genes by HGT in the host genome, a compatibility of condon usage between donor and recipient genome should be present (Medrano-Soto et al. 2004). Comparative methods (including phylogenetic and phylogenomic approaches) proved to be more sensitive and specific than surrogate methods (Poptsova and Gogarten 2007). However, although the effectiveness of surrogate methods for identifying HGT is questioned, the lack of evolutionarily related sequence data to identify HGTs from different species or taxonomic levels makes it difficult to use them on a large scale.

Half of the *Burkholderia* genes families are inferred to have experienced HGT at least once during their evolution. The process of gene gain and loss appears to be a consistent trend throughout the evolution of *Burkholderia*. In the *Burkholderia* ancestor and the ancestral branches of the major clades, a substantial number of gene acquisitions occurred. The common ancestor of *Burkholderia* had an estimated 1,335 acquired genes compared with the outgroup of other Burkholderiaceae (Zhu et al. 2011).

PCR detection of n-TASE sequences in *B. seminalis* TC3.4.2R3 and other *Burkholderia* strains suggest that the n-TASE cluster sequence is present only in the TC3.4.2R3 strain, reinforcing the notion of its horizontal acquisition; the other five *B. seminalis* strains and the compared *B. cenocepacia* H111 strain showed no correspondence in this region, demonstrated by the small size of their amplicon (> 500 bp). We examined phylogenetic trees constructed from those genes forming the n-TASE cluster that have orthologs in species that are close relatives of TC3.4.2R3 (Bcc groups strains, specifically *B. cenocepacia*) and non-*Burkholderia* strains (i.e., *Aquamicrobium* sp. and *Mumia flava*). We would predict

that if an ancestor of Bcc group strains were the recipient of HGT, then the genes in the n-TASE cluster would be present in all compared Bcc strains, which was not observed. Conversely, if the direction of transfer were from a donor to *B. seminalis*, it would be expected that all compared (and available) *B. seminalis* genomes would show the n-TASE cluster sequence. As we presented in the results, none of the *B. seminalis* strains in the IMG database have this n-TASE cluster; furthermore, the genomic DNA from five *B. seminalis* strains isolated from environmental and clinical samples showed no n-TASE on PCR analysis (Fig. 3). Therefore, we believe that the acquisition of the n-TASE cluster in the TC3.4.2R3 strain occurred in a unique event in this specific sugar cane rhizosphere-isolated endophytic strain.

A complete n-TASE cluster sequence was also found *Aquamicrobium* sp. SK-2 and *Mumia flava* MUSC201. However, based on the *Burkholderia* Genome BD database (<https://www.burkholderia.com/>), there are 17 *Burkholderia* strains presenting the n-TASE cluster (5.93%) in a 287 complete *Burkholderia* genomes database. Of these, *B. contaminans*, *B. latens*, and the majority were strains from the *B. territorri* (76%) species (Supplementary material Table S5).

The n-TASE cluster present in these species was most likely transferred from a common donor; however, the mutational index and adaptation of the n-TASE cluster after integration of HGT acquired DNA fragment evolved independently in the TC3.4.2R3, explaining its unique characteristics and low percentage of identity with other sequences. The random distribution of the n-TASE cluster could be a result of independent HGT events or frequent losses of this cluster in different ancestor strains (Fig. 2). Considering that phylogenetic trees, based on the n-TASE genes, present topologies non-congruent with the expected phylogeny for the evaluated group, we suggest that this cluster should be acquired as a result of independent events in different strains (Figs. 4 and 6).

The absence of homologous glycosyltransferase and methyltransferase genes for Bsem_02857 and Bsem_02858 within all n-TASE-containing genomes of compared strains reveals that there is no possibility of a gene duplication event, suggesting that events of gene duplication and losses could not be associated to this HGT. If there were a gene-loss event, the phylogenetic trees of the n-TASE cluster would agree with the evolutionary species phylogeny (Khaldi et al. 2008). Furthermore, a gene-loss scenario would necessitate a precise loss of a four-gene cluster in 638 *Burkholderia* strains (according to the *Burkholderia* Genome database). This four-gene cluster does not necessarily form a functional operon, and we cannot provide evidence for this. In general, genes belonging in a specific pathway are organized in clusters that are regulated in specific conditions. This kind of organization allow them to be transferred as a block enabling HGT events and acquisition of new cellular functions (Lawrence 1999; Walton 2000). The discontinuous distribution of the n-TASE cluster among *Burkholderia* species suggests that this gene cluster was maintained only in few species, which could occupy a specific niche (not described in the present study) or present better fitness in some conditions. However, a physiological and ecological studies should be carried out to identify these advantageous functions.

A monophyletic clade was observed when comparing the n-TASE cluster from TC3.4.2R3 and the cluster's homologous sequences from *Aquamicrobium* sp. SK-2 and *B. contaminans* LMG23361. *B. contaminans*

LMG23361 is the type strain of a species isolated from the milk of a dairy sheep with mastitis (Vanlaere et al. 2009). *B. contaminans* was found as a contaminant of pharmaceutical and personal care products (PPCPs) and linked to outbreaks. This strain degrades benzalkonium chloride, one of the major antiseptic formulations for PPCPs, at higher levels than the other Bcc strains (Ahn et al. 2016). An isolated of *B. contaminans* obtained from brown patch disease of lawn grass suppressive soil produce a potent antifungal compound, named occidiofungin, against a broad range of phytopathogenic fungi, including *Pythium* spp. (Lu et al. 2009). *Aquamicrobium* sp. SK-2 was isolated from a sewage treatment plant and is considered a good candidate for the bioremediation of PCB (polychlorinated biphenyl)-contaminated soil (Chang et al. 2015). Although the n-TASE containing *Burkholderia* strains are opportunistic pathogens (*B. seminalis*, *B. contaminans*, *B. cenocepacia*, *B. latens*, and *B. territori*), *Aquamicrobium* sp. has not yet been linked to any pathogenic behavior.

We are tempted to speculate that the n-TASE cluster could be involved in processes related to biotechnological applications such as antimicrobial activity or degradation of pollutants. The proteins encoded by these genes might be related to a biosynthetic pathway for these biotechnologically compelling features. In addition, this unknown biological feature had to be selected by the environment occupied by these bacteria. Identifying the functions associated with the n-TASE cluster will be necessary to fully comprehend the role of these genes in *B. seminalis* TC3.4.2R3 biology and could provide clues about the evolution of the ancestral donor of this cluster.

A mutant transposon library of *B. seminalis* TC3.4.2R3 strain was generated and screened to identify mutants compromised in orchid necrosis biocontrol activity (Araújo et al. 2016) as well as for antifungal activity (Neves 2011). More than 3,800 transposons mutants were obtained and individually tested for defective biocontrol and antifungal activity. Neves (2011) observed that the M30 mutant, truncated by the Tn5 mini-transposon at the Bsem_02858 (methyltransferase gene), showed an antifungal phenotype that differed from the wild-type strain TC3.4.2R3, and the phytopathogenic *Fusarium oxysporum* was not inhibited by the M30 mutant. Neves (2011) hypothesized the possible involvement of the methyltransferase Bsem_02858 or possibly the entire n-TASE cluster to produce antifungal compounds by TC3.4.2R3. This hypothesis is plausible; nevertheless, gene knockout experiments should be conducted for the gene function characterization and understanding of the biosynthesis pathway in which the n-TASE could take part directly or indirectly.

The n-TASE cluster has a complex history with multiple independent HGT events through *Burkholderia* and non-*Burkholderia* strains. It is not a gene cluster commonly shared within strains from the same species; instead, it occurs independently within one or a few strains from some *Burkholderia* strains. It might appear more straightforward to suggest that the n-TASE cluster always occurs as an independent HGT event within bacterial strains. We believe that if there was only one HGT event that originated all the n-TASE-containing strains, and it would be expected that the n-TASE cluster phylogenetic tree would be formed by a unique monophyletic group, which was not seen. Instead, there are different clades composed of randomly distributed strains that are not even evolutionarily related. The sporadic distribution of the n-TASE in *Burkholderia* and non-*Burkholderia* strains is reliable with a model of

independent HGT events, whereas adjacent genes order and orientation are conserved among compared strains (Fig. 2).

