UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

Phenotypic, Chemotaxonomic and Genomic Characterization of a Novel Anaerobic Acetogen, *Clostridium muelleri* sp. nov. P21^{T} Isolated from Old Hay

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By
DENA A. DOYLE
Norman, Oklahoma
2020

Phenotypic, Chemotaxonomic and Genomic Characterization of a Novel Anaerobic Acetogen, *Clostridium muelleri* sp. nov. P21^{T} Isolated from Old Hay

A THESIS APPROVED FOR THE

DEPARTMENT OF MICROBIOLOGY AND PLANT BIOLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Paul A. Lawson, Chair

Dr. Ralph S. Tanner

Dr. Elizabeth A. Karr

I dedicate this thesis to my dad who always believed in me and taught me what never giving up truly means. I miss you every day.

Acknowledgements

I would like to thank my committee members Dr. Paul A. Lawson, Dr. Ralph S. Tanner and Dr. Elizabeth A. Karr for their support and guidance. In particular, thank you Dr. Tanner, for allowing me to take on this project and for always believing in me and encouraging me. You pushed me to be a better scientist and person and for that I am forever grateful. Dr. Karr and Dr. Lawson, thank you for helping me with professional advice and for always encouraging me in my endeavors. Thank you to Dr. Kathleen Duncan for performing the DNA extraction and PCR on this organism and helping me get started with the genome portion of my work. Thank you to Dr. Hasan K. Atiyeh at Oklahoma State University for analyzing the end products of metabolism using GC. Thank you, Dr. Kara De Leon, for all of your help and guidance on figuring out Kbase and for showing me how to use Geneious for cleaning up my kmer content. I would also like to thank Neil Wofford for teaching me how to repair and build just about anything involving an anaerobic chamber and gassing manifold. To my former lab mate, Dr. Nisha Patel, thank you for your constant support whether it was teaching me new lab skills via skype, helping me get through programming errors or offering professional advice before meetings. To my 8th floor lab mates Shannon Fulton, Emily Junkins and James Floyd thank you for endlessly supporting me, helping me think through experiments and encouraging me to push forward during tough times. To my undergraduate Parker Smith, I am so proud of you and it has been incredibly rewarding watching you grow in the lab and as a person. Lastly, I would like to give a huge thank you to all of the microbiology faculty and my fellow microbiology and plant biology graduate students for all of the love and support you have shown me these past two years. I could not have done this without all of you!

Table of Contents

List of Tables

List of Figures

Abstract

Acetogens are a group of obligately anaerobic bacteria that use the Wood-Ljungdahl (acetyl-CoA) pathway for the production of acetyl-CoA from 2 moles of CO_2. Acetyl-CoA is then converted to various acids and alcohols such as ethanol, butanol, hexanol, acetic, butyric and caproic acids using CO, CO_2 and H_2 or carbohydrates as the energy source. This particular metabolic process is key for the use of acetogens in syngas fermentation. Biofuels have been used since the late 19th century with fluctuations in its use due to the efficiency and cost-effectiveness of non-renewable energy sources. In recent years, the world has seen significant traction in decreasing the global carbon footprint by increasing the use of biofuels generated from renewable energy sources. A novel, Gram-positive staining, solvent-and-acid producing, strictly anaerobic rod belonging to the genus *Clostridium*, designated P21^T, was recovered from an enrichment selecting for acetogens from a sample of old hay collected in Stillwater, Oklahoma. Genomic analysis confirmed the laboratory observations that strain P21^T contains the genes encoding for the Wood-Ljungdahl pathway for acetogenesis as well as the genes required for acid and alcohol production. Additional phenotypic and chemotaxonomic characteristics of strain P21^T showed consistencies with previous observations seen from its nearest neighbors *Clostridium carboxidivorans* P7^T, *Clostridium drakei* SL1^T and *Clostridium scatologenes* ATCC 25775^T. The physiological and genomic analysis of strain P21^T will be used to examine its future potential use for industrial purposes.

Chapter 1: Introduction to Acetogens

Historical Perspective for Studying Acetogens

The term 'acetogenesis' was coined in 1932 when Franz Fischer and colleagues discovered the metabolism of two CO_2 molecules to acetate occurring in sewage samples (1). Since then, over 100 different species of acetogens have been described from a wide array of ecosystems including soil, sediment, rumen fluid, sewage, oil field and water sources (2). In 1936, Klaas T. Wieringa isolated the first acetogen *Clostridium aceticum*, which was lost until its re-discovery in 1980 (3)(4)(5). In 1942 Francis E. Fontaine isolated the second known acetogen, *Clostridium thermoaceticum*, from horse manure (6). Consequently, *C. thermoaceticum* became the most studied acetogen and led scientists to the discovery of what we know today as the Wood-Ljungdahl pathway (also known as the acetyl-CoA pathway). Unlike other CO_2 fixing pathways, the Wood-Ljungdahl pathway, discovered by Harland G. Wood and Lars G. Ljungdahl, is a reductive, single carbon process where the carbon molecule from one mole of CO_2 is reduced to a methyl group and the carbon molecule from the other mole of CO_2 is reduced to a carbonyl group. Together, these two groups form acetyl-CoA which is further converted to acetate. Given what scientists know of the composition of early life, it is speculated that this process was the first form of metabolism on earth (7).

Diversity and Taxonomy

Any organism that produces acetic acid should not automatically be considered an acetogen. An acetogen, at its core definition, is an anaerobic microorganism that conserves energy using the Wood-Ljungdahl pathway for the conversion of two CO_2 molecules to acetyl-CoA (7). Acetogens are found throughout numerous genera which include *Acetoanaerobium*, *Acetobacterium*, *Acetohalobium*, *Acetonema*, *Bryantella*, *Caloramator*, *Clostridium*, *Eubacterium*, *Holophaga*, *Moorella*, *Natroniella*, *Natronincola*, *Oxobacter*, *Ruminococcus*, *Sporomusa*, *Syntrophococcus*, *Thermacetogenium*, *Thermoanaerobacter*, *Treponema* and numerous as-yet unclassified (2). The most predominant genera in which acetogens can be found are *Acetobacterium* and *Clostridium* (8). Throughout these acetogens, there is substantial variation in the phenotypic and genotypic features including Gram-staining negative and Gram-staining positive bacteria, cell morphologies varying from rods to cocci, and a broad range in G+C content (~20-65%) (2). As scientists have continued to isolate novel microorganisms, taxonomic shifts directly impacted the classification of acetogens. In 1994, M. D. Collins and co-workers recognized the huge diversity in members of the genus *Clostridium*, resulting in the taxonomic revision of this group of organisms that also included the reclassification of *C. thermoaceticum* to *Moorella thermoacetica* (9). This reconstruction divided the genus *Clostridium* into 19 clusters with *M. thermoacetica* placed in cluster VI: I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII, XIII, XIVa, XIVb, XV, XVI, XVII, XVIII and XIX (9). Phylogenetically, acetogens cannot be distinguished between other organisms due to the number of vastly different genera and phenotypic expressions of these microorganisms. However, acetogens themselves are classified based on which enzyme system they use for energy conservation: “Rnf-dependent” or “Ech-dependent” (10). Rnf-dependent acetogens

transfer sodium or hydrogen ions across the cytoplasmic membrane and simultaneously transfer the electrons from reduced ferredoxin to NAD+ using a ferredoxin: NAD^+ oxidoreductase. Ech-dependent acetogens lack the genes necessary for an Rnf-complex and instead contain an energy conserving hydrogenase that translocate its ions, hence Ech (10).

Applications and Use of Acetogens

Tackling climate change is more urgent than ever due to the approaching threshold of irreversible damage to the global ecosystem (11). In the 1920's, the design of combustion engines was modified for the use of petroleum instead of bioethanol, as petroleum became readily available and was a more efficient fuel source (12). The 2019 United Nations Climate Action Summit priorities included decreasing CO_2 emissions by nearly 50% in the hopes of reaching a net zero on CO_2 emissions in the next 30 years. It also mentioned rapidly accelerating the use of 100% renewable energy sources (13). So how can acetogens be used to help in the times of a climate crisis? Syngas is a mixture of CO, H_2 and CO_2 that is a byproduct of industrial flue gases as well as the gasification of various organic matter such as biomass and municipal waste (14). The syngas is injected into a bioreactor where acetogens are able to convert this gas mixture into alcohol and organic acids such as ethanol, butanol, hexanol, acetic acid, butyric acid and caproic acid. These alcohols are further distilled into bio-alcohols and the organic acids are extracted for chemical manufacturing. In the late 1980's to early 1990's, optimization of bioreactors required for large-scale syngas fermentation paved the way for industrialization of biological biofuels (15).

One of the biggest advantages of syngas fermentation is the utilization of all components in biomass, including lignin. Other methods for producing bioethanol from

biomass including consolidated bioprocessing (CBP), separate hydrolysis and fermentation (SHF), simultaneous saccharification and fermentation (SSF), and simultaneous saccharification and co-fermentation (SSCF) are unable to break down the lignin in biomass during fermentation (14). As commercial companies continue to refine the process and reduce production cost, the use of acetogens for the production of biofuels in syngas fermentation becomes a more viable alternative to non-renewable energy sources.

One organic acid naturally produced from a few bacteria is n-caproic acid. Caproic acid is primarily used as a flavor additive in food and animal feed by chemical companies and is currently being investigated for its use as a potential precursor in biofuel production (16). Caproic acid production is the result of the reverse beta-oxidation of acetic acid or butyric acid. Alcohol and acid producing bacteria, particularly *Clostridium kluyveri*, have been shown to naturally produce n-caproic acid (16). What is particularly fascinating about caproic acid is the fact that there is four times more energy (~3,452.0 kJ) in one mole of caproic acid than there is in one mole of acetic acid (~840.0 kJ) (16). Caproic acid itself is not used as a biofuel but can be converted to its alcohol form, hexanol, and be used as a fuel source. In these unprecedented times of what will happen to the world if global warming continues at its current pace, harnessing the power of bioavailable resources, including acetogens, brings forward the promising uses of bioenergy.

