Genetic Drift and Host-Adaptive Features Likely Underlie the Cladogenesis of Insect-Associated Lachnospiraceae

Arturo Vera-Ponce de Leon[1,4,†], Mathias G. Schneider[1,†], Benjamin C. Jahnes[2,†], Victoria Sadowski [iD][1], Lennel A. Camuy-Vélez[2,5], Jun Duan[3], and Zakee L. Sabree [iD][1,*]

[1]Department of Evolution, Ecology and Organismal Biology, The Ohio State University, Columbus, OH, USA

[2]Department of Microbiology, The Ohio State University, Columbus, OH, USA

[3]Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada

[4]Present address: Norwegian University of Life Sciences, Universitetstunet 3, 1433 Ås, Norway.

[5]Present address: Department of Biological Sciences, North Dakota State University, Fargo, ND, USA.

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: sabree.8@osu.edu.

Accepted: 24 May 2022

Abstract

Phylogenetic and functional group analyses of the genomes of anaerobic bacteria isolated from *Periplaneta americana* digestive tracts suggest that they represent novel Lachnospiraceae genera. PAL113 and PAL227 isolate genomes encoded short-chain fatty acid biosynthetic pathways and plant fiber and chitin catabolism and other carbohydrate utilization genes common in related Lachnospiraceae species, yet the presence of operons containing flagellar assembly pathways was among several distinguishing features. In general, PAL113 and PAL227 isolates encode an array of gene products that will enable them to thrive in the insect gut environment and potentially play a role in host diet processing. We hypothesize that the cladogenesis of these isolates can be a result of their oxygen sensitivity and reliance upon the host for dispersal and genetic drift and not necessarily a result of an ongoing mutualism.

Key words: Lachnospiraceae, cockroach, gut bacteria, cladogenesis.

Significance

Animals are unique microbial habitats that can both constrain and encourage their occupation by different bacterial species and lead to cladogenetic events. Lachnospiraceae have been detected in the digestive tracts of several vertebrate and invertebrate species, yet few have been characterized from the latter. Lachnospiraceae isolates from the globally distributed and omnivorous cockroach, *Periplaneta americana*, are anaerobic and their genomes reveal potentially host-supportive capabilities for short-chain fatty acid provisioning and complex carbohydrate degradation, and flagellar biosynthesis, which is relatively unique among known Lachnospiraceae. Multilocus phylogenetic analyses indicate that these isolates represent novel lineages, which highlights the need for further characterization of Lachnospiraceae associated with nonmammalian hosts.

Introduction

Gut microbiota are important contributors to animal host development and health (Turnbaugh et al. 2006; Huttenhower et al. 2012; Schauer et al. 2012), and omnivorous cockroaches maintain microbial diversity in their digestive tracts that represent dozens of phylum-level taxonomic groups (Bertino-Grimaldi et al. 2013; Sabree and Moran 2014; Tinker and Ottesen 2016). Bacteroidetes and the Firmicutes represent the majority of cockroach commensal gut communities (Sabree and Moran 2014) and several have yielded to conventional cultivation approaches (Vera-Ponce de León et al. 2020). Lachnospiraceae are strictly anaerobic Gram-positive bacteria that are frequently sampled from cockroach guts, including peridomestic *Periplaneta americana*. Although relatively uncommon in the environment, Lachnospiraceae species can comprise a substantial proportion of gut commensals in mammals (10–18%) and cockroaches (4%) (Meehan and Beiko 2014; Pérez-Cobas et al. 2015), and genomics have begun to suggest their impact on host and gut community health (Biddle et al. 2013). Lachnospiraceae can comprise 18% of the total microbiota in bovine ruminants, and they are capable of recalcitrant organic matter degradation and short-chain fatty acid (SCFA) production (Meehan and Beiko 2014). Complex carbohydrate metabolism in Lachnospiraceae is enabled by diverse glycoside hydrolases that can breakdown dietary fibers (pectin, hemicellulose, cellulose, and starch) for the assimilation of mono- and oligosaccharide degradative products (Biddle et al. 2013). Additionally, some Lachnospiraceae produce sulfatases and fucose-degradation enzymes that enable them to use host glycans in the absence of dietary fibers, which suggests adaptations to gut habitats (Salyers et al. 1977; Vacca et al. 2020). The cultivation and comparative genomic analysis of two novel Lachnospiraceae isolated from *P. americana* digestive tracts is reported, and mechanisms that underlie the cladogenesis of insect-associated bacteria comprising species-rich gut communities are posited for future study.