We must consider that genetic similarity is an important factor that can affect horizontal transfer rates between donor and recipient organisms. Thus, HGT is expected to occur more frequently among phylogenetically related biological species (Wolf et al. 2002; Choi et al. 2007). In a study involving 438 complete genomes of prokaryotes, only 30 HGT events were observed between distantly related taxonomic groups (Wagner and Chaux 2008). However, the mechanisms involved in these events have not been identified, although it is known that the Type IV secretion system (Juhas et al. 2008), conjugation (Weinert et al. 2009), transformation (Fall et al. 2007) or transduction (Zaneveld et al. 2008) may contribute to new genes or DNA sequences being transferred from one bacterium to another. The present study showed that the n-TASE cluster was likely transferred horizontally from a common donor strain shared with *B. seminalis* TC3.4.2R3, *Aquamicrobium* sp. SK-2 and *B. contaminans* LMG23361, most likely evolutionarily close to the genus *Burkholderia*, contribute to the gain of genes and functions that, although not described for this strain, may allow this bacterium to colonize different environments such as soil, plants and animals .

Conclusions

We investigated the evolution of n-TASE of *B. seminalis* TC3.4.2R3, aiming to identify its origin. We searched for homologs of n-TASE-containing genes in publicly available gene sequence databases. Our approach failed to identify any n-TASE homologs in a broad range of bacterial species not belonging to the genus *Burkholderia*. We detected homologs of the n-TASE cluster in a few *Burkholderia* strains that were not shared among the other strains from the same species. The GC content and CAI values analysis showed that the acquired n-TASE cluster presented similar characteristics to those observed for other *Burkholderia* species. However, based on PCR detection and phylogenetic analysis, we provided evidence that n-TASE was acquired through an HGT event. To the best of our knowledge, this paper is the first to show the evolutionary mechanism of a gene cluster most probably related to mechanisms related to interactions of *B. seminalis* TC3.4.2R3 in the environment. However, we did not study the n-TASE cluster's association with the capacity to inhibit phytopathogenic fungi. The inactivation of these genes and analysis of resulting phenotypes might elucidate the role of this four-gene cluster in the interaction of this bacterium in the environment.

Based on phylogenetic support for the hypothesis presented in the present study, we conclude that the four genes forming the n-TASE cluster in the TC3.4.2R3 strain was acquired by HGT event that occurred sporadic in *Burkholderia* species. It was not possible to determine precisely the donor strain; however, it was most probably a common donor (not necessarily related to *Burkholderia*) for the n-TASE cluster in *Aquamicrobium* sp. SK-2 and *B. contaminans* LMG233361. Based on the nucleotide composition similarity between n-TASE cluster sequences (Fig. 2), we believe that there is a common donor containing this n-TASE cluster, and its transference occurred in an independent way throughout the species. We cannot determine the most likely donor strain using the current data because of the absence of n-TASE

sequences that would allow a complete comparison. It is likely that the HGT event did not occur in all Bcc group strains and did not occur within the *B. seminalis* species but only in the TC34.2R3 strain.

Declarations

Competing Interests

The authors declare that they have no conflict of interests.

Author's Contributions

Wellington Luiz de Araújo designed and formulated the study. Sarina Tsui performed the research, prepared experiments, analyzed the data and prepared the figures and tables. Sarina Tsui and Wellington Luiz de Araújo wrote the manuscript, both authors read and approved final manuscript.

Acknowledgments

This work was supported by grant from FAPESP- Fundação de Amparo à Pesquisa do Estado de São Paulo (Processos 2016/17173-3; 2017/12510-4; 2018/04397-6). We thank Luiz Ricardo Olchanheski for performing the Integrated analysis of genomic islands analysis. WLA are productivity fellowship from CNPq. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

1. Ahn Y, Kim JM, Kweon O, Kim SJ, Jones RC, Woodling K, Costa GG, LiPuma JJ, Hussong D, Marasa BS, Cerniglia CE (2016) Intrinsic resistance of Burkholderia cepacia complex to benzalkonium chloride. *Mbio*. <https://doi.org/10.1128/mBio.01716-16>
2. Araújo FDS, Vieira RL, Molanob EPL, Máximoc HJ, Dalioc RJD, Vendraminia PH, Araújo WL, Eberlin MN (2017a) Desorption electrospray ionization mass spectrometry imaging reveals chemical defense of Burkholderia seminalis against cacao pathogens. *RSC Adv*. <https://doi.org/10.1039/C7RA03895J>
3. Araújo FDS, Araújo WL, Eberlin MN (2017b) Potential of Burkholderia seminalis TC3. 4.2 R3 as biocontrol agent against Fusarium oxysporum evaluated by mass spectrometry imaging. *J Am Soc Mass Spectrom*. <https://doi.org/10.1007/s13361-017-1610-6>
4. Batada NN, Hurst LD (2007) Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat Genet*. <https://doi.org/10.1038/ng2071>
5. Becq J, Churlaud C, Deschavanne P (2010) A benchmark of parametric methods for horizontal transfers detection. *PLoS One*. <https://doi.org/10.1371/journal.pone.0009989>
6. Bertelli C, Laird MR, Williams KP et al (2017) IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res*. <https://doi.org/10.1093/nar/gkx343>

7. Buades C, Moya A (1996) Phylogenetic analysis of the isopenicillin-N-synthetase horizontal gene transfer. *J Mol Evol* 42:537–542
8. Chain PSG, Deneff VJ, Konstantinidis KT et al (2006) *Burkholderia xenovorans* LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. *PNAS*. <https://doi.org/10.1073/pnas.0606924103>
9. Chang YC, Sawada K, Kim ES, Jung K, Kikuchi S (2015) Whole-genome sequence of *Aquamicrobium* sp. strain SK-2, a polychlorinated biphenyl-utilizing bacterium isolated from sewage sludge. *Genome Announc*. <https://doi.org/10.1128/genome>
10. Choi IG, Kim SH (2007) Global extent of horizontal gene transfer. *PNAS*. <https://doi.org/10.1073/pnas.0611557104>
11. Darsandhari S, Dhakal D, Shrestha B, Parajuli P, Seo JH, Kim TS, Sohng JK (2018) Characterization of regioselective flavonoid O-methyltransferase from the *Streptomyces* sp. KCTC 0041BP. *Enzyme Microb Technol* <https://doi.org/10.1016/j.enzmictec.2018.02.007>
12. Eber L, Vandamme P (2016) Members of the genus *Burkholderia*: good and bad guys. *F1000Research* 5 <https://doi.org/10.12688/f1000research.8221.1>
13. Fall S, Mercier A, Bertolla F, Calteau A, Gueguen L, Perrière G, Vogel TM, Simonet P (2007) Horizontal gene transfer regulation in bacteria as a “spandrel” of DNA repair mechanisms. *PloS one*. <https://doi.org/10.1371/journal.pone.0001055>
14. Fang Y, Xie GL, Lou MM, Li B, Muhammad I (2011) Diversity analysis of *Burkholderia cepacia* complex in the water bodies of West Lake, Hangzhou, China. *Res J Microbiol*. <https://doi.org/10.1007/s12275-011-0267-2>
15. Flórez LV, Kaltenpoth M (2017) Symbiont dynamics and strain diversity in the defensive mutualism between *Lagria* beetles and *Burkholderia*. *Environ Microbiol*. <https://doi.org/10.1111/1462-2920.13868>
16. Flórez LV, Scherlach K, Miller IJ, Rodrigues A, Kwan JC, Hertweck C, Kaltenpoth M (2018) An antifungal polyketide associated with horizontally acquired genes supports symbiont-mediated defense in *Lagria villosa* beetles. *Nat Commun*. <https://doi.org/10.1038/s41467-018-04955-6>
17. Flutre T, Duprat E, Feuillet C, Quesneville H (2011) Considering transposable element diversification in de novo annotation approaches. *PloS one*. <https://doi.org/10.1371/journal.pone.0016526>
18. Gonçalves PJRO, Hume CCD, Ferreira AJ, Tsui S, Brocchi M, Wren BW, Araújo WL (2019) Environmental interactions are regulated by temperature in *Burkholderia seminalis* TC3.4.2R3. *Sci Rep*. <https://doi.org/10.1038/s41598-019-41778-x>
19. Gootz TD (2010) The global problem of antibiotic resistance. *Crit Rev Immunol*. <https://doi.org/10.1615/critrevimmunol.v30.i1.60>
20. Guindon S, Perrière G (2001) Intragenomic base content variation is a potential source of biases when searching for horizontally transferred genes. *Mol Biol Evol*. <https://doi.org/10.1093/oxfordjournals.molbev.a003972>
21. Hanage WP, Fraser C, Tang J, Connor TR, Corander J (2009) Hyper-recombination, diversity, and antibiotic resistance in pneumococcus. *Science* <https://doi.org/10.1126/science.1171908>

