An Expanded Genetic Code Enables Trimethylamine Metabolism in Human Gut Bacteria

Veronika Kivenson, Stephen J. Giovannoni

ABSTRACT Cardiovascular disease (CVD) has been linked to animal-based diets, which are a major source of trimethylamine (TMA), a precursor of the proatherogenic compound trimethylamine-N-oxide (TMAO). Human gut bacteria in the genus Bilophila have genomic signatures for genetic code expansion that could enable them to metabolize both TMA and its precursors without production of TMAO. We uncovered evidence that the Bilophila demethylation pathway is actively transcribed in gut microbiomes and that animal-based diets cause Bilophila to rapidly increase in abundance. CVD occurrence and Bilophila abundance in humans were significantly negatively correlated. These data lead us to propose that Bilophila, which is commonly regarded as a pathobiont, may play a role in mitigating cardiovascular disease. Human gut microbiomes have been shown to affect the development of a myriad of disease states, but mechanistic connections between diet, health, and microbiota have been challenging to establish. The hypothesis that Bilophila reduces cardiovascular disease by circumventing TMAO production offers a clearly defined mechanism with a potential human health impact, but investigations of Bilophila cell biology and ecology will be needed to fully evaluate these ideas.

IMPORTANCE Links between trimethylamine-N-oxide (TMAO) and cardiovascular disease (CVD) have focused attention on mechanisms by which animal-based diets have negative health consequences. In a meta-analysis of data from foundational gut microbiome studies, we found evidence that specialized bacteria have and express a metabolic pathway that circumvents TMAO production and is often misannotated because it relies on genetic code expansion. This naturally occurring mechanism for TMAO attenuation is negatively correlated with CVD. Ultimately, these findings point to new avenues of research that could increase microbiome-informed understanding of human health and hint at potential biomedical applications in which specialized bacteria are used to curtail CVD development.

KEYWORDS cardiovascular disease, microbiome, molecular genetics
frequently referred to as the sole breakdown product of TMA, a subset of methano-
genetic archaea in the gut have the ability to utilize TMA by an alternative pathway (15, 16). This pathway is enabled by genetic code expansion (GCE), via insertion of the 22nd amino acid, pyrrolysine (Pyl), in place of a TAG amber codon at conserved sites of the tri-, di-, and monomethylamine methyltransferase genes, with methane as the end product (17–19). TMA metabolism that relies on proteins that use an expanded code has also been described in some bacteria from environmental settings, including symbionts of gutless marine worms (20), as well as the *Firmicutes* bacterium *Acetohalobium arabaticum* isolated from a Crimean lagoon (21). Despite its potential importance, bacterial metabolism of TMA remains largely unexplored in the gut microbiome.

A Deltaproteobacterium commonly found in the human gut, *Bilophila wadsworthia*, also has genes necessary for encoding pyrrolysine (22). First identified in an appendicitis infection in 1989, this “bile-loving,” taurine-degrading bacterium is commonly referred to as a pathobiont associated with abscesses (23, 24). Additionally, *Bilophila* is able to produce hydrogen sulfide, and these bacteria may be linked to inflammatory bowel disease, colorectal cancer, and systemic inflammation (24–28). Despite its classification as a pathobiont, *Bilophila* is also commonly present in healthy human microbiomes (29, 30). There is uncertainty about GCE and TMA metabolic pathway characteristics in these organisms; notably, whether the pyrrolysine pathway is functional in *trans* and whether the pathway is expressed. The function of this pathway is questioned (21) because of organizational differences in the pyrrolysyl-tRNA synthetase gene (required for encoding Pyl): in archaea, this synthetase is encoded by a single gene, while in *Bilophila*, the N- and C-terminal domains are encoded by two distinct genes. Second, in previous reports of Pyl pathways, the trimethylamine methyltransferase and Pyl genes are commonly adjacent to one another (21, 22), while this is not the case in *Bilophila*. Third, the TMA methyltransferase protein is prematurely truncated at the TAG codon in public *Bilophila* data sets available on the NCBI and JGI-IMG web servers. Despite the ubiquity of this bacterium, and its potential importance in modulating TMAO levels (and thus CVD risk), these ambiguities remain unresolved, and studies of the human gut microbiome make no mention of the possibility of a bacterial pathway for TMA utilization. In this study, we investigate this alternative pathway of trimethylamine metabolism in *Bilophila*, the change in abundance of this taxon in response to an animal-based diet, and the correlation between *Bilophila* abundance and CVD pathology (Fig. 1B).
RESULTS

