Manual curation and reannotation of the genomes of *Clostridium difficile* 630Δerm and *C. difficile* 630

Henning Dannheim,† Thomas Riedel,‡ Meina Neumann-Schaal,† Boyke Bunk,‡,3 Isabel Schober,‡ Cathrin Spröer,‡,3 Cynthia Maria Chibani,4 Sabine Gronow,2,3 Heiko Liesegang,4 Jörg Overmann2,3 and Dietmar Schomburg1,*

Abstract

**Purpose.** We resequenced the genome of *Clostridium difficile* 630Δerm (DSM 28645), a model strain commonly used for the generation of insertion mutants.

**Methodology.** The genome sequence was obtained by a combination of single-molecule real-time and Illumina sequencing technology.

**Results.** Detailed manual curation and comparison to the previously published genomic sequence revealed sequence differences including inverted regions and the presence of plasmid pCD630. Manual curation of our previously deposited genome sequence of the parental strain 630 (DSM 27543) led to an improved genome sequence. In addition, the sequence of the transposon Tn5397 was completely identified. We manually revised the current manual annotation of the initial sequence of strain 630 and modified either gene names, gene product names or assigned EC numbers of 57% of genes. The number of hypothetical and conserved hypothetical proteins was reduced by 152. This annotation was used as a template to annotate the most recent genome sequences of the strains 630Δerm and 630.

**Conclusion.** Based on the genomic analysis, several new metabolic features of *C. difficile* are proposed and could be supported by literature and subsequent experiments.

INTRODUCTION

*Clostridium difficile* is a facultative virulent human pathogen causing antibiotic-associated diseases. It is an anaerobic, rod-shaped, Gram-positive and spore-forming bacterium and belongs to the order *Clostridiales*. The widely used model strain 630 was isolated from a hospital patient with severe pseudomembranous colitis in Zürich (Switzerland) [1]. About 10 years ago, an erythromycin-sensitive mutant (630Δerm) was found, which lacked one of two erythromycin resistance genes [2]. This strain is now used for insertion mutant generation in *C. difficile* using the CloStron system [3]. The genome of strain *C. difficile* 630 was first sequenced in 2006 [4] and reannotated twice [5, 6]. Genome resequencing of the strain 630 (DSM 27543) with the help of long-read sequencing revealed genomic differences such as the partial excision of the transposon Tn5397, the presence of one additional rRNA gene cluster as well as a putative loss of the plasmid pCD630 [7]. Genome sequencing of strain 630Δerm verified the lack of one erythromycin resistance gene but also revealed a translocation of the conjugative transposon CTn5, several single nucleotide polymorphisms, insertions and deletions compared to the wild-type strain 630 [8]. A recent study by Collery *et al.* [9] showed that strain 630Δerm is closer to the parental strain than the similar strain 630E, supporting the importance of strain 630Δerm.
Genome annotations of model organisms have to be regularly updated to provide an accurate and relevant basis for a wide range of biochemical and bioinformatical methods. In this study, we resequenced the genome of *C. difficile* 630Δerm (DSM 28645) using a combination of single-molecule real-time (SMRT) and Illumina sequencing technology. In addition, we manually curated and reannotated the genome sequences of both strains 630Δerm and 630. For this purpose the current manual annotation of strain 630 [6] was revised and used as a template. The resulting annotations include substantial updates of the putative metabolic capabilities based on characterized enzymes from other organisms as well as completely new functions which are supported by laboratory results and literature data.

**METHODS**

**Strains, media and growth conditions**

All studies were carried out with *C. difficile* 630Δerm (DSM 28645) [2]. For genome sequencing the organism was cultivated anaerobically in Wilkins-Chalgren Anaerobe Broth (Oxoid) at 37 °C. Genomic DNA was extracted as described previously [7, 10]. For the other experiments the media and growth conditions were used as already described [11]. Due to an impaired growth after replacing proline completely by hydroxyproline in minimal defined medium (MDM), proline levels were reduced to 4.3 mM and residual proline was replaced by 13 mM hydroxyproline. As indicated, 12.9 mM histidine was added to the MDM, which was supplemented with 4.5 mM threonine to achieve higher cell densities.