Results and Discussion

PAL113 and PAL227 were cultivated from *P. americana* (supplementary fig. S1*A and B*, Supplementary Material online) and represent novel genera within the Lachnospiraceae based on several lines of evidence, including phylogenies using either 16S rRNA genes (fig. 1*A*) or concatenated orthologous proteins (fig. 1*B*). Additionally, amino acid identity (AAI) analysis supports this hypothesis as PAL113 and PAL227 were ≤74% identical to any neighboring Lachnospiraceae and ~76% identical to each other (supplementary figure S2, Supplementary Material online). Pan-Lachnospiraceae analyses indicate that PAL113 and

PAL227 genomes share many general features with other Lachnospiraceae (fig. 2*A*), but fine-scale gene content analysis indicates that their genomes are differentially enriched across 8 out of 11 clusters of orthologous gene (COG) functional categories (fig. 2*B*), which also suggests that these lineages have significantly diverged from closely related Lachnospiraceae. BLAST-based queries of the GenBank "nt" database were performed (December 2021) using PAL113 and PAL227 16S rRNA gene sequences, which yielded hits ≤96% identity to any cultured isolates, and all matches >96% identity were to 16S rRNA gene amplicons derived from insect (i.e., *Pachnoda ephippiata* and *Shelfordella lateralis*) digestive tissues. These results suggest an insect host association rather than a recent acquisition from the environment. An important caveat is that the overwhelming majority of available Lachnospiraceae genomes are from mammal-associated genera, and, therefore, a greater genomic sampling of Lachnospiraceae from non-mammalian habitats is warranted. Many Lachnospiraceae, including PAL113 and PAL227, are oxygen sensitive, which will limit their dispersal range, but they can exploit low-oxygen hindguts as a viable habitat and coprophagic behaviors to migrate between individuals. Notably, filial coprophagy in cockroaches is both common and preferred (Wada-Katsumata et al. 2015), and early stage *P. americana* nymphs acquire their gut microbiota through the consumption of maternal (vertical) and nestmate (horizontal) feces, and they must do so to develop normally (Jahnes et al. 2019, 2021). Filial coprophagy is an intraspecific host transmission modality that can restrict gene flow between bacterial populations following host association and provide the conditions for speciation to occur. Specifically, isolation within the host can lead to a genetic drift that can drive allopatric speciation, which can possibly explain the cladogenesis of these host-associated lineages (Groussin et al. 2020) independent of mutually beneficial symbioses or coevolution.

