1 **To be submitted to Journal of Bacteriology**

2 **TITLE:** Both *adhE* and a separate NADPH-dependent alcohol dehydrogenase (*adhA*) are
3 necessary for high ethanol production in *Thermoanaerobacterium saccharolyticum*

4 **AUTHORS:** Tianyong Zheng,[a, c] Daniel G. Olson,[b, c] Sean J. Murphy, [b, c] Xiongjun Shao,
5 [b, c] Liang Tian, [b, c] Lee R. Lynd [a, b, c,#]

6 Department of Biological Sciences, Dartmouth College, Hanover, NH, USA[a]; Thayer
7 School of Engineering, Dartmouth College, Hanover, NH, USA[b]; BioEnergy Science
8 Center, Oak Ridge, TN, USA[c]

9 **RUNNING TITLE:** *adhA* and *T. saccharolyticum* high ethanol production

10 #Corresponding author: Lee R. Lynd, Lee.R.Lynd@Dartmouth.edu

11

12

13

14

15

16

17

18

ABSTRACT

Thermoanaerobacterium saccharolyticum has been engineered to produce ethanol at ~90% theoretical yield and titer of 70 g/L. Its ethanol-producing ability has drawn attention to its metabolic pathways, which could potentially be transferred to other organisms of interest. Here we report that the iron-containing AdhA is important for ethanol production in the high-ethanol strain of *T. saccharolyticum* (LL1049). A single-gene deletion of *adhA* in LL1049 reduced ethanol production by ~50%, whereas multiple gene deletions of all annotated alcohol dehydrogenases except *adhA* and *adhE* did not affect ethanol production. Deletion of *adhA* in wild-type *T. saccharolyticum* reduced NADPH-linked ADH activity (acetaldehyde-reducing) by 93%.

IMPORTANCE

In this study we set out to identify the necessary alcohol dehydrogenases for high ethanol production in *T. saccharolyticum*. Based on previous work, we had assumed that *adhE* was the primary alcohol dehydrogenase. Here we show that both *adhA* and *adhE* are needed for high ethanol yield in the engineered strain LL1049. This is the first report showing *adhA* is important for ethanol production in a native *adhA* host, which has important implications for achieving higher ethanol yields in other microorganisms.

INTRODUCTION

Thermophilic bacteria are excellent candidates for producing ethanol from lignocellulosic biomass. Over the past decade, much work has been dedicated to engineering these organisms for increased ethanol production (1, 2) . In particular, *Thermoanaerobacterium saccharolyticum* has been engineered to produce ethanol at

~90% theoretical yield and titer of 70 g/L(3, 4). However, *T. saccharolyticum* cannot break down cellulose, which accounts for a half to a third of plant biomass. Due to the abundance of cellulose in lignocellulosic biomass, the ideal ethanol producer would be an organism that can rapidly solubilize cellulose. Toward this goal, we have been engineering *Clostridium thermocellum*, one of the most efficient cellulose consumers, for increased ethanol yield. Currently, the highest ethanol yield reported for *C. thermocellum* is 75% of the theoretical maximum (5). One strategy for engineering *C. thermocellum* is to transfer the genes from *T. saccharolyticum* responsible for ethanol production.

In many bacterial organisms, the last two steps of ethanol production are acetyl-CoA to acetaldehyde (ALDH reaction) and acetaldehyde to ethanol (ADH reaction) (Figure 1). While a number of alcohol dehydrogenases can catalyze the ADH reaction, the only known enzyme capable of catalyzing the ALDH reaction is AdhE, a bifunctional enzyme with both ADH and ALDH domains (Figure 1). Previously we have demonstrated that *adhE* is an important gene for ethanol production in both *C. thermocellum* and *T. saccharolyticum*: deletion of *adhE* reduced ethanol yield by >95% in both organisms (6). This was not unexpected, since AdhE proteins are essential for anaerobic ethanol production in many organisms (7–11). However, because AdhE is a bifunctional enzyme with two functional domains, the loss of ethanol formation from the *adhE* deletion is not evidence that both domains are essential for ethanol production. As presented below, there is evidence suggesting multiple enzymes could perform the ADH reaction in *T. saccharolyticum*, and AdhE may only be necessary for its ALDH function. We have previously reported the biochemical properties of the *C. thermocellum* and *T. saccharolyticum* bifunctional AdhE: wild-type AdhE is mostly NADH-linked for ADH

activity in both organisms (12). Deleting *adhE* in *C. thermocellum* eliminated >90% of ADH activity in cell extracts, suggesting that AdhE is the primary enzyme contributing to ADH activity. Interestingly, this was not the case for *T. saccharolyticum*. Deleting *adhE* eliminated only NADH-linked ADH activity in *T. saccharolyticum* cell extracts, while significant amounts of NADPH-linked ADH activity remained (6, 12). This suggested that AdhE is not the only enzyme involved in the reduction of acetaldehyde to ethanol, and at least one other NADPH-linked ADH enzyme also plays a role in maintaining cofactor and redox balances in ethanol production.

There are six proteins in *T. saccharolyticum* annotated as "alcohol dehydrogenase" and two as "aldehyde dehydrogenase" in the Pfam protein database (13) (Table 1). Since we had previously characterized the bifunctional alcohol dehydrogenase, AdhE (12), in this work we focused on understanding the roles of the other putative aldehyde dehydrogenase (ALDH) and ADH enzymes, in an effort to identify the set of genes that is both necessary and sufficient for high ethanol production in *T. saccharolyticum*. In this study we generated a series of gene deletions of putative *adh* and *aldh* genes, the results of which cast light on the importance of one *adh* other than *adhE* that contributes to high ethanol production in *T. saccharolyticum*.

MATERIALS AND METHODS

Plasmid and strain construction. Strains used in this study are listed in Table 2. Kanamycin marker deletions were made by transforming *T. saccharolyticum* cells with deletion vectors, using a protocol described previously (14). In brief, deletion vectors consist of a kanamycin marker flanked by upstream and downstream homology regions

of the target gene. Vectors were constructed from linear PCR products via Gibson assembly (15), followed by anaerobic incubation with cells at 55 °C to allow the naturally-competent *T. saccharolyticum* cells to take up the DNA construct. Cells incubated with the deletion vector were harvested at OD_{600} ~0.6 during the exponential growth phase and plated with 200 µg/ml kanamycin. The Tsac_0218 markerless deletion was made using the *pta-ack* marker removal system as described previously by Shaw et al. (16), all other markerless deletions were made using the *tdk* (thymidine kinase) marker removal system described by Shao et al. (17). Briefly, the *tdk* gene was deleted in strains LL1049 and LL1076, and the Δ*tdk* strains were used for subsequent transformations. The selection process includes a first round of positive selection based on resistance to kanamycin, followed by a second round of negative selection based on resistance to FUDR (Floxuridine) as previously described (17). Plasmids and primers used for cloning are listed in Table S1 and S2, the Genbank sequence of each plasmid is also included as supplementary information (Supplementary text S1). *Escherichia coli* competent cells (T7 Express *lysY/I*^{q}, NEB) were transformed as previously described (12), using the pEXP5-NT/TOPO (Invitrogen) plasmid backbone for AdhA protein expression.