Chapter 2: Recovery and Characterization of *Clostridium muelleri* sp. nov. P21^T from Old Hay

Abstract

Characterizing microorganisms utilizes a polyphasic approach that requires the examination of physiological, chemotaxonomic and genomic characteristics. A novel species belonging to the genus *Clostridium*, designated as strain P21^T, was isolated from old hay collected in Stillwater, Oklahoma. Identification using 16S rRNA gene sequencing revealed P21^T was phylogenetically related to the acetogen *Clostridium carboxidivorans* P7^T with 97.8% sequence similarity, demonstrating it represented a novel species for which the name *Clostridium muelleri* sp. nov. is proposed. Other closely related phylogenetic species in this clade include *Clostridium carboxidivorans* P7^T, *Clostridium drakei* SL1^T and *Clostridium scatologenes* ATCC 25775^T. Analysis of end products of metabolism using gas chromatography were performed to determine how much ethanol, butanol, hexanol, acetate, butyrate and caproic acid was produced compared to its nearest neighbor, *C. carboxidivorans* P7^T. Observations showed P21^T showed ten times the amount of acetic acid produced compared to *C. carboxidivorans* P7^T and a significant production of caproic acid, which can further be investigated for industrial uses.

Introduction

As discussed in chapter 1, acetogens are found in numerous environments, most of which are various types of soils and sediments (2). Organisms within a community maintain essential roles for the regulation of homeostasis in their ecosystem. Soil and environmental based acetogens contribute significantly to global carbon cycling, generating more than 10^{13} kg of acetic acid in a year (17). A goal of this study was to characterize *Clostridium muelleri* sp. nov. strain P21^T, a novel acetogen belonging to the genus *Clostridium*, using a polyphasic approach. The characterization of a novel organism involves the examination of phenotypic, phylogenetic and chemotaxonomic characteristics between the novel bacterium and its nearest neighbors, in addition to genotypic characteristics which will be discussed in chapter 3 (18). It is important to remember, not all methods involved in characterizing a novel species are relevant to all bacteria. Identifying the 16S rRNA pairwise similarity between a novel microorganism and its nearest neighbor is the first step in classifying its relationship to other bacteria. The use of the 16S rRNA gene as a molecular marker to track the evolution of microorganisms over time was first used by Carl Woese and colleagues in 1977, which led to the discovery of the third domain of cellular life, Archaea (19). The 16S rRNA gene contains highly conserved regions shared among all microorganisms as well as hypervariable regions which can vary tremendously across different microorganisms. It is these hypervariable regions that allow the differentiation of microbial taxa between bacteria (20). Fatty acid analysis of strain P21^T and its nearest neighbors *C. carboxidivorans* P7^T, *C. drakei* SL1^T and *C. scatologenes* ATCC 25775^T, whose fatty acid profiles had not been previously determined, revealed consistent major fatty acids of $C_{16:0}$ and $C_{16:1\ \omega7c}$/$C_{16:1\ \omega6c}$ between all four bacteria

in this clade. Metabolically significant findings of strain P21^{T} include production of acetic acid, butyric acid, caproic acid, ethanol, butanol and hexanol.

Materials & Methods

Old Hay Collection & Enrichment Composition

Samples were collected from old hay from the Oklahoma Agricultural Experiment Station at Oklahoma State University, Stillwater, Oklahoma, USA in 1998 by Dr. Ralph Tanner from the University of Oklahoma, Norman, Oklahoma (36° 7′ 5″ N, 97° 5′ 15″ W). The old hay was incubated anaerobically with ATCC medium no. 1754 (21) at 38°C with an initial pH of 5 and an atmosphere of $CO:CO_2:H_2$ (70:24:6) at a pressure of 200 kPa. The enrichment was monitored using gas chromatography for the production of alcohols and acids, particularly ethanol and acetic acid (22). Strain $P21^T$ was isolated from this enrichment using the roll tube method (23). The medium, which will be referred to as P21 medium, used for maintaining a continuous culture of $P21^T$ contained (L^{-1}): 25 ml minerals solution (24); 10 ml vitamins solution (24); 10 ml trace metals solution (24); 1.0 g BD Bacto™ yeast extract (Sparks, MD, USA, cat no. 212750); 5.0 g glucose; 10.0 g 2-(N-morpholino)ethanesulfonic acid hemisodium salt (MES), pH set to 6.3 and 7 ml of cysteine sulfide reducing agent added (24). The medium was prepared using a strictly anaerobic modified Hungate technique (25).

DNA Isolation and 16S rRNA Gene Sequencing

DNA was extracted from strain $P21^T$ using the automated Maxwell ®16 Tissue LEV Total RNA Purification Kit (Promega Corp) purification system version 4.90 modified from the manufacturer's instruction as described in Oldham et al. (26). PCR amplification of the regions V1-V9 16S rRNA gene using primers fD1 (5' AGAGTTTGATCCTGGCTCAG 3') and rP2 (5' ACGGCTACCTTGTTACGACTT 3') was performed as described by Weisburg and colleagues (27). The product was then purified using Exo-SapIT (USB Corporation) and

sent to OU's Biology Core Laboratory for sequencing using Big-Dye Terminator v3.1 (3100 Avant model, Applied Biosystems).

Phylogenetic Analysis of strain P21^T via 16S rRNA Gene Sequencing

Sequences obtained were concatenated and assembled to produce a coverage of 1,421 bp using Sequencher 5.4 (Gene Code Corporation). The phylogenetic position of strain P21^T was determined by performing a BLAST search (28). The 16S rRNA gene sequences of the nearest neighbors were downloaded from EZBiocloud (29), aligned using the MUSCLE algorithm and trimmed in MEGA (30). A neighbor-joining phylogenetic tree using Kimura 2-parameter substitution model with 1000 bootstrap replicates was built using MEGA (30)(31)(32)(33). A maximum-likelihood phylogenetic tree using Tamura-Nei substitution model with 1000 bootstrap replicates was built using MEGA to compare and confirm similar phylogenetic distances of strain P21^T to others in the genus *Clostridium* (33)(34)(30)(35).

Morphological, Physiological and Biochemical Characterization

Cells were inspected for motility, endospore production and Gram stain reaction using an Olympus CH2 phase microscope. The pH range of strain P21^T was tested from pH 4-8 using fructose as the substrate and a Good's buffer (20 g l^{-1}): HOMOPIPES (pK_a 4.6); MES (pK_a 6.0); TES (pK_a 7.2) or TAPS (pK_a 8.1) (22). The temperature range of strain P21^T was analyzed from 6-60°C using fructose as the substrate. Substrate utilization assays were performed in duplicate and determined by measuring the OD of broth cultures at 600 nm using a Spectronic 20D$^+$ spectrophotometer (Milton Roy, DE) (22). The biochemical tests nitrate reduction, gelatin hydrolysis, catalase, oxidase and Methyl-Red Voges-Proskauer

(MR-VP) reaction were performed using the methods found in the American Society for Microbiology's Methods for General and Molecular Microbiology (36) with the following modifications: P21 medium without glucose at a pH of 6.3 was used as the base medium for all biochemical tests with the exception of the MR-VP tests, following the Hungate technique (25). For the MR-VP tests, 5.0 g (L^{-1}) glucose was added and 5.0g (L^{-1}) dipotassium phosphate replaced MES in the P21 medium. For the nitrate reduction test, 5.0 g (L^{-1}) glucose replaced beef extract and peptone. Beef extract was not used in the gelatinase test.

End products of metabolism from fructose, CO:CO_2 and H_2:CO_2 as the substrate were measured from duplicate cultures grown in basal medium (22) at 38°C. These measurements were performed by colleagues at Oklahoma State University. Analysis was performed in duplicate from each of the duplicate cultures using gas chromatography (Agilent 6890 N GC, Agilent Technologies, Wilmington, DE, USA) with a flame ionization detector and DB-FFAP capillary column (Agilent Technologies, Wilmington, DE, USA). The carrier gas used was hydrogen with an initial flow rate of 1.9 ml per minute for three minutes with the flow rate increasing to 4 ml per minute with a ramping rate of 0.5 ml every two minutes. The inlet port temperature was maintained at 200°C with a split ratio of 50:1. The oven temperature was initially set to 40°C for 1.5 minutes and increased at a ramping rate of 40°C per minute until it reached 235°C. The flame ionization detector temperature was 250°C with the hydrogen rate as 40 ml per minute and the air flow rate as 450 ml per minute (37). Peak areas of fermentation products were compared to standards of ethanol, butanol, hexanol, acetate, butyric acid and caproic acid.

Chemotaxonomic Characterization

Cells for analyzing the fatty acid composition of P21^T, *C. carboxidivorans* P7^T, *C. scatologenes* ATCC 25775^T and *C. drakei* SL1^T were collected and spun down after a 24 h incubation at 38°C in the basal medium previously mentioned (22) with 5.0 g (L^{-1}) fructose. Fatty acid methyl esters (FAME) were extracted in duplicate using the Sherlock Microbial Identification System (MIDI) v6.1. and analyzed using an Agilent Technologies 6890N gas chromatograph equipped with a phenyl methyl silicone-fused silica capillary column (HP-2 25m x 0.2 mm x 0.33 mm film thickness) and a flame ionization detector with hydrogen as the carrier gas (38). Before analysis the GC column was burned off at a temperature of 300°C for two hours to rid of any dust particulates or residue that could interfere with peak quality. The initial program temperature was set at 170°C and increased 5°C per minute to reach a final temperature of 270°C (38). Fatty acids and acyl compounds were identified using the peak naming database QTSA1 (39)(40).

Results & Discussion

Phylogenetic Analysis of strain P21^T

The neighbor-joining phylogenetic tree (Fig. 1) of strain P21^T with its nearest relatives in the genus *Clostridium* yielded high bootstrap values (percentages seen on respective branches) for the placement of strain P21^T within this clade of *Clostridium*. The outgroup used for this tree was *Ethanoligenens harbinense*. A maximum-likelihood tree using a Tamura-Nei substitution model revealed overall identical arrangement of species placement (Fig. 2). Based on the phylogenetic distance, strain P21^T represents a novel species for which the name *Clostridium muelleri* is proposed in honor of German microbiologist Volker Müller for his contributions in understanding the physiology of acetogenic bacteria. The nucleotide sequence of the 16S rRNA gene was deposited with GenBank under the accession number MT176110.