22. Hanuszkiewicz A, Pittock P, Humphries F et al (2014) Identification of the flagellin glycosylation system in *Burkholderia cenocepacia* and the contribution of glycosylated flagellin to evasion of human innate immune responses. *J Biol Chem*. <https://doi.org/10.1074/jbc.M114.562603>
23. Hiramatsu K, Cui L, Kuroda M, Ito T (2001) The emergence and evolution of methicillin-resistant *Staphylococcus aureus*. *Trends Microbiol* [https://doi.org/10.1016/s0966-842x\(01\)02175-8](https://doi.org/10.1016/s0966-842x(01)02175-8)
24. Holden MTG, Seth-Smith HMB, Crossman LC et al (2009) The genome of *Burkholderia cenocepacia* J2315, an epidemic pathogen of cystic fibrosis patients. *J Bacteriol Res*. <https://doi.org/10.1128/JB.01230-08>
25. Hsiao W, Wan I, Jones SJ, Brinkman FSL (2003) IslandPath: aiding detection of genomic islands in prokaryotes. *Bioinformatics* <https://doi.org/10.1093/bioinformatics/btg004>
26. Jain R, Rivera MC, Moore JE, Lake JA (2003) Horizontal gene transfer accelerates genome innovation and evolution. *Mol Biol Evol*. <https://doi.org/10.1093/molbev/msg154>
27. Jain R, Rivera MC, Moore JE, Lake JA (1999) Horizontal gene transfer among genomes: the complexity hypothesis *PNAS* <https://doi.org/10.1073/pnas.96.7.3801>
28. Hanage WP, Fraser C, Tang J, Connor TR, Corander J (2009) Hyper-recombination, diversity, and antibiotic resistance in pneumococcus. *Science* <https://doi.org/10.1126/science.1171908>
29. Juhas M, Crook DW, Hood DW (2008) Type IV secretion systems: tools of bacterial horizontal gene transfer and virulence. *Cell Microbiol* <https://doi.org/10.1111/j.1462-5822.2008.01187.x>
30. Khaldi N, Collemare J, Lebrun MH, Wolfe KH (2008) Evidence for horizontal transfer of a secondary metabolite gene cluster between fungi. *Genome Biol*. <https://doi.org/10.1186/gb-2008-9-1-r18>
31. Koski LB, Morton RA, Golding GB (2001) Codon bias and base composition are poor indicators of horizontally transferred genes. *Mol Biol Evol*. <https://doi.org/10.1093/oxfordjournals.molbev.a003816>
32. Kuo CH, Ochman H (2008) The fate of new bacterial genes. *FEMS Microbiol Rev*. <https://doi.org/10.1111/j.1574-6976.2008.00140.x>
33. Langille MGI, Hsiao WWL, Brinkman FSL (2008) Evaluation of genomic island predictors using a comparative genomics approach. *BMC Bioinform*. <https://doi.org/10.1186/1471-2105-9-329>
34. Lawrence J (1999) Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes. *Curr Opin Genet Dev*. [https://doi.org/10.1016/s0959-437x\(99\)00025-8](https://doi.org/10.1016/s0959-437x(99)00025-8)
35. Liang D, Qiao J (2007) Phylogenetic analysis of antibiotic glycosyltransferases. *J Mol Evol* <https://doi.org/10.1007/s00239-006-0110-2> Loveridge EJ, Jones C, Bull MJ (2017)
36. Reclassification of the specialized metabolite producer *Pseudomonas mesoacidophila* ATCC 31433 as a member of the *Burkholderia cepacia* complex. *J Bacteriol* <https://doi.org/10.1128/JB.00125-17>
37. Lu SE, Novak J, Austin FW et al (2009) Occidiofungin, a unique antifungal glycopeptide produced by a strain of *Burkholderia contaminans*. *Biochemistry* <https://doi.org/10.1021/bi900814c>
38. Luvizotto DM, Marcon J, Andreote FD et al (2010) Genetic diversity and plant-growth related features of *Burkholderia* spp. from sugarcane roots. *World J Microbiol Biotechnol* 26:1829–1836

39. Mao S, Lee SJ, Hwangbo H et al (2006) Isolation and characterization of antifungal substances from *Burkholderia* sp. culture broth. *Curr Microbiol.* <https://doi.org/10.1007/s00284-005-0333-2>
40. Marri PR, Golding GB (2008) Gene amelioration demonstrated: the journey of nascent genes in bacteria. *Genome* <https://doi.org/10.1139/g07-105>
41. Medrano-Soto A, Moreno-Hagelsieb G, Vinuesa P, Christen JA, Collado-Vides J (2004) Successful lateral transfer requires codon usage compatibility between foreign genes and recipient genomes. *Mol Biol Evol.* <https://doi.org/10.1093/molbev/msh202>
42. Mendes R, Pizzirani-Kleiner AA, Araújo WL, Raaijmakers JM (2007) Diversity of cultivated endophytic bacteria from sugarcane: genetic and biochemical characterization of *Burkholderia cepacia* complex isolates. *AEM* <https://doi.org/10.1128/AEM.01222-07>
43. Neves AA (2011) Identificação de genes de *Burkholderia* sp. associados a antagonismo a microrganismos fitopatogênicos. Dissertation, University of Mogi das Cruzes
44. Panhwar QA, Naher UA, Jusop S et al (2014) Biochemical and molecular characterization of potential phosphate-solubilizing bacteria in acid sulfate soils and their beneficial effects on rice growth. *PLoS one.* <https://doi.org/10.1371/journal.pone.0097241>
45. Parke JL, Gurian-Sherman D (2001) Diversity of the *Burkholderia cepacia* complex and implications for risk assessment of biological control strains. *Annu Rev Phytopathol.* <https://doi.org/10.1146/annurev.phyto.39.1.225>
46. Poptsova MS, Gogarten JP (2007) The power of phylogenetic approaches to detect horizontally transferred genes. *BMC Evol Biol.* <https://doi.org/10.1186/1471-2148-7-45>
47. Putonti C, Luo Y, Katili C, Chumakov S, Fox GE, Graur D, Fofanov Y (2006) A computational tool for the genomic identification of regions of unusual compositional properties and its utilization in the detection of horizontally transferred sequences. *Mol Biol Evol* (2006) <https://doi.org/10.1093/molbev/msl053>
48. Ragan MA (2001) On surrogate methods for detecting lateral gene transfer. *FEMS Microbiol Lett.* <https://doi.org/10.1111/j.1574-6968.2001.tb10755.x>
49. Ravenhall M, Škunca N, Lassalle F, Dessimoz C (2015) Inferring horizontal gene transfer. *PLoS Comput Biol.* <https://doi.org/10.1371/journal.pcbi.1004095>
50. Rocha PC, Touchon M, Feil EJ (2006) Similar compositional biases are caused by very different mutational effects. *Genome Res.* <https://doi.org/10.1101/gr.5525106>
51. Rocha EPC (2008) Evolutionary patterns in prokaryotic genomes. *Curr Opin Microbiol.* <https://doi.org/10.1016/j.mib.2008.09.007>
52. Sawana A, Adeolu M, Gupta RS (2014) Molecular signatures and phylogenomic analysis of the genus *Burkholderia*: proposal for division of this genus into the emended genus *Burkholderia* containing pathogenic organisms and a new genus *Paraburkholderia* gen. nov. harboring environmental species. *Front Genet.* <https://doi.org/10.3389/fgene.2014.00429>
53. Sharp PM, Li WH (1987) The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/15.3.1281>

