Arch Microbiol
DOI 10.1007/s00203-015-1166-4

ORIGINAL PAPER

Insights from genome of *Clostridium butyricum* INCQS635 reveal mechanisms to convert complex sugars for biofuel production

Thiago Bruce[1,4] · Fernanda Gomes Leite[1] · Milene Miranda[2] · Cristiane C. Thompson[2] · Nei Pereira Jr.[3] · Mariana Faber[3] · Fabiano L. Thompson[2]

Received: 26 May 2015 / Revised: 9 October 2015 / Accepted: 22 October 2015

Abstract *Clostridium butyricum* is widely used to produce organic solvents such as ethanol, butanol and acetone. We sequenced the entire genome of *C. butyricum* INCQS635 by using Ion Torrent technology. We found a high contribution of sequences assigned for carbohydrate subsystems (15–20 % of known sequences). Annotation based on protein-conserved domains revealed a higher diversity of glycoside hydrolases than previously found in *C. acetobutylicum* ATCC824 strain. More than 30 glycoside hydrolases (GH) families were found; families of GH involved in degradation of galactan, cellulose, starch and chitin were identified as most abundant (close to 50 % of all sequences assigned as GH) in *C. butyricum* INCQS635. KEGG metabolic pathways reconstruction allowed us to verify possible routes in the *C. butyricum* INCQS635 and *C. acetobutylicum* ATCC824 genomes. Metabolic pathways for ethanol synthesis are similar for both species, but alcohol dehydrogenase of *C. butyricum* INCQS635 and *C. acetobutylicum* ATCC824 was different. The genomic repertoire of *C. butyricum* is an important resource to underpin future studies towards improved solvents production.

Keywords Genomics · Carbohydrate-active enzymes · Biofuel · *Clostridium butyricum* · Ion Torrent

Communicated by Jorge Membrillo-Hernández.

Electronic supplementary material The online version of this article (doi:10.1007/s00203-015-1166-4) contains supplementary material, which is available to authorized users.

✉ Thiago Bruce
thbruce@gmail.com

✉ Fabiano L. Thompson
fabianothompson1@gmail.com

[1] Faculdade de Tecnologia e Ciências, Laboratory of Environmental Biotechnology, Salvador, Brazil

[2] Laboratory of Microbiology and SAGE-COPPE, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

[3] Laboratory of Bioprocesses Development, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

[4] Department of Biotechnology, Federal University of Bahia, Salvador, Brazil

Introduction

The genus *Clostridium* consists of more than 100 bacterial species with well-established ability to consume hexose and pentose sugar monomers released from woody material. One of the forms of these sugars is metabolised through fermentation. Acetone–butanol–ethanol (ABE) is one of the oldest known industrial fermentations (Dürre 1998). The market demand for solvents is expected to dramatically increase since ethanol, butanol and acetone can now be produced economically from low-cost biomass. An immense amount of biomass is lost as residue worldwide. This waste could be reused in the production of biofuels such as ethanol or solvents with great industrial interest, such as ABE. Since the 1950s, industrial ABE fermentation has declined continuously. Among organic solvents, butanol emerges as a promising alternative fuel. Compared to ethanol (19.6 MJ/l), butanol (27 MJ/l) has high energy content, low miscibility with water and low volatility with the potential to meet the need of sustainable and green energy systems and can be transported through existing pipeline infrastructure (Jang et al. 2012; Li et al. 2014). Furthermore, the energy content of butanol is similar to that of gasoline (32 MJ/l), approximately 11 times the vapour pressure (4 mmHg at 20 °C) less than that of ethanol

Published online: 02 November 2015

(45 mmHg at 20 °C). Finally, butanol can replace gasoline without any modification of the current vehicle and engine technologies (Cheng et al. 2012). However, there is a lack of information concerning the enzymatic potential of several *Clostridium* strains with industrial applications.

Microorganisms of interesting metabolic properties include non-pathogenic bacteria of the genus *Clostridium*, particularly *C. acetobutylicum*, and *C. butyricum*. Unlike *C. butyricum* strain 5521, which had its genome sequenced because of its clinical importance, since it is characterised by being a toxigenic strain, this work was focused on the sequencing of *C. butyricum* INCQS635 strain. Main motivation to *C. butyricum* INCQS635 genome sequencing is related to their biotechnological potential. This strain has been well established by its ability to convert by-product of biodiesel production (glycerol) to 1,3-propanediol. The genetic repertoire was investigated to elucidate the regulation of metabolic processes involved (Raynaud et al. 2011, 2003). Experiments showed strain INCQS635 ranging between 0.56 and 0.57 g per g of glycerol consumed, close to the maximum theoretical value of 0.60 g g^{-1}, and butyric acid was the predominant organic acid with maximum end concentration of 8.8 g L^{-1} (Chatzifragkou et al. 2011; Zeng et al. 1994). Added to these findings, the ability widely distributed among *Clostridium* of converting sugars into organic acids by ABE fermentation points this strain as promising from biotechnological point of view. So it is urgent to deepen the knowledge about the strain INCQS635 for better understanding of their genetic and metabolic potential. The *C. butyricum* INCQS635 potential set meets biorefinery concept, where products of an industrial stage are used as raw material, mainly residues, for generation of marketable products.

Microbial enzymes represent a promising alternative for the degradation of lignocellulosic material present in plant biomass and the production of biofuels (Lynd et al. 2002). The major pathways leading to solvent formation are already known (Gheshlaghi et al. 2009). However, the potential of *Clostridium* strains is not fully understood. The genomic characterisation of clostridia used in industrial applications is still limited (Gheshlaghi et al. 2009; Hou et al. 2013; Lee et al. 2012; Lütke-Eversloh and Bahl 2011). Among the 200 known *Clostridium* species, only about 20 have had their genomes completely sequenced, and most are human pathogens (see GOLD—Genomes Online Database) (Pagani et al. 2012) with no industrial applications for solvent production. The regulatory mechanisms controlling solventogenesis have begun to emerge with the analysis of *C. acetobutylicum* (Nölling et al. 2001).

The genome sequencing of a new *C. butyricum* strain (INCQS635) will be useful, as it will broaden our knowledge on the genomic diversity related to EBA fermentation. The aim of this study was to perform a genomic analysis of strain INCQS635 with focus on potential to convert complex sugars to organic solvents.

Methods

Clostridium strains and whole-genome sequences

The *Clostridium butyricum* INCQS635 strain was acquired from the collection of the National Institute of Quality Control in Health (INCQS–FIOCRUZ, Brazil). The *C. butyricum* INCQS635 strain corresponds to VPI1718, NCIMB 8082 and ATCC860. The strain was delivered in vials, stored in the lyophilised form. Reactivation was performed with modified reinforced clostridial broth medium for 24 h at 37 °C in anaerobiosis. The genome was sequenced in this study and deposited in the RAST server (ID: 1492.7) for comparison with two other strains of *Clostridium*, and whole-genome shotgun project was deposited at DDBJ/EMBL/GenBank under the accession number JRMA00000000.

The genomes of *C. butyricum* 5521 strain (first genome of *C. butyricum* released in GenBank) and *C. acetobutylicum* were obtained from GenBank, deposited under the accession numbers ABDT01 and AE001438, respectively. The genomes were downloaded and re-annotated as described below in the section focused on genome annotation.