**Sequencing and genome sequence curation of strain *C. difficile* 630Δerm**

Genome sequencing was carried out on the PacBio RSII (Pacific Biosciences) using two SMRT-Cells and P5 chemistry. Genome assembly was performed with the ‘RS_HGAP_Assembly.3’ protocol included in the SMRT Portal version 2.3.0, utilizing 50,517 postfiltered reads with an average read length of 10,879 bp. One complete chromosomal contig was obtained and trimmed, circularized and adjusted to dnaA (chromosomal replication initiator protein, CDIF630erm_00001) as the first gene. For confirmatory purposes, two additional SMRT-Cells were sequenced using P5 (20,715 postfiltered reads with an average read length of 11,233 bp) and the newest P6 chemistry with even longer read lengths (31,299 postfiltered reads with an average read length of 13,231 bp). A final genome quality of QV60 was determined during resequencing using the RS_BridgeMapper.1 protocol in SMRT Portal. In addition, genome sequencing was carried out on a Genome Analyzer GAIIx (Illumina) in a 112 bp paired-end single-indexed run, resulting in 2.5 million paired-end reads. Quality improvement was achieved with the Burrows-Wheeler Aligner [12] mapping the Illumina reads onto the chromosome obtained by PacBio sequencing (hybrid assembly) followed by subsequent automatic detection of sequencing errors by VarScan (http://varscan.sourceforge.net) and GATK Consensus calling (https://software.broadinstitute.org/gatk/).

The resulting consensus sequence was further manually approved by visual comparison of the read alignments on PacBio and Illumina nucleotide resolution by using the Integrative Genomics Viewer [13, 14]. In case of nucleotide variants the majority fraction was called for the final consensus sequence. Single nucleotide variants, insertions and deletions were identified by using the genome comparison tool Mauve [15–17] and the previously published genome of strain 630Δerm [8] as reference. For verification of putatively inverted genomic regions, a PCR amplification and sequencing (Method S1, available in the online Supplementary Material) with specific primers (Table S1) was performed.

**Genome sequence curation of strain 630 and recovery of Tn5397**

The genome sequence of strain 630 (DSM 27543, accession no. CP010905.1; [7]) was manually curated according to the protocol already described for *C. difficile* 630Δerm. For final sequence verification the sequence was also manually compared to reference sequences of strain 630 (accession no. AM180355.1; [4]), 630Δerm (accession no. LN614756.1; [8]) and 630Δerm (DSM 28645).

The Tn5397-encoding sequence of strain 630 (DSM 27543) [7] was recovered from a small number of reads compared to the chromosomal coverage. For the assembly, the available reads were mapped on the Tn5397-encoding sequence (accession no. AF333235.1; [18]). The final genome quality of QV60 was obtained as described for strain 630Δerm. The putative excision of Tn5397 in strain 630 (DSM 27543) was verified by colony-PCR (Fig. S1 and Method S2) using specific primers (Table S1).

**Verification of the presence of plasmid pCD630**

For confirmation of the putative loss or presence of pCD630 in the strains 630 (DSM 27543) and 630Δerm (DSM 28645), the available PacBio sequencing reads of both strains were mapped separately on pCD630 (accession no. AM180356.1; [4]) as reference using the RS_BridgeMapper.1 protocol in SMRT Portal. The Illumina reads were mapped onto the obtained pCD630 plasmid sequence using the Burrows-Wheeler Aligner [12].

**Reannotation**

To generate a state-of-the-art annotation of the resequenced *C. difficile* strain 630Δerm, the latest manual annotation [4–6] of the wild-type strain *C. difficile* 630 (accession no. AM180355.1) was used. All coding sequences (CDSs) were analysed with BLASTp [19], BrEPS [20] and InterProScan [21] were used for pattern and Hidden Markov model based function predictions. Product names, gene names or EC number information were updated whenever necessary based on new findings or nomenclature changes. Significant alterations of the annotation were justified in the GenBank file and documented in the corresponding inference qualifiers. The resulting annotation of the wild-type was transferred to strain 630Δerm (DSM 28645) using Prokka 1.8 [22] with the genome of strain 630 used as genus database. The programmed ribosomal frameshift of *larC* and
translation exceptions for selenocysteines were included manually. Selenocysteine incorporation was based on previous annotations of strain 630 [5, 6] and automatic predictions [22]. Additional CDSs were predicted using the same methods as mentioned before. Notes and inferences were transferred based on manually corrected orthologue predictions. Within the annotation transfer orthologues genes in strain 630Δerm received the same locus tag number as in strain 630 (accession no. CP010905.1; [7]), e.g., both dnaA genes are numbered with 00 001. In addition, the annotation of the wild-type (accession no. CP010905.1) was updated based on the annotation of strain 630Δerm using the locus tag numbers.