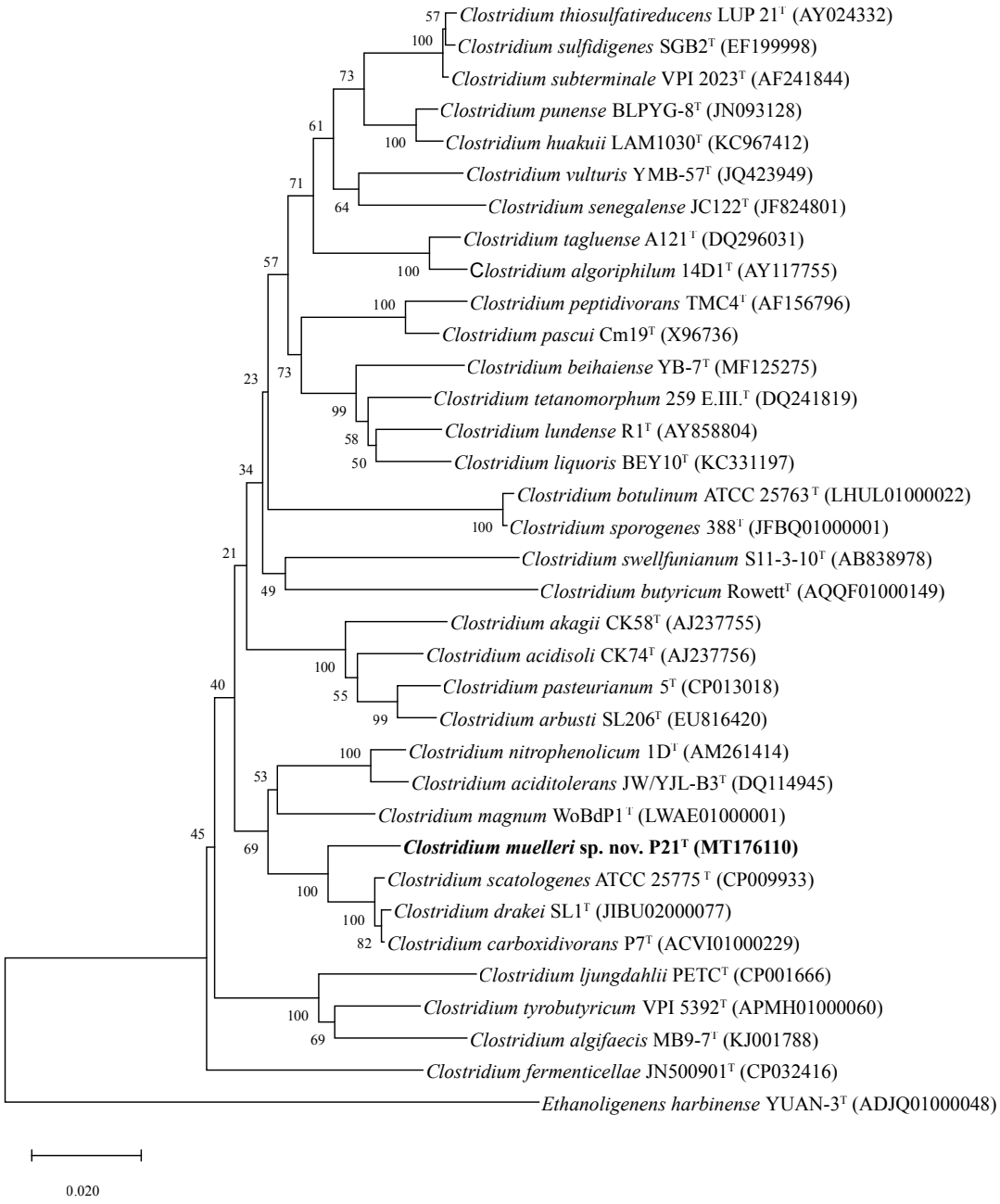

Fig. 1. Neighbor-Joining tree showing the phylogenetic analysis of *Clostridium muelleri* sp. nov. P21^{T} within the genus *Clostridium*.

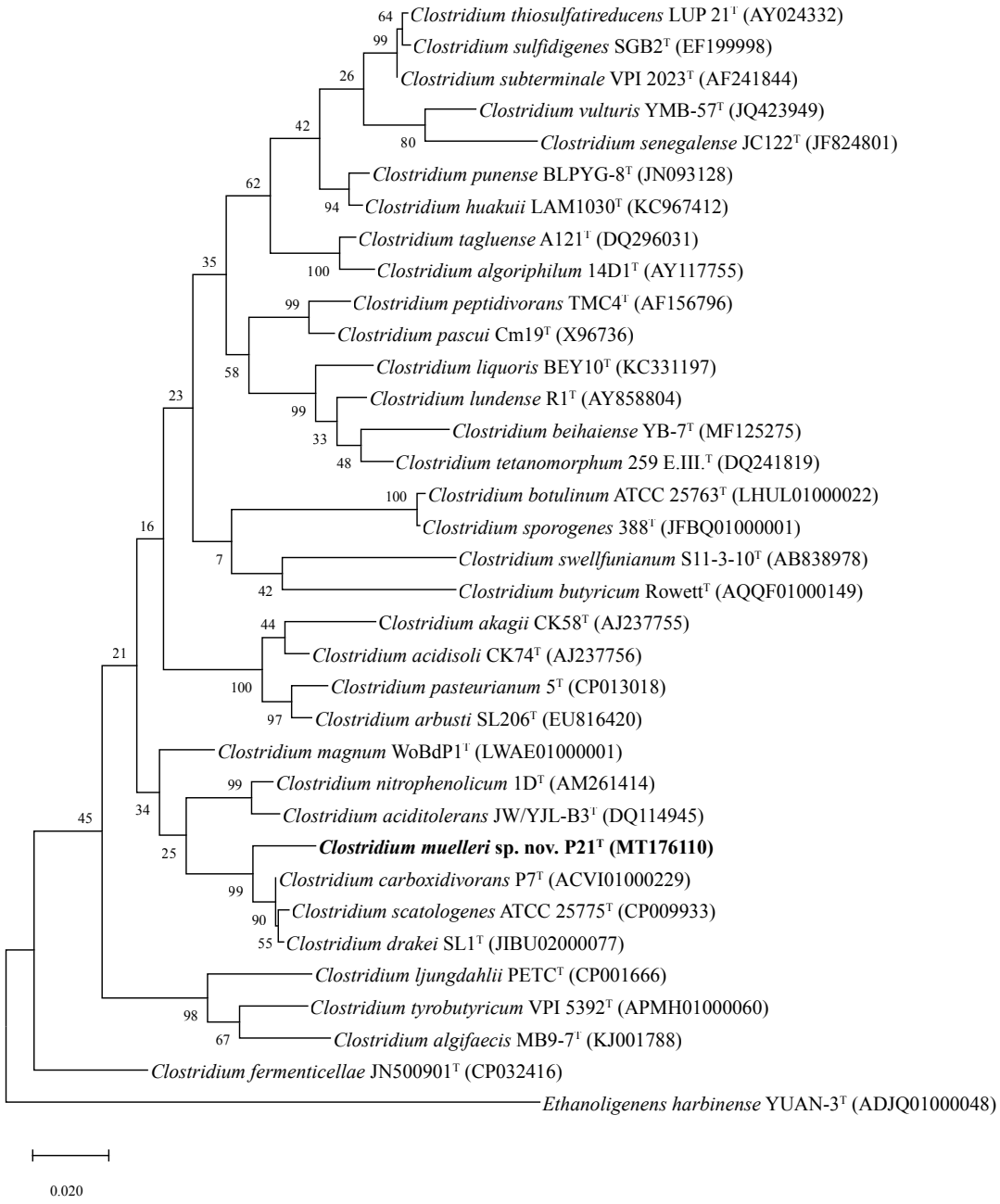

Fig. 2. Maximum-Likelihood tree showing the phylogenetic analysis of *Clostridium muelleri* sp. nov. P21^T within the genus *Clostridium*.

Morphological, Physiological and Biochemical Characterization of strain P21^T

Strain P21^T is a Gram-stain positive, non-spore forming, non-motile, strictly anaerobic rod that grows with cells arranged in singles and pairs, with chains also present, and a cell size of 0.4 x 3-6 μm. Table 1 shows substrate utilization of strain P21^T with its nearest neighbors *C. carboxidivorans* P7^T, *C. drakei* SL1^T and *C. scatologenes* ATCC 25775^T. All species grew on glucose, fructose, galactose, xylose, mannose, ribose, sucrose, cellobiose, cellulose, histidine, serine, choline, Casamino acids and glutamate. All species did not grow on raffinose, melibiose, D-arabinose, methanol, formate, amygdalin, vanillate, succinate, ferulate and trimethoxybenzoate. Compared to the nearest neighbors, strain P21^T grew on lactose and threonine. Strain P21^T did not grow on trehalose, rhamnose, melezitose, xylan, sorbitol, mannitol, propanol, butanol, ethanol, methanol, glycerol, inositol, syringate, betaine, fumarate, malate, cysteine, gluconate, asparagine, threonine, alanine, aspartate, glutamine, arginine, tryptophan, methionine, leucine, lysine, phenylalanine and valine. It should be noted the utilization results for lactate, pyruvate, L-arabinose, citrate, starch, maltose, fucose and 2-propanol will be collected and published after this thesis will be published (manuscript in preparation).

Confirmation of lactose utilization over a 15-day period is shown in Fig. 2. After inspecting the final pH for potential acid shock, an interesting question is why strain P21^T was unable to utilize propanol or ethanol unlike its nearest phylogenetic neighbors. The temperature growth range for strain P21^T is 20-45°C with an optimum temperature of 30-40°C. Strain P21^T was routinely cultured at 38°C. Growth was observed at a pH range of 5.0-8.5 with an optimum pH of 6.5.