Ion Torrent-based genome sequencing

Clostridium butyricum INCQS635 DNA was extracted using the Pitcher method (Pitcher et al. 1989). The Ion Torrent sequencing was performed as described by Quail et al. (2012) with minor modifications described as follows. Library preparation was carried out using the Ion Plus Fragment Library kit, with 1 µg of DNA diluted in 50 µl of low TE. The DNA was fragmented by using the BioRuptor® sonication system as described in Ion Plus Fragment Library kit protocol. End repair, adapter ligation, nick repair and amplification (8 cycles) were also performed as described in the Ion Plus Fragment Library protocol. Size selection after adapter ligation was performed using 2 % agarose gel, with collection between 300 bp and 400 bp. The quality and concentration of the libraries were determined by means of the Agilent 2100 Bioanalyser (Agilent Technologies) and the associated High Sensitivity DNA kit (Agilent Technologies), and by the Applied Biosystems® 7500 Real-Time PCR system with the Ion library Quantisation kit using TaqMan®. The amount of library required for template preparation was calculated using the template dilution factor calculation described in the protocol. Emulsion PCR and enrichment steps were carried out in the

Ion OneTouch™ 200 Template kit v2. Ion sphere particle quality assessment was carried out as outlined in this protocol. Ion Torrent sequencing was undertaken using 314 chips. The Ion PGM™ 200 Sequencing kit was used for all sequencing reactions, following the recommended protocol. The Torrent suite 1.5 was used for preliminary analyses. The reads obtained were assembled using De novo Assembly in the Ion Torrent Platform based on the MIRA software (Chevreux et al. 2004). The genomes were annotated automatically using the RAST server (Rapid Annotations Using Subsystems Technology) (Aziz et al. 2008).

Annotation of *Clostridium butyricum* INCQS635 genome sequences

To assess the genetic potential related to the production of metabolic products with biotechnological interest as organic solvents, the genomes were annotated by three different strategies. It was then possible to present an overview of the metabolic pathways committed to the formation of these products. The first entry was made through the subsystems approach through the RAST server (Overbeek et al. 2014). The functional classification of proteins from genes coding was calculated by projecting the FIGfam, based on similarity of searches through BlastX.

Genomes were also annotated for the identification of conserved domains present in genes encoding proteins involved in carbohydrate metabolism. Those proteins code for carbohydrate-active enzymes (so-called CAZYmes). Genomic sequences were compared with sequences deposited on the CAZY database (Cantarel et al. 2009) by using the CAZYmes Analysis Tool kit (e-value 10^{-5}) (Park et al. 2010). This specialised database describes families of enzymes that degrade, modify or create glycosidic linkages.

The metabolic pathways have been reconstructed by the genes based on orthology analysis using the KEGG database (Aoki-Kinoshita and Kanehisa 2007), available on the RAST server. This allowed the reconstruction of metabolic pathways involved in ethanol, acetone and butanol production. The reconstruction in parallel allowed us to identify possible specific metabolic pathways, as well as those shared between species. From these reconstructions, it was possible to identify each of the sequences committed to each step of the metabolic pathways and their EC number, as well as the degree of homology between the species evaluated. The EC numbers, as well as the genes, were confirmed by searching the GenBank database using the BLAST Nucleotide tool (Mega Blast).

Solvent production by *C. butyricum* INCQS635

The strain was activated in activation medium (peptone 10 g/L; beef extract 10 g/L; yeast extract 3 g/L; sodium chloride 5 g/L; cysteine 0.5 g/L; sodium acetate 3 g/L; 0.5 g/L; glucose 5 g/L). Then, 0.001 % rezasurina (marker of O_2) was inoculated into culture medium in the proportion 1:50 and grown at 37 °C/24 h.

Culture was propagated twice in fresh medium (10:50 ratio) to be transferred to the production medium (0.215 % $MgSO_4$, 0.001 % $MnSO_4$, 0.001 % sodium acetate, 0.002 % iron sulphate, 5 % glucose). Growth rates (0, 72 and 96 h) were monitored to determine solvent production.

Analysis was performed by using high-performance liquid chromatography (HPLC—Shimadzu 20A). Ionic exchange column was used. Agilent HiPlex H 7.7 × 300 mm, 8-µm column was used, which presents a stationary phase composed of cationic resin consisting of sulphonated styrene–divinylbenzene copolymer, and mobile phase was 0.005 mol H_2SO_4/L: cationic resin formed by sulfonated styrene–divinylbenzene (flux: 0.6 mL/min at 45 °C). Purified molecules used as reference were 99.8 % acetic acid (Sigma-Aldrich) and 99 % n-butyric acid (Vetec).

Results

General features of *Clostridium* genomes

Table 1 shows the numbers related to the *Clostridium* genomes. Right side shows general information regarding genomes annotated using RAST technology. The genome sequencing of *C. butyricum* INCQS635 generated 345 contigs. The genome comprised a total size of 4,428,945 bp; 87 RNAs and 5771 protein-coding sequences were identified. From this total, 2626 were classified into subsystems 376. The *C. butyricum* 5521 genome was composed of 4,540,699 bp and 123 contigs. It showed a total of 137 RNAs and 4031 protein-coding sequences, of which 1692 were recorded in 353 subsystems. *C. acetobutylicum* ATCC824 had the smallest genome equivalent to 3,381,828 bp and consisted of 3672 contigs. We identified 364 subsystems; only one RNA gene sequence was found in this strain, but a previous work was able to identify 107 RNA genes and a total of 3608 protein-coding sequences distributed in 1563 subsystems.

Right side shows general information regarding genomes annotated using RAST technology. On the left side of Table 1, we dissected information about the coding genes for each genome. A total of 46 and 42 % of the sequences were identified in subsystems in *C. butyricum* INCQS635 and *C. butyricum* 5521. Approximately 44 % of the total protein-coding sequences from *C. acetobutylicum* ATCC824 belonged to hypothetical and non-hypothetical component subsystems. Sequences identified as coding sequences were divided into two main groups according to subsystem (set of related functional roles) assignment.

Table 1 General features for tree genomes of fermentative *Clostridium* lineages annotated by subsystems technology

General information for C. butyricum INCQS635		*Number of coding sequences in C. butyricum INCQS635*	
Domain	Bacteria	In subsystem	
Taxonomy	Bacteria, Firmicutes, Clostridia, Clostridiales, Clostridiaceae, Clostridium, Clostridium butyricum INCQS365	Hypothetical Not-hypothetical	101 2525
Size	4,428,945 bp	Not in subsystem	
Number of contigs (with PEGs)	345	Hypothetical	1609
Number of subsystems	376	Not-hypothetical	1536
Number of coding sequences	5771	Total	5771
Number of RNAs	87		
General information for Clostridium butyricum 5521		*Number of coding sequences in C. butyricum* 5521	
Domain	Bacteria	In subsystem	
Taxonomy	Bacteria, Firmicutes, Clostridia, Clostridiales, Clostridiaceae, Clostridium, Clostridium butyricum 5521	Hypothetical Not-hypothetical	61 1631
Size	4,540,099 bp	Not in subsystem	
Number of contigs (with PEGs)	123	Hypothetical	1181
Number of subsystems	353	Not-hypothetical	1158
Number of coding sequences	4031	Total	4031
Number of RNAs	137		
General information for Clostridium acetobutylicum ATCC 824		*Number of coding sequences in C. acetobutylicum ATCC 824*	
Domain	Bacteria	In subsystem	
Taxonomy	Bacteria, Firmicutes, Clostridia, Clostridiales, Clostridiaceae, Clostridium, Clostridium acetobutylicum, Clostridium acetobutylicum ATCC 824	Hypothetical Not-hypothetical	62 1501
Size	3,381,828 bp	Not in subsystem	
Number of contigs (with PEGs)	3672	Hypothetical	957
Number of subsystems	364	Not-hypothetical	1088
Number of coding sequences	3608	Total	3608
Number of RNAs	107		

Clostridium butyricum INCQS635 was sequenced and annotated in this work. *Clostridium butyricum* ATCC824 and *Clostridium acetobutylicum* were sequenced previously and re-annotated for this work

Here, we found a fraction of sequences which are not yet assigned under any established subsystem, referred as "Not in subsystem", and sequences assigned under already established subsystems, referred as "In subsystem". Moreover, into the subsystems, coding sequences were assigned according to the comparison of sequences deposited on databases in order to identify hypothetical functional role and divided into "Hypothetical" (gene is functionally coupled to at least one other protein-encoding gene that has been assigned a function, but all of these coupled neighbours are hypothetical) and "Not-hypothetical" (gene is functionally coupled to at least one other protein-encoding gene that has been assigned a function that is considered non-hypothetical).