To explore the potential for *Bilophila* TMA metabolism, we retranslated the protein-coding sequences from publicly available *Bilophila* genomes (Table S1) using a custom translation table with the TAG codon as a readthrough rather than a stop codon (see Materials and Methods). We found an in-frame TAG codon at a conserved position in the TMA methyltransferase, corresponding to the pyrrolysine residue of this protein in archaeal methanogens (17) and bacteria (20, 21) (Fig. 2 and Table S2). All genes known to have functions specific to pyrrolysine production were also identified in the *Bilophila* genomes: PylB, PylC, and PylD and the pyrrolysyl-tRNA synthetase N and C termini, as well as the Pyl-specific tRNA, with the corresponding CUA anticodon (Fig. 2; Fig. S1).

One exception was that *Bilophila wadsworthia* ATCC 49260 is missing one component: a Pyl tRNA was not located. In addition to the Pyl-containing TMA methyltransferase, accessory genes for the TMA methyltransferase pathway, including *ramA* (the activating gene) and the TMA methyltransferase corrinoid protein, are encoded in all of the *Bilophila* genomes (Fig. 2 and Table S3). The dimethylamine methyltransferase gene is not encoded in any of the genomes, and a gene similar to monomethylamine methyltransferase, while present, does not have the conserved in-frame TAG codon that is a hallmark of this gene. In summary, metabolic inference from genomic data indicates that *Bilophila* have the potential to convert TMA to dimethylamine (DMA) in the human gut.
The TMA-utilizing pathway we observed may benefit *Bilophila* by enabling them to use TMA directly and also TMA produced metabolically within the cell from two precursor compounds, choline and glycine betaine (Fig. 2). Briefly, *Bilophila* genomes encode the glycylation radical choline-TMA lyase and its associated activating protein (CutC and CutD, respectively), which convert choline to TMA (31). Glycine betaine (sources of which includes choline (32) and possibly carnitine (33)) can be converted to TMA via the selenocysteine (Sec)-containing glycine betaine reductase (GRD) pathway (34). The GRD pathway and the full set of machinery required for the noncanonical amino acid selenocysteine, encoded with a repurposed UGA stop codon, are also present in the *Bilophila* genomes (Table S3). Multiple copies of proteins belonging to the betaine/carnitine/choline family of transporters are also encoded in the *Bilophila* genomes. We conclude that the genomic data indicate that *Bilophila* has the ability to use choline and glycine betaine, converting these compounds to TMA, and subsequently to DMA. In doing so, these bacteria may deplete precursor compounds and TMA that would otherwise be available for host hepatic processes, thereby reducing or circumventing production of TMAO (Fig. 2).

Next, we asked whether the pyrrolysine pathway is functional and whether GCE-enabled TMA metabolism is active in *Bilophila* from the gut environment. To do so, we reexamined data sets from a recent human gut metatranscriptomic study (35) and a mouse model system study (36) (Table S1). Applying a minimum threshold of 98% amino acid sequence identity with annotated *Bilophila* proteins, we identified expression of the TMA methyltransferase and pyrrolysine machinery proteins in both human fecal and mouse cecum samples (Fig. 2 and Tables S4 and S5). The fraction of TMA consumed via this bacterial metabolic process in the human gut microbiome remains uncertain, but expression data support the conclusion that this metabolic process is active.

**DISCUSSION**

This survey of published microbiome sequence data uncovered evidence that bacteria in the genus *Bilophila* use genetic code expansion in the human gut to produce a TMA methyltransferase. We hypothesized that this mechanism could be used to compete with other TMA-utilizing processes, potentially decreasing the production of TMAO from the proatherogenic precursor trimethylamine. To explore this hypothesis, we reexamined additional publicly available data (Table S1) to determine whether this naturally occurring mechanism for TMAO attenuation is correlated with CVD. In a recent study describing the gut microbiome in atherosclerotic cardiovascular disease, Jie et al. (29) report that *Bilophila* is one of the 20 most abundant genera in the samples examined for this project. Their data also show that the abundance of *Bilophila* is significantly enriched in the microbiomes of individuals in the healthy/control group (n = 187) compared to the CVD group (n = 218). Second, in a study describing a rapid diet-induced change in the human gut microbiome, David et al. (37) reveal that *Bilophila* significantly increase in abundance in response to an animal-based diet compared to a plant-based diet. Finally, in a study detailing the transmission of atherosclerosis susceptibility via gut microbial transplanation, Gregory et al. (38) show that mice with certain taxa have increased TMAO levels and develop atherosclerotic lesions and postulate that this is a microbiome-dependent, transmissible trait. *Bilophila* is one of only six taxa that are significantly enriched in both the cecal and fecal microbiome of the healthy group compared to the mice that developed atherosclerotic lesions. The observations we report may challenge the widely held idea that members of this taxon act exclusively as pathobionts; their role in the microbiome and human health may be context dependent, and their potential to mitigate the impacts of animal products on CVD warrants further study.