Table 1. Substrate utilization of strain P21^T and nearest neighbors

Substrate	P21^T	*Clostridium carboxidivorans* P7^T	*Clostridium drakei* SL1^T	*Clostridium scatologenes* ATCC 25775^T
Pectin	-	+	-	+
Trehalose	-	+	-	-
Sorbitol	-	-	-	+
Mannitol	-	+	+	-
Glycerol	-	+	+	+
Gluconate	-	-	+	+
Fumarate	-	-	+	+
Malate	-	-	+	-
Ethanol	-	+	+	+
Propanol	-	+	+	+
Butanol	-	+	+	-
Syringate	-	+	+	-
Glutamine	-	-	-	+
Arginine	-	-	+	+
Aspartate	-	+	+	+
Alanine	-	+	+	+
Asparagine	-	+	+	+
Betaine	-	+	+	+
Rhamnose	-	+	+	+
Melezitose	-	+	+	+
Inositol	-	+	+	+
Lactose	+	-	-	-

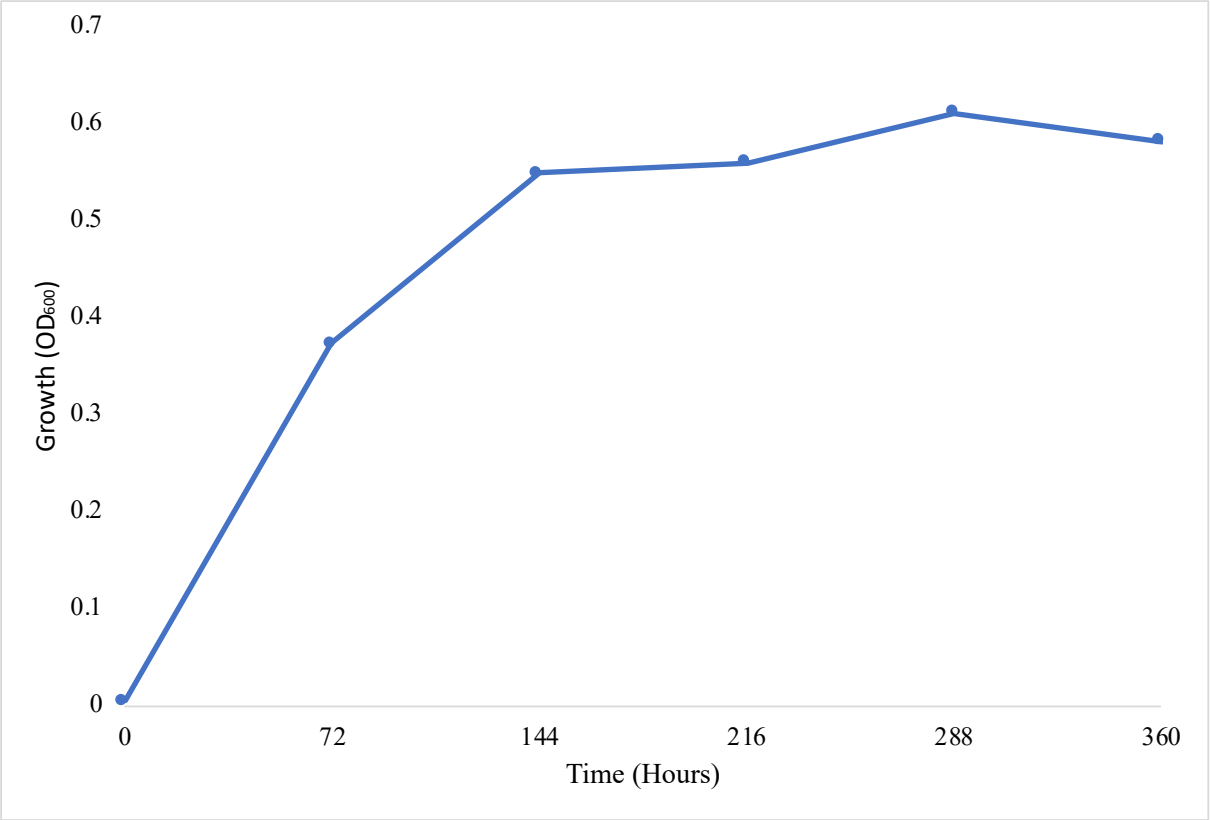

Fig. 3: The utilization of lactose by strain P21^{T} over the course of a 15-day time period.

The biochemical tests for strain P21^T are consistent with its nearest neighbors in the genus *Clostridium* (Table 2). Comparative information used for Table 2 was found in Liou et al. and Suresh et al. (22)(41). It is gelatinase negative indicating its inability to break down gelatin, nitrate reductase, catalase and oxidase negative and negative for the Voges-Proskauer test but positive for the Methyl Red test.

Acetogens can reduce acids such as acetic, butyric and caproic acid through the corresponding alcohols such as ethanol, butanol and hexanol in order to increase the pH that has been lowered due to the production of the aforementioned acids. GC analysis performed on strain P21^T and *C. carboxidivorans* P7^T grown on CO:CO_2 revealed strain P21^T produced ten times more acetic acid than *C. carboxidivorans* P7^T (Fig. 4). Unlike *C. carboxidivorans* P7^T , strain P21^T produced caproic acid, indicating strain P21^T must be performing reverse beta-oxidation, which is rare and has also been observed to occur naturally in *Clostridium kluyveri* (16). Hexanol production by *C. carboxidivorans* P7^T has been documented previously (42) which exemplifies how end product amounts can vary depending on culture conditions. Metabolic end products of strain P21^T were analyzed further under growth conditions of CO:CO_2, H_2:CO_2 and fructose. Under all three conditions, acetic, butyric and caproic acids were all produced, with acetic acid production being predominant (Table 3). The highest production of acetic acid and butyric acid occurred under CO:CO_2 conditions, while the highest caproic acid produced occurred when fructose was the growth substrate. Alcohol production under CO:CO_2, H_2:CO_2 and fructose showed ethanol is the only alcohol produced under all three conditions, with the highest amount occurring under H_2 (Table 3). Butanol was not produced under any of the conditions, while hexanol was produced under CO:CO_2 and H_2:CO_2.

Table 2. Biochemical tests of strain P21^T and nearest neighbors and *Clostridium nitrophenolicum* 1D^T.

Biochemical Tests:	P21^T	*Clostridium carboxidivorans* P7^T	*Clostridium drakei* SL1^T	*Clostridium scatologenes* ATCC 25775^T	*Clostridium nitrophenolicum* 1D^T
Nitrate Reduction	-	-	-	-	-
Gelatin Hydrolysis	-	-	-	-	+
Methyl Red	+	+	+	+	+
Voges-Proskauer	-	-	-	-	-
Catalase	-	-	-	-	-
Oxidase	-	-	-	-	-

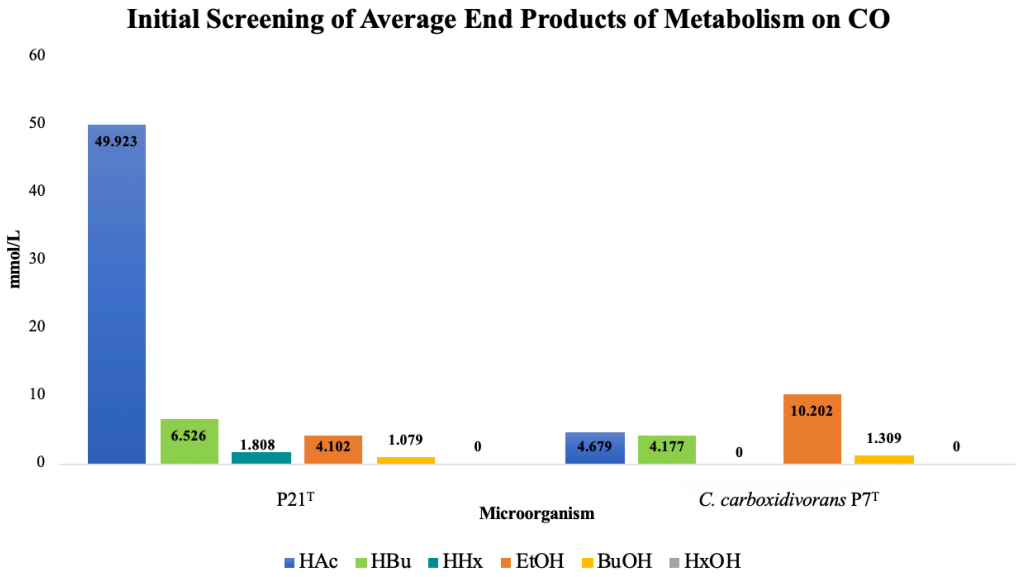

Fig. 4. Comparison of end products of metabolism on carbon monoxide for strain P21^{T} and its nearest neighbor *C. carboxidivorans* P7.

Table 3: Strain P21^T average acid and alcohol end products of metabolism in mmol/L on carbon monoxide, hydrogen and fructose.

End Product	$CO:CO_2$	$H_2:CO_2$	Fructose
Acetate	24.4	10.3	22.2
Butyrate	5.5	3.4	5.4
Caproate	2.7	2.0	3.5
Ethanol	3.1	11.3	0.7
Butanol	0.0	0.0	0.0
Hexanol	0.3	1.5	0.0

Chemotaxonomic Characterization

Fatty acid profiles were previously lacking in the literature for *C. carboxidivorans* P7^T, *C. drakei* SL1^T and *C. scatologenes* ATCC 25775^T. The fatty acid profiles for strain P21^T, nearest neighbors and *C. nitrophenolicum* is shown in Table 3 with the predominant products in bold and values less than 1% are not reported. Comparative values for *C. nitrophenolicum* used in table 3 was found in Suresh et al. (41). Major fatty acids (>10%) for strain P21^T included $C_{16:0}$ (41.5), $C_{16:1\ \omega7c}$/$C_{16:1\ \omega6c}$ (10.0) and cyclo $C_{17:0}$/$C_{18:0}$ (17.3). *C. scatologenes* ATCC 25775^T, *C. drakei* SL1^T and *C. carboxidivorans* P7^T had similar major fatty acid profiles with *C. carboxidivorans* P7^T having $C_{16:0}$ (51.0) and $C_{16:1\ \omega7c}$/$C_{16:1\ \omega6c}$ (21.8). *C. scatologenes* ATCC 25775^T having $C_{16:0}$ (39.9) and $C_{16:1\ \omega7c}$/$C_{16:1\ \omega6c}$ (23.4) and *C. drakei* SL1^T having $C_{16:0}$ (40.4) and $C_{16:1\ \omega7c}$/$C_{16:1\ \omega6c}$ (25.0). Minor fatty acids for strain P21^T are $C_{13:1}$ (4.1), $C_{14:0}$ (3.5), unknown $C_{14:969}$ (8.8), $C_{15:1\ \omega8c}$ (4.4), $C_{17:0}$ (6.0) and $C_{18:3\ \omega6c}$ (4.5). *C. carboxidivorans* P7^T showed similar minor fatty acids with $C_{14:0}$ (2.8), unknown $C_{14:969}$ (3.8), $C_{15:1\ \omega8c}$ (5.3), $C_{17:0}$ (2.3), unknown $C_{17:360}$ (1.6) and $C_{18:3\ \omega6c}$ (1.5). The main difference between the minor fatty acid profiles for all four of these bacteria are the presence or absence of $C_{13:1}$, unknown $C_{17:360}$ and $C_{18:3\ \omega6c}$. Unlike the other bacteria in this group, *C. scatologenes* ATCC 25775^T contained all minor fatty acids with $C_{13:1}$ (2.3), $C_{14:0}$ (5.2), unknown $C_{14:969}$ (8.9), $C_{15:1\ \omega8c}$ (7.8), $C_{17:0}$ (2.5), cyclo $C_{17:0}$/$C_{18:0}$ (7.4), unknown $C_{17:360}$ (1.6) and $C_{18:3\ \omega6c}$ (1.1). Lastly, the minor fatty acids for *C. drakei* SL1^T are $C_{13:1}$ (1.9), $C_{14:0}$

(3.1), unknown $C_{14:969}$ (4.8), $C_{15:1\ \omega 8c}$ (9.7), $C_{17:0}$ (3.9), cyclo $C_{17:0}$/$C_{18:0}$ (9.6) and unknown $C_{17:360}$ (1.5).