PAL113 and PAL227 encode many carbohydrate utilization enzymes (CAZYmes) observed in Lachnospiraceae genomes, and these will enable them to metabolize host dietary and structural carbohydrates likely encountered in cockroach guts (fig. 3). Glycosyl hydrolase (GH) genes involved in cellulose/hemicellulose catabolism were present, but only one of the genes encoding canonical cellulose catabolic enzymes (i.e., GH3, GH5, and GH9; Naas et al. 2018) was present (fig. 3), suggesting that PAL113 and PAL227 may be unable to directly use cellulose as a carbon source. The PAL113 and PAL227 genomes encode six copies of the GH13 alpha-amylase gene, which suggests that they can use starch as a carbon source, which is a feature of Lachnospiraceae (Vacca et al. 2020). Notably, PAL113 and PAL227 are capable of pectin (e.g., PL1) and polygalacturonan (e.g., GH28) utilization, which was uncommon among the Lachnospiraceae genomes analyzed in this
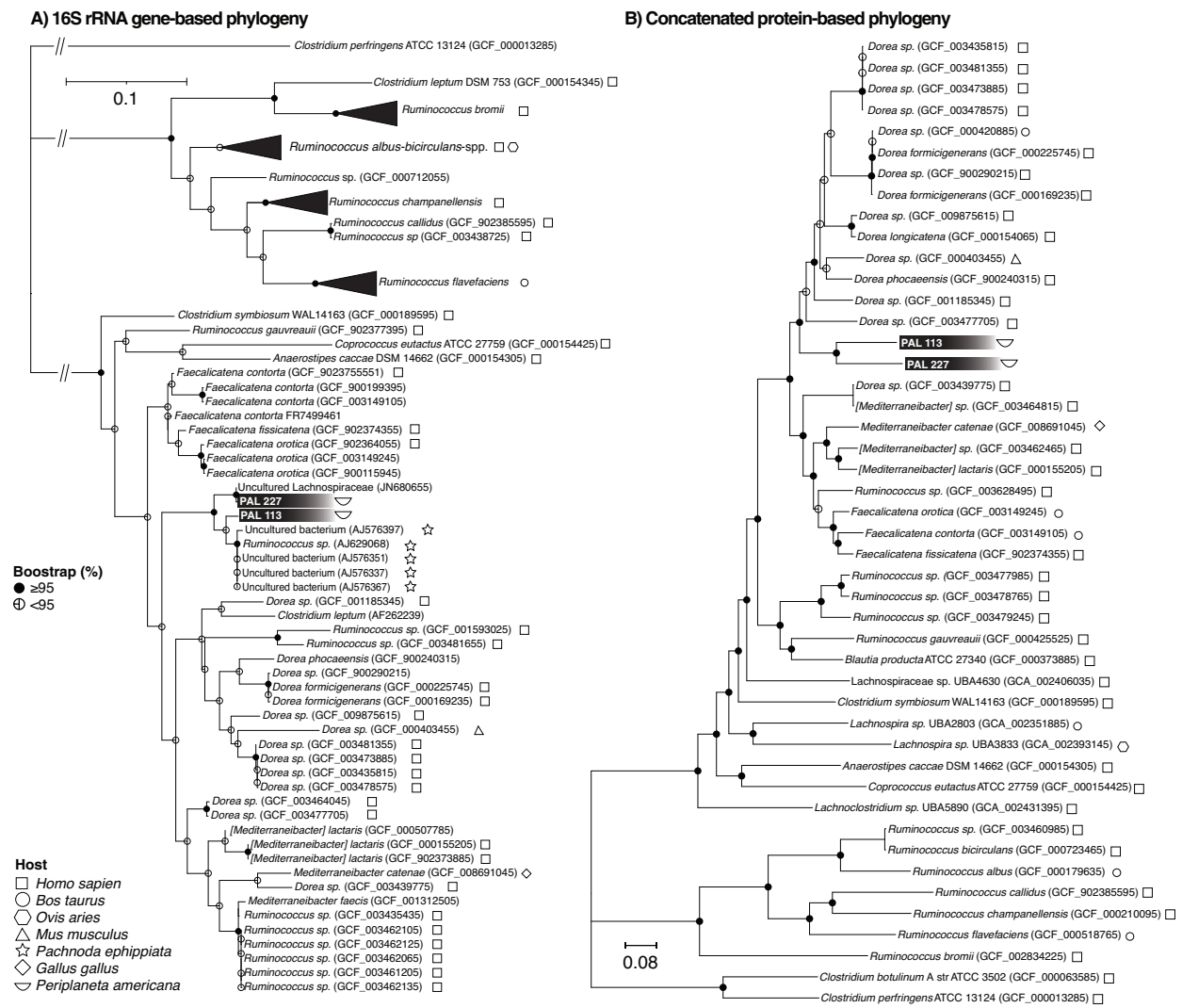
FIG. 1.—Cladogenesis of *P. americana* Lachnospiraceae isolates. Maximum-likelihood phylogenetic trees of 16S rRNA genes (*A*) and concatenated single-copy orthologous proteins (*B*). *Clostridium perfringens* ATCC_1324 was used as an outgroup for both phylograms. Scale bars represent estimated sequence divergence.

work, and niche specialization may explain the presence of PAL113 and PAL227, given that omnivorous cockroach diets include fruits (Bell and Adiyodi 1982). GH28 and PL1 are essential for homogalacturonan catabolism (Luis et al. 2018) in mammalian gut-dwelling *Bacteroides thetaiotaomicron*, suggesting that PAL227 can also catabolize homogalacturonans. In addition to enzymes for plant fiber degradation, PAL113 and PAL227 encoded canonical chitin (i.e., GH18) and mucin (i.e., GH109) catabolism genes, which was common in most of the gut-dwelling Lachnospiraceae species (fig. 3). Chitin is commonly abundant in insect tissues, and it is a main component of the protective gut peritrophic matrix (Tellam et al. 1999), which is functionally analogous to the mucus layer in vertebrates, including being comprised of macromolecules (i.e., peritrophins) with mucin-like domains (Terra 2001; Erlandson

et al. 2019). Genes encoding the chitinase GH18 have also been detected in the genomes of Bacteroidetes isolated from *P. americana* (Vera-Ponce de León et al. 2020), suggesting that these enzymes may be host-adaptive by enabling these gut-dwelling bacteria to successfully colonize insect digestive tissues and use host glycans.