Functional profile based on subsystems annotation

The identified sequences were divided into 26 subsystems (Fig. 1a; Supplementary 1). The subsystems with higher contribution accounted for 8–40 % of the identified sequences. *C. butyricum* exhibit a greater contribution of the carbohydrates subsystems (18 % for strain INCQS635 and 20 % for strain 5521, compared with 15 % contribution in the genome of *C. acetobutylicum* ATCC824). The carbohydrate subsystem included fermentation processes for ethanol, butanol and acetone production (Fig. 1b). Fermentation subsystems included mixed acids fermentation, butanol biosynthesis, lactate fermentation, butanediol and acetoin metabolism and a subunit of acetolactate synthase

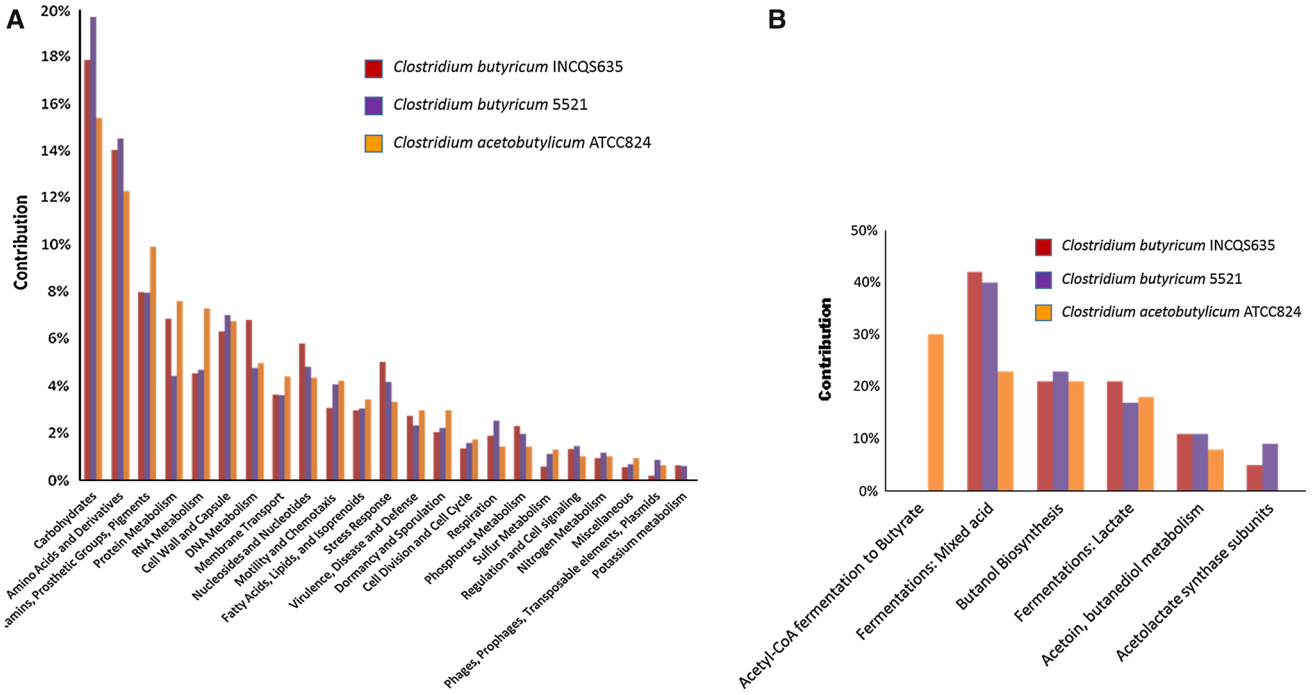

Fig. 1 Subsystems category contribution for each sequenced genome. Assigned sequences were distributed along 28 subsystems. **a** A higher contribution was observed for the carbohydrate subsystem in lineages. **b** Fermentation subsystems found in carbohydrate metabolism were mixed acids fermentation, biosynthesis of butanol, lactate fermentation, butanediol and acetoin metabolism and a subunit of acetolactate synthase, in the case of *C. butyricum*, and fermentation of acetyl-CoA to butyrate in *C. acetobutylicum*

in the case of *C. butyricum*, and fermentation of acetyl-CoA to butyrate for *C. acetobutylicum*. Fermentation contribution for *C. butyricum* INCQS635 reached 10 % of the sequences involved in the carbohydrate subsystem followed by *C. butyricum* 5521 (15 %) and *C. acetobutylicum* ATCC824 (12 %).

In *C. butyricum* INCQS635, 42 % of the sequences were involved in the fermentation of mixed acids, 21 % in the biosynthesis of butanol, another 21 % in the fermentation of lactate, 11 % in butanediol and acetoin metabolism and 5 % in subunits of acetolactate synthase. In *C. butyricum* 5521, 40 % of the sequences were involved in fermentation working for mixed acids fermentation, 23 % in butanol biosynthesis, 17 % in lactate fermentation, 11 % in butanediol and acetoin metabolism and 9 % in subunits of acetolactate synthase. In *C. acetobutylicum* ATCC824, 30 % of sequences assigned to the fermentation subsystem were involved in the fermentation of acetyl-CoA butyrate, 23 % in the fermentation of mixed acids, 21 % in the biosynthesis of butanol, 18 % in the fermentation of lactate and 8 % in the metabolism of butanediol and acetoin. The second greatest contribution was found for amino acids subsystems (12–15 %). In *C. butyricum* 5521, this contributed 15 %, followed by *C. butyricum* INCQS635 (14 %) and *C. acetobutylicum* ATCC824, which reached 12 % of the sequences identified.

Subsystems with an intermediate contribution represented 3–10 % of the identified sequences. Subsystems such as cofactors, prosthetic groups and vitamins showed 8 % in *C. butyricum* INCQS635 followed by *C. butyricum* 5521 (10 %) and *C. acetobutylicum* ATCC824. The protein metabolism represented 7 % of contribution in *C. butyricum* INCQS635, 4 % in *C. butyricum* 5521 and 8 % in *C. acetobutylicum* ATCC824. The RNA metabolism subsystem contributes to 5 % of *C. butyricum* (both) and 7 % of *C. acetobutylicum* ATCC824. DNA metabolism reached a contribution of 7 % in *C. butyricum* INCQS635 and 5 % in other organisms. The cell wall and capsule subsystem presented a 6 % contribution in *C. butyricum* INCQS635 and 7 % in the other two genomes. The DNA metabolism subsystem presented a 7 % contribution of *C. butyricum* INCQS635 and 5 % in *C. butyricum* 5521 and *C. butyricum acetobutylicum* ATCC828. Membrane transport subsystems had a contribution of 4 % in all of the analysed genomes. The nucleosides and nucleotides subsystem has a 6 % contribution in *C. butyricum* INCQS635, 5 % in *C. butyricum* 5521 and 4 % in *C. acetobutylicum* ATCC824. The response to a cellular stress subsystem represented 5 % of the sequences in *C. butyricum* INCQS635, 4 % in *C. butyricum* 5521 and 3 % in *C. acetobutylicum* ATCC824. Motility and chemotaxis subsystems showed a 3 % contribution in *C. butyricum* INCQS635 and 4 % in other organisms.