**MATERIALS AND METHODS**

Microbiome data selection and accession numbers. Genomic, metagenomic, and metatranscriptomic data sets used for this study were accessed on 15 January 2020 from publicly available sequencing projects (see Table S1 in the supplemental material). The *Bilophila* genomes analyzed are reference...
genomes from the Human Microbiome Project (39) as well as the type strain for this genus, from the Refseq database (40). All of the available genomes from the genus Bilophila were examined. BioProject accession no. PRJEB33885 was excluded because it includes a duplicate of a previously published genome (included in accession no. PRJNA41963). Paired metagenomic-metatranscriptomic data were accessed from a large cohort study of adult men (35), and additional metatranscriptomic data were used from a mouse model system (36). Metatranscriptomic and metagenomic read data were obtained using the SRA-toolkit (https://github.com/ncbi/sra-tools) v2.9.1 data via the fastq-dump option.

Genomic analysis and implementation of a custom protein translation table. Predicted proteins in each genome were initially identified using Prodigal (42) v.2.6.3 using single genome mode, and functional annotation was determined using Hmmer (43) v.3.1, with the hmmscan option (1E−10 cutoff), with top hits only, against the Pfam (44) database v.31 and Tigrfam (45) database v.15.0.

For genomes in which enzymes were identified for the synthesis of pyrrolysine, we applied an alternate protein prediction procedure that reassigned TAG codons from stop to readthrough. To accomplish this task, we modified the source code for Prodigal by adding a custom translation table that has TAG readthrough and retains all three canonical bacterial start codons. Proteins with in-frame stop codons in the relevant genes were then manually inspected to determine whether the region containing and following the stop codon was conserved in comparisons to homologues containing pyrrolysine at a similar position (see Fig. S2 in the supplemental material). The modified code and documentation for the modified translation table are freely available at https://github.com/VeronikaKivenson/Prodigal.

Protein searches for the full-length TMA methyltransferase amino acid sequence from Bilophila were performed on the NCBI and JGI-IMG webservers on 1 July 2020. For searching and identifying putative selenoproteins, the bSECIS (46) webserver was used and results were inspected and compared with previously identified selenoproteins. Protein sequence alignment was performed using Muscle (47), with Geneious 2020.1 used for visualization of the alignments. The Aragorn (48) web server was used to locate the tRNA sequences from each genome and to determine the secondary structure of Sec- and Pyl-specific tRNAs, as well as their corresponding anticodon sequences. tRNAscan-SE (49) v.1.23 was also used to search for the tRNAs for Sec. Chemdraw and Biorender software were used to create figures.

Metatranscriptomic data analysis. Preliminary mapping of metatranscriptomic sequence data to Bilophila genomes was done using Bowtie2 (50) v.2.3.4.1. This approach identified samples in which these Bilophila bacteria were present and active at detectable levels. Next, metatranscriptomic reads were coassembled using SRA data from select samples belonging to each BioProject (Table S4) using Megahit (51) v.1.1.1 with default parameters. Functional annotation and identification of stop codon readthrough were performed as described earlier. In addition to Prodigal, FragGeneScan was used to identify partial genes (52). The computational biology data processing and analysis workflow were completed using the Extreme Science and Engineering Discovery Environment (XSEDE) Bridges resource at the Pittsburgh Supercomputing Center (53, 54).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.2 MB.
FIG S2, PDF file, 0.1 MB.
TABLE S1, PDF file, 0.04 MB.
TABLE S2, PDF file, 0.04 MB.
TABLE S3, PDF file, 0.03 MB.
TABLE S4, PDF file, 0.1 MB.
TABLE S5, PDF file, 0.05 MB.