The SCFA biosynthesis pathways were encoded by PAL113 and PAL227 genomes, indicating that they may play important roles in recycling fucose-containing host glycoproteins and producing host-assimilable propionate and butyrate from dietary metabolites such as succinate, pyruvate, and rhamnose (fig. 4). Fucose is common in the host glycans of insect and mammalian tissues and in insect peritrophic membrane glycans (Liu et al. 2019), whereas rhamnose is common in some plants and produced by bacteria, but it is not common in host gut tissues
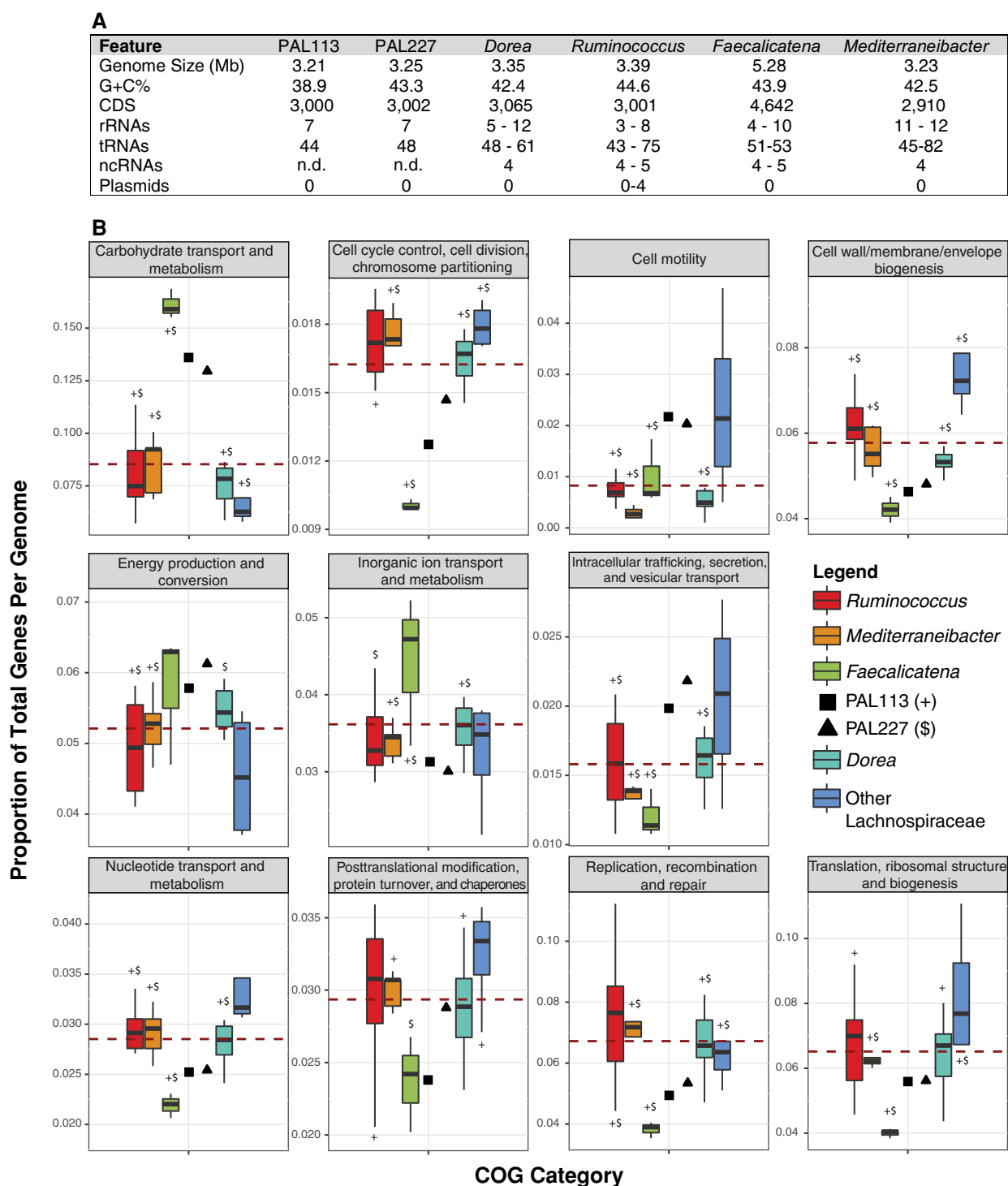
A