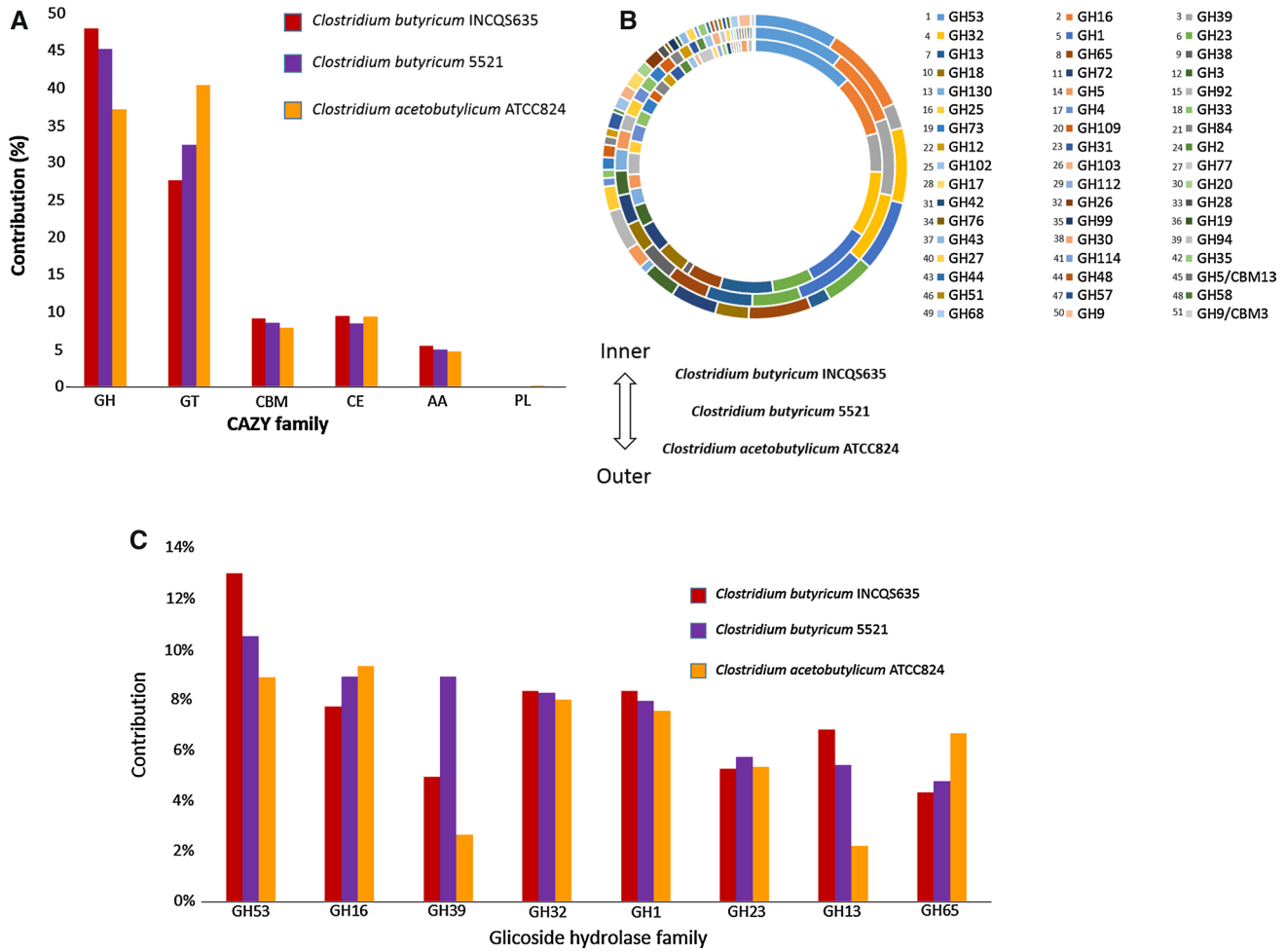

Fig. 2 Carbohydrate-active enzymes (CAZYmes) distribution. **a** Distribution of major protein families involved in carbohydrate metabolism (*GH* glycoside hydrolases, *GT* glycosyl transferase, *CE* carbohydrate esterase, *CBM* carbohydrate-binding domain, *AA* auxiliary activities and *PL* polysaccharide lyase); **b** the *circles* represent the contribution of all GH families found for each genome. The *inner circle* represents GH families found in *C. butyricum* INCQS635 genome, followed by *C. butyricum* 5521 in the middle, and the *outer circle* represents families of GH in *C. acetobutylicum* ATCC824 genome. *Sequential numbers* represent position on contribution ranking; **c** contribution of the most abundant families of GH found in genomes. The first eight families, which contributed at least to 50 % of all sequences assigned as GH, were considered most abundant

Carbohydrate-active enzymes and glycoside hydrolase contribution

We found sequences distributed in different subfamilies of CAZYmes (Fig. 2a). We identified 672 coding sequences (11.6 % contribution in relation to the number of coding sequences) that represented CAZYmes in *C. butyricum* INCQS635 sequences followed by 605 (16.7 %) in *C. acetobutylicum* and 693 (17.2 %) in *C. butyricum* 5521. The distribution of these sequences in CAZYmes families was evaluated in the three strains (Fig. 2a). Most enzymes found were classified into the large family of glycoside hydrolases (GH), varying between 37 and 48 % among the species. From 27 to 40 % of the enzymes were classified as belonging to the glycosyltransferases (GT), and around 8 % were carbohydrate esterases (EC). The auxiliary activities (AA) family range a smaller contribution around 4–5 %.

Clostridium butyricum INCQS635 presented 323 (48 %) sequences belonging to the GH family, 186 (27 %) to the GT family, 64 (9.5 %) to the CE enzyme family, 62 (9.2 %) to CBM and 37 (5.5 %) to AA. In *C. butyricum* 5521 genome, 314 (45.3 %) sequences were assigned as GH family, 225 (32.4 %) as GT, 59 (8.5 %) as CE, 60 (8.7 %) as CBM and 35 (5 %) as AA. Finally, we found in *C. acetobutylicum* ATCC824 genome a total of 245 sequences (40.4 %) for GT and 225 (37 %) for GH, in contrast to *C. butyricum* strains whose family of GH was most abundant. It was followed by 57 (9.4 %) sequences for CE, 48 (8.6 %) for carbohydrate-binding modules, 4.7 % for AA and only

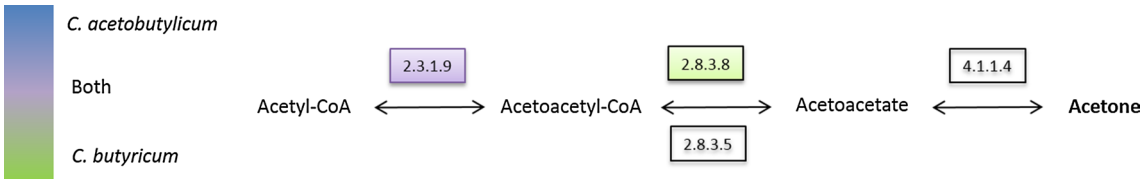

Fig. 3 Comparative metabolic pathways for acetone synthesis present in *Clostridium* genomes. *Arrows* indicate direction of metabolic steps, and *boxes* show EC number of key enzymes for each pathway. *Box colours* indicate the presence of sequences in *C. acetobutylicum* (*blue*), *C. butyricum* (*green*) and present in both species (*purple*). *Blank box* indicates the absence of sequences related to the step (colour figure online)

0.1 % for PL. Any sequence assigned as PL was found in *C. butyricum* genomes.

Sequences assigned as GH were distributed along 31, 29 and 34 families (*C. butyricum* INCQS635, *C. butyricum* 5521 and *C. acetobutylicum* ATCC824, respectively) (Fig. 2b). A high number of sequences were concentrated in a few number of families, and a higher variety of families present low contribution of sequences. The eight most abundant GH families contributed to approximately 50 % of all sequences.

Figure 2c presents the contribution of the most frequent GH families found in the analysed genomes. GH53 and GH16 families were the most abundant in *C. butyricum* INCQS635, *C. butyricum* 5521 and *C. acetobutylicum* ATCC824 genomes (13, 11 and 9 %, respectively). We observed a higher contribution of families GH39 and GH13 in *C. butyricum* strains in comparison with *C. acetobutylicum* ATCC824. Family GH39 contributes to 5 and 9 % in *C. butyricum* INCQS635 and 5521 genomes, against 3 % in *C. acetobutylicum* ATCC824. Family GH13 contributed to 7 and 5 % in *C. butyricum* (INCQS and 5521, respectively) and only 2 % in *C. acetobutylicum* ATCC824 genome. On the other hand, family 65 presented tendency to be higher in *C. acetobutylicum* ATCC824 (7 %) in comparison with *C. butyricum* (5 % for INCQS and 4 % for 5521).

