博士論文要約

博士の専攻分野の名称： 博　士（農学）　　氏名　Isaiah Youhak Song

学　位　論　文　題　名

Physiological and genomic characteristics of *Eubacterium* sp. c-25 and their implications on the diversity of deoxycholic acid producers in the human gut
（*Eubacterium* sp. c-25 の生理学的およびゲノムの特徴とヒト腸内のデオキシコール酸生産菌の多様性における意義）

Introduction

Bile acids are secreted into the human gut to aid in the digestion of lipids and fat-soluble vitamins. A small amount of liver-derived primary bile acids are exposed to the biotransformative actions of the colonic microbiota and result in the formation of secondary bile acids (SBAs) through 7α-dehydroxylation, which involves the removal of a hydroxy group at the C-7 position of primary bile acids. The two most common primary bile acids produced in humans are cholic acid (CA) and chenodeoxycholic acid (CDCA), which are transformed into their respective SBAs: deoxycholic acid (DCA) and lithocholic acid (LCA). DCA, the SBA derived from CA, is a regulator of the gut microbiota and has been implicated in diseases such as colonic and hepatic cancer. For these reasons, DCA is closely intertwined with the maintenance of gut health and is thus an important topic of biomedical research.

It must be emphasized that DCA and other SBAs cannot be produced by human cells. All SBAs in the human body are formed by the concerted effort of several enzymes produced by members of the gut microbiota. However, very few DCA-producing microbes have been identified to-date, with the pool of DCA producers limited to several species belonging to or closely related to the genus *Clostridium*. Considering the high efficiency of the CA-to-DCA conversion process, in which virtually all of the CA that enters the colon is converted to DCA (Ridlon et al. 2006), it is difficult to imagine that such a small number of bacteria represented by these species are solely responsible for producing DCA. This discrepancy is compounded by recent studies suggesting that SBA-producing bacteria may range from ~0.0001% of the colonic microbiota (Ridlon et al. 2006) to almost 1% (Rath et al. 2018, Vital

et al. 2019). Therefore, additional studies are necessary in order to determine the extent of DCA-producer numbers and diversity in the human gut.

Eubacterium sp. c-25 is a DCA-producing bacteria isolated and briefly studied in the 1980s (Hirano et al. 1981) that is poorly understood in the context of modern-day analytical techniques and knowledge of bile acid transformation. This human-fecal isolate was thought to be a DCA producer that was morphologically dissimilar to known *Clostridium* species, but mentions of this strain have largely disappeared in the recent literature. As a potentially novel DCA-producing species, we were interested in acquiring and studying this strain to see what insight it could provide into the community of bile acid-metabolizing microbes in the gut. Overall, this study aims to characterize the physiology and genome of *Eubacterium* sp. c-25 against known DCA producers and extrapolate the data to explore the diversity of intestinal DCA producers and identify additional species with the potential to 7α-dehydroxylate bile acids.

Comparative physiological and genomic characterization of *Eubacterium* sp. c-25 and *Clostridium scindens*

Eubacterium sp. c-25 morphology was first observed by scanning electron microscopy (SEM). Images of c-25 cells at multiple time points showed that this strain consistently exhibited a unique filamentous morphology that is not found in other DCA producers. While c-25 shared the same bacillus-like shape found in other DCA producers, c-25 cells were observed to form very long end-to-end linked chains, with the single-cell morphology being exceedingly rare. This is very unusual both in DCA producers and *Clostridium* species, (De Vos et al. 209), illustrating the morphological uniqueness of c-25.