Feature	PAL113	PAL227	*Dorea*	*Ruminococcus*	*Faecalicatena*	*Mediterraneibacter*
Genome Size (Mb)	3.21	3.25	3.35	3.39	5.28	3.23
G+C%	38.9	43.3	42.4	44.6	43.9	42.5
CDS	3,000	3,002	3,065	3,001	4,642	2,910
rRNAs	7	7	5 - 12	3 - 8	4 - 10	11 - 12
tRNAs	44	48	48 - 61	43 - 75	51-53	45-82
ncRNAs	n.d.	n.d.	4	4 - 5	4 - 5	4
Plasmids	0	0	0	0-4	0	0

B

Fig. 2.—(*A*) General genomic characteristics for isolates PAL113 and PAL227 compared with near-neighbors in four genera. Values are based on genomes used in functional analysis. Genome sizes (megabasepairs, Mb), G + C%, and the number of protein coding sequences (CDS) are given as averages for the available *Dorea* (n = 15), *Ruminococcus* (n = 12), *Faecalicatena* (n = 3), and *Mediterraneibacter* (n = 5) species for each genus. (*B*) A comparison of COGs between the *P. americana* isolates and other Lachnospiraceae genera. The proportion of genes is averaged within each group for each COG category. The overall mean across groups, excluding PAL113 and PAL227, is represented by a dashed line. Differences are marked between PAL113 (+) or PAL227 ($) and the mean proportion of other groups in the same category when they are greater than 2.5×, the group standard deviation.
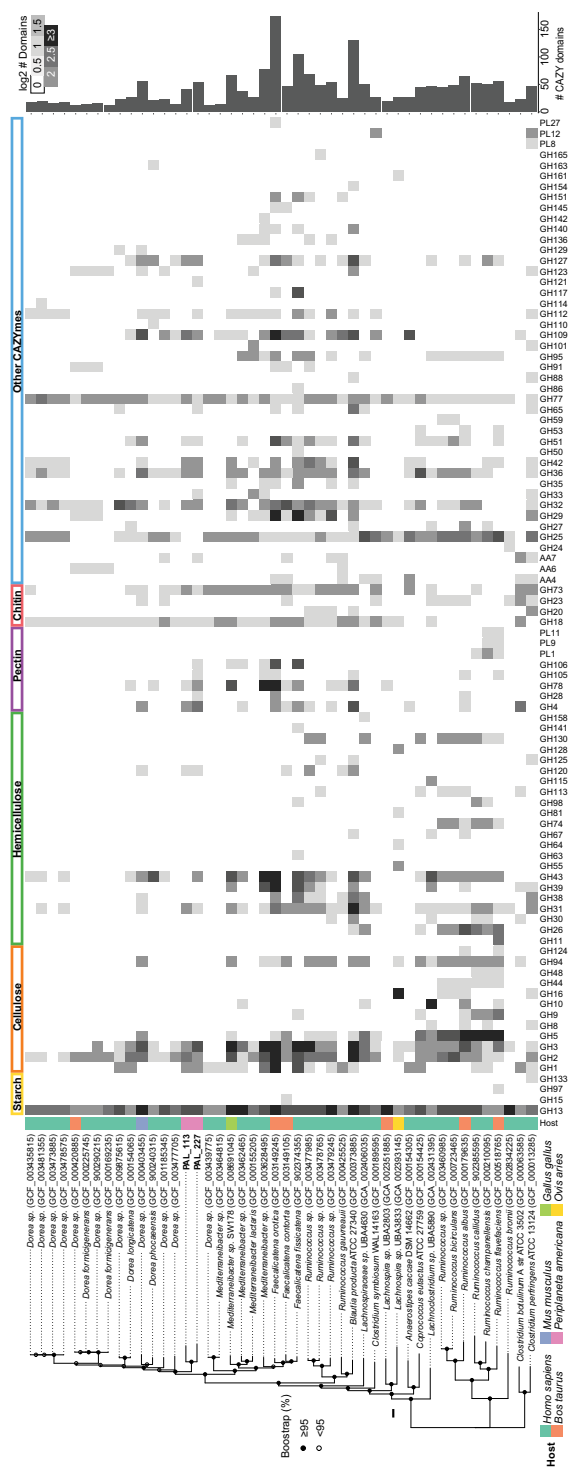
FIG. 3.—Carbohydrate active enzyme (CAZy) coding the gene profile of PAL113, PAL227, and other Lachnospiraceae genomes. Heatmap shows the log2-normalized number of CAZy domains encoded in the genome. Putative polysaccharide targets are indicated at the top of each set of CAZy domains based on DRAM annotations and descriptions. Multilocus phylogeny is from figure 1. The colors next to the tree tip indicate the isolate source. The bar plot indicates the total number of CAZy encoding domains annotated in each genome.