54. Suárez-Moreno ZR, Caballero-Mellado J, Coutinho BG, Mendonça-Previato L, James EK, Venturi V (2012) Common features of environmental and potentially beneficial plant-associated Burkholderia. *Microb Ecol.* <https://doi.org/10.1007/s00248-011-9929-1>
55. Vanlaere E, van der Meer JR, Falsen E, Salles JF, Brandt E, Vandamme P (2008) *Burkholderia sartisoli* sp. nov., isolated from a polycyclic aromatic hydrocarbon-contaminated soil. *Int J Syst Evol Microbiol.* <https://doi.org/10.1099/ijs.0.65451-0>
56. Vanlaere E, Baldwin A, Gevers D, Henry D, Brandt ED, LiPuma JJ, Mahenthiralingam E, Speert DP, Dowson C, Vandamme P (2009) Taxon K, a complex within the *Burkholderia cepacia* complex, comprises at least two novel species, *Burkholderia contaminans* sp. nov. and *Burkholderia lata* sp. nov. *Int J Syst Evol Microbiol.* <https://doi.org/10.1099/ijs.0.001123-0>
57. Vial L, Groleau MC, Dekimpe V, Déziel E (2007) *Burkholderia* diversity and versatility: an inventory of the extracellular products. *J Microbiol Biotechnol* 17:1407–1429
58. Vial L, Chapalain A, Groleau MC, Déziel E (2011) The various lifestyles of the *Burkholderia cepacia* complex species: a tribute to adaptation. *Environ Microbiol.* <https://doi.org/10.1111/j.1462-2920.2010.02343.x>
59. Videira PA, Garcia AP, Sá-Correia I (2005) Functional and topological analysis of the *Burkholderia cenocepacia* priming glucosyltransferase BceB, involved in the biosynthesis of the cepacian exopolysaccharide. *J Bacteriol.* <https://doi.org/10.1128/JB.187.14.5013-5018.2005>
60. Vos M, Hesselman MC, Beek TAT, van Passel MWJ, Eyre-Walker A (2015) Rates of lateral gene transfer in prokaryotes: high but why? *Trends Microbiol.* <https://doi.org/10.1016/j.tim.2015.07.006>
61. Waack S, Keller O, Asper R et al (2006) Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinform.* <https://doi.org/10.1186/1471-2105-7-142>
62. Andreas W, la Chaux ND (2008) Distant horizontal gene transfer is rare for multiple families of prokaryotic insertion sequences. *Mol Genet Genom.* <https://doi.org/10.1007/s00438-008-0373-y>
63. Walton JD (2000) Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: an hypothesis. *Fungal Genet Biol.* <https://doi.org/10.1006/fgbi.2000.1224>
64. Wang Y, Xiao M, Geng X, Liu J, Chen J (2007) Horizontal transfer of genetic determinants for degradation of phenol between the bacteria living in plant and its rhizosphere. *Appl Microbiol Biotechnol.* <https://doi.org/10.1007/s00253-007-1187-2>
65. Wang Y, Li H, Zhao W, He X, Chen J, Geng X, Xiao M (2010) Induction of toluene degradation and growth promotion in corn and wheat by horizontal gene transfer within endophytic bacteria. *Soil Biol Biochem.* <https://doi.org/10.1016/j.soilbio.2010.03.002>
66. Weinert LA, Welch JJ, Jiggins FM (2009) Conjugation genes are common throughout the genus *Rickettsia* and are transmitted horizontally. *Proc R Soc B.* <https://doi.org/10.1098/rspb.2009.0875>
67. Wolf YI, Rogozin IB, Grishin NV, Koonin EV (2002) Genome trees and the tree of life. *Trends Genet* [https://doi.org/10.1016/s0168-9525\(02\)02744-0](https://doi.org/10.1016/s0168-9525(02)02744-0)
68. Yabuuchi E, Kosako Y, Oyaizu H, Yano I, Hotta H, Hashimoto Y, Ezaki T, Arakawa M (1992) Proposal of *Burkholderia* gen. nov. and transfer of seven species of the genus *Pseudomonas* homology group

II to the new genus, with the type species *Burkholderia cepacia* (Palleroni and Holmes 1981) comb. nov." *Microbiol Immunol* <https://doi.org/10.1111/j.1348-0421.1992.tb02129.x>

69. Zaneveld JR, Nemergut DR, Knight R (2008) Are all horizontal gene transfers created equal? Prospects for mechanism-based studies of HGT patterns. *Microbiology* <https://doi.org/10.1099/mic.0.2007/011833-0>

70. Zhu B, Zhou S, Lou M, Zhu J, Li B, Xie G, Jin GL, Mot RD (2011) Characterization and inference of gene gain/loss along burkholderia evolutionary history. *Evol Bioinform.* <https://doi.org/10.4137/EBO.S7510>