Subsequently, anaerobic cultures of c-25 in GAM medium were prepared at pH values ranging from 6 to 9 in order to test for CA-to-DCA formation in environments reflecting the various physiochemical conditions of the human gastrointestinal tract (Thursby and Juge 2017). These *in vitro* cultures showed successful conversion of CA to DCA at a peak rate of ~50%, verifying the strain's 7α-dehydroxylation ability. When compared to similarly prepared cultures of reference and comparison strains *C. scindens* G10 and *C. scindens* ATCC 35704, which were able to rapidly convert 80-90% of CA to DCA in ideal conditions, the conversion yield of c-25 was noticeably lower. However, all three strains exhibited a preference for pH 8 for maximum growth and DCA production, with pH 7 being the next optimal pH condition. This preference for slightly alkaline pH perhaps illustrates one

physiochemical condition of the gastrointestinal tract that is conducive for DCA producer colonization and, by extension, DCA formation. Gastrointestinal pH tends to fluctuate between studies and individuals, but a pH of 8 is slightly higher than average pH readings in the colon as described by several studies in the literature (Evans et al. 1988, Fleming et al. 2014, Koziolek et al. 2015). This suggests that elevated colonic pH conditions such as in certain cases of Crohn's disease and ulcerative colitis (Press et al. 1998) may result in higher levels of DCA formation.

In order to determine c-25's genetic basis for its ability to 7α-dehydroxylate CA, the whole genome of c-25 was sequenced and closed using a combination of Illumina MiSeq (Illumina, Inc., San Diego, CA, USA) and Nanopore MinION (Oxford Nanopore Technologies, Ltd., Oxford, UK) technologies. The two genomes were combined using the Unicycler v0.4.7 hybrid assembly pipeline (Wick et al. 2017) and annotated with DFAST (Tanizawa et al. 2018). The genome was revealed to consist of a 3,042,110 bp circular chromosome containing 2,893 coding sequences. Using the orthologous grouping algorithm of OrthoFinder (Emms and Kelly 2015, Emms and Kelly 2019) and manual gene mapping using Artemis (Carver et al. 2012), bile acid-inducible (*bai*) genes necessary for 7α-dehydroxylation were identified on the basis of predicted orthology. That is, comparative proteomic analysis of c-25 and other DCA producer genomes was conducted in order to detect genes in c-25 that were predicted to be orthologous to known *bai* genes. According to what is currently known, *bai* genes are organized in an operon, with each of the genes encoded in a single polycistronic mRNA, as characterized in *C. scindens.* Each of these genes, from *baiA* to *baiH,* are known to be necessary and sufficient for 7α-dehydroxylation of CA to DCA (Funabashi et al. 2020) and were thus the targets of identification in c-25. The DCA producer genomes included in the OrthoFinder analysis were: *Eubacterium* sp. c-25, *C. scindens* ATCC 35704, *Clostridium hylemonae, Peptacetobacter hiranonis,* and *C. scindens* G10. As a result of the analysis, it was found that the c-25 genome contains likely orthologues of all of the known *bai* genes, albeit at low amino acid sequence identities ranging from ~35-55% with the exception of BaiH at 83%. However, the unusual arrangement of the hypothesized *bai* genes into multiple clusters differed from the orderly *bai* operon observed in *C. scindens* and other DCA producers. This is an important exception to the *bai* operon paradigm and has never been observed in any DCA producers aside from c-25. Additionally, the c-25 genome did not appear to possess an orthologue of *baiG*, which is believed to be responsible for intracellular uptake of bile acids (Mallonee and Hylemon

1996). Instead, an MFS-family transporter was located in close proximity to the other predicted *bai* genes, suggesting that this gene performs a similar role as a functional homologue of *baiG*. We referred to this gene as the "*baiG*-like" gene. The presence of this *baiG*-like gene is also unique to c-25 and has crucial implications on the unrevealed diversity of intestinal DCA producers. Overall, it is unclear how the lack of a *baiG* orthologue and differences in sequence identity manifest in the DCA production phenotype. While c-25 showed a far lower CA-to-DCA conversion rate compared to *C. scindens*, whether the genes themselves are the limiting factor remains to be discovered.