ACKNOWLEDGMENTS

This work was supported by a grant from the Simons Foundation (SFARI 649176 to V.K.). This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. Specifically, it used the Bridges system, which is supported by NSF award number ACI-1445606, at the Pittsburgh Supercomputing Center (PSC), through allocation TG-DEB170007.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

We thank Grace Deitzler and Maude David for helpful discussion of the gut microbiome data sets.

We declare that we have no competing interests.

The project was conceived by V.K. and S.J.G. V.K. performed the bioinformatic data analysis, and interpretation was performed by V.K. and S.J.G. V.K. and S.J.G. wrote the manuscript.
34. Andreesen JR. 1994. Glycine metabolism in anaerobes. Antonie Van Leeuwenhoek 66:223–237. https://doi.org/10.1007/BF00871641.

35. Abu-Ali GS, Mehta RS, Lloyd-Price J, Mallick H, Branck T, Ivey KL, Drew DA, DuLong C, Rimm E, Izard J, Chan AT, Huttenhower C. 2018. Meta-transcriptome of human faecal microbial communities in a cohort of adult men. Nat Microbiol 3:356–366. https://doi.org/10.1038/s41564-017-0084-4.

36. Natividad JM, Lamas B, Pham HP, Michel M-L, Rainteau D, Bridonneau C, da Costa G, van Hylckama Vlieg J, Sovran B, Chamignon C, Planchais J, Richard ML, Langella P, Veiga P, Sokol H. 2018. Biphlopia wadsworthia aggravates high fat diet induced metabolic dysfunctions in mice. Nat Commun 9:2802. https://doi.org/10.1038/s41467-018-05249-7.

37. David LA, Maurice CF, Carmody RN, Gootenberg DB, Turnbaugh PJ. 2014. Diet rapidly and reproducibly alters the human gut microbiome. Nature 505:559–563. https://doi.org/10.1038/nature12820.

38. Gregory JC, Buffa JA, Org E, Wang Z, Levison BS, Zhu W, Wagner MA, Bennett BJ, Li L, DiDonato JA, Lusis AJ, Hazen SL. 2015. Transmission of atherosclerosis susceptibility with gut microbial transplantation. J Biol Chem 290:5647–5660. https://doi.org/10.1074/jbc.M114.618249.

39. Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res 39:W29–W37. https://doi.org/10.1093/nar/gkr367.

40. Tatusova T, Ciufo S, Fedorov B, O’Neill K, Tolstoy I. 2014. RefSeq microbial genomes database: new representation and annotation strategy. Nucleic Acids Research 42:D553–D559. https://doi.org/10.1093/nar/gkt1274.

41. Reference deleted.

42. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Reference genomes from the human microbiome. Science 328:994–999. https://doi.org/10.1126/science.1183605.

43. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.

44. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. 2004. The Pfam protein families database. Nucleic Acids Res 32:D138–D141. https://doi.org/10.1093/nar/gkh121.

45. Haft DH, Selengut JD, White O. 2003. The TIGRFAMs database of protein families. Nucleic Acids Res 31:371–373. https://doi.org/10.1093/nar/gkg128.

46. Zhang Y, Gladyshev VN. 2005. An algorithm for identification of bacterial selenocysteine insertion sequence elements and selenoprotein genes. Bioinformatics 21:2580–2589. https://doi.org/10.1093/bioinformatics/bti400.

47. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.

48. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16. https://doi.org/10.1093/nar/gkh152.

49. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964. https://doi.org/10.1093/nar/25.5.955.

50. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.

51. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033.

52. Rho M, Tang H, Ye Y. 2010. FragGeneScan: predicting genes in short and error-prone reads. Nucleic Acids Res 38:e191. https://doi.org/10.1093/nar/gkq747.

53. Towns J, Cockerill T, Dahan M, Foster I, Gaither K, Grimshaw A, Hazelwood V, Lathrop S, Lifka D, Peterson GD, Roskies R, Scott JR, Wilkins-Diehr N. 2014. XSEDE: accelerating scientific discovery. Comput Sci Eng 16:62–74. https://doi.org/10.1109/MCSE.2014.80.

54. Nystrom NA, Levine MJ, Roskies RZ, Scott JR. 2015. Bridges: a uniquely flexible HPC resource for new communities and data analytics, p 1–8. In Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure. Association for Computing Machinery, New York, NY.