Media and growth conditions. For transformations and enzyme assays, *T. saccharolyticum* strains were grown anaerobically to exponential phase (OD_{600}~0.6) in CTFUD media as previously described (12); for fermentation end-product analysis, all strains were cultivated on MTC-6 defined medium. MTC-6 medium was prepared as previously described (and called "modified MTC medium" instead of "MTC-6") (18), filter sterilized immediately after preparation into sterilized serum bottles, and purged of oxygen using twenty 45-second cycles of ultra high purity N_2 gas and vacuum. Bottles

were released of residual pressure inside an anaerobic chamber using a sterile filter and needle prior to inoculation. All media used for fermentation purposes were inoculated within two hours of preparation. All fermentations were incubated for 72 hours, in an orbital shaking incubator (Innova 4080 Incubator Shaker, New Brunswick Scientific) held at 55°C and 180 RPM. Fermentations were performed in 125 ml glass serum bottles with a working volume of 50 ml, and were inoculated with 2% of frozen cell culture.

E. coli cells used for protein expression were grown in Lysogeny Broth (LB) media.

Fermentation end-product analysis. All fermentation data were calculated from biological replicates and reported in Table S3 and Table S4.

Gas Chromatography

Gas composition was analyzed using an SRI 310C gas chromatograph (SRI Inc.), utilizing a nitrogen carrier gas at a flow rate of 8.2 ml/minute. Oven temperature was maintained at 151°C, and the current at 80 mA. Triplicate injections of 99% hydrogen and 10% carbon dioxide calibration gasses (MESA Specialty Gasses & Equipment) were used for standard curves, using 100 µl and 500 µl injections respectively. Single 500 µl gas samples from cultures were then analyzed for H_2 and CO_2 concentration. The absolute H_2 and CO_2 measurements were determined after correction for pressure differences. Cultures were then stored in the dark at 4°C overnight, prior to further analysis.

TOCN (Total Organic Carbon and Nitrogen) analysis

TOCN analysis was used to determine pellet carbon (pellet C), which is a proxy for cell biomass production. 10 ml samples were removed from fermentation bottles by syringe with constant mixing of the contents for homogeneity and divided into 1ml aliquots. These aliquots were centrifuged at 15,000 rpm for five minutes; pellets were used for TOCN analysis and supernatant for HPLC analysis. Pellets were washed twice with 950 µl of deionized water.

Washed cell pellets were resuspended via vigorous vortexing and were added to the TOCN vials filled with 19.5 ml deionized water. A single standard for the quantification of total organic carbon and nitrogen was prepared by adding 1 ml of a 5 g/L glycine solution to a vial also filled with 19.5 ml deionized water. TOCN analysis was performed using an organic carbon/nitrogen analyzer with a liquid module and automatic sample injection (ASI-V/TOC-V/TNM-1/SSM-5000, Shimadzu Scientific Instruments Inc.). Eluent consisted of 6.67 ml concentrated H_2SO_4 diluted in 10 L deionized water. Ultrazero compressed air (AI UZ300, Airgas) was used as the carrier gas, at a pressure of 200 kPA and flow rate of 150 ml/min.

HPLC (High Pressure Liquid Chromatography) analysis

700 µl supernatant resulting from preparation of cell pellets was acidified with 35 µl 10% H_2SO_4 solution. These were mixed gently by inversion, and then passed through a Costar SpinX HPLC 0.2 um nylon microcentrifuge filter (Corning, Inc.) via centrifugation at 15,000 rpm for five minutes.

Fermentation products were quantified using a Waters 2695 Separations Module, fitted with an additional UV detector (for the accurate quantification of pyruvate). Injection

volumes were 40 µl, and product concentrations were calculated from standard curves. All peaks were reviewed manually for errors.

Enzymatic assays. Cells were harvested and lysed anaerobically as previously described (12). ADH reactions (acetaldehyde reduction direction) were carried out anaerobically at pH 7.0 in a reaction mixture containing 0.2 mM NADH or NADPH, 20 mM acetaldehyde, 100 mM Tris-HCl, 5 µM $FeSO_4$ and cell extract or purified protein solution. The final volume was 1000 µl, the assay temperature was 55°C, and the assay was started by the addition of acetaldehyde. ADH activity was monitored and activity was calculated as previously described (12). One unit of activity (U) is equal to the formation of 1 µmol of product per minute. Specific activities are expressed as U/mg of protein.

Protein expression and purification. AdhA protein was expressed in *E. coli* with an N-terminus His-tag and purified by affinity chromatography. 100 ml *E. coli* cell culture was grown to OD_{600} ~ 0.6 before induction of protein expression with IPTG (isopropyl β-D-1-thiogalactopyranoside) at 0.4 mM final concentration. Cells were induced for 2 hours before harvesting, after which pellets were stored at -80°C. Anaerobic protein expression was performed similarly, however cell cultures were transferred to an anaerobic serum bottle (purged with nitrogen) prior to IPTG induction to maintain anaerobic conditions. After induction, cells were cultured for 2 hours before harvesting, and pellets were stored anaerobically at -80°C.

Protein purification was carried out using affinity columns (Ni-NTA spin columns, Qiagen) as previously described (12): aerobic purification was performed under normal

174 atmospheric conditions (i.e. outside the anaerobic chamber), and anaerobic protein
175 purification was performed inside an anaerobic chamber with oxygen maintained at
176 levels below 5 parts per million. Purified proteins were visualized and their sizes were
177 confirmed by SDS-PAGE.

178 **RESULTS**

179 **Deletion of *adhA* reduced ethanol production in high-ethanol strain LL1049.**
180 Individual deletions of putative alcohol and aldehyde dehydrogenases in the wild type
181 (LL1025) and high ethanol (LL1049) strains are shown in Figure 2A and 2B. In the wild-
182 type strain, *adhE* was the only gene whose deletion affected ethanol production, in
183 agreement with previous reports (6). Deleting the other five putative *adh* and *aldh* genes
184 had no significant effect on ethanol yield; the resulting strains grew to high cell densities
185 similar to their parent strain (wt) (See pellet C data in Table S3). However, in the high-
186 ethanol strain (LL1049), deleting *adhA* reduced ethanol yield by ~50% (see comparison
187 between strains LL1287 and LL1049 in Figure 2B). Furthermore, the resulting strain
188 LL1287 grew poorly. Deleting the other *adh* and *aldh* genes in the high-ethanol strain
189 had no significant effect on ethanol yield or cell growth. We were unable to delete *adhE*
190 in the high-ethanol strain LL1049 (marked by asterisks in Figure 2) because ethanol is the
191 only significant organic end product of this strain; that is, producing ethanol is the
192 organism's only way to balance redox levels via NAD(P) cofactors and generate the ATP
193 needed for growth.