In order to confirm the identity of these supposed *bai* genes, several *bai* genes in c-25, specifically *baiB, baiCD,* and *baiH*, were selected for measurement of *in vitro* expression levels with and without addition of CA substrate. These three were specifically selected in order to account for each *bai* gene cluster in c-25 as well as encompass the entire length of the *bai* operon in *C. scindens.* Gene-specific primers were constructed and expression was measured by qRT-PCR relative to a reference *recA* gene. This process was also conducted with *C. scindens* G10 in parallel, for comparative purposes. As a result, it was found that all three were upregulated in the presence of CA substrate in c-25, supporting their involvement in 7α-dehydroxylation. Interestingly, the c-25 *baiB* and *baiCD* genes showed exceedingly high transcriptional upregulation that eclipsed what was observed in *C. scindens*, though the implications of these observations are unclear.

Exploration of *Eubacterium* sp. c-25 phylogeny and identification of additional 7α-dehydroxylating bacteria

After verifying the DCA-production capabilities and genomic evidence of *bai* genes in c-25, the amino acid sequence of BaiB from c-25 was used for a BLASTp search of strains that could share the same *bai* gene characteristics as c-25 and also produce DCA. Three additional strains were identified possessing genes sharing 74-77% sequence identity by using this method: *Sporofaciens musculi* (obese mouse cecal isolate; Rasmussen et al. 2021), *Dorea* sp. AF36-15AT (human fecal isolate; Zou et al. 2019), and *Dorea* sp. AM58-8 (human fecal isolate; Zou et al. 2019). Genomic analyses revealed that these three strains possessed *bai* genes that were arranged almost identically to c-25. In order to investigate the phylogenetic linages of these strains of interest and whether they followed an observable phylogenetic pattern, a 16S rDNA phylogenetic tree was constructed using MEGA X (Kumar et al. 2018), implementing the Tamura-Nei model (Tamura and Nei 1993) with 1000 bootstraps and

rooting at the midpoint. The tree included c-25, the top 16S BLASTn hits for c-25, the aforementioned unconfirmed DCA producers, and known DCA producers. The results suggested that none of the known and predicted DCA producers followed a strict evolutionary lineage, as they were scattered throughout the phylogenetic tree. Surprisingly, the closest relative to c-25, *Lachnoclostridium phocaeense* (Brahimi et al. 2017), did not seem to possess *bai* genes. This raises questions as to how *bai* genes are inherited among DCA-producing microbes. The *bai* gene arrangements known so far can be divided into two "archetypes": the *bai* operon of *C. scindens,* and the *bai* gene clusters of *Eubacterium* sp. c-25. Judging by the phylogenetic tree, it is difficult to assume vertical transmission of *bai* genes from a common ancestor due to the widespread phylogenetic diversity, perhaps indicating the possibility of horizontal gene transfer. Further investigation is needed to determine how DCA producers evolved in the human gut environment and how the ability to produce DCA is conferred upon intestinal microbes.

Assuming that DCA producers are, in fact, phylogenetically more diverse than currently believed, we aimed to identify certain marker genes that could be reliably indicative of 7α-dehydroxylation ability. To do so, the available genomes of strains identified in the 16S phylogenetic tree were screened using OrthoFinder in order to find genes that were ubiquitously present in DCA producers while being absent in non-DCA producers. It was found that the only genes following these criteria were *bai* genes (or other genes related to 7α/β-dehydroxylation). Specifically, non-DCA producers consistently lacked orthologues of *baiE, baiI,* and *barB*. The *baiI* gene is thought to encode a 7β-dehydratase (Ridlon et al. 2006) involved in the formation of DCA from ursocholic acid, an epimer of CA. The *barB* gene is thought to play a role in transcriptional regulation of the *bai* operon, but no studies have been able to verify this supposed function as of this writing. Due to the direct role of *baiE* in the process of 7α-dehydroxylating CA to DCA (Funabashi et al. 2020), this gene was the obvious candidate for further study, as it is strongly likely that *baiE* is critical for the conversion of CA to DCA while also satisfying the aforementioned criteria for 7α-dehydroxylation marker genes. To address one of the possible concerns of this gene, which was the low amino acid sequence identity, the protein structures of BaiE from c-25 and G10 were predicted using AlphaFold (Jumper 2021) and compared by pairwise distance-matrix alignment using DALI (Holm 2020) in order to infer functional homology by structural similarity. Both structures were very similar in this regard despite only sharing 55.0% sequence identity, lending support for their predicted functional similarity in

7α-dehydroxylation. It could be reasonably assumed that the genes held identical roles in their respective organisms.