Finally, families GH32 and GH1 presented similar contribution in the three genomes (8 %), and family GH23 contributed in a similar way (5–6 %). The contribution of following GH families was considered low. We observe discrete shifts related to contribution along the three genomes (Fig. 2b).

Comparative metabolic pathways for acetone synthesis

The flow chart for acetone metabolism shows the EC numbers related to the sequences found in the genomes analysed (Fig. 3). Enzymes highlighted are those presented in sequenced genomes and come from genetic sequences organised in ECs (supplementary 2). The enzyme which acts to convert acetyl-CoA to acetoacetyl-CoA is present in all genomes analysed. It is an acetyl-CoA acetyltransferase, and the reaction is represented by EC 2.3.1.9. Acetoacetyl-CoA may act as a substrate to follow intermediary reaction to produce acetoacetate.

In a first view, the metabolic pathways related to acetone metabolism suggest conversion of acetoacetyl-CoA to acetate by enzymes classified under EC 2.8.3.5 (from KEGG reconstruction), regarding synthesis and degradation of ketone bodies. However, sequences related to EC 2.8.3.5 were not found in the *C. butyricum* INCQS635, *C. butyricum* 5521 or *C. acetobutylicum* ATCC825 genomes. We found the conversion of acetoacetyl-CoA to acetoacetate by enzymes related to EC 2.8.3.8 (acetyl-CoA/acetoacetyl-CoA-transferase) accessing propanoate metabolism in both *C. butyricum* strains. Five different sequences were identified in the *C. butyricum* INCQS635 and two sequences in *C. butyricum* 5521 genome.

In *C. acetobutylicum* ATCC825 genome were identified sequences assigned under EC 2.3.1.9. In contrast, the conversion of acetoacetyl-CoA to acetoacetate by EC 2.8.3.8-related enzymes is carried out by acetate CoA-transferase, as observed in which is not present in the metabolic route map performed using RAST annotation for this genome.

The key enzymes for acetone production promote the decarboxylation of acetoacetate-producing acetone, as characterised by EC 4.1.1.4. However, the enzyme acetoacetate decarboxylase was not found by using RAST annotation method.

Comparative metabolic pathways for butanol synthesis

C. acetobutylicum ATCC824 is known for its ability to produce butanol. For this reason, the metabolic model was used as a reference for comparison with *C. butyricum* strains in this work. The comparative flow chart exposes pathways shared between strains, as well as exposing singularities of each one. Butanol flow chart metabolism from acetyl-CoA is presented in Fig. 4. In general, main differences between species are the presence of the EC 2.8.3.8 in *C. butyricum* strains, while sequences identified under this EC number were not found in *C. acetobutylicum* genome sequences. In contrast, a key enzyme involved in conversion of butanal to butanol, under EC 1.1.1.-, was present only in *C.*

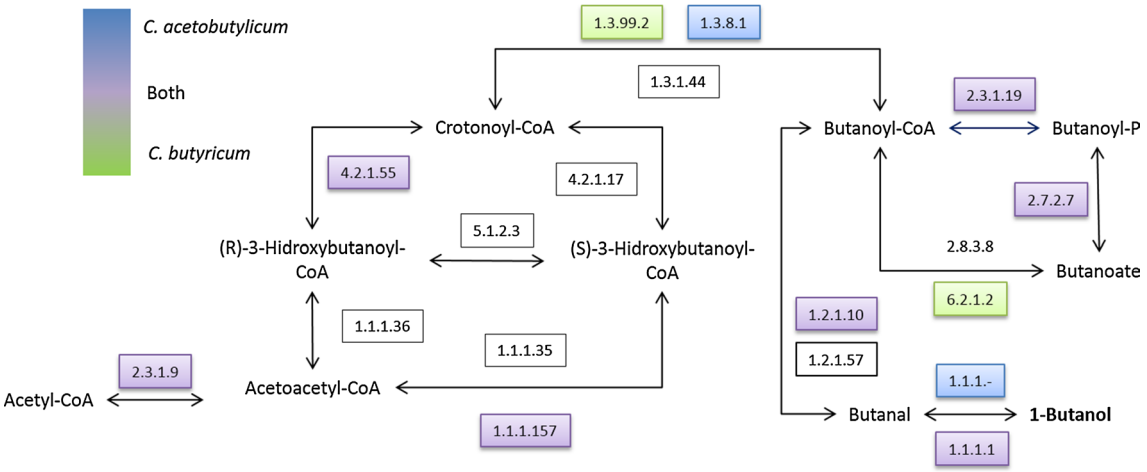

Fig. 4 Comparative metabolic pathways for butanol synthesis present in *Clostridium* genomes. *Arrows* indicate direction of metabolic steps, and *boxes* show EC number of key enzymes for each pathway. *Box colours* (*blue*) indicate the presence of sequences in *C. acetobutylicum* (*blue*), *C. butyricum* (*green*) and present in both species. Key enzyme for butanol production in *C. acetobutylicum* is present in a plasmid (colour figure online)

acetobutylicum ATCC824 genome sequences, and sequences assigned under EC 1.1.1.1 are present in both species.

Furthermore, genomic analysis identified different EC numbers involved in conversion of crotonoil-CoA to butanoyl-CoA for such species. We found sequences assigned under EC 1.3.99.2 in *C. butyricum* genomic sequences. On the other hand, the sequences identified as catalysts for this in *C. acetobutylicum* ATCC824 were assigned under EC 1.3.8.1. It suggests that *C. butyricum* and *C. acetobutilicum* ATCC824 may use different strategies to synthesise butanoyl-CoA.

The pathways highlighted were identified in sequenced genomes and organised based on ECs assignment (Supplementary 3). Six EC groups were found in both species. The sequences assigned under EC 2.3.1.9 are responsible for the conversion of two molecules of acetyl-CoA to acetoacetyl-CoA by releasing coenzyme A by acetyl-CoA acetyltransferase. The EC 1.1.1.157 enzymes promote the conversion of acetoacetyl-CoA produced in (S)-3-hydroxybutanoyl-CoA in an NADPH-dependent reaction, by the enzyme classified as 3-hydroxybutyryl-CoA dehydrogenase. The product of this reaction, or its isomer (we did not found sequences related to one-step isomerisation process in the species analysed), is converted into crotonoil-CoA by CoA-butyryl dehydratase (EC 4.2.1.55) and followed by butanoyl-CoA production by dehydrogenases activity (EC 1.3.99.2 for *C. butyricum* or 1.3.8.1 for *C. acetobutylicum*). Following step synthesises butanal by acetaldehyde dehydrogenase activity (EC 1.2.1.10) (in this case, annotation suggested dual role as acetaldehyde-CoA/alcohol dehydrogenase). Finally, following the metabolic pathway reconstruction, butanal may be used as substrate to produce butanol by NADH-dependent butanol dehydrogenase activity (EC 1.1.1.-).

The alcohol dehydrogenase was identified as a key enzyme for production of butanol in the *C. butyricum* genome. When compared with each other, sequences assigned under EC 1.1.1.- in both genomes showed the lowest similarity in comparison with the other sequences identified as related to butanol biosynthesis (Supplementary 4). The metabolic model obtained from KEGG tool for *C. acetobutylicum* ATCC824 also demonstrated the presence of a key enzyme in the production of butanol, an NADH-dependent alcohol dehydrogenase (EC 1.1.1.1). Thus, this enzyme catalyses the conversion of aldehyde butanal into a primary alcohol, as butanol.

Comparative metabolic pathways for ethanol synthesis in *Clostridium*

The flow chart shows the metabolic pathways present in the *C. butyricum* strains in comparison with *C. acetobutylicum* ATCC824 genome (Fig. 5). The enzymes listed in Supplementary 5, according to their ECs numbers, correspond to the reactions highlighted in flow chart. The metabolic pathways comprising the glycolysis process include the ethanol production.