194 **The role of *adhE* and *adhA* for high yield ethanol production**. Having
195 investigated deletions of putative *adh* and *aldh* genes individually, we now sought to

196 investigate them in combination to control for epistatic effects. The parent strain used for
197 these deletions was LL1328, which is strain LL1049 with the *tdk* gene deleted (necessary
198 for making markerless deletions). The *tdk* deletion did not have any effect on ethanol
199 production, as expected (Figure 2, LL1328 vs LL1049). Figure 2C shows the results of
200 multiple *adh* deletions. In strain LL1332, Tsac_0218, Tsac_2222, Tsac_0285, and
201 Tsac_1049 have all been deleted, leaving only *adhE* and *adhA* as potential alcohol
202 dehydrogenase genes. This strain had an ethanol yield (0.45 g/g) similar to that of its
203 parent strain LL1328 (0.41 g/g), showing no reduction in ethanol and indicating that
204 Tsac_0218, 2222, 0285 and 1049 are not required for high yield ethanol production and
205 that *adhA* and *adhE* are, in fact, sufficient for high yield ethanol production.

206 When we deleted *adhA* in the strain with the 4 other *adh* deletions (Figure 2C,
207 LL1333 vs LL1332), we saw a large decrease in ethanol production, confirming the
208 importance of *adhA* in this strain.

209 Additionally, we considered the possibility that the "alcohol dehydrogenase"
210 proteins annotated in the Pfam database may not be an exhaustive list, in which case
211 another unidentified *adh* gene could also be necessary for ethanol production. To
212 investigate this possibility, we deleted both *adhE* and *adhA* in the wild type strain
213 background (Table 2, LL1025 → LL1076 → LL1334 → LL1335), generating strain
214 LL1335. If another functional *adh* gene exists, there should be remaining ADH
215 enzymatic activity in strain LL1335. As shown in Table 3, wild-type *T. saccharolyticum*
216 exhibits high ADH activity with both NADH and NADPH cofactors. Deletion of *adhE*
217 (strain LL1334) reduced NADH-linked ADH activity by 98%, but NADPH-linked ADH
218 activity remained high. On the contrary, deletion of *adhA* (strain LL1286) did not

219 significantly affect NADH-linked ADH activity, but reduced NADPH-linked activity by
220 93%. Deletion of both *adhE* and *adhA* eliminated NADH-linked activity and reduced 93%
221 of NADPH-linked activity (LL1335). This strongly suggests that AdhA and AdhE are the
222 two main alcohol dehydrogenases involved in ethanol production in *T. saccharolyticum*.
223 Furthermore, this suggests that ADH activity from AdhA is NADPH-linked.

224 **Complementation of *adhE* and *adhA* deletions.** In order to confirm the
225 functions of *adhE* and *adhA* and gauge the effect of secondary mutations, we attempted
226 to complement the *adhA* and *adhE* deletions by introducing these genes back onto the
227 chromosome. In strains LL1287 (single gene deletion) and LL1333 (multiple gene
228 deletions) where *adhA* had been deleted, transformation efficiency was extremely low.
229 After multiple attempts, we were able to insert the *adhA* gene in strain LL1333 but not in
230 strain LL1287— transformations with LL1333 had a single colony while transformations
231 with LL1287 had none. This colony became the resulting strain LL1402 with *adhA*
232 introduced back into its original locus, producing ethanol at a yield of 0.19 g/g (Table S4).
233 This was a two-fold increase in ethanol production comparing to its parent strain LL1333,
234 whose ethanol yield was 0.09 g/g. However, reintroducing *adhA* only restored ethanol
235 production to 43% of the pre-deletion strain (LL1332). This suggested the accumulation
236 of secondary mutations as a result of deleting *adhA* and other alcohol dehydrogenase
237 genes.

238 Similarly, *adhE* was reintroduced to strain LL1334 (Δ*adhE*). Transformation
239 efficiency of strain LL1334 was also low, yielding only five colonies with the desired
240 *adhE* insertion: strains LL1403-LL1407. All five colonies were able to restore ethanol
241 production to different degrees, with ethanol yields ranging from 0.14 g/g to 0.22 g/g

(Table S4). Strain LL1403 had the highest ethanol yield 0.22 g/g, restoring ethanol production to 88% of the wild-type strain.

AdhA is an oxygen-insensitive NADPH-linked alcohol dehydrogenase. To further characterize the biochemical properties of AdhA, we expressed and purified *T. saccharolyticum* AdhA in *E. coli* (Figure S2). Purified AdhA had high NADPH-linked ADH activity (6.8 ± 1.7 U/mg) and negligible amounts of NADH-linked ADH activity (0.3 ± 0.6 U/mg), which confirms the result of the gene deletion studies (see previous section).

In addition, we noted that it was unnecessary to perform anaerobic protein expression and purification for AdhA, as was done previously for AdhE (12). Unlike AdhE, which was extremely sensitive to oxygen, AdhA appeared to be insensitive to the presence of oxygen. It had comparable ADH activity under anaerobic and aerobic expression conditions: 6.9 ± 1.9 U/mg and 6.8 ± 1.7 U/mg, respectively.

DISCUSSION

AdhA proteins in various organisms. Keshav et al. first adopted the name *adhA* in 1990 to describe a gene that encodes a zinc alcohol dehydrogenase in *Zymomonas mobilis* (19). Analysis of the amino acid sequence of the *Z. mobilis* AdhA showed that it contained an "alcohol dehydrogenase GroES-like domain" (ADH_N) and a "zinc-binding dehydrogenase domain" (ADH_zinc_N) (Figure 3). Alcohol dehydrogenases in general have been extensively studied in various organisms for many years. However, many of these proteins were given the name "AdhA" regardless of their homology to the zinc-binding *Z. mobilis* AdhA. In an effort to clarify the differences among the "AdhA"

proteins reported in literature and avoid confusion, we compiled a non-exhaustive list of AdhAs from literature according to Pfam protein domains and families. AdhAs that have sequence-similarity to the *Z. mobilis* AdhA include those from *Corynebacterium glutamicum* (20, 21), *Lactococcus lactis* (22), *Picrophilus torridus* (23), and *Rhodococcus erythropolis* (24). "AdhA" was also used to describe the *Pyrococcus furiosus* short-chain dehydrogenase (adh_short) (25, 26) (Figure 3), and the PQQ-dependent alcohol dehydrogenases from *Frateuria aurantia* (27) (Figure 3) and *Acetobacter pasteurianus* (28, 29). In many cases, the name "AdhA" simply implies that it was the first alcohol dehydrogenase studied in a given organism.

