Required gene set for autotrophic growth of *Clostridium autoethanogenum*

Craig Woods[1], Christopher M. Humphreys[1], Claudio Tomi-Andrino[1,2,3], Anne M. Henstra[1], Michael Köpke[4], Sean D. Simpson[4], Klaus Winzer[1], Nigel P. Minton[1,5]

[1] Clostridia Research Group, BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), Biodiscovery Institute, School of Life Sciences, The University of Nottingham, Nottingham, NG7 2RD, UK

[2] Centre for Analytical Bioscience, Advanced Materials and Healthcare Technologies Division, School of Pharmacy, The University of Nottingham, Nottingham, NG7 2RD, UK

[3] BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), School of Mathematical Sciences, The University of Nottingham, Nottingham, NG7 2RD, UK

[4] LanzaTech Inc., 8045 Lamon Avenue, Suite 400, Skokie, IL, USA

[5] NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham, Nottingham, NG7 2RD, UK.

Corresponding author: Professor Nigel Minton (email: nigel.minton@nottingham.ac.uk)

Running Title: Required gene set for CO growth in an acetogen

ABSTRACT

The majority of the genes present in bacterial genomes remain poorly characterised with up to one third of those that are protein encoding having no definitive function. Transposon insertion sequencing represents a high-throughput technique that can help rectify this deficiency. The technology, however, can only be realistically applied to those species in which high rates of DNA-transfer can be achieved. Here we have developed a number of approaches that overcome this barrier in the autotrophic species *Clostridium autoethanogenum* using a *mariner-based* transposon system. The inherent instability of such systems in the *Escherichia coli* conjugation donor due to transposition events was counteracted through the incorporation of a conditionally lethal *codA* marker on the plasmid backbone. Relatively low frequencies of transformation of the plasmid into *C. autoethanogenum* were circumvented through the use of a plasmid that is conditional for replication coupled with the routine implementation of an Illumina library preparation protocol that eliminates plasmid-based reads. A transposon library was then used to determine the essential genes needed for growth using carbon monoxide as a sole carbon and energy source.

IMPORTANCE. Although microbial genome sequences are relatively easily determined, assigning gene function remains a bottleneck. Consequently, relatively few genes are well characterised, leaving the function of many as either hypothetical or entirely unknown. High-throughput, transposon sequencing can help remedy this deficiency, but is generally only applicable to microbes with efficient DNA-transfer procedures. These exclude many microorganisms of importance to humankind either as agents of disease or as industrial process organisms. Here we developed approaches to facilitate transposon-insertion sequencing in the acetogen *Clostridium autoethanogenum*, a chassis being exploited to

convert single-carbon waste gases, CO and CO_2, into chemicals and fuels at an industrial scale. This allowed the determination of gene essentiality under heterotrophic and autotrophic growth providing insights into the utilisation of CO as a sole carbon and energy source. The strategies implemented are translatable and will allow others to apply transposon-insertion sequencing to other microbes where DNA-transfer has until now represented a barrier to progress.

INTRODUCTION

Although microbial genome sequences are relatively easily determined, assigning gene function remains a bottleneck. Consequently, relatively few genes are well characterised, leaving the function of many as either hypothetical or entirely unknown. Thus, even the Syn 3.0 minimal genome retains 149 genes (32%) of unknown function [1]. A greater understanding of gene functionality can be gleaned through the deployment of high throughput transposon sequencing. This technique is characterised by the simultaneous Illumina sequencing of the site of transposon insertion in pooled mutant libraries using a sequencing primer specific to the transposon-chromosomal junction. If the library consists of a sufficiently high number of unique insertions then the required gene set for the growth conditions used can be inferred since unrepresented or underrepresented genes are likely to be essential. There are several names for this type of approach, including the first four all published in 2009: TraDIS [2], HITS [3], Tn-Seq[4], and INSeq [5]. All of these techniques aim to identify the position and quantity of transposon mutants and are collectively referred to as Transposon insertion sequencing (TIS) [6].

The deployment of TIS typically is largely dependent on high frequency DNA transfer. This excludes its application to many microbial species. Anaerobic bacteria, and in particular

members of the genus *Clostridium*, are of both medical and industrial importance but generally display low rates of DNA transfer. This has limited the exploitation of TIS in this grouping where to date TIS has only been applied [7] to the pathogen *Clostridioides difficile* (formerly *Clostridium difficile*). One group of bacteria with increasing importance are the anaerobic acetogens, typified by *Clostridium autoethanogenum.* Acetogens possess the Wood-Ljungdahl pathway (WLP), or reductive acetyl-CoA pathway, which allows the fixation of CO and CO_2 [8]. Suggested to be the earliest autotrophic pathway [9], it is the most energy efficient of the seven known carbon fixation pathways since it conserves energy while all others require its input [10]. Reducing equivalents needed for metabolic processes are obtained either from H_2 or CO using hydrogenases or CO dehydrogenase (CODH), respectively. Carbon is fixed via the Eastern branch of the pathway where, through a series of cobalamin and tetrahydrofolate-dependent reactions, CO_2 is reduced to a methyl group. The methyl group from the Eastern branch is then combined with CO to form acetyl-CoA which is the root of subsequent anabolic reactions [11-14].

While the majority of acetogens synthesize acetate as the sole fermentation product some, typified by *C. autoethanogenum,* naturally produce industrially relevant compounds as 2,3-butanediol and ethanol, the latter on a commercial scale [15]. Commercial efforts to extend the product range further are ongoing with isopropanol being a notable example [16]. *C. autoethanogenum* is one of the best understood autotrophic acetogens with a manually annotated genome [17,18] and has been subjected to transcriptomic and proteomic analysis [19].

In the current study we sought to maximise the benefit of available *C. autoethanogenum* genome data through implementation of TIS. However, as DNA transfer into *C. autoethanogenum* is only possible at relatively low frequencies, a number of essential modifications to the procedure were required. Specifically, the use of a conditional replicon

and an inducible orthogonal expression system to control production of transposase allows the controlled generation of a large mutant library from a small number of initial transconjugant colonies. Additionally, the incorporation of I-SceI recognition sequences into the delivery vehicle provided a mechanism to eliminate those mini-transposon sequences still present on autonomous copies of the plasmid during the transposon mutant library preparations stage. These adaptations have allowed a thorough genetic analysis of the WLP in *C. autoethanogenum* and, for the first time, the determination of the essential gene set required for growth on CO as a sole carbon and energy source.

RESULTS AND DISCUSSION

Control of transposition. A fundamental requirement of an effective transposon-delivery system is that transposition should preferentially take place in the target strain and not in the donor strain. A previously described clostridial system exploited the *Clostridioides difficile* alternate sigma factor TcdR [20] and one of the only two promoters it recognises, the P_{tcdB} promoter of the toxin B gene (*tcdB*). By generating a derivative of *Clostridium acetobutylicum* in which the TcdR-encoding *tcdR* gene was inserted into the genome at the *pyrE* locus, any subsequently introduced gene that was placed under the control of the P_{tcdB} promoter is expressed. We postulated that this system would be ideal for tightly regulating the transposase.