Table 4. Fatty acid profiles of strain P21^T and nearest neighbors.
[a]Predominant products in bold.

Fatty Acid[a]	P21^T	*Clostridium carboxidivorans* P7^T	*Clostridium scatalogenes* ATCC 25775^T	*Clostridium drakei* SL1^T	*Clostridium nitrophenolicum* 1D^T
$C_{13:1}$	4.09	-	2.26	1.92	-
$C_{14:0}$	3.51	2.84	5.23	3.13	**10.0**
Unknown $C_{14:969}$	8.76	3.75	8.94	4.8	-
$C_{15:1\ \omega8c}$	4.35	5.32	7.76	9.73	-
$C_{16:0}$	**41.54**	**51.03**	**39.89**	**40.42**	**28.0**
$C_{16:1\ \omega7c}$/ iso-$C_{15:0}$ 2-OH	-	-	-	-	**10.8**
$C_{16:1\ \omega7c}$/ $C_{16:1\ \omega6c}$	**10.01**	**21.77**	**23.37**	**25.03**	-
$C_{17:0}$	6.01	2.29	2.49	3.87	-
Cyclo$C_{17:0}$/ $C_{18:0}$	**17.26**	**9.82**	7.41	9.57	-
iso-$C_{17:1}$I/ anteiso B	-	-	-	-	**23.1**
Unknown $C_{17:360}$	-	1.63	1.57	1.52	-
$C_{18:3\ \omega6c}$	4.47	1.54	1.08	-	-

Formal Description of *Clostridium muelleri* sp. nov. P21^T

Clostridium muelleri (muel'ler.i. N.L. n.). Gram-positive, non-motile, non-spore forming rods occurring in singles, pairs and in chains. Obligately anaerobic with an optimum growth temperature of 30-40°C and optimum growth pH of 6.5. Growth autotrophically with $CO:CO_2$ and $H_2:CO_2$ or chemoorganotrophically with glucose, galactose, fructose, xylose, mannose, ribose, sucrose, lactose, cellobiose, cellulose, histidine, serine, choline, Casamino acids and glutamate. Growth was not observed with trehalose, rhamnose, melezitose, raffinose, D-arabinose, melibiose, xylan, sorbitol, mannitol, propanol, butanol, ethanol, methanol, glycerol, inositol, syringate, betaine, formate, amygdalin, vanillate, fumarate, malate, succinate, ferulate, trimethoxybenzoate, cysteine, gluconate, asparagine, threonine, alanine, aspartate, glutamine, arginine, tryptophan, methionine, leucine, lysine, phenylalanine and valine. End products of metabolism include ethanol, butanol, hexanol, acetate, butyrate and caproic acid. It showed a positive Methyl Red test but produced a negative result for Voges-Proskauer, nitrate reduction, gelatin hydrolysis, catalase and oxidase tests. The DNA G+C content is 29.5%. The type strain, P21^T, was isolated from old hay from Stillwater, Oklahoma, USA.

Conclusion & Future Directions

Due to unforeseen circumstances involving the COVID-19 pandemic, the colony morphology and growth rate of strain P21^T will be examined upon the reopening of the university and will be included in peer-reviewed publications resulting from this work. It is unclear why strain P21^T did not utilize ethanol or propanol like its nearest neighbors. It is hypothesized strain P21^T had lost its phenotype for this as a result of continuous cultivation in the lab lasting over a decade. Similarly, while all other type 1 *Clostridium* produce endospores and are motile, strain P21^T did not express this phenotype under various conditions, leading to a speculation of a potential loss in phenotypic expression. The metabolic ends products for strain P21^T revealed its potential use in syngas fermentation by the production of acetic acid, butyric acid, caproic acid, ethanol and butanol. Future investigations involve altering varying medium components for the optimization of caproic acid and hexanol production for industrial uses.

Chapter 3: Genome Sequencing and Annotation of *Clostridium muelleri* sp. nov. P21^T

Abstract

The genome of *Clostridium muelleri* sp. nov. P21^T was sequenced using an Illumina MiSeq, assembled and further analyzed to examine the metabolic potential including properties involved in acetogenesis and acid-alcohol fermentation pathways. The genome size was 5.6 Mb and was assembled into 67 contigs. After annotation, 1,818 open reading frames were functionally assigned with the highest percentages contributing to amino acid biosynthesis and degradation (19%), protein metabolism (14%), carbohydrate metabolism and fermentation (12%) and cofactors, vitamins, prosthetic groups biosynthesis (8%). Strain P21^T had all of the genes required for converting CO_2 to acetyl-CoA using the Wood-Ljungdahl pathway. Additionally, strain P21^T contained all genes required to convert acetyl-CoA to acetic acid, butyric acid, ethanol and butanol. The presence of genes encoding for all eight subunits of the Rnf-complex and lack of genes for Ech (energy conserving hydrogenase) confirmed the classification of strain P21^T as a Rnf-dependent acetogen.

Introduction

The era of prokaryotic genome sequencing began in 1995 with the bacterium *Haemophilus influenzae* (43). Since its inception, what used to be an extremely laborious, high cost process is now a more feasible, rapid process that has led to significant increases in available genomic data (44). Genomic analysis has contributed enormously to the understanding of the physiological potential of microorganisms, of which laboratory-based experiments previously not implemented can be performed. Genome sequencing and bioinformatics as a whole, while not 100% perfect, has provided a means of collecting information about a microorganism; many of these in-silico approaches are replacing time and labor-intensive laboratory-based assays.

A goal of this study was to assemble and annotate the genome of *Clostridium muelleri* sp. nov. P21^T to confirm laboratory-based findings that strain P21^T is an acetogen that uses the Wood-Ljungdahl pathway to convert two CO_2 molecules to acetyl-CoA, as well as classify enzymatically its energy conservation as Rnf-dependent or Ech-dependent. Genome sequencing of acetogens is important for better understanding what enzymes are responsible for the increased or decreased production of acids and alcohols. The genes encoding these enzymes have been used in recombinant microorganisms for industrial production of the aforementioned products (45). After comparing the end products of metabolism using various carbon sources observed using GC to the genomic information, it was confirmed strain P21^T contained all genes encoding the required enzymes for the Wood-Ljungdahl pathway as well as the fermentation pathways for acetic acid, butyric acid, ethanol and butanol.

Methods & Materials

[1]DNA Extraction and Library Prep

A 10% sterile salt solution consisting of molecular-biology grade reagents 5 M NaCl, 0.5 M EDTA, 1 M Tris and ultra-pure water was made to wash cells before performing the DNA extraction on P21^T. A 10 ml culture of strain P21^T was grown in basal medium (22) and centrifuged at 6000 x g for 20 minutes. 1 ml of the 10% sterile salt solution was added to the pellet, mixed, then removed into 1.5 ml microcentrifuge tube and centrifuged 10,000 x g for 3 minutes. The wash was repeated two more times for a total of three washes. 100 µl of RLA lysis buffer (Maxwell ®16, Promega Corp) was added to the pellet and placed in a freezer until DNA extraction was performed with the automated Maxwell ®16 Tissue LEV Total RNA Purification Kit (Promega Corp) purification system version 4.90 modified from the manufacturer's instructions as described in Oldham et al. 2012 (26). The DNA was sent to Oklahoma Medical Research Foundation (OMRF) Next Generation Sequencing Core (Oklahoma City, Oklahoma) where the sample was sequenced using the MiSeq Reagent Kit v3 on an Illumina MiSeq.

Data Analysis & Processing

The paired-end sequencing reads were retrieved from Illumina's BaseSpace sequence hub. The sequence was trimmed and adapters were removed using Trimmomatic v0.36 (46) on the Department of Energy Systems Biology Knowledgebase (KBase) data platform (47). Parameters set include using TruSeq3-PE and TruSeq2-PE and NexteraPE-PE adapter files

[1] Dr. Kathleen Duncan at the University of Oklahoma performed the DNA extraction and sent the DNA to OMRF for library prep.

with seed mismatches set to 2, a palindrome clip threshold set to 30 and a simple clip threshold set to 10. Adapters were also manually removed using Cutadapt v1.18 (48). Low complexity reads were filtered out using entropy with PRINSEQ v0.20.4 (49). The quality filtered paired reads were then assembled into contigs using SPAdes v3.13.0 (50) with the minimum contig base pair length set to 500.

Quality Control Checks for Genome Completeness & Contamination

Four quality control checks were performed on this genome using CheckM v1.0.18 (51), FastQC v0.11.5 (Bioinformatics Group at the Babraham Institute, UK. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) Bowtie2 v2.3.2 (52) and QUAST v4.4 (53). CheckM was performed to determine if contamination by non-P21^T DNA occurred during the DNA extraction, sample library prep or the Illumina run. FastQC was run to ensure all quality filtering parameters were met and kmer content was at an appropriate number as well as confirming the removal of all adapters that were inserted for the Illumina run itself. Bowtie2 was used to map the adapter removed sequencing reads back to the assembled genome to ensure the sequences were quality filtered correctly and the contigs were properly formed based on the given gene loci. The assembled genome was processed through QUAST to provide the number of contigs, predicted genes and uncalled bases as well as providing the G + C content and other contig metrics.