(Giraud and Naismith 2000; Jiang et al. 2021). Although both isolates encode complete propionate biosynthesis pathways (supplementary fig. S4, Supplementary Material online), PAL227 encodes a complete butyrate pathway, whereas genes encoding the last step in the conversion of butanoyl-CoA to butyrate PAL113 were not detected. Propionate production from rhamnose can allow for symbiont-mediated metabolism of bacterial sugars, whereas propionate production from fucose can enable the recycling of host peritrophic membrane glycans. The position of the rhamnose-degradation genes adjacent to the propionate operon in PAL113 may suggest that they are co-transcribed and that rhamnose is a preferred substrate for propionate production in this isolate. *P. americana* diets often include plant-derived ingredients, providing PAL113 and PAL227 with ample substrate. Given that peritrophic matrix-associated fucoses are also likely abundant in the insect gut, PAL113 or PAL227 may use fucose-derived substrates liberated by other *P. americana* gut bacteria, including several fucose-degrading Bacteroidetes species (Vera-Ponce de León et al. 2020), but neither encode fucose-degrading enzymes. Finally, complete canonical flagellar biosynthesis pathways in the PAL113 and PAL227 genomes, which may facilitate host colonization and penetration of insect peritrophic matrices (Kajikawa et al. 2016, 2018), were unexpected as they were atypical across Lachnospiraceae genomes (supplemental fig. S3, Supplementary Material online).