The acetate produced from the acetyl-CoA can be converted into acetaldehyde by the enzyme aldehyde dehydrogenase (EC 1.2.1.3), which is found in *C. butyricum* INCQS653. Acetaldehyde is then converted to ethanol through EC 1.1.1.1, which includes a key alcohol dehydrogenase, found in the genome using RAST annotation method. In the *C. butyricum* INCQS635 and *C. butyricum* 5521 genomes, EC 1.1.1.1 consists of three different sequences. In the *C. acetobutylicum* ATCC824 genome, the

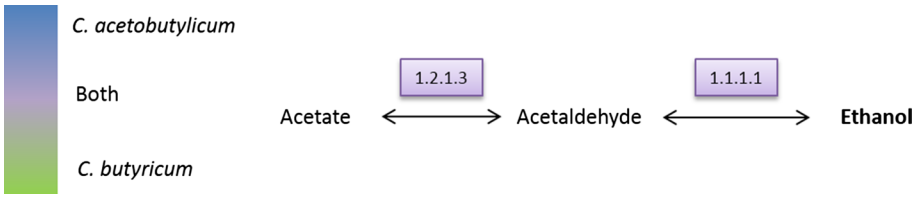

Fig. 5 Comparative metabolic pathways for ethanol synthesis present in *Clostridium* genomes. *Arrows* indicate direction of metabolic steps, and *boxes* show EC number of key enzymes for each pathway. *Box colours* indicate the presence of sequences *in C. acetobutylicum* (*blue*), *C. butyricum* (*green*) and present in both species (*purple*) (colour figure online)

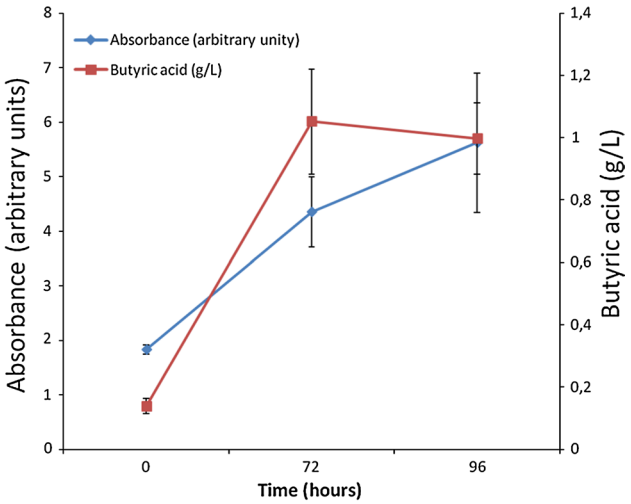

Fig. 6 Solvent production by *Clostridium butyricum* INCQS635 and growth curve. Organic solvents such as butyric and acetic acid were found to be closely related to lag phase

enzymes aldehyde dehydrogenase (EC 1.2.1.3) and alcohol dehydrogenase (EC 1.1.1.1) were identified in the metabolic pathways for ethanol production.

Organic solvent production

Butyric acid production by *C. butyricum* INCQS635 was monitored for 96 h by means of HPLC (Fig. 6). Experiment intervals were determined according to bacterial growth. Absorbance was monitored during the lag phase, when primary metabolites are produced as organic solvents. After 72 h, *C. butyricum* INCQS635 produced more than 1 g L, suggesting maximum production.

Discussion

Polysaccharides potential degradation

According to previous study, *Clostridium acetobutylicum* ATCC824 has more than 90 genes encoding enzymes involved in the degradation of carbohydrate (CAZYmes), including at least 14 distinct families of glycoside hydrolases (Nölling et al. 2001). Our annotation based on conserved domains allowed the identification of a variety of glycoside hydrolases in all strains. In the present study, at least 29 different families of glycoside hydrolases were found in *C. butyricum* and *C. acetobutylicum* strains. Our findings identified a versatile repertoire of CAZYmes in *C. butyricum* INCQS635 and suggest a metabolic potential able to convert a variety of carbohydrates from interesting natural sources which can be applied in industrial processes. The most abundant family of GH indicates the utilisation of new residual natural sources (subexplored so far) which should be explored to provide a pavement for strategies to use new raw materials for production of chemicals products. Here, we found in *C. butyricum* INCQS635 genome GH families involved in degrading of abundant polysaccharides such as galactan, cellulose, starch and chitin.

Based on high abundance of families GH53 and GH16 would be expected potential to convert marine biomass into chemicals by *C. butyricum* INCQS635. Those families comprise enzymes with β-1,4-galactanase activity. Marine biomass represents a promising natural resource due to the low-cost production (based on solar energy, water and CO_2 during the photosynthesis) and high capability to produce biomass and low lignin content in its composition in comparison with terrestrial plants (Delattre et al. 2011; Sanderson 2011). Macroalgae are rich in galactan and cellulose and can be submitted to saccharification process to obtain ethanol (Cho et al. 2012, 2014). In bacteria, GH53 β-1,4-galactanase genes have been found as part of gene clusters devoted to galactan utilisation (i.e. *Bifidobacterium longum*, *Bacillus licheniformis* and *Cellvibrio japonicus*) (Braithwaite et al. 1997; Hinz et al. 2005; Le Nours et al. 2009).

Regarding cellulose utilisation, we found GH1 among the most abundant families. Family GH1 group enzymes are applied for biomass hydrolysis, an important step in the conversion of lignocellulosic feedstocks to fuels and chemicals (Lynd et al. 2008). The β-glicosidase is one of the three main groups of enzymes involved in cellulose

deconstruction. Family GH1 presents encoding genes which plays an important role in the cleavage of the bonds of cellobiose molecule (a reducing sugar composed by a glucose dimer). Cellulases have represented the third largest group of industrial enzymes on market value (Wilson 2009). It is expected to occur even greater expansion of the market focused in cellulases, since it can be used to convert lignocellulosic biomass to sugars in large scale (Kumar et al. 2008). Other family found in *C. butyricum* INCQS635 genome which may be involved in lignocellulosic biomass conversion is GH39. β-xylosidases from family GH39 are able to attack xylan slowly to liberate xylose (Cantarel et al. 2009). Xylose is related to hemicellulose structure in plants connecting cellulose to lignin molecules. The breakdown of hemicellulose is accomplished by the synergistic action of xylosidases and xylanases. Hemicellulose is usually treated as a secondary stream due to the lack of efficient fermentation of hemicellulosic sugars to ethanol (FitzPatrick et al. 2010). There are reports regarding producing intracellular β-xylosidade by *C. thermocellum*, *B. pumilus*, *B. subitilis* and *Acetovibrio cellulolyticus* (Paulova et al. 2015).

Starch is another source of sugar, but usually used, converted to ethanol by enzymes from family GH13, found in high contribution in *C. butyricum* INCQS635 genome. It is the major glycoside hydrolase family, acting on substrates containing α-glucoside linkages. Some of the GH13 members use a multiple attack involving several glycoside bond cleavages (Kramhøft et al. 2005). The GH13 enzymes have a wide range of different preferred substrates and products. Proteins related to GH13 are applied to bioethanol production. Pullulanase and isoamylase belong to the GH13 family with similar sequences, catalytic mechanisms and three-dimensional folds (beta/alpha 8-barrel structure). Starch debranching enzymes can hydrolyse the alpha-1,6-glucosidic bonds at the branch sites of starch and improve raw material utilisation and production efficiency in the starch fermentation. α-Amylase-coding genes from *C. butyricum* can be a future target for the genetic improvement of industrial strains.

Finally, the high contribution of family GH23 in the *C. butyricum* INCQS635 genome suggests the genetic potential to deal with chitin and chitooligosaccharides. The glycoside hydrolases of this family are lytic transglycosylases, and among them there are the lysozymes which are enzymes capable of acting on chitooligosaccharides. Recently, new insights into how enzymes accomplish this have been obtained from studies on the enzymatic conversion of chitin which has been used for second generation bioethanol (Eijsink et al. 2008). It can be obtained from residual material of crustaceans processing by aquaculture industry and fungal biomass production. Despite being the most abundant biopolymers in terrestrial and marine environments, chitinolytic enzyme costs are currently a major limiting factor for its commercial use. *Clostridium* species identified in the mammalian gut are able to use chitin from the diet (Koppová et al. 2012). Chitinases of *Clostridium paraputrificum* J4 were identified in excretomes (Šimůnek et al. 2012). Chitin-degrading *Clostridium* may play an important role related to the processing of food additives such as chitin/chitosan and controlled drug release medicine (Vernazza et al. 2005). Other species, such as *C. thermocellum,* are able to convert chitin by using the multi-modular non-cellulosomal endo-1,3(4)-*β*-glucanase (Dvortsov et al. 2009).