Figures

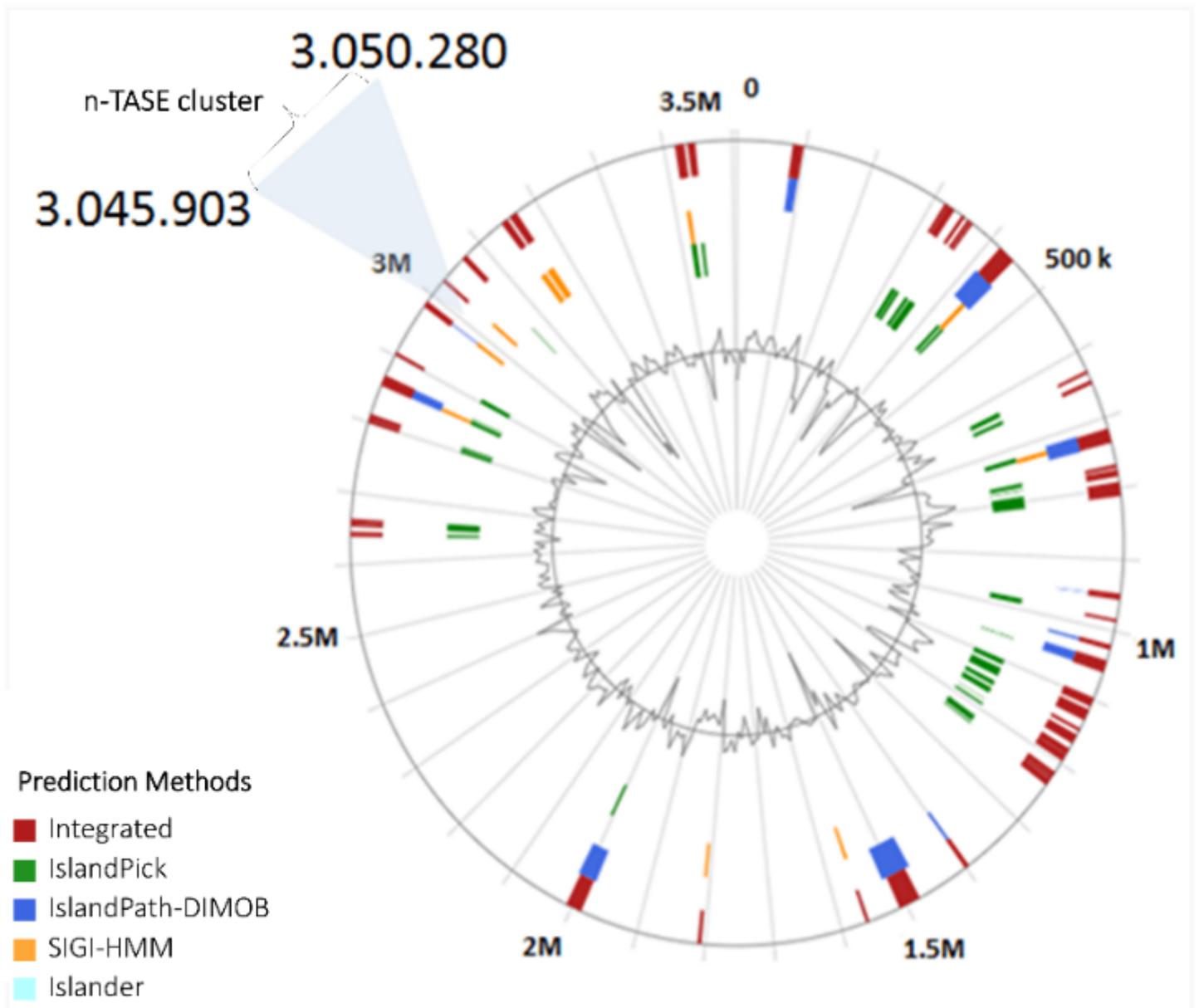


Figure 1

Prediction of genomic islands on chromosome 1 of *B. seminalis* TC3.4.2R3, circular visualization of the GIs with colored blocks according to the prediction tool; IslandPick (green), IslandPath-DIMOB (blue), SIGI-HMM (orange), Islander (blue-turquoise) as well as the integrated results in red. The region of the n-TASE cluster is located between 3,045,903 to 3,050,280 (Bsem_02857; Bsem_02858; Bsem_02859 and Bsem_02860)

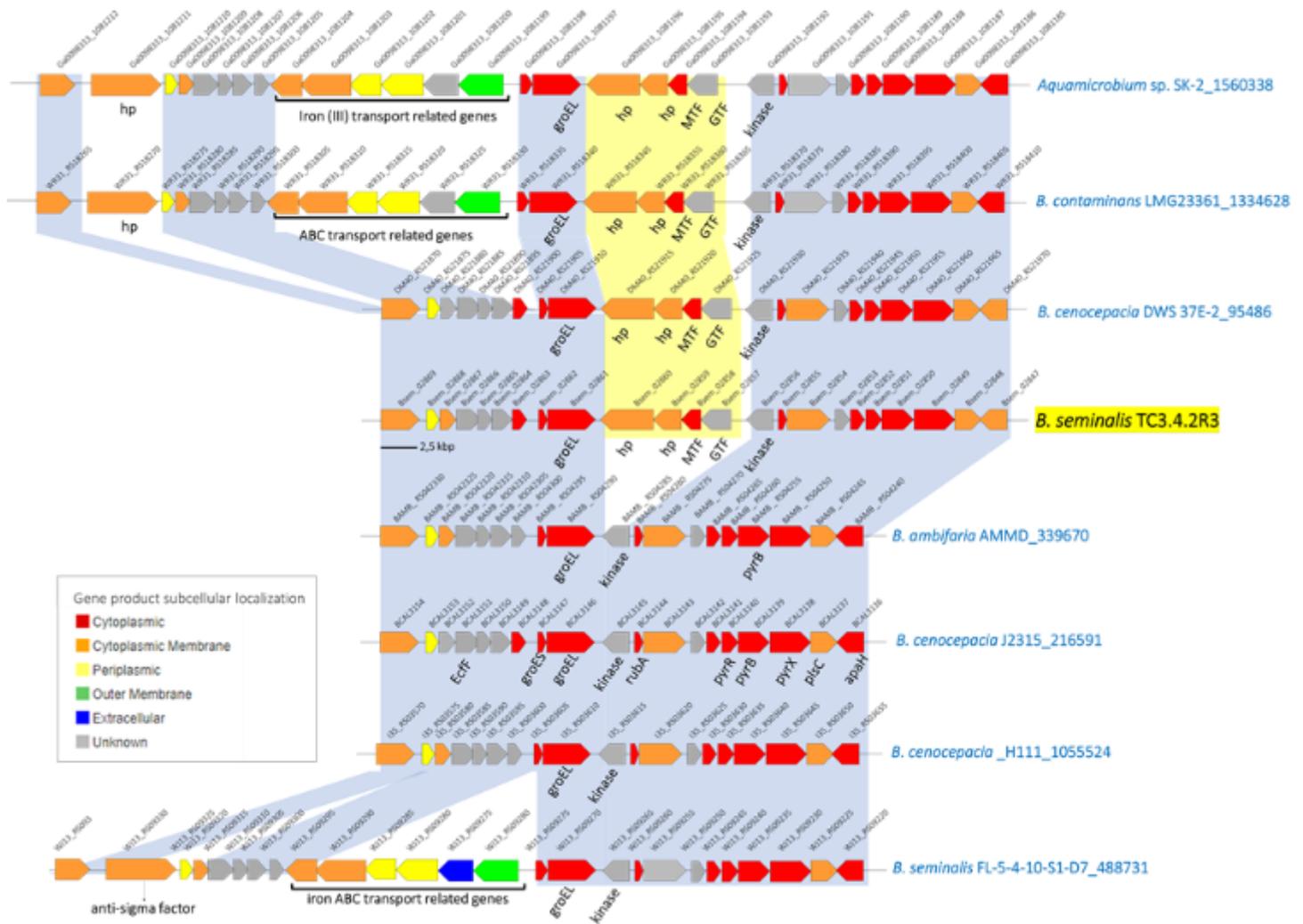


Figure 2

Comparison of the common 23,720 bp encoding regions of TC3.4.2R3 (comprising genes from locus tag Bsem_02847 to Bsem_02869) in *Aquamicrobium* sp. SK-2, *B. contaminans* LMG23361, *B. cenocepacia* DWS37E-2, *B. seminalis* TC3.4.2R3, *B. ambifaria* AMMD, *B. cenocepacia* J2315, *B. cenocepacia* H111 and *B. seminalis* FL-5-4-10-S1-D7. The NCBI taxon ID is written beside the strain name. The homologous genes for all compared strains are marked with the same background color (light blue for adjacent genes and yellow for n-TASE cluster genes) and linked together, while non-homologous genes were left blank. Gene colors were represented as their gene product subcellular localization based on the genetic maps in the 'Burkholderia Genome BD database' (Represented in the square on the left size). The extra six genes insertion composition is underlined in black and its function are described below. All compared strains has their locus tags identification (according to the Burkholderia Genome BD

database) depicted above the gene representation. The most commonly described genes are also represented below the gene presentation: groEL, molecular chaperone GroEL; kinase, pyrimidine kinase; hp, hypothetical protein; MTF, methyltransferase; GTF, glycosyltransferase; EcfF, sigma factor 70; groES, co-chaperonin GroES; pyrR, bifunctional pyrimidine regulatory protein uracil phosphoribosyltransferase; pyrB, aspartate carbamoyltransferase catalytic subunit; pyrX, dihydroorotase; plsC, 1-acyl-sn-glycerol-3-phosphate acyltransferase; and apaH, diadenosine tetraphosphatase