(Figure 1)

For exploitation in *C. autoethanogenum*, further control was engineered into the system by placing expression of *tcdR* (Fig. 1) under the control of a lactose inducible promoter P_{bgaL} previously shown to be functional in the closely related *Clostridium ljungdahlii* [21].

Accordingly, the P_{bgaL} promoter, together with the necessary *bgaR* which encodes a transcriptional regulator, was positioned 5′ to the *tcdR* gene and the DNA module created (*bgaR*-P_{bgaL}::*tcdR*) integrated into the *C. autoethanogenum* chromosome at the *pyrE* locus using ACE (Allele Coupled Exchange) [21-24]. This involved restoring a uracil-requiring Δ*pyrE* mutant strain to prototrophy concomitant with genomic insertion of *bgaR*-P_{gaL}::*tcdR* using the ACE plasmid pMTL-CH20lactcdR. Successful mutant generation was confirmed by PCR analysis and Sanger sequencing of the amplified DNA and the resulting strain was termed *C. autoethanogenum* C24.

To confirm that TcdR production could be controlled by the addition of exogenous lactose in strain C24, the *Clostridium perfringens catP* reporter gene encoding a chloramphenicol acetyltransferase (CAT) was cloned downstream of the P_{tcdB} promoter on an appropriate clostridial shuttle vector. Regulation of the reporter gene was shown to be dependent on the addition of the lactose inducer (Figure 2). We therefore chose to use this expression system to create our transposon library by placing the transposase under the control of P_{tcdB} on the transposon-delivery plasmid pMTL-CW20 (Fig. 3).

(Figure 2)

A second feature of pMTL-CW20 designed to control unwanted transposition was based on the provision of a promoter-less copy of the *E. coli codA* gene encoding cytosine deaminase to prevent premature transposition in the donor strain. Use of the previously described transposon-delivery vector pMTL-YZ14 [25] was characterised by inconsistent frequencies of transfer to the clostridial recipient and/or to variation in the effectiveness with which transposon mutants were generated once in *C. autoethanogenum*. These inconsistencies appeared to correlate with spontaneous plasmid rearrangements in the donor, as evidenced by unexpected DNA fragment profiles on agarose gels of diagnostic digests of the isolated

plasmid DNA (Fig. 3). This was assumed to be due to transposition of the mini transposon from pMTL-YZ14 while in *E. coli* either into the genome or, as transposition into closed circular autonomous plasmids is preferred, into alternative positions in the vector backbone. The cut and paste nature of the transposition event would mean that plasmids would be generated that either no longer carried a mini-transposon or which had been affected in their maintenance or ability to transfer. Similar instabilities have been noted elsewhere [26].

Cytosine deaminase catalyses conversion of 5-fluorocytosine (5-FC) to the toxic product 5-fluorouracil (5-FU) which ultimately blocks DNA and protein synthesis. On the plasmid pMTL-CW20, *codA* is separated from its P_{thl} promoter (derived from the thiolase gene of *Clostridium acetobutylicum*) by the *catP* mini-transposon. Excision of the mini-transposon as a consequence of its transposition leads to expression of *codA*, a lethal event in the presence of exogenously supplied 5-FC. The addition of this feature to pMTL-CW20 improved the reproducibility with which the plasmid was transferred to *C. autoethanogenum* and appeared to prevent the occurrence of plasmid rearrangements (Fig. 3).

Removal of plasmid-based reads. The use of suicide vectors for transposon delivery is reliant on high frequencies of DNA transfer. Our initial attempts in *C. autoethanogenum* using a delivery vehicle lacking a Gram-positive replicon yielded just 5 transposon mutants from 3 independent conjugations. To overcome this low frequency of mutant generation, a conditional replicon was utilised which has been described previously[25]. To further remove any residual plasmid from the sequencing library I-SceI recognition sites were incorporated into pMTL-CW20 which provided a mechanism for removal of plasmid reads at the sequencing library stage. After adapter-ligation an I-SceI restriction is used to cleave the site between the adapter and the library primer binding site on the transposon, making those fragments originating from plasmid DNA unsuitable templates for the subsequent PCR

amplification step as described in a similar strategy [27]. Since there is no I-SceI recognition site in the *C. autoethanogenum* genome, transposon-insertion sites in the genome will be identified as usual. In the initial transposon library grown on rich medium, 0.2% of reads mapped to the transposon-delivery plasmid. This compares favourably with a study on *Clostridiodes difficile* which used a replicative vector where 48% of the reads in the initial rich medium library mapped to the delivery plasmid [7].

(Figure 3)

Generation of transposon library and growth in autotrophic conditions. Approximately 1.3 million colonies were pooled from 125 transposon selection plates and inoculated into 200 mL of rich medium (YTF) supplemented with thiamphenicol and IPTG. After 24 h of growth, genomic DNA was extracted from this culture and -80°C freezer stocks were made. This first genomic DNA extraction was used to determine the required gene set for growth on rich media, 100,065 unique insertion sites were found from this sample. The genome length divided by the number of unique insertions is 43.49. Subsequently the freezer stocks were used to restore the mutant pool into a defined medium (PETC) supplemented with pyruvate as the carbon source. The PETC culture was used to inoculate a 1.5 L bioreactor containing fermentation medium which lacked a carbon source. The sole carbon and energy source after the inoculation was provided by CO gas sparged into the bioreactor with a gradual increase of CO. The pyruvate was quickly used up as shown by HPLC data (Supplementary Table S3) and *C. autoethanogenum* instead relied upon fixation of CO. The PETC medium provided no supplementary amino acids and instead relied on the native biosynthesis pathways of *C. autoethanogenum*. Vitamin requirements were met via the addition of Wolfe's vitamin solution.

Samples for HPLC analysis of metabolites and for possible genomic DNA extraction and TIS analysis were taken on a daily basis. Samples from 72, 144, 168, 336, and 360 h of growth were used for sequencing. These sequencing data were used to determine the required gene set for growth using CO in a defined medium. An insertion index value of lower than 0.0013 was the cut-off for essentiality. Ultimately, the samples from 336 h and 360 h were used to determine the gene set required for growth on CO, these represent the endpoint of the reactor fed batch culture. The reactor endpoint was sequenced revealing 66,524 unique insertion sites.