As a test of *baiE* as a DCA-producer marker gene, the BaiE amino acid sequence from c-25 and *C. scindens* G10 were used as queries in a BLASTp search in an attempt to identify additional DCA producers in reference genomic databases. From these searches, three additional species were detected as potential DCA producers, with each species possessing a unique derivation of the *bai* operon that was not observed in other species investigated in this study. Notably, these newly identified species appeared to possess *bai* genes arranged in a similar manner as *C. scindens*, but some of the genes deviated significantly from the typical *bai* operon model, which has not been observed in past DCA producers. While the discovery of additional DCA producer candidates is promising, the reliability of *baiE* as a marker gene for 7α-dehydroxylation is called into question. Having discovered a unique *bai* gene arrangement in c-25, it is very possible that other unknown *bai* gene arrangements exist in unidentified intestinal DCA producers of the human gut. However, the results of BaiE BLASTp only resulted in the discovery of *bai* operon-derived *bai* gene arrangements. Additional analysis of DCA-producer screening methodology is necessary for robust identification of microbes harboring these unique genotypic derivations.

Conclusion

As a result of these studies, we were able to characterize the genome and physiology of *Eubacterium* sp. c-25, compare it to other DCA producers, and discover a novel *bai* gene arrangement that is indicative of undiscovered DCA producer diversity. It is clear by its physiology and genomic characteristics that c-25 is an atypical DCA producer that is representative of the unrealized DCA producer community in the human gut. At this point, the next step is to verify findings experimentally by acquiring all of the newly identified candidate DCA producers and test whether they are able to produce DCA from CA. This is critical for supporting the main hypotheses and conclusions of our studies, as successful *in vitro* conversion is necessary to verify whether the identified bai gene orthologues function as *bai* genes in the predicted DCA producers. Additionally, when DCA production activity has been confirmed in candidate DCA producers, a population analysis of human fecal samples measuring abundance of known and newly discovered DCA producers may be prudent to determine their prevalence in the human gut environment.

Future studies in bile acid metabolism and DCA producer diversity are necessary in

understanding how the gut microbiota composition is involved in defining the bile acid profile in humans. Although large-scale metabolomic and metagenomic analyses are becoming relatively commonplace, including in bile acid studies, mechanistic studies are still necessary in order to elucidate the specific interactions and genetic requirements for DCA production at a more focused level. In a biomedical context, several diseases are associated with elevated DCA levels such as liver and colon cancer (Kitazawa et al. 1990, Yoshimoto et al. 2013, Cao et al. 2017). We also know that DCA production is exclusively catalyzed by gut microbes harboring the necessary enzymes (Ridlon et al. 2006). However, it is difficult to develop preventative measures against carcinogenic bile acids such as DCA, as the specific taxa responsible for producing them have not been clearly determined. We hope that we can bridge the gap between bile acid metabolic mechanisms and the diverse gut microbial community in providing a holistic understanding of bile acid transformation in the human gut and its application to the maintenance of gastrointestinal health.

In conclusion, this study revealed that *Eubacterium* sp. c-25 is a phylogenetically unique DCA producer possessing a novel arrangement of *bai* genes. While experimental evidence of DCA formation in the newly identified, unconfirmed DCA producers is necessary to verify *in silico* findings, the discovery of c-25 and its shared *bai* gene arrangement sets a precedent for the plausible existence of other genotypically non-traditional 7α-dehydroxylating bacteria and further implies that DCA producer diversity in the human gut is greater than expected.