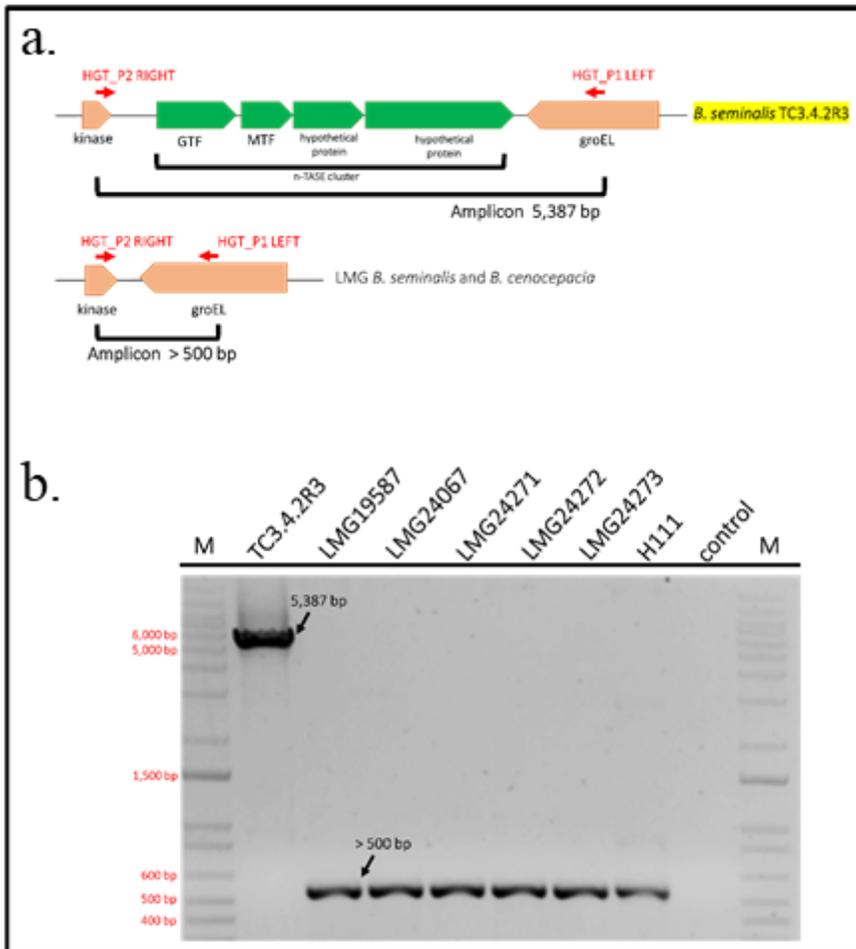


Figure 3

PCR detection of the n-TASE cluster in *Burkholderia* strains. (A) Representation for the genetic region containing the n-TASE cluster in TC3.4.2R3 and representation of other *Burkholderia* strains region (absence of n-TASE cluster region). Primers position, expected amplicon position and size were represented. MTF, methyltransferase; GTF, glycosyltransferase; groEL, molecular chaperone GroEL; kinase, pyrimidine kinase. (B) PCR amplification using DNA templates derived from *B. seminalis* TC3.4.2R3, five other *B. seminalis* strains (LMG19587, LMG24067, LMG24271, LMG24272 and LMG24273) and *B. cenocepacia* H111. Lanes: M, 1 kb plus DNA ladder; control, negative control made with dH₂O. All samples were run in the same gel. Black arrow indicates the PCR product size obtained from TC3.4.2R3 (5,387 bp) and for all other *Burkholderia* strains (> 500 bp)

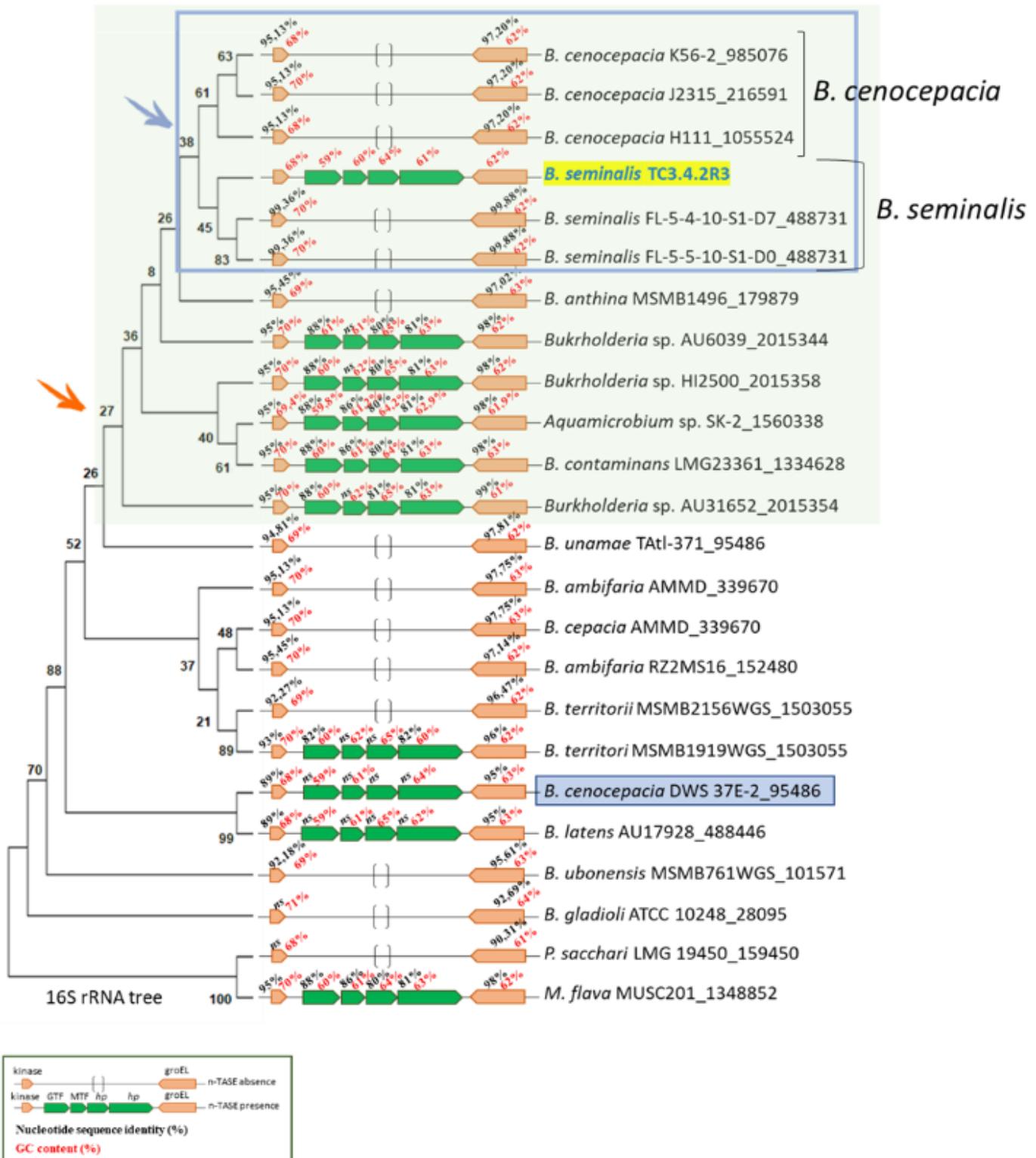


Figure 4

Neighbor-joining tree for 16S rRNA (left) for all compared *Burkholderia* strains obtained from the IMG database (containing n-TASE homologous sequence and not). The numbers at each node were the bootstrap values. The nucleotide composition, GC content (red) and nucleotide identity (black) are positioned at the bottom of the gene representation. Genetic representation corresponds to a pyrimidine kinase and a chaperonin GROEL genes (both in orange color) and the four genes composing n-TASE

cluster (GTF, glycosyltransferase; MTF, methyltransferase and two hp, hypothetical proteins) represented in green color. ns, “no significant similarity found” with sequences. Orange arrow indicates the phylogenetic node which n-TASE appears in different strains, including TC3.4.2R3. Blue arrow indicates the closest strains to TC3.4.2R3 (*B. cenocepacia* and *B. seminalis* – depicted with a blue square) and the absence of n-TASE in these strains