**Analysis of intracellular compounds**

Cells were grown to the mid-exponential growth phase and harvested by centrifugation (10 min, 8000 g) under anoxic conditions using gas-tight polypropylene tubes (TPPs). The supernatant was removed and the cells were immediately harvested by centrifugation (10 min, 8000 g) prior to drying in a vacuum concentrator (SpeedVac, Labconco) prior to GC/MS measurements and data processing were done as described previously [11]. Significant changes in metabolite levels were calculated by non-parametric Kruskal–Wallis test [24]; P-value <0.01 using Benjamini–Yekutieli correction [25] to control the false discovery rate.

CoA derivatives were extracted and analysed as described earlier [26]. Statistically significant altered metabolite abundances were determined with the non-parametric Kruskal–Wallis test as described above.

**Enzymatic assays**

The enzymatic 4-hydroxyproline assays including crude extracts were performed as follows: Cells were grown in presence of 4-hydroxyproline (Sigma-Aldrich), harvested in the mid-exponential growth phase and resuspended in 100 mM Tris, pH 7.2, supplemented with 1 mg ml⁻¹ lysozyme and incubated on ice for 60 min. Cells were sonicated twice with 2 min pulse and 2 min cooling. The enzymatic assay was performed with 60 µl crude extract in a total volume of 600 µl in 12 mM phosphate buffer, pH 7.0, supplemented with 60 µM (NH₄)₂SO₄, 35.4 µM CaCl₂, 19.8 µM MgCl₂, 10.1 µM MnCl₂, 0.8 µM CoCl₂ and 4.2 mM NADPH or NADH. The reaction was started by the addition of 7.6 mM hydroxyproline, for 15 min at 37 °C and finally stopped by the addition of 300 µl chloroform. Samples were vigorously mixed for 5 min and centrifuged at 14 000 g and 4 °C for 10 min. Fifty microlitres of the polar phase was dried in a vacuum concentrator (SpeedVac, Labconco) prior to GC/MS analysis.

**RESULTS AND DISCUSSION**

**Comparison of the genome sequences of C. difficile 630Δerm**

In comparison (Fig. S2) to the recently published genome sequence of laboratory strain C. difficile 630Δerm [8] we detected three chromosomal single nucleotide variants [30942 (A→C; CDIF630erm_00035), 3548102 (A→T; CDIF630erm_03372) and 3905079 (C→A; CDIF630erm_03681)]. These differences affect a 16S rRNA, a glucose-like phosphotransferase system transporter of unknown function (K→I) and magnesium-transporting ATPase (G→V). We cannot predict any modification of the biological function caused by these mutations. Additionally, we identified three chromosomal non-coding inversions (Fig. 1, Table 1) and validated them by PCR and sequencing. Strain 630 [4] contains the third inversion in the same orientation and with the same sequence orientation (Table 1). Compared to the published genome sequence of strain 630Δerm [8], the third inverted region showed several single nucleotide variants (Fig. 1), so the orientation in the latter genome sequence remains unclear. Thus, we were able to confirm inversions already reported for other C. difficile strains including strain 630 [27, 28] to occur frequently in closely related strains.

Interestingly, screening of all available PacBio as well as Illumina reads revealed the presence of the plasmid pCD630 in strain 630Δerm (accession no. CP016319.1). Whereas the plasmid was not described in the recently published complete genome sequence of 630Δerm [8], pCD630 was detected in the first sequencing study of strain 630 [4], but was not found after resequencing [7].