The above AdhA proteins are not homologous to the *T. saccharolyticum* AdhA that we report in this study in terms of protein sequence and domains. The *T. saccharolyticum* AdhA consists of an "iron-containing alcohol dehydrogenase domain" (Fe-ADH); similar AdhAs (>60% protein sequence identity) include those from *Thermoanaerobacter brockii* (30–32), *Moorella* sp. (33) and *Thermoanaerobacter mathranii* (34). We suggest that, for clarification purposes, future studies involving AdhA should specify a distinct protein feature, for example iron-containing AdhA or PQQ-dependent AdhA.

In addition, it is worth noting that the Fe-ADH family (which includes the *T. saccharolyticum* AdhA) encompasses a large group of proteins, some of which are not named alcohol dehydrogenases and have less than 30% protein sequence identity with *T. saccharolyticum* AdhA. These additional Fe-ADH family members include: *E. coli* FucO (29% identity)(35, 36), *Klebsiella pneumoniae* 1,3-propanediol dehydrogenase (29% identity) (37), and *E. coli* YqhD (24% identity)(38) (Figure 3). The ADH domain of *T.*

saccharolyticum AdhE is also an iron-containing ADH with 30% protein sequence identity to *T. saccharolyticum* AdhA (Figure 3).

Oxygen tolerance of AdhA. Protein sequence alignments between *E. coli* AdhE, *T. saccharolyticum* AdhE and *T. saccharolyticum* AdhA showed that the AdhE proteins had a Glu568 residue, while the *T. saccharolyticum* AdhA had a Lys residue at the corresponding position (Figure S1). Glu568 has been shown to be the critical residue responsible for oxygen sensitivity in *E. coli* AdhE: Holland-Staley et al. reported that the Glu568Lys mutation "produced AdhE that was active under both aerobic and anaerobic conditions" (39). This difference in amino acids at position 568 may explain the difference in oxygen sensitivity between AdhE and AdhA.

Identification of cryptic NADPH-ADH activity in *T. saccharolyticum*. Previously it was shown that deletion of *adhE* in *T. saccharolyticum* nearly eliminated NADH-linked ADH activity, but only slightly reduced NADPH-linked ADH activity (6). Furthermore, in the high-ethanol-producing strain, LL1049, levels of NADPH-ADH in the *cell extract* were higher compared to the wild type strain (LL1025), whereas levels of NADPH-ADH from *purified* AdhE were lower, compared to the wild type (12). The combination of these observations suggested that there was a cryptic *adh* gene responsible for the NADPH-ADH activity that was distinct from *adhE*. Here we have presented evidence that this cryptic NADPH-ADH gene is, in fact, *adhA*.

Cofactor specificity and ethanol production in *T. saccharolyticum*. Why was *adhA* important for ethanol production only in the high-ethanol-producing strain (LL1049) and relatively unimportant in the wild type strain (LL1025)? This may be explained by the

309 different NAD(P)H cofactors involved in ethanol production. Previously we have shown
310 that wild-type *T. saccharolyticum* (strain LL1025) and strain LL1049 appear to utilize
311 different cofactors in the pyruvate-to-ethanol pathway (40). Pyruvate:ferredoxin
312 oxidoreductase (PFOR) has been shown to be the enzyme responsible for the oxidation of
313 pyruvate to acetyl-CoA in *T. saccharolyticum* (18)(41), coupled with the reduction of
314 ferredoxin. In order to maintain electron balance, the cell uses a ferredoxin:$NAD(P)^+$
315 oxidoreductase (FNOR) to transfer electrons from reduced ferredoxin to $NAD(P)^+$. In the
316 high-ethanol-producing strain, NADPH-FNOR activity is increased relative to the wild
317 type (40), which presumably leads to increased NADPH production. In order to use the
318 electrons from NADPH for ethanol production, high levels of NADPH-ADH activity are
319 needed. Previously we have shown that the high-ethanol-producing strain (LL1049) has
320 mutations in its *adhE* gene that confer NADPH-ADH activity, but the newly acquired
321 NADPH-ADH activity in the LL1049 AdhE is lower than the ADH activity in wt AdhE
322 (12). In this study we observed that without AdhA, the NADPH-ADH activity from AdhE
323 is only sufficient to produce ethanol at about 40% of the theoretical maximum yield (strain
324 LL1287). It is only when the gene *adhA* is also present (strain LL1332) that high ethanol
325 yields (~90% of theoretical) are observed. Thus NADPH-ADH activity of AdhA is needed
326 to balance the NADPH-FNOR activity (presumably from NfnAB) in strain LL1049. The
327 proximity of these on the genome suggests that their cofactor compatibility may not be a
328 coincidence, and that they may, in fact, be an operon.

329 Additionally, acetaldehyde is very reactive and ultimately toxic at thermophilic
330 temperatures (42). Therefore, it is possible that the NADPH-linked ADH activity from the
331 mutated AdhE is not sufficient for effective acetaldehyde conversion, and AdhA is needed

to serve as an additional acetaldehyde scavenger.

Importance of the iron-containing AdhA in ethanol production. Alcohol dehydrogenases have long been utilized in engineering strategies for increased biofuel production. Among them, AdhE has been widely studied for ethanol production (7–11), the zinc-binding AdhA was used in isobutanol production (22, 43), and YqhD was used to increase isobutanol (22) and propanediol production (44). To date, there has only been one study where the iron-containing AdhA was utilized to increase bioethanol production (26). Basen et al. inserted the *adhA* gene from *Thermoanaerobacter* strain X514 (Figure 3) into the hyperthermophilic archaeon *P. furiosus*, which has a unique aldehyde ferredoxin oxidoreductase (AOR) enzyme that reduces acetate to acetaldehyde, but no native alcohol dehydrogenase. Wild-type *P. furiosus* does not produce ethanol, however, as a result of the *adhA* insertion, the engineered strain was able to generate ethanol from acetaldehyde, and produced >20 mM ethanol from 5 g/L cellobiose (26), demonstrating that the iron-containing AdhA can be used for ethanol production. Here we show that *adhA* is important for ethanol production in a native *adhA* host. Although strain LL1049 was engineered for increased ethanol production, none of the genetic modifications targeted *adhA*, and in fact, the role of *adhA* in this strain was largely unknown until this report. The ability of engineered strains of *T. saccharolyticum* to produce ethanol at high yield (90% of theoretical) and titer (70 g/L) has drawn attention to its metabolic pathways. Uncovering the role of *adhA* has shed light on the ethanol production pathway in *T. saccharolyticum* and provided ideas for improving ethanol production in other organisms, such as *C. thermocellum*.