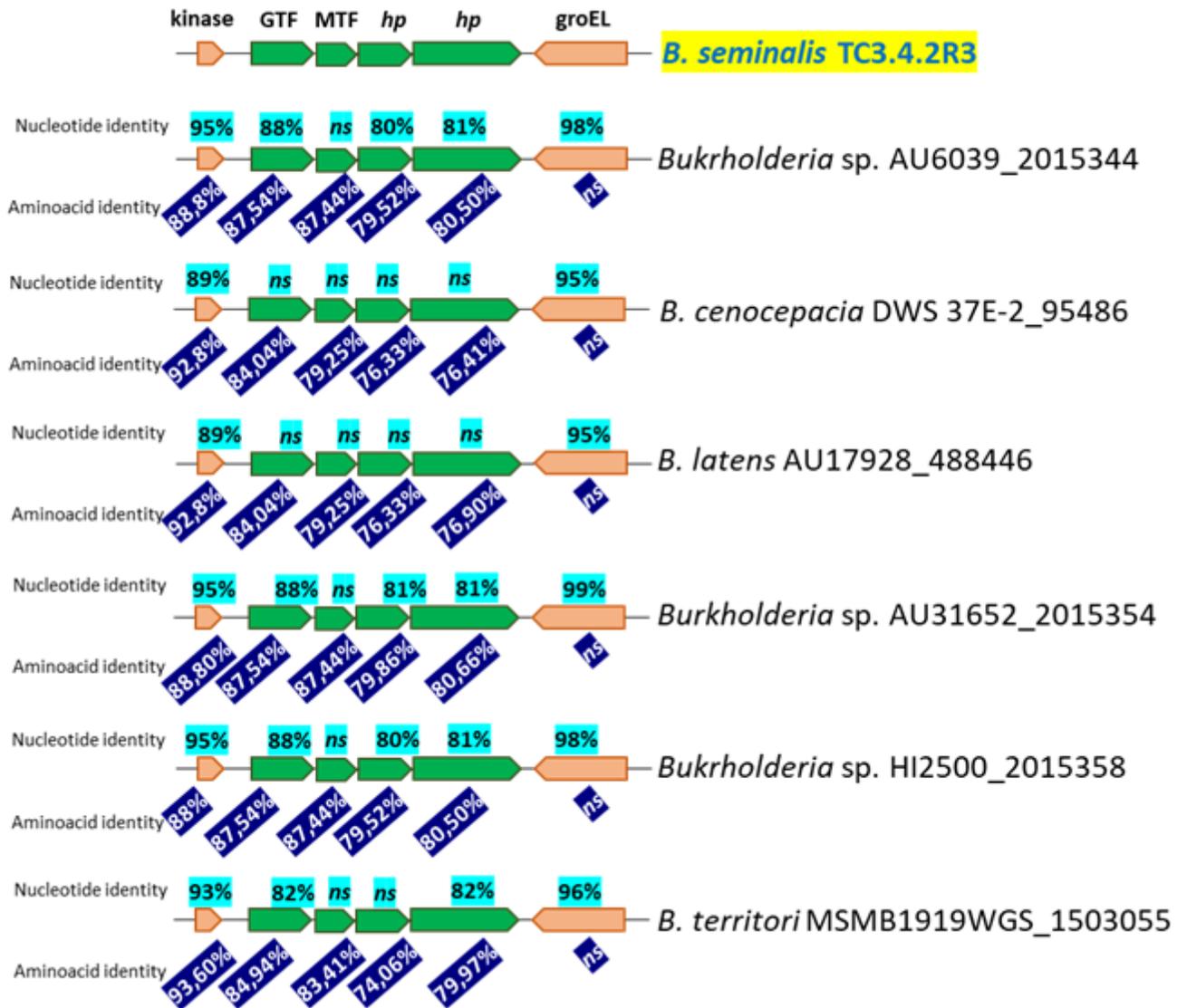


Figure 5

Nucleotide (clear blue) and amino acid (dark blue) sequences identity comparison for n-TASE containing strains, which showed a “no significant similarity found” in at least one of the n-TASE genes in the previous nucleotide sequence similarity search (Figure 4). ns, “no significant similarity found” with sequences

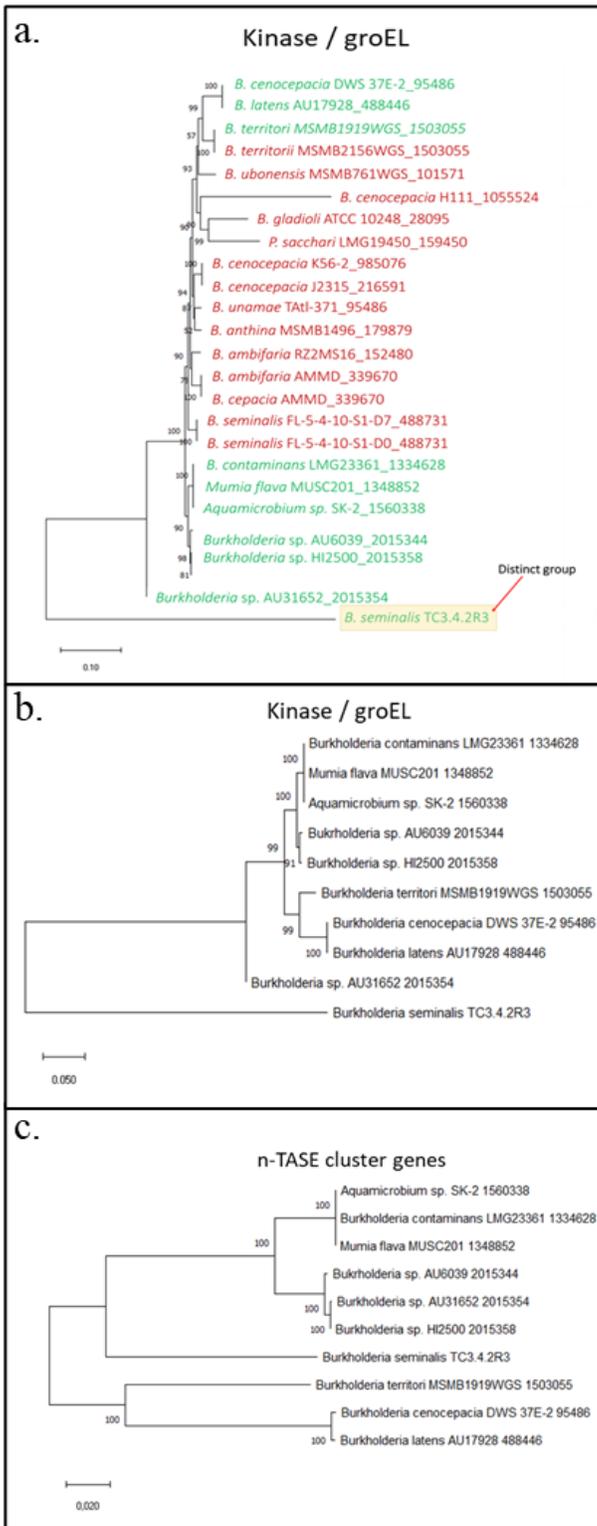


Figure 6

Neighbor-joining trees for concatenated sequences of (A) pyrimidine kinase and chaperonin GROEL from all compared strains. Strains with the n-TASE cluster are represented in green color, while strains without the n-TASE cluster are represented in red color; (B) pyrimidine kinase and chaperonin GROEL strains with the n-TASE cluster; (C) genes from the n-TASE cluster