Acetone, butanol and ethanol metabolic pathways

C. butyricum strain DSM10702 showed the presence of 12 coding sequences involved in the production of butanol, while 17 coding sequences were found in *C. butyricum* INCQS635 (Xin et al. 2013). However, no genes have been identified coding for all stages of this pathway. Four basic enzymes are involved in the metabolic pathway responsible for butyril-CoA synthesis from acetyl-CoA (Gheshlaghi et al. 2009). These are the tiolases, 3-hydroxybutyryl-CoA, and crotonases butyril-CoA dehydrogenase. The tiolases are also called acetyl-CoA C-acetyltransferase, and crotonases are a superfamily which includes enoyl-CoA hydratase. We demonstrate the possible pathways used by the species *C. butyricum* and *C. acetobutylicum*. The final step of butanol production is performed by an alcohol dehydrogenase/aldehyde dehydrogenase belonging to EC 1.1.1.1. This enzyme catalyses the butanal conversion to butanol, a primary alcohol. Some singularities encountered in each pathway include the presence of EC 2.8.3.8 in *C. butyricum* INCQS635 and converting crotonoil-CoA into butanoyl-CoA, which is performed by different EC numbers in these species. With the data obtained by analysing the RAST system, it was not possible to identify the gene sequence of the last enzyme in the butanol pathway in the *C. butyricum* INCQS635 genome, although this sequence has been identified in the *C. acetobutylicum* ATCC824 genome (Nölling et al. 2001). A comparison of the metabolic pathways of these species suggests that EC 1.2.1.10, the penultimate enzyme of the pathway, has a dual role in *C. butyricum* INCQS635. This enzyme might exert the function of aldehyde dehydrogenase, alcohol dehydrogenase and then completing the metabolic pathway of butanol production in this organism. Low similarity between the genes identified suggests that different species of *C. butyricum* could perform butanol synthesis using similar but not identical pathways. An attempt to increase the butanol/acetone ratio through overexpression of the alcohol–aldehyde dehydrogenase gene and down-regulation of coenzyme A transferase using antisense RNA showed that the butanol

level was drastically reduced in the resultant mutant, while the ethanol concentration was 23-fold higher than the control (the highest ethanol level was reported in *C. acetobutylicum* (Tummala et al. 2003). Although progress has been made, the use of molecular engineering was not able to construct a hyper-butanol-producing industrial strain. As butanol is highly toxic to the solventogenic clostridia, metabolic engineering of various microorganisms not typically associated with butanol production, but resistant to butanol toxicity, has been suggested. It reinforces the importance of studies such as the present work. Our findings for ethanol pathway analysis showed that *C. butyricum* INCQS635 and *C. acetobutylicum* ATCC824 are able to synthesise both solvents using an identical pathway. On the other hand, *C. butyricum* 5521 showed different groups of possible enzymes involved in ethanol synthesis.

The complete acetone metabolic pathways were not found in this study. Our annotation did not identify sequences related to a key step which converts acetoacetate to acetone. It could be a bias from subsystem annotation since it is well established that *C. butyricum* and *C. acetobutylicum* are able to produce acetone by EBA fermentation (Li et al. 2014). A key step for acetone formation in *C. acetobutylicum* is performed by an acetoacetate decarboxylase encoded by the *amyP* gene. This gene is located on the pSOL1 megaplasmid that carries all of the genes involved in the final steps of solvent formation (Nölling et al. 2001; Sabathé et al. 2002) and which could have been missed in the present study. Butyric acid production identified by HPLC analysis in this work reinforces those findings since the formation of acetone and butanol by *C. acetobutylicum* is induced at the end of the exponential growth phase, during which the organism produces acids in a typical butyric acid fermentation (Jones and Woods 1986; Lee et al. 2008). Butyric acid formation by *C. butyricum* INCQS635 can be considered a relevant finding at the bioenergy point of view because butanol is chemically synthesised from butyric acid, currently. Previously, *C. butyricum* W5 grew on glucose producing hydrogen and carbon dioxide, along with a diversity of acids (lactate, acetate and butyrate) and a small amount of ethanol, among which butyrate was the dominant by-product (Cai et al. 2010). Indeed, this also suggests that the *C. butyricum* INCQS635 strain is able to produce hydrogen. The butyrate formation pathway competes during hydrogen production, consuming more NADH than other pathways and, consequently, reducing the yield of hydrogen (Kumar et al. 2001; Saint-Amans et al. 2001). Inactivation of a beta-hydroxybutyryl-CoA dehydrogenase in *C. butyricum* eliminated the butyrate formation pathway, resulting in a significant increase in ethanol production and a decrease in the H_2 yield compared with the *C. butyricum* W5 wild-type strain. However, a deficient strain showed increased H_2 production, under low partial pressure of H_2, followed by a decrease in ethanol production, indicating that H_2 production by *C. butyricum* may compete for NADH with the ethanol formation pathway (Cai et al. 2011; Hallenbeck 2009). However, successful genetic manipulation of *C. butyricum* has not been widely reported in the literature (González-Pajuelo et al. 2005).

Concluding remarks

The genomic repertoire of *C. butyricum* INCQS635 is a promising resource to underpin future studies towards improved solvent production. Moreover, genomic insights suggest that *C. butyricum* INCQS635 is able to produce chemicals by using the organic matter rich in polysaccharides such as galactan, cellulose, starch and chitin.

Acknowledgments The authors thank CNPq, CAPES and FAPERJ.

References

Aoki-Kinoshita K, Kanehisa M (2007) Gene annotation and pathway mapping in KEGG. In: Bergman N (ed) Comparative genomics. Humana Press, New York City, pp 71–91

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M et al (2008) The RAST Server: rapid annotations using subsystems technology. BMC Genom 9:75

Braithwaite KL, Barna T, Spurway TD, Charnock SJ, Black GW, Hughes N, Lakey JH, Virden R, Hazlewood GP, Henrissat B et al (1997) Evidence that galactanase A from pseudomonas fluorescens subspecies cellulosa Is a retaining family 53 glycosyl hydrolase in which E161 and E270 are the catalytic residues. Biochemistry 36:15489–15500

Cai G, Jin B, Saint C, Monis P (2010) Metabolic flux analysis of hydrogen production network by *Clostridium butyricum* W5: effect of pH and glucose concentrations. Int J Hydrog Energy 35:6681–6690

Cai G, Jin B, Saint C, Monis P (2011) Genetic manipulation of butyrate formation pathways in *Clostridium butyricum*. J Biotechnol 155:269–274

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucl Acids Res 37:D233–D238

Chatzifragkou A, Aggelis G, Komaitis M, Zeng A-P, Papanikolaou S (2011) Impact of anaerobiosis strategy and bioreactor geometry on the biochemical response of *Clostridium butyricum* VPI 1718 during 1,3-propanediol fermentation. Bioresour Technol 102:10625–10632

Cheng C-L, Che P-Y, Chen B-Y, Lee W-J, Lin C-Y, Chang J-S (2012) Biobutanol production from agricultural waste by an acclimated mixed bacterial microflora. Appl Energy 100:3–9

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, Suhai S (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res 14:1147–1159

Cho JK, Kim SY, Lee DH, Kim B, Jung JW (2012) Method for producing biofuel using marine algae-derived galactan. Google Patents

Cho JK, Kim SY, Lee DH, Kim B, Jung JW (2014) Method for producing biofuel using marine algae-derived galactan. Google Patents