Gene Prediction and Annotation

Basic Rapid Ribosomal RNA Predictor (Barrnap) v0.9 (https://github.com/tseemann/barrnap), which uses Hidden Markov Model's (HMM) for each

type of rRNA gene, was used to predict the location of the 5S rRNA, 16S rRNA and 23S rRNA's in strain P21^T's genome. It should be noted, RNAmmer (54) is another commonly used option to predict the location of rRNA's in a genome. However, RNAmmer relies on HMMER version 2 which resulted in complications getting RNAmmer to work on OU's OSCER supercomputer since version 3 is now being used. Therefore, Barrnap was used due to its reliability in predicting rRNA gene location. tRNAscan-SE v.2.0.5 (55) and Infernal v1.1.3 (56)i, covariance model programs for secondary structure characteristics, were used to predict the location of tRNA's in strain P21^T's genome. The genome assembly was annotated using Rapid Annotation using Subsystem Technology (RAST) v.0.1.1 (57)(58)(59) to assign metabolic functions for all open reading frames (ORF).

To confirm the presence of genes involved in the Wood-Ljungdahl pathway, metabolic fermentation pathways for acids and alcohols, and the Rnf-complex, the genome assembly was ran against the protein sequences of each gene responsible for these pathways using BLAST (28). The protein sequences were taken from the UniProtkb database, specifically from *C. carboxidivorans* P7^T (60).

Genome Deposition

The genome data of P21^T was deposited with the National Center for Biotechnology Information (NCBI) resulting in a Biosample accession number of SAMN14366063, BioProject ID PRJNA613251, Sequence Read Archive (SRA) accession number SRR11451959 with a locus tag prefix designated as HBE96, and the whole genome sequencing (WGS) project was given an accession number of JABBNI010000000.

Results & Discussion

Quality Control Checks for Genome Completeness & Contamination

The genome assembly of strain P21^T using SPAdes v.3.13.0 initially resulted in the assembly of 136 contigs. After contig inspection using Geneious Prime v.2020.1.1 (https://www.geneious.com), 69 contigs assembled had a depth of coverage less than 2X. There was a clear jump where contig coverage went from 2X to 120X which resulted in the removal of these contigs. The final number of contigs assembled with a minimum coverage of ~100X was 67 with a G+C content of 29.45%, N50 of 223,637 and a depth of coverage equaling ~ 405X (Table 4). CheckM showed the genome was 99.31% complete with 3.4% adapter contamination, primarily showing near the end of the genome. rRNAs for strain P21^T using Barrnap predicted (n=8) 5S rRNAs, (n=1) 16S rRNA and (n=2) 23S rRNAs. tRNAscan-SE predicted a total of 81 tRNAs, which falls in the typical range for tRNAs in this cluster of *Clostridium*. Of the 67 contigs for strain P21^T, three of them showed coverage >1,000X. Contig 56 was 1,478 bp long with 1,175X coverage; contig 59 was 1,331 bp long with 1,181X coverage and contig 61 was 1,053 bp long with 1,086X coverage. The potential reasoning for such large coverage could be due to transposases within the genome. Genome statistics comparing strain P21^T and *C. carboxidivorans* P7^T (NCBI: txid536227) (61) are provided in Table 5. When looking at the RAST annotation, a significant number of proteins were assigned to invasion and intracellular resistance as well as resistance to antibiotics and toxic compounds. Additionally, proteins were assigned to functions involving transposable elements including a listeria pathogenicity island LIPI-1. These transposable elements could have left physical scars on the genome, causing the coverage for these contigs to be

significantly higher. Further antibiotic screening will occur in the future to see if the genotype matches its phenotype.

Table 5. Genome statistics comparing strain P21^T and *C. carboxidivorans* P7^T after genomic assembly.

Genome Statistics	P21^T	*C. carboxidivorans* P7^T
# Reads	7,670,130	NA
Genome Length	5,645,749 bp	5,650,800 bp
# Contigs (>= 0 bp)	67	151
# Contigs (>= 1000 bp)	62	NA
# Contigs (>= 10000 bp)	43	NA
# Contigs (>= 100000 bp)	20	NA
Largest Contig	389,214 bp	215,023 bp
N50	223,637 bp	NA
GC (%)	29.45	29.7
Genome Coverage	405X	40X
Coding DNA Sequences	5,382	5,325
rRNAs	5S n=8	5S n=2
	16S n=1	16S n=1
	23S n=2	23S n=1
tRNAs	84	49

Gene Prediction and Annotation

Approximately 1,782 of the predicted proteins of strain P21^T were assigned to known functions. Of these, the percentage of what functional categories these are designated to can be seen in Fig. 5 with amino acids biosynthesis and degradation, protein metabolism and carbohydrate metabolism and fermentation showing the highest values. As mentioned previously, motility and spore formation were not observed in the lab for strain P21^T. However, the annotation using RAST shows strain P21^T contains the majority of the genes involved in motility and spore formation. Based on these results we can imply that it has all of the necessary genes to perform these functions, however, this would have to be tested and confirmed by transcriptomics. The BLAST results in Tables 5-6 confirm the presence of genes required for the Wood-Ljungdahl pathway and pathways to produce various acids and alcohols showed strain P21^T to possess all of the necessary genes for acetogenesis and end product formation (Figs. 6-7). As previously mentioned in chapter 1, acetogens can be enzymatically classified based on energy conservation using either the Rnf complex or the Energy-conserving hydrogenase (Ech). The genome assembly and gene sequences for these enzyme complexes were compared on BLAST. The results (Table 7) show genes present for all six subunits of the Rnf complex while no genes were present for the energy conserving hydrogenase.

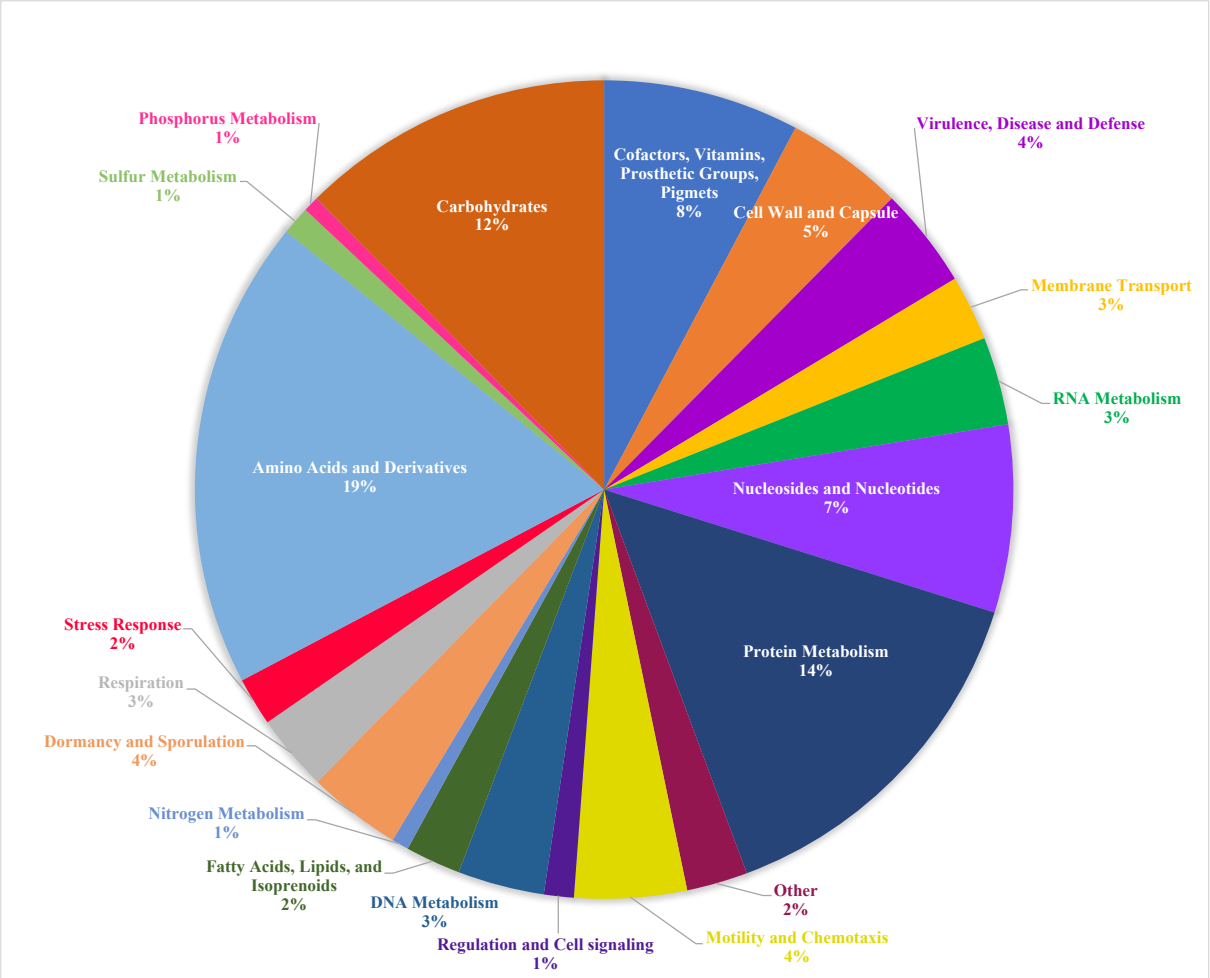

Fig. 5: Strain P21^{T} Open Reading Frames based on functional category percentages from the SEED Viewer through RAST

Table 6: Strain P21^T Blast output of sequence identity match to *C. carboxidivorans* genes for the Wood-Ljungdahl pathway.