Functions of essential genes in heterotrophic compared with autotrophic conditions. The functions of candidate essential genes for growth in rich medium and minimal medium with CO as the carbon and energy source were compared using the KEGG database as summarised in Table 1. There were 439 genes (11%) identified as candidate essential genes out of a total of 4059 genes in the genome for heterotrophic growth on the rich medium YTF where fructose and yeast extract serves as a carbon and energy source (Supplementary Table S1). This is comparable with the number of genes in the Syn3.0 genome and close to the 404 reported in *Clostridiodes difficile* [1,7]. As expected, genes involved in fundamental biological processes such transcription, translation, DNA replication and cell division are common in the rich media essential gene list. Eighteen of the twenty common amino acids have clearly annotated tRNA synthetases which appear essential except for tyrosine and asparagine. Tyrosine appears to exhibit redundancy via CLAU_1290 (*tyrZ)* and CLAU_1635. There is only one annotated asparagine tRNA synthetase *(asnB)* but it seems likely that there is another present (CLAU_2687) and that together they provide functional redundancy meaning that both genes are found to be non-essential. CLAU_2687 is currently annotated as a tRNA synthetase class II but is most likely to be an asparagine-specific tRNA synthetase. Another explanation for the non-essential status of the asparagine tRNA synthetase could be that *C.*

autoethanogenum uses a mechanism common to many bacterial and archaeal taxa which entirely lack an asparagine tRNA synthetase. These taxa rely on a non-discriminating aspartic acid tRNA synthetase followed by an amidotransferase to generate asparagine-tRNAs [31].

The candidate essential gene list for rich medium calls into question several of the annotations in the *C. autoethanogenum* genome. For instance, CLAU_0265 which is annotated as a small acid-soluble spore protein is required on rich medium despite that fact that sporulation should not have been required in the library preparation process. . The gene must, therefore, have an additional or alternative role. Much functional genomics work has yet to be performed on *C. autoethanogenum* since there are 44 rich medium essential genes annotated as hypothetical proteins (Supplementary Table S1).

(Table 1)

In total, 758 genes (19% of the genome) were predicted to be required for autotrophic growth by the endpoint of the CO-fed reactor (Supplementary Table S1). This includes all of the 'core' gene set which were also required on rich medium and all of the genes required to grow on minimal medium lacking amino acids. The core gene set was predicted to be comprised of 439 genes. This means that 319 genes are likely to be required for the synthesis of all amino acids and utilisation of CO as a carbon and energy source. As vitamins were provided, their biosynthetic pathways were not expected to be represented, similarly nitrogen and sulphur were supplied in the medium as ammonium chloride and sodium sulphide, respectively. All genes and their predicted essentiality status in each experimental condition are presented in Supplementary Table S1. Comparing two predicted required gene lists from different times in an experiment is an imperfect method of deducing condition-specific genes, but the data is nevertheless extremely informative.

(Figure 4)

Essential genes of the Wood-Ljungdahl (WLP) pathway. In order to grow using CO as a sole carbon and energy source it is necessary for *C. autoethanogenum* to use two molecules of CO to form one molecule of acetyl-CoA. Acetyl- CoA consists of a methyl group, a carbonyl group and the CoA cofactor. The methyl group is supplied by the action of the bifunctional CODH enzyme which oxidises CO to CO_2, this CO_2 molecule then follows the Eastern branch of the WLP before being combined with another CO molecule and CoA by the acetyl-CoA synthase (ACS). It was therefore expected that all the genes involved in the WLP would be required for growth on CO. A complete WLP was indeed found in the list of essential genes and has been mapped out in Fig. 4. The WLP was not required during heterotrophic growth despite the fact that it is utilised during heterotrophic growth to fix CO_2 released during glycolysis using the reducing equivalents generated by glycolysis [24].

(Figure 5)

To generate ATP, *C. autoethanogenum* is reliant on generating a transmembrane electrochemical gradient via the intrinsically important [32] Rnf complex and a membrane integral ATP synthase [19, 33]. The Rnf complex of *C. autoethanogenum* is encoded by the region CLAU_3144-CLAU_3149. A comparison of the insertion sites found in heterotrophic and autotrophic conditions for this region is shown in Fig. 5. With exception of *rnfB,* all of the encompassed genes were found to be essential for growth on CO (Fig.5), confirming previous observations that inactivation of these genes in either *C. ljungdahlii* and *Acetobacterium woodii* curtailed growth on $H_2 + CO_2$ [34,35]. Despite *rnfB* being above the insertion index threshold for essentiality on CO, it is significantly under-represented when the data obtained from cells grown on pyruvate is compared with CO (log_2 fold change = -2.86, p-value = 1.16E-11). It may be the case that *rnfB* encodes a non-essential component of the

complex which aids functionality but is not required for it. In Methanogens RnfB has been characterized as an entry point for electrons to the Rnf complex [36].

The importance of Nfn for autotrophic growth. In order to further verify the calling of gene essentially under specific conditions using our parameters, a candidate gene was selected for directed CRISPR mutagenesis. The *nfn* gene (CLAU_1539) encodes an electron-bifurcating ferredoxin-dependent transhydrogenase, responsible for the production of NADPH from NADH and Fd^{2-}, thus recycling NAD+. Our TIS data analysis found that the *nfn* gene was non-essential when *C. autoethanogenum* was grown on rich medium or when grown on minimal medium with pyruvate, but when autotrophic conditions were used the gene was essential. This suggested that a directed CRISPR knockout mutant should be obtainable while the culture is grown under heterotrophic conditions but should fail to survive when transferred to autotrophic conditions. A CRISPR in-frame deletion mutant of *nfn* (Δ*nfn*) was created which was viable on rich media, and on minimal medium with pyruvate as a carbon source but was unable to grow when CO was the sole carbon and energy source.

Initially the Δ*nfn* strain was characterised in serum bottles, using minimal PETC media and either 10 mM of sodium pyruvate, or 1 bar of CO in the headspace, as the carbon and energy source. Serum bottles were inoculated with 1 ml (1:50 inoculum) of a late exponential culture grown in the anaerobic cabinet on minimal media with fructose as a carbon source. The cultures grown on pyruvate grew similarly to the wild type control, however, no evidence of growth was evident when CO was used instead of pyruvate as the carbon and energy source. This inability to utilise CO as a carbon and energy source was further demonstrated on a larger scale using a fed-batch CSTR experiment, whereby a 1.5 L culture was inoculated with 150 ml of an early exponential culture grown on minimal media and pyruvate. The pH was

281 controlled with NaOH and H_2SO_4, and sparged through continual addition of nitrogen at a
282 rate of 60 ml/min. At the time of inoculation 5 mM of sodium pyruvate was added to the
283 culture. Once an OD_{600} of approximately 0.3 had been reached, CO was introduced at a rate
284 of 10 ml/min. In the case of the wild type culture, the strain was able to adapt to the CO
285 carbon and energy source and after 48 h the OD continued to increase after the pyruvate had
286 been depleted. In the case of the *nfn* mutant, the culture was not able to adapt to utilising CO,
287 and the optical density rapidly declined following depletion of the pyruvate (data not shown).