Journal of Bacteriology

JB

Journal of Bacteriology

ACKNOWLEDGMENTS

We thank Johannes P. Van Dijken for his useful comments and discussions of metabolism. We thank the Mascoma Corporation for their gift of the *T. saccharolyticum* strains LL1076 (a.k.a. M3223) and LL1049 (a.k.a. M1442 or MO1442).

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. The manuscript has been authored by Dartmouth College under contract no. DE-AC05-00OR22725 with the U.S. Department of Energy

REFERENCES

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

2. **Olson DG**, **Sparling R**, **Lynd LR**. 2015. Ethanol production by engineered thermophiles. Curr. Opin. Biotechnol. **33**:130–141.

3. **Shaw AJ**, **Podkaminer KK**, **Desai SG**, **Bardsley JS**, **Rogers SR**, **Thorne PG**, **Hogsett DA**, **Lynd LR**. 2008. Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield. Proc. Natl. Acad. Sci. U. S. A. **105**:13769–13774.

4. **Herring CD**, **Kenealy WR**, **Joe Shaw A**, **Covalla SF**, **Olson DG**, **Zhang J**, **Ryan Sillers W**, **Tsakraklides V**, **Bardsley JS**, **Rogers SR**, **Thorne PG**, **Johnson JP**, **Foster A**, **Shikhare ID**, **Klingeman DM**, **Brown SD**, **Davison BH**, **Lynd LR**, **Hogsett DA**. 2016. Strain and bioprocess improvement of a

thermophilic anaerobe for the production of ethanol from wood. Biotechnol. Biofuels **9**:125.

5. **Tian L**, **Papanek B**, **Olson DG**, **Rydzak T**, **Holwerda EK**, **Zheng T**, **Zhou J**, **Maloney M**, **Jiang N**, **Giannone R**, **Hettich R**, **Guss A**, **Lynd L**. 2016. Simultaneous achievement of high ethanol yield and titer in *Clostridium thermocellum*. Biotechnol. Biofuels **(in press)**:1–11.

6. **Lo J**, **Zheng T**, **Hon S**, **Olson DG**, **Lynd LR**. 2015. The bifunctional alcohol and aldehyde dehydrogenase gene, *adhE*, is necessary for ethanol production in *Clostridium thermocellum* and *Thermoanaerobacterium saccharolyticum*. J. Bacteriol. **197**:JB.02450–14.

7. **Membrillo-Herna J**, **Echave P**, **Cabiscol E**, **Tamarit J**, **Ros J**, **Lin ECC**. 2000. Evolution of the *adhE* gene product of *Escherichia coli* from a functional reductase to a dehydrogenase. J. Biol. Chem. **275**:33869–33875.

8. **Peng H**, **Wu G**, **Shao W**. 2008. The aldehyde/alcohol dehydrogenase (AdhE) in relation to the ethanol formation in *Thermoanaerobacter ethanolicus* JW200. Anaerobe **14**:125–127.

9. **Pineda E**, **Encalada R**, **Olivos-García A**, **Néquiz M**, **Moreno-Sánchez R**, **Saavedra E**. 2013. The bifunctional aldehyde–alcohol dehydrogenase controls ethanol and acetate production in *Entamoeba histolytica* under aerobic conditions. FEBS Lett. **587**:178–184.

10. **Boxma B**, **Voncken F**, **Jannink S**, **van Alen T**, **Akhmanova A**, **van Weelden**

SWH, **van Hellemond JJ**, **Ricard G**, **Huynen M**, **Tielens AGM**, **Hackstein JHP**. 2004. The anaerobic chytridiomycete fungus *Piromyces* sp. E2 produces ethanol via pyruvate:formate lyase and an alcohol dehydrogenase E. Mol. Microbiol. **51**:1389–1399.

11. **Atteia A**, **van Lis R**, **Mendoza-Hernández G**, **Henze K**, **Martin W**, **Riveros-Rosas H**, **González-Halphen D**. 2003. Bifunctional aldehyde/alcohol dehydrogenase (ADHE) in chlorophyte algal mitochondria. Plant Mol. Biol. **53**:175–188.

12. **Zheng T**, **Olson DG**, **Tian L**, **Bomble YJ**, **Himmel ME**, **Lo J**, **Hon S**, **Shaw AJ**, **van Dijken JP**, **Lynd LR**. 2015. Cofactor specificity of the bifunctional alcohol and aldehyde dehydrogenase (AdhE) in wild-type and mutant *Clostridium thermocellum* and *Thermoanaerobacterium saccharolyticum*. J. Bacteriol. **197**:2610–2619.

13. **Finn RD**, **Coggill P**, **Eberhardt RY**, **Eddy SR**, **Mistry J**, **Mitchell AL**, **Potter SC**, **Punta M**, **Qureshi M**, **Sangrador-Vegas A**, **Salazar GA**, **Tate J**, **Bateman A**. 2015. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. **44**:gkv1344.

14. **Shaw AJ**, **Hogsett DA**, **Lynd LR**. 2010. Natural competence in *Thermoanaerobacter* and *Thermoanaerobacterium* species. Appl. Environ. Microbiol. **76**:4713–4719.

15. **Gibson DG**, **Young L**, **Chuang R**, **Venter JC**, **Iii CAH**, **Smith HO**, **America N**. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases

419 **6**:12–16.

420 16. **Shaw AJ**, **Covalla SF**, **Hogsett DA**, **Herring CD**. 2011. Marker removal system
421 for *Thermoanaerobacterium saccharolyticum* and development of a markerless
422 ethanologen. Appl. Environ. Microbiol. **77**:2534–6.

423 17. **Shao X**, **Zhou J**, **Olson DG**, **Lynd LR**. 2016. A markerless gene deletion and
424 integration system for *Thermoanaerobacter ethanolicus*. Biotechnol. Biofuels
425 **9**:100.

426 18. **Zhou J**, **Olson DG**, **Lanahan AA**, **Tian L**, **Murphy SJ-L**, **Lo J**, **Lynd LR**. 2015.
427 Physiological roles of pyruvate ferredoxin oxidoreductase and pyruvate formate-
428 lyase in *Thermoanaerobacterium saccharolyticum* JW/SL-YS485. Biotechnol.
429 Biofuels **8**:138.

430 19. **Keshav KF**, **Yomano LP**, **An H**, **Ingram L**. 1990. Cloning of the *Zymomonas*
431 *mobilis* structural gene encoding alcohol dehydrogenase I (*adhA*): sequence
432 comparison and expression in *Escherichia coli*. J. Bacteriol. **172**:2491–2497.

433 20. **Arndt A**, **Eikmanns BJ**. 2007. The alcohol dehydrogenase gene *adhA* in
434 *Corynebacterium glutamicum* is subject to carbon catabolite repression. J.
435 Bacteriol. **189**:7408–7416.