glycosyltransferase

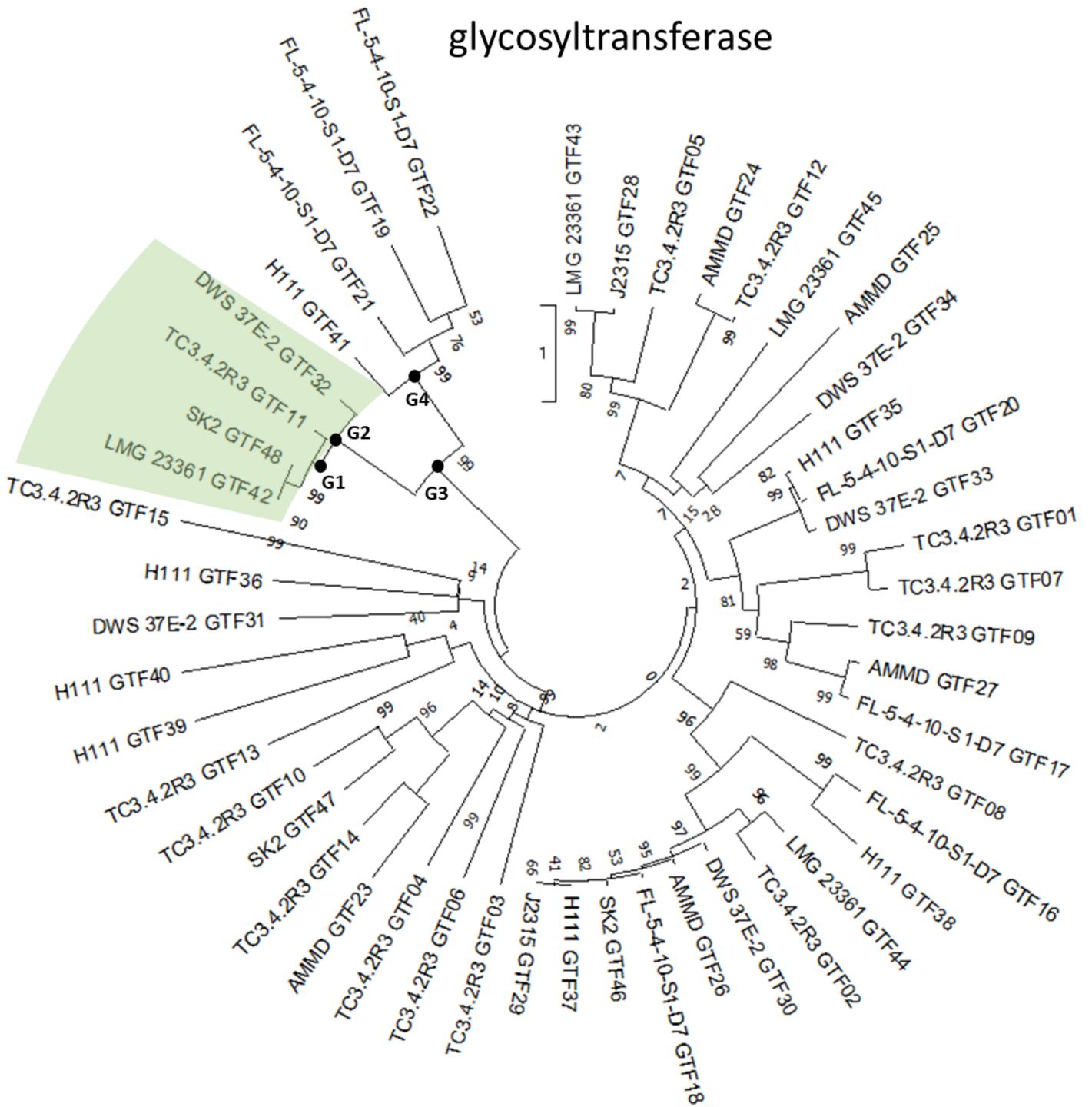


Figure 7

Maximum likelihood tree for the glycosyltransferase selected sequences. A distinct clade with Bsem_02857 (TC3.4.2R3_GTF11) gene sequence (highlighted in green color on the left size) is formed, all other glycosyltransferase genes selected from TC3.4.2R3 strain did not clustered with Bsem_02857. Bootstrap percentages are shown for all nodes

methyltransferase

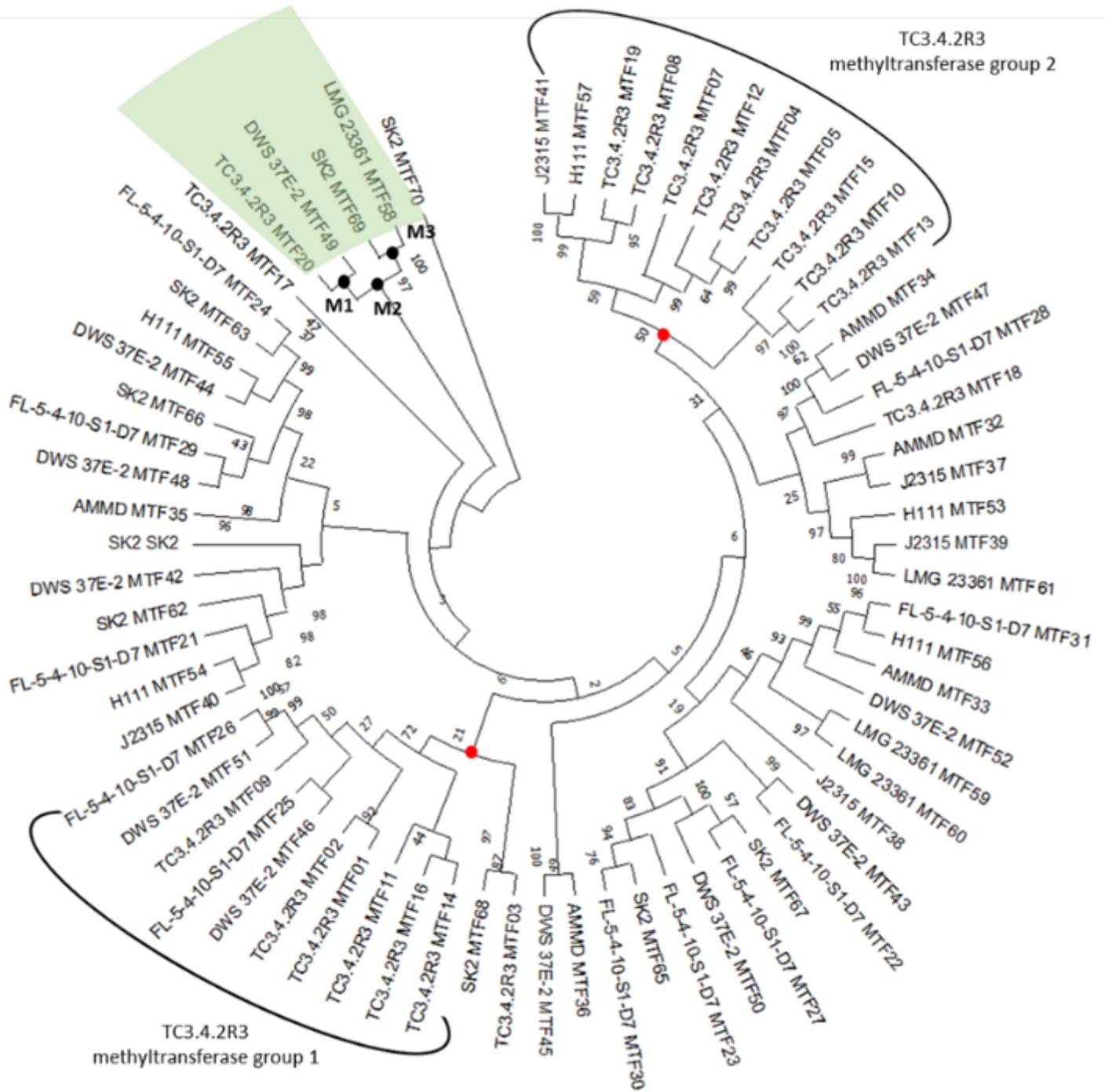


Figure 8

Maximum likelihood tree for the methyltransferase selected sequences. A distinct clade with Bsem_02858 (TC3.4.2R3_MTF20) gene sequence (highlighted in green color on the left size) is formed, all other methyltransferase genes selected from TC3.4.2R3 strain did not clustered with Bsem_02858. There are two distinct groups of methyltransferase genes from the TC3.4.2R3, they are depicted with a

black line and identified as TC3.4.2R3 methyltransferase group 1 and group 2. Bootstrap percentages are shown for all nodes

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [renamed032e5.docx](#)