Gene	Protein	Location	Query Coverage	E-value	Percent Identity
Ccar_0158	Carbon monoxide dehydrogenase	Contig 1: 389,214 bp	100%	0.00	91.04%
Ccar_0159	4Fe-4S ferredoxin iron-sulfur binding domain protein	Contig 1: 389,214 bp	100%	1.00E-87	88.31%
Ccar_0160	FAD-dependent pyridine nucleotide-disulfide oxidoreductase	Contig 1: 389,214 bp	99%	0.00	76.98%
Ccar_1225	Formate dehydrogenase H	Contig 3: 315,338 bp	100%	0.00	94.97%
Ccar_2671	Formate dehydrogenase chain D	Contig 8: 263,718 bp	9%	1.20	47.37%
Ccar_3255	Formate-tetrahydrofolate ligase	Contig 7: 275,055 bp	99%	0.00	91%
Ccar_3253	Bifunctional methenyl-tetrahydrofolate cyclohydrolase/methylene-tetrahydrofolate dehydrogenase (NADP+)	Contig 7: 275,055 bp	100%	0.00	96.49%
Ccar_3251	Methylene-tetrahydrofolate reductase	Contig 7: 275,055 bp	100%	0.00	96.91%
Ccar_3246	5-methyltetrahydrofolate corrinoid iron sulfur protein methyltransferase	Contig 7: 275,055 bp	100%	4.00E-171	96.93%
Ccar_3247	Acetyl-CoA synthase corrinoid iron-sulfur protein large subunit	Contig 7: 275,055 bp	100%	0.00	97.78%
Ccar_3248	Acetyl-CoA synthase corrinoid iron-sulfur protein small subunit	Contig 7: 275,055 bp	100%	0.00	97.09%
Ccar_3257	Carbon monoxide dehydrogenase	Contig 7: 275,055 bp	100%	0.00	92.06%
Ccar_3245	CO dehydrogenase/Acetyl-CoA synthase complex, beta subunit	Contig 7: 275,055 bp	100%	0.00	90.96%

Table 7: Strain P21^T Blast output of sequence identity match to *C. carboxidivorans* genes for acid and alcohol producing metabolic fermentation pathways.

Gene	Protein	Location	Query Coverage	E-value	Percent Identity
Ccar_3989	Phosphate acetyltransferase	Contig 8: 263,718 bp	95%	0.00	94.34%
Ccar_3990	Acetate kinase	Contig 8: 263,718 bp	100%	0.00	75.19%
Ccar_5153	Aldehyde-alcohol dehydrogenase	Contig 2: 363,825 bp	100%	0.00	94.06%
Ccar_0530	Acetyl-CoA acetyltransferase	Contig 6: 279,607 bp	100%	0.00	90.82%
Ccar_0529	3-hydroxybutyryl-CoA dehydrogenase	Contig 6: 279,607 bp	100%	1.00E-165	88.61%
Ccar_0528	3-hydroxybutyryl-CoA dehydratase	Contig 6: 279,607 bp	100%	1.00E-144	84.56%
Ccar_0531	Acyl-CoA dehydrogenase domain protein	Contig 6: 279,607 bp	100%	0.00	89.97%
Ccar_0247	Phosphate butyryltransferase	Contig 17: 141,930 bp	99%	1.00E-173	87.75%
Ccar_0248	Butyrate kinase	Contig 17: 141,930 bp	100%	0.00E+00	87.64%
Ccar_0742	Aldehyde dehydrogenase	Contig 7: 275,055 bp	88%	2.00E-43	29.64%
Ccar_0005	Iron-containing alcohol dehydrogenase	Contig 11: 207,686 bp	100%	0.00E+00	91.24%
Ccar_0017	Iron-containing alcohol dehydrogenase	Contig 11: 207,686 bp	100%	0.00E+00	85.79%
Ccar_0559	Coenzyme A transferase	Contig 20: 103,158 bp	98%	0.00E+00	77.17%

Table 8: Strain P21^{T} blast output of sequence identity match to genes encoding for the Rnf Complex and Energy-conservation hydrogenase (Ech).

Gene	Protein	Location	Query Coverage	E-value	Percent Identity
rnfG	Proton-translocating ferredoxin: NAD+ oxidoreductase complex subunit G	Contig 11: 207,686 bp	95.00%	2.00E-69	61.02%
rnfB	Proton-translocating ferredoxin: NAD+ oxidoreductase complex subunit B	Contig 11: 207,686 bp	94%	2.00E-114	65.06%
rnfC	Proton-translocating ferredoxin: NAD+ oxidoreductase complex subunit C	Contig 11: 207,686 bp	94%	0.00E+00	75.52%
rnfD	Proton-translocating ferredoxin: NAD+ oxidoreductase complex subunit D	Contig 11: 207,686 bp	100%	3.00E-150	68.88%
rnfA	Proton-translocating ferredoxin: NAD+ oxidoreductase complex subunit A	Contig 11: 207,686 bp	89%	4.00E-84	77.06%
rnfE	Proton-translocating ferredoxin: NAD+ oxidoreductase complex subunit E	Contig 11: 207,686 bp	100%	8.00E-81	61.97%
MSTHT_2302	Energy-conserving hydrogenase (Ferredoxin) Subunit E	Contig 9: 232,394 bp	94%	2.00E-31	27.03%

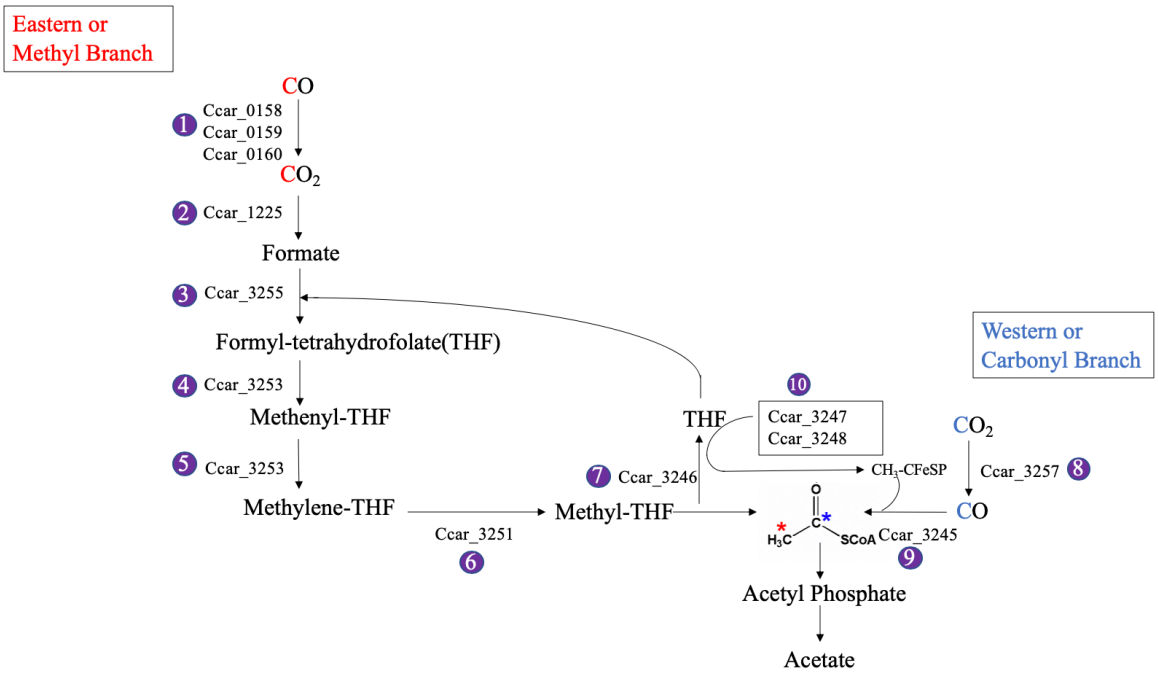

Fig. 6. Gene locations within the Wood-Ljungdahl pathway.

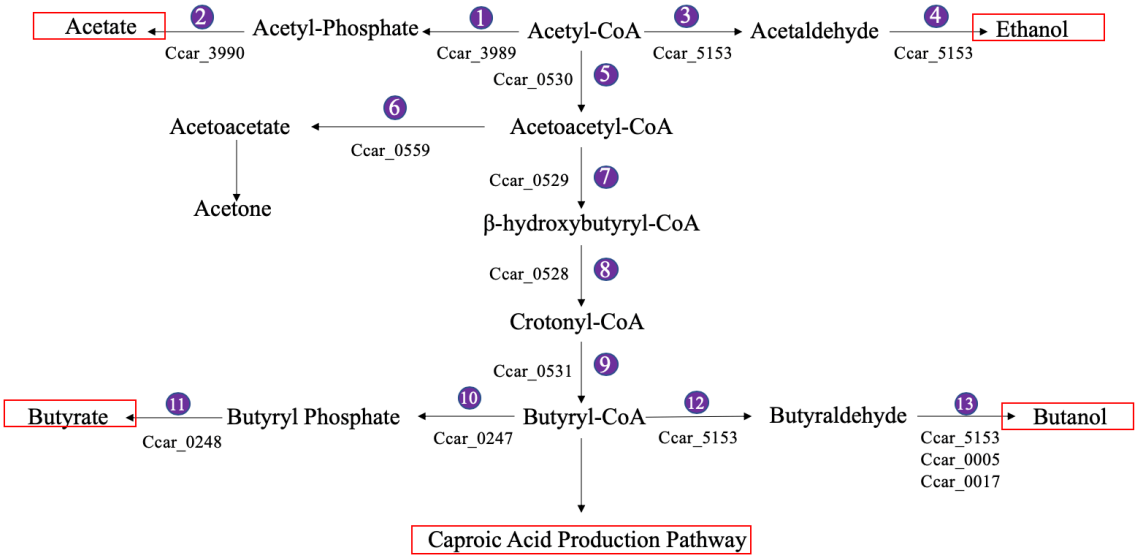

Fig. 7. Gene locations for metabolic fermentation pathways of acetyl-CoA.

Concluding Remarks & Future Investigations

Genes involved in the synthesis of caproic acid and hexanol were not investigated due to the limited understanding of what genes are involved in the reverse beta oxidation pathways for microorganisms to convert butyric or acetic acid into caproic acid with ethanol as the electron donor. The genomes of all organisms provide the opportunity to examine physiological potential within a system. In particular, genomes of acetogens have been used to identify particular genes and transcription factors involved in the upregulation or downregulation of ethanol production (62) . Future investigations of strain $P21^T$'s genome could yield possible clues to better understanding the mechanism in caproic acid production and how that can be further implemented in other organisms. Of the 1,782 functionally assigned ORF, 71 were assigned to antibiotic and toxic compound resistance as well as invasion and intracellular resistance. Future antibiotic screenings will take place in order to confirm the phenotypic expression of these genes.