Delattre C, Fenoradosoa TA, Michaud P (2011) Galactans: an overview of their most important sourcing and applications as natural polysaccharides. Braz Arch Biol Technol 54:1075–1092

Dürre P (1998) New insights and novel developments in clostridial acetone/butanol/isopropanol fermentation. Appl Microbiol Biotechnol 49:639–648

Dvortsov IA, Lunina NA, Chekanovskaya LA, Schwarz WH, Zverlov VV, Velikodvorskaya GA (2009) Carbohydrate-binding properties of a separately folding protein module from β-1,3-glucanase Lic16A of Clostridium thermocellum. Microbiology 155:2442–2449

Eijsink VGH, Vaaje-Kolstad G, Vårum KM, Horn SJ (2008) Towards new enzymes for biofuels: lessons from chitinase research. Trends Biotechnol 26:228–235

FitzPatrick M, Champagne P, Cunningham MF, Whitney RA (2010) A biorefinery processing perspective: treatment of lignocellulosic materials for the production of value-added products. Bioresour Technol 101:8915–8922

Gheshlaghi R, Scharer JM, Moo-Young M, Chou CP (2009) Metabolic pathways of clostridia for producing butanol. Biotechnol Adv 27:764–781

González-Pajuelo M, Meynial-Salles I, Mendes F, Andrade JC, Vasconcelos I, Soucaille P (2005) Metabolic engineering of *Clostridium acetobutylicum* for the industrial production of 1,3-propanediol from glycerol. Metab Eng 7:329–336

Hallenbeck PC (2009) Fermentative hydrogen production: principles, progress, and prognosis. Int J Hydrogen Energy 34:7379–7389

Hinz SWA, Pastink MI, van den Broek LAM, Vincken J-P, Voragen AGJ (2005) Bifidobacterium longum Endogalactanase Liberates Galactotriose from Type I Galactans. Appl Environ Microbiol 71:5501–5510

Hou X, Peng W, Xiong L, Huang C, Chen X, Chen X, Zhang W (2013) Engineering *Clostridium acetobutylicum* for alcohol production. J Biotechnol 166:25–33

Jang Y-S, Malaviya A, Cho C, Lee J, Lee SY (2012) Butanol production from renewable biomass by clostridia. Bioresour Technol 123:653–663

Jones DT, Woods DR (1986) Acetone-butanol fermentation revisited. Microbiol Rev 50:484–524

Koppová I, Bureš M, Šimůnek J (2012) Intestinal bacterial population of healthy rats during the administration of chitosan and chitooligosaccharides. Folia Microbiol 57:295–299

Kramhøft B, Bak-Jensen KS, Mori H, Juge N, Nøhr J, Svensson B (2005) Involvement of individual subsites and secondary substrate binding sites in multiple attack on amylose by barley α-amylase. Biochemistry 44:1824–1832

Kumar N, Ghosh A, Das D (2001) Redirection of biochemical pathways for the enhancement of H2 production by *Enterobacter cloacae*. Biotechnol Lett 23:537–541

Kumar R, Singh S, Singh O (2008) Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. J Ind Microbiol Biotechnol 35:377–391

Le Nours J, De Maria L, Welner D, Jørgensen CT, Christensen LLH, Borchert TV, Larsen S, Lo Leggio L (2009) Investigating the binding of β-1,4-galactan to Bacillus licheniformis β-1,4-galactanase by crystallography and computational modeling. Proteins Struct Funct Bioinform 75:977–989

Lee SY, Park JH, Jang SH, Nielsen LK, Kim J, Jung KS (2008) Fermentative butanol production by clostridia. Biotechnol Bioeng 101:209–228

Lee J, Jang Y-S, Choi SJ, Im JA, Song H, Cho JH, Seung DY, Papoutsakis ET, Bennett GN, Lee SY (2012) Metabolic Engineering of *Clostridium acetobutylicum* ATCC 824 for isopropanol-butanol-ethanol fermentation. Appl Environ Microbiol 78:1416–1423

Li J, Baral N, Jha A (2014) Acetone–butanol–ethanol fermentation of corn stover by Clostridium species: present status and future perspectives. World J Microbiol Biotechnol 30:1145–1157

Lütke-Eversloh T, Bahl H (2011) Metabolic engineering of *Clostridium acetobutylicum*: recent advances to improve butanol production. Curr Opin Biotechnol 22:634–647

Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS (2002) Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev 66:506–577

Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, Hamilton R, Himmel M, Keller M, McMillan JD, Sheehan J et al (2008) How biotech can transform biofuels. Nat Biotech 26:169–172

Nölling J, Breton G, Omelchenko MV, Makarova KS, Zeng Q, Gibson R, Lee HM, Dubois J, Qiu D, Hitti J et al (2001) Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. J Bacteriol 183:4823–4838

Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M et al (2014) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42:D206–D214

Pagani I, Liolios K, Jansson J, Chen I-MA, Smirnova T, Nosrat B, Markowitz VM, Kyrpides NC (2012) The Genomes OnLine Database (GOLD) v. 4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res 40:D571–D579

Park BH, Karpinets TV, Syed MH, Leuze MR, Uberbacher EC (2010) CAZymes Analysis Toolkit (CAT): web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. Glycobiology 20:1574–1584

Paulova L, Patakova P, Branska B, Rychtera M, Melzoch K (2015) Lignocellulosic ethanol: technology design and its impact on process efficiency. Biotechnol Adv 33(6 Pt 2):1091–1107. doi:10.1016/j.biotechadv.2014.12.002

Pitcher DG, Saunders NA, Owen RJ (1989) Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. Lett Appl Microbiol 8:151–156

Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13:341

Raynaud C, Sarçabal P, Meynial-Salles I, Croux C, Soucaille P (2003) Molecular characterization of the 1,3-propanediol (1,3-PD) operon of *Clostridium butyricum*. Proc Natl Acad Sci 100:5010–5015

Raynaud C, Lee J, Sarçabal P, Croux C, Meynial-Salles I, Soucaille P (2011) Molecular characterization of the glycerol-oxidative pathway of *Clostridium butyricum* VPI 1718. J Bacteriol 193:3127–3134

Sabathé F, Croux C, Cornillot E, Soucaille P (2002) amyP, a reporter gene to study strain degeneration in *Clostridium acetobutylicum* ATCC 824. FEMS Microbiol Lett 210:93–98

Saint-Amans S, Girbal L, Andrade J, Ahrens K, Soucaille P (2001) Regulation of carbon and electron Flow in *Clostridium butyricum* VPI 3266 grown on glucose-glycerol mixtures. J Bacteriol 183:1748–1754

Sanderson K (2011) Lignocellulose: a chewy problem. Nature 474:S12–S14

Šimůnek J, Koppová I, Tiščenko G, Dohnálek J, Dušková J (2012) Excretome of the chitinolytic bacterium *Clostridium paraputrificum* J4. Folia Microbiol 57:335–339

Tummala SB, Junne SG, Papoutsakis ET (2003) Antisense RNA downregulation of coenzyme A transferase combined with alcohol-aldehyde dehydrogenase overexpression leads to predominantly alcohologenic *Clostridium acetobutylicum* fermentations. J Bacteriol 185:3644–3653

Vernazza CL, Gibson GR, Rastall RA (2005) In vitro fermentation of chitosan derivatives by mixed cultures of human faecal bacteria. Carbohydr Polym 60:539–545

Wilson DB (2009) Cellulases and biofuels. Curr Opin Biotechnol 20:295–299

Xin B, Tao F, Wang Y, Gao C, Ma C, Xu P (2013) Genome sequence of *Clostridium butyricum* strain DSM 10702, a promising producer of biofuels and biochemicals. Genome Announc 1(4):e00563–13. doi:10.1128/genomeA.00563-13

Zeng AP, Ross A, Biebl H, Tag C, Günzel B, Deckwer WD (1994) Multiple product inhibition and growth modeling of *Clostridium butyricum* and klebsiella pneumoniae in glycerol fermentation. Biotechnol Bioeng 44:902–911