**Manual curation of the genome sequence of C. difficile 630**

We manually curated seven single nucleotide positions and one repetitive genomic region in the published genome sequence of C. difficile 630 (Table 2) [7]. Furthermore, the Tn5397-encoding sequence was missing in the majority of the chromosomal sequence of strain 630 [7] and was completely recovered from very few reads mapping on the Tn5397-encoding sequence (accession no. AF333235.1). The excision was confirmed by PCR and sequencing (Fig. S1). The comparison of the transposon sequences (Fig. S3) led to the discovery of 33 single nucleotide variants, 1 insertion and 3 deletions [18] (Table S2). By detailed comparison, we observed that all other available Tn5397-sequences from C. difficile 630 and 630Δerm are identical to our sequence (Fig. S3).

**Reannotation**

The published manual genome annotation of C. difficile 630 [6] was completely manually revised taking into account the current state of knowledge. This led to a modification of the predicted function for 601 gene products. In addition, the information for 1577 CDSs was updated reflecting current use of nomenclature or by adding additional information...
items like EC values or others (see Fig. 2 and Table S3). The number of (conserved) hypothetical proteins was reduced by 152 to 314. This updated annotation was used as a template for the annotation of our sequences from \textit{C. difficile} 630 and 630\textit{Derm}. The annotation of strain 630 \textit{Derm} contains 3782 CDS including 289 hypothetical proteins. 115 CDS of the initial strain 630 [5] are not existent in our annotation of 630\textit{Derm} mainly because a different gene caller software was used, but none of them (except the expected \textit{ermB1}) had a known specific function. Compared to the 2010 annotation [5], 90 additional genes of the initial strain 630 sequence [4] were detected including \textit{larC} (\textit{Ni}^{2+}-carrier, CDIF630erm_00317), \textit{rbfA} (ribosome-binding factor A, CDIF630erm_01466) and a gene for a putatively selenocysteine-containing glycine/sarcosine/betaine reductase complex protein (CDIF630erm_01932). The revised annotation of strain 630 presented in this study contains 3771 CDS, 3762 are shared by both strains (630 and 630\textit{Derm}). The main differences are due to the deletion of the \textit{erm} region in the mutant [2] and the missing Tn5397-encoding sequence in the strain 630.

The reannotation led to a number of insights into the putative metabolic features of the pathogen \textit{C. difficile}. It includes additional experimental results on proteins involved in pathways like pentose metabolism [29] or posttranslational modification of flagellin [30]. The new annotation also includes gene functions previously only characterized in other organisms like the Re-citrate synthase (EC 2.3.3.3) [31] or the aspartate decarboxylase (EC 4.1.1.12) [32].

**Table 1. Inversions in \textit{C. difficile} strains 630, 630\textit{Derm} and selected reference strains**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Sequence</th>
<th>Inversion 1</th>
<th>Inversion 2</th>
<th>Inversion 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>630</td>
<td>AM180355.1 [4]</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>630\textit{Derm}</td>
<td>CP010905.2 ([7] and this study)</td>
<td>Identical</td>
<td>Identical</td>
<td>Inverted</td>
</tr>
<tr>
<td>ATCC 9689</td>
<td>LN614756.1 [8]</td>
<td>Identical</td>
<td>Identical</td>
<td>Unclear</td>
</tr>
<tr>
<td>Z31</td>
<td>CP013196.1 [54]</td>
<td>Inverted</td>
<td>Inverted</td>
<td>Inverted</td>
</tr>
<tr>
<td>CD196</td>
<td>NC_013315.1 [28]</td>
<td>Inverted</td>
<td>Inverted</td>
<td>Inverted</td>
</tr>
</tbody>
</table>

**Fig. 1.** Alignment of all three regions of possible inversions in strain 630\textit{Derm}. Whereas the sequence of this study is shown in each line 1 and inverted in line 2, the reference sequence of strain 630\textit{Derm} [8] is each shown in line 3. The exact positions of each inversion refer to the region of the coloured bases in blue. Predicted inversion sites are marked by red boxes. Blue shading indicates positions of conserved bases.
Bile acid metabolism