436 21. **Kotrbova-Kozak A**, **Kotrba P**, **Inui M**, **Sajdok J**, **Yukawa H**. 2007.
437 Transcriptionally regulated *adhA* gene encodes alcohol dehydrogenase required for
438 ethanol and *n*-propanol utilization in *Corynebacterium glutamicum* R. Appl.
439 Microbiol. Biotechnol. **76**:1347–1356.

22. **Atsumi S**, **Wu TY**, **Eckl EM**, **Hawkins SD**, **Buelter T**, **Liao JC**. 2010. Engineering the isobutanol biosynthetic pathway in *Escherichia coli* by comparison of three aldehyde reductase/alcohol dehydrogenase genes. Appl. Microbiol. Biotechnol. **85**:651–657.

23. **Hess M**, **Antranikian G**. 2008. Archaeal alcohol dehydrogenase active at increased temperatures and in the presence of organic solvents. Appl. Microbiol. Biotechnol. **77**:1003–1013.

24. **Peng X**, **Taki H**, **Komukai S**, **Sekine M**, **Kanoh K**, **Kasai H**, **Choi SK**, **Omata S**, **Tanikawa S**, **Harayama S**, **Misawa N**. 2006. Characterization of four *Rhodococcus* alcohol dehydrogenase genes responsible for the oxidation of aromatic alcohols. Appl. Microbiol. Biotechnol. **71**:824–832.

25. **van der Oost J**, **Voorhorst WGB**, **Kengen SWM**, **Geerling ACM**, **Wittenhorst V**, **Gueguen Y**, **de Vos WM**. 2001. Genetic and biochemical characterization of a short-chain alcohol dehydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus*. Eur. J. Biochem. **268**:3062–3068.

26. **Basen M**, **Schut GJ**, **Nguyen DM**, **Lipscomb GL**, **Benn RA**, **Prybol CJ**, **Vaccaro BJ**, **Poole FL**, **Kelly RM**, **Adams MWW**. 2014. Single gene insertion drives bioalcohol production by a thermophilic archaeon. Proc. Natl. Acad. Sci. **111**:17618–17623.

27. **Trček J**, **Matsushita K**. 2013. A unique enzyme of acetic acid bacteria, PQQ-dependent alcohol dehydrogenase, is also present in *Frateuria aurantia*. Appl. Microbiol. Biotechnol. **97**:7369–7376.

462 28. **Yuan Y**, **Feng F**, **Chen L**, **Yao Q**, **Chen K**. 2014. Surface display of *Acetobacter*
463 *pasteurianus* AdhA on *Bacillus subtilis* spores to enhance ethanol tolerance for
464 liquor industrial potential. Eur. Food Res. Technol. **238**:285–293.

465 29. **Quintero Y**, **Poblet M**, **Guillamón JM**, **Mas A**. 2009. Quantification of the
466 expression of reference and alcohol dehydrogenase genes of some acetic acid
467 bacteria in different growth conditions. J. Appl. Microbiol. **106**:666–674.

468 30. **McMahon M**, **Mulcahy P**. 2002. Bioaffinity purification of NADP(+)-dependent
469 dehydrogenases: studies with alcohol dehydrogenase from *Thermoanaerobacter*
470 *brockii*. Biotechnol. Bioeng. **77**:517–27.

471 31. **Oestreicher EG**, **Pereira DA**, **Pinto GF**. 1996. Steady-state kinetic mechanism of
472 *Thermoanaerobium brockii* alcohol dehydrogenase: A study of discrimination
473 between alternative kinetic models. J. Biotechnol. **46**:23–31.

474 32. **Lamed R**, **Zeikus JG**. 1980. Ethanol production by thermophilic bacteria:
475 relationship between fermentation product yields of and catabolic enzyme
476 activities in *Clostridium thermocellum* and *Thermoanaerobium brockii*. J.
477 Bacteriol. **144**:569–578.

478 33. **Inokuma K**, **Nakashimada Y**, **Akahoshi T**, **Nishio N**. 2007. Characterization of
479 enzymes involved in the ethanol production of *Moorella* sp. HUC22-1. Arch.
480 Microbiol. **188**:37–45.

481 34. **Yao S**, **Mikkelsen MJ**. 2010. Identification and overexpression of a bifunctional
482 aldehyde/alcohol dehydrogenase responsible for ethanol production in

Thermoanaerobacter mathranii. J. Mol. Microbiol. Biotechnol. **19**:123–133.

35. **Montella C**, **Bellsolell L**, **Perez-Luque R**, **Badia J**, **Baldoma L**, **Coll M**, **Aguilar J**. 2005. Crystal structure of an iron-dependent group III dehydrogenase that interconverts L-lactaldehyde and L-1,2-propanediol in *Escherichia coli*. J. Bacteriol. **187**:4957–4966.

36. **Extance J**, **Crennell SJ**, **Eley K**, **Cripps R**, **Hough DW**, **Danson MJ**. 2013. Structure of a bifunctional alcohol dehydrogenase involved in bioethanol generation in *Geobacillus thermoglucosidasius*. Acta Crystallogr. Sect. D Biol. Crystallogr. **69**:2104–2115.

37. **Marçal D**, **Rêgo AT**, **Carrondo MA**, **Enguita FJ**. 2009. 1,3-Propanediol dehydrogenase from *Klebsiella pneumoniae*: decameric quaternary structure and possible subunit cooperativity. J. Bacteriol. **191**:1143–1151.

38. **Jarboe LR**. 2011. YqhD: A broad-substrate range aldehyde reductase with various applications in production of biorenewable fuels and chemicals. Appl. Microbiol. Biotechnol. **89**:249–257.

39. **Holland-Staley CA**, **Lee K**, **Clark DP**, **Cunningham PR**. 2000. Aerobic activity of *Escherichia coli* alcohol dehydrogenase is determined by a single amino acid. J. Bacteriol. **182**:6049–54.

40. **Lo J**, **Zheng T**, **Olson DG**, **Ruppertsberger N**, **Tripathi SA**, **Guss AM**, **Lynd LR**. 2015. Deletion of *nfnAB* in *Thermoanaerobacterium saccharolyticum* and its effect on metabolism. J. Bacteriol. **197**:2920–2929.