288 **Assessment of metabolic modelling-derived gene essentiality.** Experimentally confirmed
289 gene essentiality for growth on minimal medium supplemented with CO was compared
290 against the predicted essentiality calculated from a metabolic model of *C. autoethanogenum*
291 by means of Flux Balance Analysis (FBA) [37,38] (Supplementary Table S2). To that end, the
292 confusion matrix was generated (Table 2) to calculate Matthew's correlation coefficient
293 (MCC), a robust metric ranging from -1 to +1 used to evaluate binary classifications (such as
294 essential or non-essential gene) [39]. A MCC = 0.34 was obtained which compares
295 unfavourably to a similar study in *E. coli* where a value of 0.69 was reported [40,41] (for the
296 well-curated model *i*JO1366, and by using thermodynamics and multi-omics constraints).
297 The lower value for *C. autoethanogenum* is a measure of the comparatively limited genome
298 annotation in this organism. However, Table 2 shows that the model does have predictive
299 power with 87 genes being correctly predicted to be essential.

300 (Table 2)

301

302 **Essentiality in the Wood-Ljungdahl pathway (WLP).** The generation of the methyl group
303 from CO first requires its oxidation to CO_2 by a CODH. *C. autoethanogenum* possesses three

such enzymes genome, namely CLAU_1578/CLAU_1579 (*acsA*), CLAU_2924 (CAETHG_3005) (*cooS1*) and CLAU_3807 (CAETHG_3899) (*cooS2*) [18]. Their independent interruption via ClosTron mutagenesis [42] showed that autotrophic growth was only abolished in the *acsA* knockout mutant suggesting it is the only CODH required for growth on either CO or CO_2. Our data validates this conclusion with only *acsA* being predicted as required for growth on CO but not on YTF. The remaining putative CODH genes were required in neither condition and there was no substantial change in insertion index between YTF and CO conditions. The conditionally essential CODH encoded by *acsA* has an internally translated stop codon (TGA) not found in the equivalent genes of related organisms and can alternatively be thought of as two ORFs (CLAU_1578 and CLAU_1579) although it appears that CLAU_1579 makes no separate product. It has been demonstrated that *acsA* can be translated either as a 44 kDa protein or as a 69 kDa protein depending on whether the TGA internal stop codon is the end of translation or whether it causes the incorporation of a selenocysteine residue [24]. It appears from our data that both ORFs are required under autotrophic conditions. Thus, the 44 kDa protein alone does not appear to be sufficient for autotrophy and cells apparently require the 69 kDa protein to be autotrophic.

There are three putative formate dehydrogenases in the *C. autoethanogenum* genome encoded by CLAU_0081, CLAU_2712/CLAU_2713 (*fdhA*), and CLAU_2907. Of these, *fdhA* alone appears to be essential only on CO while the remaining two genes are required in neither tested condition. The most important formate dehydrogenase is therefore *fdhA* which is found in a complex with an NADP-specific electron-bifurcating [FeFe]-hydrogenase (Hyt) [43]. Two of the three putative formate dehydrogenases are selenoenzymes which may be higher efficiency than the cysteine-containing analogues, it is therefore tempting to speculate that the non-selenoenzyme formate dehydrogenase may be present as a backup for low selenium conditions [44]. However, it appears from our data that neither CLAU_0081 nor CLAU_2907

could provide sufficient activity in the *fdhA* mutants for them to not be outcompeted causing *fdhA* to appear essential under autotrophic conditions.

The steps from formate to methyl-THF are catalysed by the products of CLAU_1572-CLAU_1576 which all appear to be required for growth on CO. CLAU_1574 and CLAU_1576 additionally appear to also be required for growth on the rich medium.

The methyl group of methyl-THF is transferred to the Corrinoid Iron-Sulfur Protein (CoFeSp) cofactor before being combined with the carbonyl group supplied by another molecule of CO by the action of the ACS (acetyl-CoA synthase). The ACS is encoded by the region CLAU_1566-70 in which CLAU_1566, CLAU_1568, and CLAU_1569 were only required on CO whereas CLAU_1567 and CLAU_1570 were required on CO and on YTF.

Essentiality in the metabolism from acetyl-CoA. There are four main carbon compounds at the end points of metabolism for *C. autoethanogenum*: acetate, ethanol, 2,3 – butanediol, and lactate. The route to acetate from acetyl-CoA proceeds through acetyl phosphate; catalysed by the enzymes phosphate acetyltransferase (encoded by *pta*) and acetate kinase (encoded by *ackA*). Both *pta* and *ackA* were found to be essential when growing on rich medium. The step from acetyl phosphate to acetate regenerates ATP and so this pathway may be required for energy generation.

However, *pta* has been knocked out in *C. ljungdahlii* [45-47] where it significantly impaired growth rates and acetate formation. The *C. ljungdahlii pta* knockout may be viable only because of a second putative phosphate acetyltransferase (WP_063556670.1) which is annotated as a bifunctional enoyl-CoA hydratase/phosphate acetyltransferase and has no homolog in *C. autoethanogenum*. The bifunctional enoyl-CoA hydratase/phosphate acetyltransferase may be producing sufficient ATP for cells to be viable. The absence of an

alternative phosphate acetyltransferase in *C. autoethanogenum* is likely the cause of the essentiality of *pta* in our data.

The route to ethanol can proceed from acetyl-CoA either straight to acetaldehyde and then to ethanol or via acetate, then acetaldehyde and finally ethanol. The more direct route from acetyl-CoA to acetaldehyde is catalysed by an acetaldehyde dehydrogenase (EC 1.2.1.10) which could be encoded by an estimated five genes within the *C. autoethanogenum* genome (CLAU_1772, CLAU_1783, CLAU_3204, CLAU_3655, CLAU_3656) none of which appear to be required in either growth condition. This could represent redundancy between these genes which further knockout studies could aim to confirm, or it could be that this route to ethanol is not required. The alternative route to ethanol via acetate is similar in that there are two predicted genes (CLAU_0089 and CLAU_0099) which could encode an aldehyde ferredoxin oxidoreductase (AOR; EC 1.2.7.5) but neither of them appear to be essential in either growth condition. In this case the result is best explained by a lack of biological necessity for this reaction since it has been shown that a double AOR knockout strain was still viable autotrophically [24].

There are two candidate genes encoding pyruvate synthase enzymes for formation of pyruvate from acetyl-CoA (CLAU_0896 and CLAU_2947) of which only CLAU_2947 appears to be required; this is true in both growth conditions. All of the genes encoding functions for the pathways leading to lactate and 2,3-butanediol appear non-essential. In the case of the conversion of acetolactate to acetoin and in the production of lactate utilising NADH there appears to be only one gene encoding the relevant functions (CLAU_2851 and CLAU_1108 respectively); in these cases redundancy is unlikely to be the reason for their non-essential status meaning it is more likely these are unnecessary biological routes.