We identified a gene (CDIF630erm_00131) with 57 % protein sequence identity to the previously characterized NADP-dependent 7α-hydroxysteroid dehydrogenase of Clostridium sordellii [33] whose metabolic products were found in bile acid supplemented culture of C. difficile ATCC 9689 [34]. In addition, we annotated a putative bile acid-CoA:aminino N-acyltransferase (CDIF630erm_02171) which probably is responsible for the reported unusual splitting of taurocholic acid but not glycocholic acid [34]. Cholate, glycocholate and taurocholate stimulate germination of C. difficile spores and do not inhibit growth, whereas deoxycholate also leads to sporulation but inhibits growth [35, 36]. Thus, the 7α-dehydroxylation of bile acids by Clostridium scindens [37] was suggested as the reason for its inhibiting effect on C. difficile growth [38, 39]. The oxidation of the 7α-hydroxy-group might therefore prevent 7α-dehydroxylation and thus inhibit the growth of C. difficile.

Lactate metabolism

Previous annotations of C. difficile included an L-lactate dehydrogenase (EC 1.1.1.27) and a D-lactate dehydrogenase (EC 1.1.1.28). However, the latter enzyme has been shown to be a (R)-2-hydroxyisocaproate dehydrogenase (EC 1.1.1.345) [40]. In addition, we detected homologues of three subunits of the D-lactate dehydrogenase of Acetobacterium woodii [41] which catalyses the parallel oxidation of ferredoxin and lactate under reduction of two molecules NAD⁺, enabling the degradation of D-lactate under physiological conditions. Direct upstream, the genome of C. difficile harbours a gene for a lactate racemase (EC 5.1.2.1) which was previously characterized from Lactobacillus plantarum [42]. We could also find genes for the lactate racemase accessory proteins LarB, LarC and LarE. In previous annotations of C. difficile 630 the gene larC encoding for a Ni²⁺ carrier was annotated as a hypothetical protein disrupted by a frameshift mutation and therefore as a pseudogene [4–6, 8] or two hypothetical proteins [7]. Interestingly, larC of L. plantarum also includes a frameshift, but LarC could be shown to be the functional form of the protein. In this context the authors suspected a programmed ribosomal (−1) frameshift [42]. In fact, an insertion of an additional adenine in the stretch of 10 adenines in the larC1 genes in both organisms would lead to two similar and full-length LarC proteins of 37 % sequence identity. We therefore assume larC to be a fully functional gene with a programmed ribosomal frame shift.

Detoxification of reactive oxygen species

C. difficile harbours highly homologous genes for all three terminal rubredoxin-dependent catalytic subunits flavodiiron protein A (EC 1.6.3.-), desulfoferdoxin (EC 1.15.1.2) and reverse rubrerythrin (EC 1.11.1.-) of the reactive oxygen species detoxification complex characterized from Clostridium acetobutylicum [43]. No highly homologous gene could be found for the NADH-rubredoxin oxidoreductase (EC 1.18.1.1). The only found rubredoxin domain (InterPro: IPR024935) is joined with a domain, which is homologue to the 2-oxoglutarate-aminating [44] flavin mononucleotide-binding domain of glutamate synthase. The protein therefore might have rubredoxin-dependent glutamate dehydrogenase activity. The corresponding gene is part of an operon (CDIF630erm_00944-00947) also consisting of a putative rubrerythrin, the peroxide-responsive