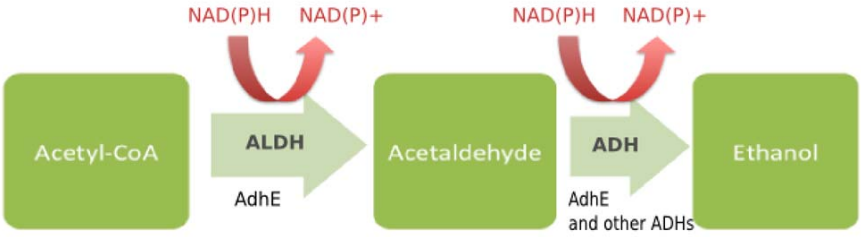
41. **Shaw AJ**, **Jenney FE**, **Adams MWW**, **Lynd LR**. 2008. End-product pathways in the xylose fermenting bacterium, *Thermoanaerobacterium saccharolyticum*. Enzyme Microb. Technol. **42**:453–458.

42. **Ma K**, **Adams MWW**. 1999. An unusual oxygen-sensitive , iron- and zinc-containing alcohol dehydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus*. J. Bacteriol. **181**:1163–1170.

43. **Blombach B**, **Riester T**, **Wieschalka S**, **Ziert C**, **Youn JW**, **Wendisch VF**, **Eikmanns BJ**. 2011. *Corynebacterium glutamicum* tailored for efficient isobutanol production. Appl. Environ. Microbiol. **77**:3300–3310.

44. **Nakamura CE**, **Whited GM**. 2003. Metabolic engineering for the microbial production of 1,3-propanediol. Curr. Opin. Biotechnol. **14**:454–459.

45. **Mai V**, **Lorenz WW**, **Wiegel J**. 1997. Transformation of *Thermoanaerobacterium* sp. strain JW/SL-YS485 with plasmid pIKM1 conferring kanamycin resistance. FEMS Microbiol. Lett. **148**:163–167.



520

521 **Figure 1** Enzymes and cofactors involved in the *T. saccharolyticum* acetyl-CoA to

522 ethanol pathway. Enzymatic reactions and enzymes are in black and cofactors are in red.

523 Note that in the ethanol production pathway, the physiological direction of the ALDH

524 reaction (acetaldehyde dehydrogenase, EC 1.2.1.10) is acetyl-CoA reducing. However,

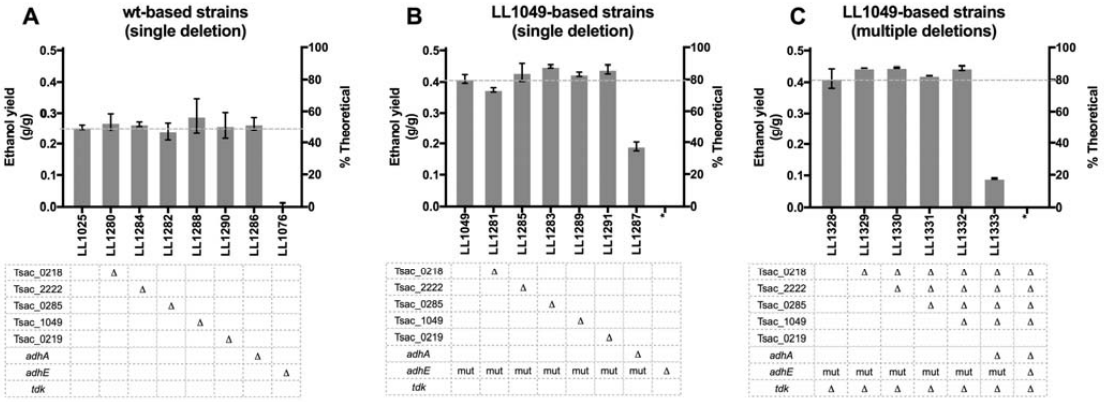
525 we continue to use the abbreviation ALDH to describe the reaction due to its familiarity.

526 Similarly, the physiological direction of the ADH (alcohol dehydrogenase, EC 1.1.1.1) is

527 acetaldehyde reducing.

528

529



530

531 **Figure 2** (A) and (B) show ethanol yields of single-gene deletion strains. (C) shows
532 ethanol yields of multiple gene deletions. Ethanol yield is reported in units of grams
533 ethanol per gram cellobiose consumed (g/g), and the theoretical maximum yield of
534 ethanol from cellobiose is 0.51 g/g. The asterisk * in (B) and (C) indicates *adhE* could
535 not be deleted in strain LL1049, the high-ethanol strain of *T. saccharolyticum*. The dotted
536 lines in the graphs indicate the level of ethanol yield in parent strains (first column of
537 each panel). The genotype of each strain is shown in the table below each graph, where
538 blanks indicate the wt alleles, Δ indicates disruption of the gene either by replacement
539 with a *kan* marker or a markerless deletion, and "mut" refers to the G544D mutation in
540 AdhE that occurred in strain LL1049 (see reference (12)). Full genotypes of each strain
541 are listed in Table 2.

542

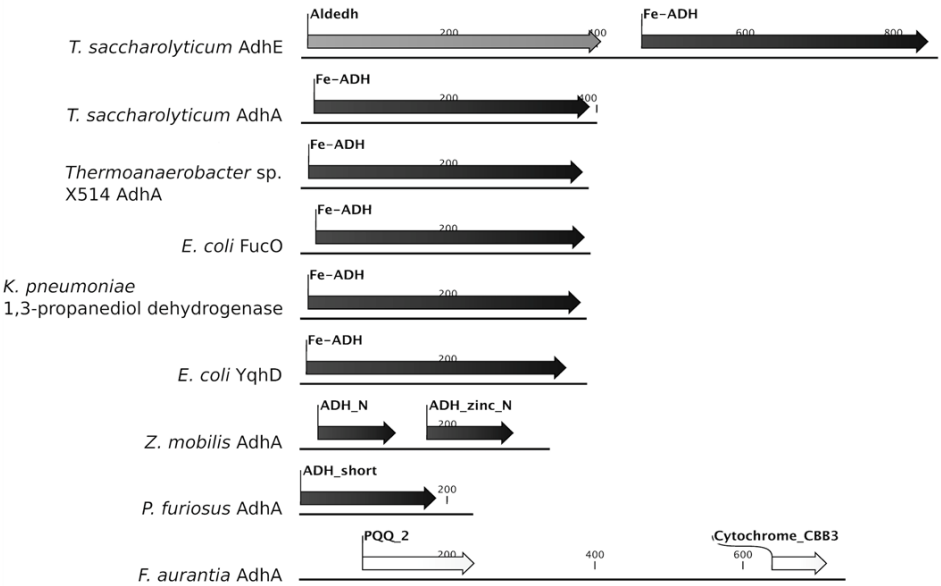


Figure 3 Sequence comparison of proteins that share homology with *T. saccharolyticum* AdhA and other AdhA proteins in the literature. The black horizontal lines indicate the amino acid sequence of each protein, on which the sequence length is marked in 200 a.a. units. The length of the *T. saccharolyticum* AdhA is 400 a.a.; the *T. saccharolyticum* AdhE is 860 a.a. with the Fe-ADH domain from 460-847. Each sequence is annotated to-scale with the appropriate Pfam protein domain/family. ADH-like domains are color-coded in black, ALDH domains in grey and other domains in white. Fe-ADH: iron-containing alcohol dehydrogenase, ADH_N: alcohol dehydrogenase GroES-like domain, ADH_zinc_N: zinc-binding dehydrogenase domain, adh_short: short-chain dehydrogenase family, PQQ_2: PQQ-like domain, Cytochrome_CBB3: Cytochrome C oxidase, cbb3-type, subunit III.