The successful application of TIS to *C. autoethanogenum* has provided a wealth of information on gene essentiality in this industrially important acetogen and represents the most thorough analysis of its kind performed to date in clostridia. The essentiality status of all *C. autoethanogenum* genes can now be consulted (in Supplementary Table S1) before directed knockouts are attempted. Overall, our findings highlight that TIS represents a powerful functional genomics tool which can be applied to less genetically tractable organisms using the methods applied here. Presented data allows a confident determination of the Wood-Ljungdahl pathway genes of *C. autoethanogenum* and opens up future avenues of investigation into the genes which are essential for autotrophic growth with no obvious reason as to why.

MATERIALS AND METHODS

Microbiology. *E. coli* DH5alpha (NEB) was used for all for cloning and sExpress [47] as a conjugal donor. Strains were cultured at 37 °C in LB broth with appropriate antibiotic supplementation and 5-FC in a shaking incubator or on LB agar in a static incubator. *C. autoethanogenum* was cultured and manipulated in an anaerobic workstation (Don Whitley) with an atmosphere of 80% nitrogen, 10% carbon dioxide and 10% hydrogen at 37 °C. The three media used, were YTF (Table S4-S7), ATCC Medium 1754 (Table S8-S10) and Fermentation medium (Table S11-S13). YTF medium was composed of yeast extract (10 g/L), tryptone (16/L), fructose (10 g/L), NaCl (0.2 g/L), H_3BO_3 (100 µg/L), $MnCl_2.4H_2O$ (230 µg/L), $FeCl_2.4H_2O$ (78 µg/L), $CoCl_2.6H_2O$ (103 µg/L), $NiCl_2.6H_2O$ (602 µg/L), $ZnCl_2$ (78 µg/L), $CuSO_4.5H_2O$ (50 µg/L), $AlK(SO_4)_2.12H_2O$ (50 µg/L), Na_2SeO_3 (58 µg/L), Na_2WO_4 (53 µg/L), $Na_2MbO_4.2H_2O$ (52 µg/L). p-aminobenzoate (114 µg/L), riboflavin (104 µg/L), thiamine (200 µg/L), nicotinate (206 µg/L), pyridoxin (510 µg/L), pantothenate (104 µg/L),

cyanocabalamin (78 µg/L), d-biotin (22 µg/L), folate (48 µg/L), lipoate (50 µg/L). The fermentation medium contained $MgCl_2.6H_20$ (0.5 g/L), $CaCl_2.2H_20$ (0.37 g/L), KCl (0.15 g/L), NaCl (0.12 g/L), 85% H_3P0_4 (0.38 ml/L), NH_4Cl (1 g/L), $CoCl_2$ (476 ng/L), HBO_4 (124 ng/L), $MnCl_2.4H_2O$ (396 ng/L), $NaMoO_4.2H_2O$ (484 ng/L), Na_2SeO_3 (346 ng/L), $FeCl_2.4H_2O$ (3.87 µg/L), $NiCl_2.6H_2O$ (238 ng/L), $ZnCl_2$ (138 ng/L).

Plasmids were transferred from sExpress to *C. autoethanogenum* as detailed in Woods et al., 2019 [48]. Briefly, this involved mixture of the donor and recipient cultures together and incubation on a non-selective YTF plate for 20 h at 37 °C before harvesting and plating onto selective YTF agar. Antibiotic selection for transposon plasmids was performed using chloramphenicol (25 µg/ml) and erythromycin (500 µg/ml) in *E. coli* or thiamphenicol (15 µg/ml) and clarithromycin (6 µg/ml) in *C. autoethanogenum*. Kanamycin (50 µg/ml) was used to select for the sExpress donor strain. D-cycloserine (250 µg/ml) was used to counter-select the sExpress donor strain. Fluorocytosine (FC) was supplemented at 30 µg/ml and IPTG at a concentration of 1 mM. Plasmid pMTL-CW20 may be sourced from www.plasmidvectors.com.

DNA manipulations. Genomic DNA purifications were performed using bacterial gDNA extraction kits from Sigma Aldrich. Plasmid DNA was purified with mini-prep kits from NEB. Screening PCRs were performed using DreamTaq polymerase (Thermo Fisher). Oligonucleotides were synthesised by Sigma Aldrich. Sanger sequencing was performed by Source Bioscience.

Mutant generation using CRISPR. A CRISPR in-frame deletion vector was designed as previously described using the pMTL40000 CRISPR vector series [49]. In this case we employ the trCas9 nickase variant under control of the P_{fdx} promoter from *C. sporogenes* ferredoxin gene, a unique sgRNA (ATATCCATTAAGAATATGTT**)** under control of the constitutive

P_{araE} promoter of the *C. acetobutylicum araE* gene targeting *nfn*, and a homologous recombination cassette to allow the precise in-frame deletion of *nfn*. Following vector assembly, the construct was transferred to wild type *C. autoethanogenum* by conjugation using sExpress as the *E. coli* donor strain [48]. Following two rounds of selection on thiamphenicol and D-cycloserine, to select for recipient strains harbouring the CRISPR vector and counterselect the *E. coli* donor cells, respectively, a colony PCR screen was performed on resultant colonies, amplifying from the genomic locus flanking the regions selected for homologous recombination. The screen revealed that the *nfn* knockout mutant was indeed present in the population, and the strain was sub-cultured for storage and preparation of genomic DNA. Sanger sequencing from a high-fidelity PCR product confirmed the precise in-frame deletion of *nfn*.

Assessment of transposon vectors. Transposon delivery vectors were transferred to *C. autoethanogenum* via conjugal transfer from sExpress and selected for on YTF agar plates supplemented with clarithromycin and D-cycloserine. Colonies were harvested from selection plates by flooding with PBS and the entire cell suspension was serially diluted and spread onto YTF agar plates supplemented with either clarithromycin and IPTG, thiamphenicol and IPTG or clarithromycin to determine the transposition frequency and plasmid-retention in the presence of IPTG.

Transposon library creation. The transposon-delivery vector pMTL-CW20 was transformed into an *E. coli* conjugative donor strain sExpress [48] which was used to transfer the plasmid into *C. autoethanogenum* C24. Twelve conjugations were performed simultaneously producing a total of around 81,000 transconjugant colonies on YTF agar supplemented with D-cycloserine and clarithromycin. All transconjugants were pooled and plated onto YTF agar plates supplemented with IPTG, lactose and thiamphenicol to select for

transposon mutants and incubated for 72 hours. Transposon mutant colonies were then harvested and inoculated to YTF broth supplemented with IPTG and thiamphenicol. The rich medium sequencing samples were taken from this liquid phase which was used to inoculate PETC pyruvate medium. The PETC pyruvate culture was allowed to reach stationary phase before being used as inoculum for the bioreactor where CO served as the carbon and energy source; a DNA sample was taken from PETC pyruvate at the point of bioreactor inoculation. Samples were taken from the bioreactor to check for the presence of pyruvate, monitor the OD and to serve as DNA samples for identification of insertion sites.