Table 2. Alterations on the genome sequence of C. difficile 630

<table>
<thead>
<tr>
<th>Position</th>
<th>Sequence</th>
<th>Type</th>
<th>CP010905.1</th>
<th>CP010905.2</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>740284</td>
<td>TAAGAGCAAGAGCAAGAGC→T</td>
<td>Deletion</td>
<td>CDIF630_00744</td>
<td>CDIF630_00744</td>
<td>Hypothetical protein</td>
</tr>
<tr>
<td>941699</td>
<td>CA→C</td>
<td>Deletion</td>
<td>CDIF630_00902</td>
<td>CDIF630_00902</td>
<td>Putative N-acetylmuramoyl-l-alanine amidase</td>
</tr>
<tr>
<td>982428</td>
<td>CA→C</td>
<td>Deletion</td>
<td>CDIF630_00940</td>
<td>CDIF630_00940</td>
<td>ATP-binding cassette-type transport system, multidrug-family ATP-binding protein</td>
</tr>
<tr>
<td>1108604</td>
<td>A→AG</td>
<td>Insertion</td>
<td>CDIF630_01086</td>
<td>CDIF630_01086</td>
<td>Putative phage protein</td>
</tr>
<tr>
<td>1110701</td>
<td>CA→C</td>
<td>Deletion</td>
<td>CDIF630_01088</td>
<td>CDIF630_01088</td>
<td>Putative phage lipoprotein</td>
</tr>
<tr>
<td>1112738</td>
<td>CA→C</td>
<td>Deletion</td>
<td>CDIF630_01092</td>
<td>CDIF630_01092</td>
<td>Putative phage tail tape measure protein</td>
</tr>
<tr>
<td>1113838</td>
<td>CA→C</td>
<td>Deletion</td>
<td>CDIF630_01467</td>
<td>CDIF630_01467</td>
<td>Putative RNA-binding protein</td>
</tr>
<tr>
<td>150695</td>
<td>CA→C</td>
<td>Deletion</td>
<td>CDIF630_02739</td>
<td>CDIF630_02739</td>
<td>Two-component sensor histidine kinase, sporulation-associated spo0A</td>
</tr>
<tr>
<td>2861067</td>
<td>GT→G</td>
<td>Deletion</td>
<td>CDIF630_02739</td>
<td>CDIF630_02739</td>
<td>Two-component sensor histidine kinase, sporulation-associated spo0A</td>
</tr>
</tbody>
</table>

Fig. 2. Venn diagram showing the differences of the annotation presented here the recent annotation [6] of the initial sequence [4] of C. difficile 630. The annotation of 2179 genes was changed. Circle with continuous line product name different circle with dotted line gene name different circle with dashed line EC numbers different.
repressor (PerR) and the desulfoferrodoxin. Thus the predicted rubredoxin-dependent glutamate dehydrogenase could supply the oxygen detoxification enzymes with the necessary electrons for reduction. The necessary glutamate can either be supplied by the transamination of amino acids (EC 2.6.1.-) during their degradation or by the common glutamate dehydrogenase (EC 1.4.1.2). This might explain the previously detected but inexplicable dependency of C. difficile to its glutamate dehydrogenase (EC 1.4.1.2) for \( \text{H}_2\text{O}_2 \) resistance [45].

**Degradation of 4-hydroxyproline**

We detected a gene for pyrroline-5-carboxylate reductase (EC 4.2.1.-, CDIF630erm_03581) clustered together with genes for a glyoxyl radical enzyme activase (CDIF630erm_03583) and a protein sharing 38 % identity with the already characterized [46, 47] B12-independent glycerol dehydratase from Clostridium butyricum (ATD26_17630, CDIF630erm_03582). While the main catalytic residues H164, C433, E435 and G763 (C. butyricum numbering) of the dehydratase are conserved in the C. difficile enzyme, the binding pocket residues H281, S282, Y339, D447 and Y640 are not. This indicates a similar reaction mechanism but a different substrate. A dehydration of 4-hydroxyproline would result in (S)-1-pyrroline-5-carboxylate, convertible to L-proline by the pyrroline-5-carboxylate reductase. In fact, a similar mechanism for the dehydration of 4-hydroxyproline compared with the glycerol dehydration seems possible including an imine–enamine tautomerism instead of a keto–enol tautomerism. The corresponding and so far unknown pathway including the proposed 4-hydroxyproline dehydratase is shown in Fig. 3(a). We could find similar genomic structure including a protein of 66 % sequence identity to the proposed 4-hydroxyproline dehydratase in the genome of Clostridium sporogenes, which is known to use 4-hydroxyproline as hydrogen acceptor [48]. In Oscillibacter valericigenes an operon consisting of genes encoding for a symporter, an L-glutamate 5-semialdehyde dehydrogenase (EC 1.2.1.88), a glyoxyl radical enzyme activase and a CDIF630erm_03582-homologue (68 % protein sequence identity) is present. This could represent a complete 4-hydroxyproline degradation pathway to glutamate via (S)-1-pyrroline-5-carboxylate, supporting the proposed function of CDIF630erm_03582. Cell lysates of C. difficile cultures in 4-hydroxyproline supplemented MDM were incubated with 4-hydroxyproline and NADPH or NADH. The assay showed a strictly NADPH-dependent increase of proline and decrease of 4-hydroxyproline (Fig. 3b). This result is in line with the hypothesized pathway above. The presence of the intermediate (S)-1-pyrroline-5-carboxylate could not be verified by GC/MS analysis due to the instability of the molecule [49, 50].