References

1. **Fischer F, Lieske R, Winzer K.** Biologische Gasreaktionen. II. U¨ ber die Bildung von Essigsa¨ure bei der biologischen Umsetzung von Kohlenoxyd und Kohlens¨aure mit Wasserstoff zu Methan. *Biochem*. 1932; 245: 2–12.

2. **Drake HL, Gößner AS, Daniel SL.** Old acetogens, new light. *Ann N Y Acad Sci*. 2008; 1125: 100–128.

3. **Gottschalk G, Braun M.** Notes: Revival of the name *Clostridium aceticum*. *Int J Syst Evol Microbiol.* 1981; 31: 476–476.

4. **Braun M, Mayer F, Gottschalk G**. *Clostridium aceticum* (Wieringa), a microorganism producing acetic acid from molecular hydrogen and carbon dioxide. *Arch Microbiol.* 1981; 128: 288–293.

5. **Adamse AD**. New isolation of *Clostridium aceticum* (Wieringa). *Antonie van Leeuwenhoek*. 1980; 46: 523–531.

6. **Fontaine FE, Peterson WH, McCoy E, Johnson MJ**. A new type of glucose fermentation by *Clostridium thermoaceticum* n. sp. *J Bacteriol*. 1942; 43: 701–715.

7. **Drake HL**. Acetogenesis. New York, NY: Chapman & Hall; 1994.

8. **Drake HL, Küsel K, Matthies C**. Acetogenic Prokaryotes. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. The prokaryotes: Prokaryotic physiology and biochemistry. Berlin, Heidelberg: Springer; 2013. p. 3–60.

9. **Collins MD, Lawson PA, Willems A, Cordoba JJ, Fernandez-Garayzabal J, Garcia P, et al.** The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. *Int J Syst Evol Microbiol*. 1994; 44: 812–826.

10. **Schuchmann K, Müller V**. Energetics and application of heterotrophy in acetogenic bacteria. *Appl Environ Microbiol*. 2016; 82: 4056–4069.

11. **Grimm NB, Chapin FS, Bierwagen B, Gonzalez P, Groffman PM, Luo Y, et al.** The impacts of climate change on ecosystem structure and function. *Front Ecol Environ*. 2013; 11: 474–482.

12. **Luque R, Herrero-Davila L, Campelo JM, Clark JH, Hidalgo JM, Luna D, et al.** Biofuels: a technological perspective. *Energy Environ Sci*. 2008; 1: 542-564.

13. **United Nations Secretary General.** Report of the Secretary-General on the 2019 Climate Action Summit and the way forward in 2020 [Internet]. 2019 [cited May 2020]. Available from: https://www.un.org/en/climatechange/reports.shtml

14. **Devarapalli M, Atiyeh HK.** A review of conversion processes for bioethanol production with a focus on syngas fermentation. *Biofuel Res J*. 2015; 2: 268–280.

15. **Stoll IK, Boukis N, Sauer J.** Syngas fermentation to alcohols: reactor technology and application perspective. *Chemie Ingenieur Technik*. 2020; 92: 125–136.

16. **Cavalcante W de A, Leitão RC, Gehring TA, Angenent LT, Santaella ST**. Anaerobic fermentation for n-caproic acid production: A review. *Process Biochem*. 2017; 54: 106–119.

17. **Ragsdale SW**. Enzymology of the Wood–Ljungdahl pathway of acetogenesis. *Ann N Y Acad Sci*. 2008; 1125: 129–136.

18. **Rainey FA. Oren A**. Methods in Microbiology: Taxonomy of Prokaryotes. London: Academic Press; 2011.

19. **Woese CR, Fox GE.** Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *PNAS*. 1977; 74: 5088–5090.

20. **Van de Peer Y, Chapelle S, De Wachter R.** A quantitative map of nucleotide substitution rates in bacterial rRNA. *Nucleic Acids Res*. 1996; 24: 3381–3391.

21. **Tanner RS, Miller LM, Yang D.** *Clostridium ljungdahlii* sp. nov., an acetogenic species in clostridial rRNA homology group I. *Int J Syst Evol Microbiol*. 1993; 43: 232–236.

22. **Liou JS-C, Balkwill DL, Drake GR, Tanner RS**. *Clostridium carboxidivorans* sp. nov., a solvent-producing clostridium isolated from an agricultural settling lagoon, and reclassification of the acetogen *Clostridium scatologenes* strain SL1 as *Clostridium drakei* sp. nov. *Int J Syst Evol Microbiol*. 2005; 55: 2085–2091.

23. **Hungate RE**. A roll-tube method for cultivation of strict anaerobes. *Methods Microbiol*. 1969; 3D: 117-132.

24. **Tanner RS.** Cultivation of bacteria and fungi. In: Manual of Environmental Microbiology. 3rd Edition. Washington, D.C: ASM Press; 2007. p. 69–78.

25. **Balch WE, Wolfe RS.** New approach to the cultivation of methanogenic bacteria: 2-mercaptoethanesulfonic acid (HS-CoM)-dependent growth of Methanobacterium ruminantium in a pressureized atmosphere. *Appl Environ Microbiol*. 1976; 32: 781–791.

26. **Oldham AL, Drilling HS, Stamps BW, Stevenson BS, Duncan KE**. Automated DNA extraction platforms offer solutions to challenges of assessing microbial biofouling in oil production facilities. *AMB Express*. 2012; 2: 60-71.

27. **Weisburg WG, Barns SM, Pelletier DA, Lane DJ.** 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol.* 1991; 173: 697–703.

28. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ**. Basic local alignment search tool. *J Mol Biol*. 1990; 215: 403–410.

29. **Yoon S-H, Ha S-M, Kwon S, Lim J, Kim Y, Seo H, et al.** Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol*. 2017; 67: 1613–1617.

30. **Stecher G, Tamura K, Kumar S**. Molecular evolutionary genetics analysis (MEGA) for macOS. *Mol Biol Evol*. 2020; 37: 1237–1239.

31. **Kimura M.** A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*. 1980; 16: 111–120.

32. **Saitou N, Nei M.** The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987; 4: 406–425.

33. **Felsenstein J**. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 1985; 39: 783–791.

34. **Felsenstein J**. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J Mol Evol*. 1981; 17: 368–376.

35. **Tamura K, Nei M.** Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol.* 1993; 10: 512–526.

36. **Reddy CA, Beveridge TJ, Breznak JA, Marzluf GA, Schmidt TM, Snyder LR**. Methods for General and Molecular Microbiology, Third Edition. Washington, DC: ASM Press. American Society of Microbiology; 2007.

37. **Liu K, Atiyeh HK, Tanner RS, Wilkins MR, Huhnke RL**. Fermentative production of ethanol from syngas using novel moderately alkaliphilic strains of *Alkalibaculum bacchi. Bioresour Technol*. 2012; 104: 336–341.

38. **Patel NB, Tito RY, Obregón-Tito AJ, O'Neal L, Trujillo-Villaroel O, Marin-Reyes L, et al**. *Ezakiella peruensis* gen. nov., sp. nov. isolated from human fecal sample from a coastal traditional community in Peru. *Anaerobe*. 2015; 32: 43–48.

39. **Sasser M**. Identification of bacteria by gas chromatography of cellular fatty acids. *MIDI Inc.* 1990; 1-7.

40. **Kämpfer P, Kroppenstedt RM.** Numerical analysis of fatty acid patterns of coryneform bacteria and related taxa. *Can J Microbiol*. 1996; 42: 989–1005.

41. **Suresh K, Prakash D, Rastogi N, Jain RK**. *Clostridium nitrophenolicum* sp. nov., a novel anaerobic p-nitrophenol-degrading bacterium isolated from a subsurface soil sample. *Int J Syst Evol Microbiol*. 2007; 55: 1886–1890.

42. **Phillips JR, Atiyeh HK, Tanner RS, Torres JR, Saxena J, Wilkins MR, et al.** Butanol and hexanol production in *Clostridium carboxidivorans* syngas fermentation: Medium development and culture techniques. *Bioresour Technol.* 2015; 190: 114–121.

43. **Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, et al**. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*. 1995; 269: 496–512.

44. **Whitman W**. The need for change: embracing the genome. *Methods Microbiol.* 2014; 41: 1–12.

45. **Schiel-Bengelsdorf B, Durre P**. Pathway engineering and synthetic biology using acetogens. *Febs Letters.* 2012; 586: 2191–2198.

46. **Bolger AM, Lohse M, Usadel B.** Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014; 30: 2114–2120.

47. **Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al**. KBase: The United States Department of Energy systems biology knowledgebase. *Nat Biotechnol*. 2018; 36: 566–569.

48. **Martin M.** Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal.* 2011; 17: 10–12.

49. **Schmieder R, Edwards R.** Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011; 27: 863–864.

50. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al.** SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012; 19: 455–477.

51. **Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW.** CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. 2015; 25: 1043–1055.

52. **Langmead B, Salzberg SL**. Fast gapped-read alignment with Bowtie2. *Nat Methods*. 2012; 9: 357–359.

53. **Gurevich A, Saveliev V, Vyahhi N, Tesler G.** QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013; 29: 1072–1075.

54. **Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW**. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res*. 2007; 35: 3100–3108.

55. **Chan PP, Lowe TM.** tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods Mol Biol*. 2019; 1962: 1–14.

56. **Nawrocki EP, Eddy SR.** Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*. 2013; 29: 2933–2935.

57. A**ziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al.** The RAST Server: rapid annotations using subsystems technology. *BMC Genom*. 2008; 9: 75-90.

58. **Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, et al.** RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep*. 2015; 5: 8365-8370.

59. **Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al.** The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res*. 2014; 42: D206-214.

60. **The UniProt Consortium**. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. 2019; 47: D506–515.

61. **Bruant G, Lévesque M-J, Peter C, Guiot SR, Masson L.** Genomic analysis of carbon monoxide utilization and butanol production by *Clostridium carboxidivorans* strain P7T. *PLoS ONE*. 2010; 5: e13033-13044.

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