Bioinformatics and metabolic modelling. Experimentally confirmed essential genes were compared against metabolic modelling-derived essentiality. Lists of essential genes were generated using the BioTraDIS toolkit approach as previous described [50]. A genome-scale model (GSM) of CO-fed *C. autoethanogenum* was handled using the COBRA Toolbox in MATLAB R2016b to predict gene essentiality [37,51]. Briefly, the wild type model was subjected to Flux Balance Analysis (FBA) by selecting the maximization of the biomass yield as the objective function [38]. A gene is deemed essential when knocking it out made the biomass reaction carry no flux [52]. Finally, Matthew's correlation coefficient was used as a metric to assess the quality of the GSM predictions, where "1" is a perfect correlation between experimental and predicted gene essentiality, "0" no correlation, and "-1" perfect anti-correlation [40]. The model and the scripts are available in GitHub (https://github.com/SBRCNottingham/C_auto_essentiality).

Sequencing and bioinformatics. Sequencing library preparation was performed as an amplicon library using a splinkerette adapter [50]. Genomic DNA was fragmented to an average of 400 bp using a covaris sonicator followed by bead purification using NEB sample preparation beads at a ratio of 1.5X beads to sample. Fragmented DNA was end repaired and A-tailed using the NEB Ultra II library preparation kit. Splinkerette adapters were ligated onto the end of A-tailed fragments with reagents from the Ultra II library preparation kit. A 1X bead purification was performed before an I-SceI digest step to cleave plasmid DNA between the library primer and P7 primer. Another 1X bead purification was performed before PCR amplification of the transposon junctions using KAPA HiFi polymerase. An initial denaturing step of 95°C for two minutes was followed by 20 rounds of 95°C for 20 sec, 61°C for 30 sec then 72°C for 30 sec before a final extension of two min at 72°C was performed.

PCR products with a size range of 250-500 bp were gel extracted from a low-melt agarose gel using the NEB monarch gel extraction kit. Gel extracted products were analysed on an Agilent bioanalyser using a DNA 1000 chip and quantified via Qubit and qPCR. Two separate runs were performed on an Illumina MiSeq.

Raw sequences were trimmed before filtering for reads which contain the expected transposon tag. The transposon tag was removed from reads which could then be mapped to the *C. autoethanogenum* genome to identify insertion sites. The BioTraDIS analysis pipeline was used for these steps and for subsequent analysis [50]. Reads in the final 10% of the gene were omitted from the analysis. Reads which mapped to multiple locations on the genome were randomly mapped between those locations.

Data availability. Raw sequencing data has been deposited with the NCBI, accession numbers SRR16990784-SRR16990788 and SRR17285607-SRR17285609.

ACKNOWLEDGEMENTS

This work was funded by the Biotechnology and Biological Sciences Research Council [grant numbers BB/L502030/1, BB/K00283X/1, BB/L013940/1]. NPM acknowledges funding from LanzaTech as part of BB/L502030/1. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Sequencing was performed by Deep Seq (University of Nottingham) with thanks to Nadine Holmes, Matthew Carlile, and Victoria Wright.

Author Contributions

Craig Woods contributed conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, software, validation, writing, reviewing, and editing the manuscript. Christopher M. Humphreys contributed conceptualization, investigation, methodology, project administration, supervision, visualization, writing (in minority), and reviewing & editing the manuscript. Claudio Tomi-Andrino contributed conceptualization, data curation, formal analysis, investigation, software, and reviewing & editing the manuscript. Anne M. Henstra and Klaus Winzer contributed supervision and reviewing & editing the manuscript. Michael Köpke and Séan D. Simpson reviewed and edited the manuscript. Nigel P. Minton contributed conceptualization, funding acquisition, methodology, project administration, supervision, and reviewing & editing the manuscript.

Competing Interests

MK and SDS are employees of LanzaTech, a for profit with commercial interest in clostridial gas fermentation.

REFERENCES

1. Hutchison III CA, Chuang R-Y, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, Gill J, Kannan K, Karas BJ, Ma L, Pelletier JF, Qi Z-Q, Richter RA, Strychalski EA, YoSuzuki L, Tsvetanova B, Wise KS, Smith HO, Glass JI, Merryman C, Gibson DG, Venter JC. 2016. Design and synthesis of a minimal bacterial genome. *Science* **351**:6253.

2. Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, Wain J, Parkhill J, Turner AK. 2009. Simultaneous assay of every *Salmonella Typhi* gene using one million transposon mutants. *Genome Res* **19**:2308–2316.

3. Gawronski JD, Wong SMS, Giannoukos G, Ward DV, Akerley BJ. 2009. Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. *Proc Natl Aca Sci USA* **106:**16422–16427.

4. van Opijnen, T., Bodi, K. L., & Camilli, A. 2009. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature Methods*, 6(10). 5. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, Knight R, Gordon JI. 2009. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **6**:279–89.

6. Chao MC, Abel S, Davis BM, Waldor MK. 2016. The design and analysis of transposon insertion sequencing experiments. *Nat Rev Microbiol* **14**:119–128.#

7. Dembek M, Barquist L, Boinett CJ, Cain AK, Mayho M, Lawley TD, Fairweather NF, Fagan RP. 2015. High-throughput analysis of gene essentiality and sporulation in *Clostridium difficile. mBio.* **6(2):**e02383.

8. Ragsdale SW, Pierce E. 2008. Acetogenesis and the Wood-Ljungdahl pathway of CO_2 fixation. *Biochim Biophys Acta - Proteins Proteomics* **1784:**1873–1898.

9. Brock TD. 1989. Evolutionary relationships of the autotrophic bacteria, p 499-512. In Schlegel H, Bowien B (ed), *Autotrophic Bacteria*. Science Tech Publishers, Madison.

10. Fast AG, Papoutsakis ET. 2012. Stoichiometric and energetic analyses of non-photosynthetic CO_2-fixation pathways to support synthetic biology strategies for production of fuels and chemicals. *Curr Opin Chem Eng* **1:**380–395.

11. Drake HL, Daniel SL. 2004. Physiology of the thermophilic acetogen *Moorella thermoacetica*. *Res Microbiol* **155:**869–883.

12. Ragsdale SW. 2008. Enzymology of the Wood-Ljungdahl pathway of acetogenesis. Ann N Y Acad Sci **1125:**129–136 doi: 10.1196

13. Shin J, Song Y, Jeong Y, Cho BK. 2016. Analysis of the core genome and pan-genome of autotrophic acetogenic bacteria. *Front Microbiol* **7:**1531.

14. Müller V, Pal Chowdhury N, Basen M. 2018. Electron Bifurcation: A Long-Hidden Energy-Coupling Mechanism. *Ann Rev Microbiol* **72:**331-353.