---

**Fig. 3.** Degradation of 4-hydroxyproline to proline. (a) Proposed degradation pathway of 4-hydroxyproline. (b) Cell extract assay of 4-hydroxyproline. White, without cell extract; grey, with cell extract. Peak areas were normalized on ribitol as internal standard and the relative proportion of the specific quantification ion. The detected proline of the blank is a contamination of the 4-hydroxyproline as purchased 4-hydroxyproline was shown to contain residual proline amounts.
Degradation of histidine

No genes for known histidine degradation pathways could be found in the genome of *C. difficile* 630 and 630Δerm. In contrast, cultures of *C. difficile* 630Δerm in histidine and threonine supplemented MDM showed detectable amounts of intracellular 4-imidazolacacetate while the overall metabolism remained largely unchanged in comparison to the reference cultures (MDM plus threonine). Moreover, HPLC/MS analysis of CoA-esters showed a peak for a CoA derivative with the corresponding exact mass to imidazolyl-acetyl-CoA. The results of the GC/MS and HPLC/MS analyses are listed in Table S4. Thus, experimental data indicate a role of histidine as a substrate for an oxidative Stickland pathway. The *C. difficile* genome contains genes for aromatic-amino-acid aminotransferases (EC 2.6.1.57) and the indolepyruvate ferredoxin oxidoreductase (EC 1.2.7.8). Indolepyruvate ferredoxin oxidoreductases of other organisms have been shown to be quite promiscuous but imidazole pyruvate has never been tested [52, 53]. Consequently, most likely histidine is fermented with the same set of enzymes as the aromatic amino acids. Histidine has been already identified as an oxidative Stickland substrate in *C. sporogenes* but these findings were based on the weak reduction of an indicator or the release of the amino group of co-incubated glycine [48]. Interestingly, neither genes of the histidine degradation pathway to glutamate, nor the indolepyruvate ferredoxin oxidoreductase (EC 1.2.7.8) could be found in *C. sporogenes*. To our knowledge, we present here the first case of histidine degradation using the classical oxidative Stickland pathway.

CONCLUSION

In this paper we present the resequenced and manually curated genome sequence of *C. difficile* strain 630Δerm (DSM 28645) and the manually curated genome sequence of *C. difficile* strain 630 (DSM 27543) and the sequence of transposon Tn5397. In addition, an inversion of three short genomic regions could be verified in *C. difficile* 630Δerm. The reannotation of both strains significantly expands the knowledge about the metabolic potential of *C. difficile*. Apart from findings included in the annotation based on literature, completely new gene functions were proposed. This includes two new Stickland-like amino acid degradation pathways for the degradation of 4-hydroxyproline and histidine which were experimentally supported. In addition, genes for bile acid metabolism and oxygen stress resistance were annotated. All these findings can now be used as a starting point for in-depth function analysis (e.g. by mutant studies and/or protein characterization) and systems biology approach.

Funding information
This work (H.D., T.R., M.N.S.) was funded by the Federal State of Lower Saxony, Niedersächsisches Vorab (WWZN2889/3215).

Acknowledgements
We thank Sabine Kaltenhäuser, Carolin Pilke, Simone Severitt and Nicole Heyer for technical assistance.

Conflicts of interest
The authors declare that there are no conflicts of interest.

Ethical statement
No ethical considerations apply.

References

2. Hussain HA, Roberts AP, Mullany P. Generation of an erythromycin-sensitive derivative of *Clostridium difficile* strain 630 (630Δerm) and demonstration that the conjugative transposon Tn916ΔE enters the genome of this strain at multiple sites. *J Med Microbiol* 2005;54:137–141.
8. Van Eijk E, Anvar SY, Browne HP, Leung WY, Frank J et al. Complete genome sequence of the *Clostridium difficile* laboratory strain 630Δerm reveals differences from strain 630, including translocation of the mobile element Cm1. *BMC Genomics* 2015;16:31.