556 **TABLE 1** Description of deleted genes

Gene locus	Gene name	Seq ID [a]	Length (a.a.)	Gene product description
Tsac_0218	/	I3VRV9	404	Alcohol dehydrogenase zinc-binding domain protein
Tsac_2222	/	I3VXI1	373	Iron-containing alcohol dehydrogenase
Tsac_0285	/	I3VS20	390	Iron-containing alcohol dehydrogenase
Tsac_1049	/	I3VU69	394	Iron-containing alcohol dehydrogenase
Tsac_2087	*adhA*	I3VX46	400	Iron-containing alcohol dehydrogenase
Tsac_0416	*adhE*	I3VSF1	860	Aldehyde-alcohol dehydrogenase
Tsac_0219	/	I3VRW0	468	Aldehyde dehydrogenase

557

558 [a] Seq ID refers to protein sequence ID in Pfam.

559 **TABLE 2** Strains used in this study

Strain name	Description	cloning plasmids [a]	Source or reference
LL1025	Wild-type *T. saccharolyticum* strain JW/SL-YS485	/	Mai et al. (45) Genbank No. CP003184
LL1049	Evolved *T. saccharolyticum* Δ(*pta-ack*) Δ*ldh* Δ*or796 ure metE* Δ*eps* strain with mutation G544D in AdhE, high-ethanol-producer.	/	a.k.a. M1442. Herring et al. (4) Genbank No. SRA233073
LL1076	Δ*adhE*::(*pta-ack kan*)	/	Mascoma Corp. Genbank No. SRS731880
LL1280	Wild-type *T. saccharolyticum* Δtsac_0218::*kan* strain	pTZvec001	this study
LL1281	LL1049 *T. saccharolyticum* Δtsac_0218::*kan* strain	pTZvec001	this study
LL1282	Wild-type *T. saccharolyticum* Δtsac_0285::*kan* strain	pTZvec002	this study
LL1283	LL1049 *T. saccharolyticum* Δtsac_0285::*kan* strain	pTZvec002	this study
LL1284	Wild-type *T. saccharolyticum* Δtsac_2222::*kan* strain	pTZvec003	this study
LL1285	LL1049 *T. saccharolyticum* Δtsac_2222::*kan* strain	pTZvec003	this study
LL1286	Wild-type *T. saccharolyticum* Δ*adhA*::*kan* strain	pTZvec004	this study
LL1287	LL1049 *T. saccharolyticum* Δ*adhA*::*kan* strain	pTZvec004	this study
LL1288	Wild-type *T. saccharolyticum* Δtsac_1049::*kan* strain	pTZvec005	this study
LL1289	LL1049 *T. saccharolyticum* Δtsac_1049::*kan* strain	pTZvec005	this study
LL1290	Wild-type *T. saccharolyticum* Δtsac_0219::*kan* strain	pTZvec006	this study
LL1291	LL1049 *T. saccharolyticum* Δtsac_0219::*kan* strain	pTZvec006	this study
LL1328	LL1049 *T. saccharolyticum* Δ*tdk* strain	pTZvec007	this study

LL1329	LL1049 *T. saccharolyticum* Δ*tdk*Δtsac_0218 strain	pTZvec008, pTZvec009	this study
LL1330	LL1049 *T. saccharolyticum* Δ*tdk*Δtsac_0218Δtsac_2222 strain	pTZvec008, pTZvec009, pTZ eco012	this study
LL1331	LL1049 *T. saccharolyticum* Δ*tdk*Δtsac_0218Δtsac_2222Δtsac_0285 strain	pTZvec008, pTZvec009, pTZ eco012, pTZeco014	this study
LL1332	LL1049 *T. saccharolyticum* Δ*tdk*Δtsac_0218Δtsac_2222Δtsac_0285Δtsac_1049 strain	pTZvec008, pTZvec009, pTZ eco012, pTZeco014, pTZeco013	this study
LL1333	LL1049 *T. saccharolyticum* Δ*tdk*Δtsac_0218Δtsac_2222Δtsac_0285Δtsac_1049Δ*adhA*::*kan* strain	pTZvec008, pTZvec009, pTZ eco012, pTZeco014, pTZeco013, pTZvec004	this study
LL1334	*T. saccharolyticum* Δ*adhE* strain	pTZvec010	this study
LL1335	*T. saccharolyticum* Δ*adhE*Δ*adhA*::*kan* strain	pTZvec010, pTZvec004	this study
LL1402	LL1049 *T. saccharolyticum* Δ*tdk*Δtsac_0218Δtsac_2222Δtsac_0285Δtsac_1049Δ*adhA*::*adhA*$^{+}$*erm* strain	pTZvec008, pTZvec009, pTZ eco012, pTZeco014, pTZeco013, pTZvec004, pTZvec011	this study
LL1403	*T. saccharolyticum* Δ*adhE*::*adhE*$^{+}$*kan* colony 8	pTZvec010, pTZvec012	this study
LL1404	*T. saccharolyticum* Δ*adhE*::*adhE*$^{+}$*kan* colony 1	pTZvec010, pTZvec012	this study
LL1405	*T. saccharolyticum* Δ*adhE*::*adhE*$^{+}$*kan* colony 3	pTZvec010, pTZvec012	this study

LL1406	*T. saccharolyticum* Δ*adhE*::*adhE*$^{+}$*kan* colony 5	pTZvec010, pTZvec012	this study
LL1407	*T. saccharolyticum* Δ*adhE*::*adhE*$^{+}$*kan* colony 7	pTZvec010, pTZvec012	this study

[a] plasmid information included in Table S1, plasmid sequences can be found in Genbank 1955854 (Submission ID); primers used for plasmid construction included in Table S2.

564 **TABLE 3** ADH activity of Δ*adhE* and Δ*adhA* strains

		ADH activity (U/mg)			
		NADH		NADPH	
LL1025	wild-type	2.64	± 0.35[a]	0.83	± 0.08
LL1334	Δ*adhE*	0.06	± 0.01	0.64	± 0.21
LL1286	Δ*adhA*	2.19	± 0.49	0.06	± 0.01
LL1335	Δ*adhE* Δ*adhA*	0.01	± 0.01	0.06	± 0.06

565

566 [a] The average of two biological replicates is shown with their standard deviations.

567