15. Köpke M, Simpson SD. 2020. Pollution to products: recycling of 'above ground' carbon by gas fermentation. *Curr Opin Biotechnol* **65:**180–189.

16. Köpke M, Simpson S, Liew F. Chen W. 2012. Fermentation Process For Producing isopropanol Using a Recombinant Microorganism. US patent 936,586,8B2.

17. Humphreys CM, McLean S, Schatschneider S, Millat T, Henstra AM, Annan FJ, Breitkopf R, Pander B, Piatek P, Rowe P, Wichlacz AT, Woods C, Norman R, Blom J, Goesman A, Hodgman C, Barrett D, Thomas NR, Winzer K, Minton NP. 2015. Whole genome sequence and manual annotation of *Clostridium autoethanogenum*, an industrially relevant bacterium. *BMC Genomics* **16:**1–10.

18. Brown SD, Nagaraju S, Utturkar S, De Tissera S, Segovia S, Mitche W, Land ML, Dassanayake A, Köpke M. 2014. Comparison of single-molecule sequencing and hybrid approaches for finishing the genome of *Clostridium autoethanogenum* and analysis of CRISPR systems in industrial relevant Clostridia. *Biotechnol Biofuels* **7:**40.

19. Marcellin E. Behrendorff JB, Nagaraju S, DeTissera S, Segovia S, Palfreyman RW, Daniell J, Licona-Cassani C, Quek L, Speight R, Hodson MP, Simpson SD, Mitchell WP, Köpke M, Nielsena LK. 2016. Low carbon fuels and commodity chemicals from waste gases-systematic approach to understand energy metabolism in a model acetogen. *Green Chem* **18:**3020-3028.

20. Minton NP, Ehsaan M, Humphreys CM, Little GT, Baker J, Henstra AM, Liew F, Kelly ML, Sheng L, Schwarz K, Zhang Y. 2016. A roadmap for gene system development in *Clostridium*. *Anaerobe* **41**:104–112.

21. Banerjee A, Leang C, Ueki T, Nevin KP, Lovley, DR. 2014. Lactose-inducible system for metabolic engineering of *Clostridium ljungdahlii*. *Appl Environ Microbiol* **80**:2410–2416.

22. Heap JT, Ehsaan M, Cooksley CM, Ng YK, Cartman ST, Winzer K, Minton NP. 2012. Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Res* **40** doi:10.1093

23. Ng YK, Ehsaan M, Philip S, Collery MM, Janoir C, Collignon A, Cartman ST, Minton NP. 2013. expanding the repertoire of gene tools for precise manipulation of the *Clostridium difficile* genome: allelic exchange using *pyrE* alleles. *PLoS One* **8**:e56051

24. Liew F, Henstra AM, Köpke M, Winzer K, Simpson SD, Minton NP. 2017. Metabolic engineering of *Clostridium autoethanogenum* for selective alcohol production. *Metab Eng* **40**:104–114.

25. Zhang Y, Grosse-Honebrink A, Minton NP. 2015. A Universal *Mariner* Transposon System for Forward Genetic Studies in the Genus *Clostridium*. *PLoS One* **10**:e0122411.

26. Dempwolff F, Sanchez S, Kearns DB. 2020. TnFLX: A third-generation mariner-based transposon system for *Bacillus subtilis*. *Appl Environ Microbiol.* **86:**e02893-19

27. Bossé JT, Li Y, Leanse LG, Zhou L, Chaudhuri RR, Peters SE, Wang J, Maglennon GA, Holden MTG, Maskell DJ, Tucker AW, Wren BW, Rycroft AN, Langford PR. 2018. Rationally designed mariner vectors to allow functional genomic analysis of *Actinobacillus pleuropneumoniae* and other bacteria by transposon-directed insertion-site sequencing (TraDIS). *bioRxiv* **433086**

28. Heap JT, Pennington OJ, Cartman ST, Minton NP. 2009. A modular system for *Clostridium* shuttle plasmids. *J Microbiol Methods* **78**:79–85.

29. Purdy D, O'Keeffe TA, Elmore M, Herbert M, McLeod A, Bokori-Brown M, Ostrowski A, Minton NP. 2002. Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Mol Microbiol* **46**:439–452.

30. Lampe DJ, Akerley BJ, Rubin EJ, Mekalanos JJ, Robertson HM. 1999. Hyperactive transposase mutants of the Himar1 mariner transposon. *Proc Natl Acad Sci USA* **96**:11428–11433.

31. Iwasaki W, Sekine S, Kuroishi C, Kuramitsu S, Shirouzu M, Yokoyama S. 2006. Structural Basis of the Water-assisted Asparagine Recognition by Asparaginyl-tRNA Synthetase. *J Mol Biol* **360**:329–342.

32. Schuchmann K, Müller V. 2014. Autotrophy at the thermodynamic limit of life: A model for energy conservation in acetogenic bacteria. *Nat Rev Microbiol* **12**:809–821.

33. Köpke M, Held C, Hujer S, Liesegang H, Wiezer A, Wollherr A, Ehrenreich A, Liebl W, Gottschalk G, Dürre P. 2010. *Clostridium ljungdahlii* represents a microbial production platform based on syngas. *Proc Natl Acad Sci* **107**:13087–13092.

34. Tremblay P, Zhang T, Dar SA, Leang C, Lovely DR. 2012. The Rnf complex of *Clostridium ljungdahlii* is a proton-translocating NAD+ oxidoreductase essential for autotrophic growth. *mBio* **4**:1–8.

35. Westphal L, Wiechmann A, Baker J, Minton N P, Müller V. 2018. The Rnf complex is an energy-coupled transhydrogenase essential to reversibly link cellular NADH and ferredoxin pools in the acetogen *Acetobacterium woodii*. *J Bacteriol* **200:**e00357-18.

36. Suharti S, Wang M, de Vries S, Ferry JG. 2014. Characterization of the RnfB and RnfG Subunits of the Rnf Complex from the Archaeon *Methanosarcina acetivorans*. *PLoS One* **9**:e97966.

37. Norman ROJ, Millat T, Schatschneider S, Henstra AM, Breitkopf R, Pander B, Annan FJ, Piatek P, Hartman HB, Poolman MG, Fell DA, Winzer K, Minton NP, Hodgman C. 2019. Genome-scale model of *C. autoethanogenum* reveals optimal bioprocess conditions for high-value chemical production from carbon monoxide. *Eng Biol* **3**:32–40.

38. Orth JD, Thiele I, Palsson BO. 2010. What is flux balance analysis? *Nat Biotechnol* **28**:245–248.

39. Chicco D, Tötsch N, Jurman G. 2021. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Min* **14**:13.

40. Salvy P, Hatzimanikatis V. 2020. The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models. *Nat Commun* **11**:1–17.

41. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, Palsson BØ. 2011. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism-2011. *Mol Syst Biol* **7**:535.

42. Liew FM, Martin ME, Tappel RC, Heijstra BD, Mihalcea C, Köpke M. 2016. Gas Fermentation-A flexible platform for commercial scale production of low-carbon-fuels and chemicals from waste and renewable feedstocks. *Front Microbiol* **7** doi:10.3389.

43. Wang S, Huang H, Kahnt J, Mueller AP, Köpke M, Thauer RK. 2013. NADP-Specific electron-bifurcating [FeFe]-hydrogenase in a functional complex with formate dehydrogenase in *Clostridium autoethanogenum* grown on CO. *J Bacteriol* **195**:4373–4386

44. Köpke M, Mihalcea C, Liew F, Tizard JH, Ali MS, Conolly JJ, Al-Sinawi B, Simpson SD. 2011. 2,3-Butanediol production by acetogenic bacteria, an alternative route to chemical synthesis, using industrial waste gas. *Appl Environ Microbiol* **77**:5467–5475.

45. Huang H, Chai C, Li N, Rowe P, Minton NP, Yang S, Jiang W, Gu Y. 2016. CRISPR/Cas9-Based Efficient Genome Editing in *Clostridium ljungdahlii*, an Autotrophic Gas-Fermenting Bacterium. *ACS Synth Biol* **5**:1355–1361.

46. Latif H, Zeidan AA, Nielsen AT, Zengler K. 2014. Trash to treasure: Production of biofuels and commodity chemicals via syngas fermenting microorganisms. *Curr Opin Biotechnol* **27**:79–87.

47. Lo J, Humphreys JR, Jack J, Urban C, Magnusson L, Xiong W, Gu Y, Ren ZJ, Maness PC. 2020. The Metabolism of *Clostridium ljungdahlii* in Phosphotransacetylase Negative Strains and Development of an Ethanologenic Strain. *Front Bioeng Biotechnol* **8**:560726.

48. Woods C, Humphreys CM, Rodrigues RM, Ingle P, Rowe P, Henstra AM, Köpke M, Simpson SD, Winzer K, Minton NP. 2019. A novel conjugal donor strain for improved DNA transfer into *Clostridium* spp. *Anaerobe* **59**:184–191.

49. Ingle P, Groothuis D, Rowe P, Huang H, Cockayne A, Kuehne SA, Jiang W, Gu Y, Humphreys CM, Minton NP. 2019. Generation of a fully erythromycin-sensitive strain of *Clostridioides difficile* using a novel CRISPR-Cas9 genome editing system. *Sci Rep* **9**:8123.

50. Barquist L, Mayho M, Cummins C, Cain AK, Boinett CJ, Page AJ, Langridge GC, Quail MA, Keane JA, Parkhill J. 2016. The TraDIS toolkit: Sequencing and analysis for dense transposon mutant libraries. *Bioinformatics* **32**:1109–1111.

51. Schellenberger J, Que R, Fleming RM, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, Kang J, Hyduke DR, Palsson BØ. 2011. Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0. *Nat Protoc* **6**:1290–1307.

52. Gatto G, Miess H, Schulze A, Nielsen J. 2015. Flux balance analysis predicts essential genes in clear cell renal cell carcinoma metabolism. *Sci Rep* 9:10738.

687 **Figure 1.** TcdR-mediated orthogonal expression. In *C. autoethanogenum* C24, *tcdR* is under
688 the control of the lactose-inducible promoter system *bgaR*-P_{bgaL} from *C. perfringens*. In this
689 way the P_{tcdB} promoter can be induced indirectly via the inducible expression of *tcdR* from
690 the chromosome.

691

692 **Figure 2.** Chloramphenicol acetyl transferase (CAT) assay of lactose-inducible orthogonal
693 system. Expression from P_{tcdB} was quantified using a CAT assay. Three plasmids were
694 conjugated into *C. autoethanogenum* P_{bgaL}*_tcdR* (*C. autoethanogenum* C24) with each
695 plasmid harbouring *catP* under the control of either P_{tcdB}, P_{thl} (positive control), or no
696 promoter (negative control). The strain harbouring the P_{tcdB} plasmid was tested with and
697 without the addition of 10 mM lactose while the remaining plasmids were tested without
698 lactose. The data shown is the result of biological triplicates with error bars showing the
699 standard deviation.

700

701 **Figure 3.** Transposon delivery plasmid pMTL-CW20. A) pMTL-CW20 is based on the
702 pMTL-YZ14 plasmid described in [25] using components from the plasmid modular transfer
703 series outlined in [28] as well as the *codA* from *E. coli* and I-SceI recognition sites. Replication
704 occurs in *E. coli* via the pUC ColE1 origin of replication and the plasmid can be transferred
705 to clostridial recipients using the *oriT* from RK2 [29]. In clostridial hosts the plasmid is
706 conditionally replicative where the presence of IPTG is the non-permissive condition.
707 Transposition is achieved via a hyperactive Himar1 variant [30] which mobilises a mini-
708 transposon containing the *catP* gene which confers chloramphenicol and thiamphenicol
709 resistance. A Rho-independent terminator downstream of the *fdx* gene of *Clostridium*
710 *sporogenes* resides upstream of *catP*. B) A verified pMTL-YZ14 plasmid was used to
711 transform *E. coli* Top10 and transformant colonies used to inoculate overnight cultures.

712 Plasmids prepared from overnight cultures were extracted and treated with SbfI. Movement
713 of the transposon into various other parts of the vector was found to have occurred (2, 4, 6,
714 and 7) while only lanes 3 and 5 exhibited the expected band pattern. ThermoFisher 1kb+ plus
715 ladder is in lane 1 C) An analogous procedure using EcoRV was later followed using pMTL-
716 CW20 instead of pMTL-YZ14. In this case all six plasmids exhibited the expected band
717 pattern (lanes 2-7) with ThermoFisher 1kb+ ladder in lanes 1 and 8.
718

719 **Table 1.** Functions of essential genes. Number of *C. autoethanogenum* essential genes for
720 various KEGG functional categories on the rich medium YTF and the minimum fermentation
721 medium with CO as a carbon and energy source.

722 **Figure 4.** Essential Genes of the Wood-Ljungdahl pathway. Route from CO to acetate
723 showing the expected gene/locus tag for each step. Each of the locus tags listed was required
724 for growth on CO.

725 **Figure 5**: Insertions in the Rnf complex region. Number of reads detected along the genomic
726 region encoding the Rnf complex for heterotrophic and autotrophic conditions. Insertions are
727 relatively abundant in heterotrophic conditions implying importance for the complex under
728 autotrophic conditions.

729 **Table 2. Confusion matrix for gene essentiality comparison.** TP = true positive, FP = false
730 positive, FN = false negative, and TN = true negative. A perfect correlation (MCC = +1)
731 would require TP = 179, FP = 0, FN = 0, and TN = 353.

732

733

734

735