Overall, PAL113 and PAL227 are broadly equipped with a diverse array of enzymes that will enable them to exploit their host's diet and inhabit their digestive tissues and/or benefit their host by contributing to complex carbohydrate degradation and producing valuable SCFAs. *P. americana* are generalist feeders with diets that are rich in difficult-to-digest plant fibers, and they harbor a highly diverse gut bacterial community (Bertino-Grimaldi et al. 2013; Sabree and Moran 2014; Tinker and Ottesen 2016; Vera-Ponce de León et al. 2020). Further evidence is needed to ascertain if PAL113 and PAL227 contribute to overall host dietary processing and nutrient provisioning. It is tempting to hypothesize from the PAL113 and PAL227 genomes that these isolates may be in a mutually beneficial relationship with *P. americana* because of the observed cladogenesis and the presence of genes that encode putatively host gut-adaptive and host-beneficial functions. An equally plausible hypothesis is that cladogenesis is the result of genetic drift and population bottlenecks due to these oxygen-sensitive isolates relying upon host behaviors (i.e., coprophagy) for dispersal. A polyspecific inoculum of *P. americana* commensals ($n = 13$), which included PAL113 and PAL227, dampened energy stress responses in gnotobiotic *P. americana* when compared with germ-free *P. americana* (Vera-Ponce de León et al. 2021), and these results
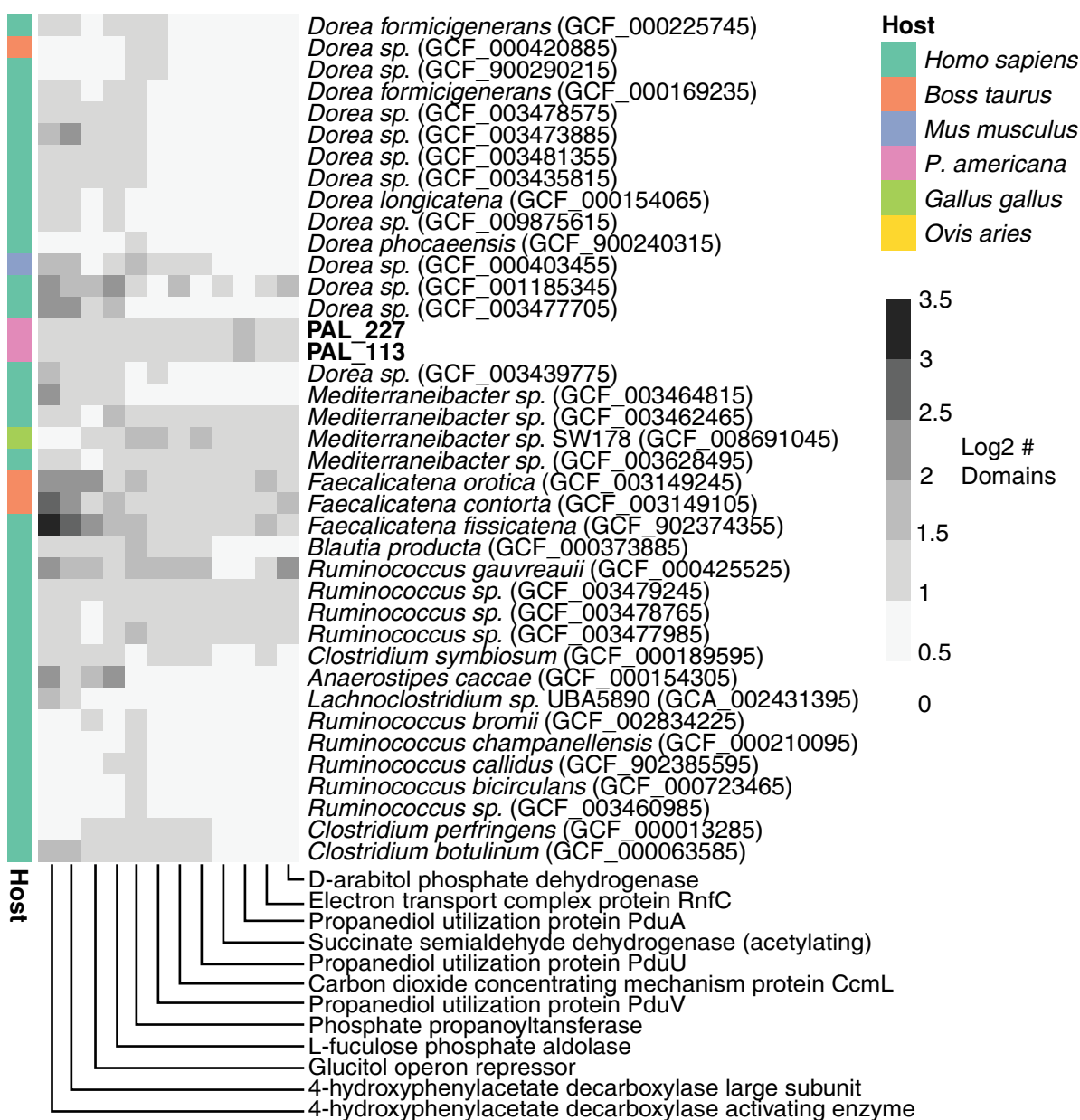
FIG. 4.—Propanediol utilization (PDU) operon in PAL227 and PAL113 encodes for key proteins in the production of short-chain fatty acids. The heatmap of gene frequency is based on the annotations of *P. americana* isolate genomes and publicly available Lachnospiraceae genomes. The colors in the rows indicate the host from which each genome was isolated. Complete PDU operon maps of PAL113 and PAL227 are illustrated in supplementary figure S4 (Supplemental Materials online).

suggest that they may be able to contribute to host health by assisting with dietary processing. The *P. americana* gut microbial community includes taxa spanning several phyla, and further genomic and experimental investigations, especially using gnotobiotic insects (e.g., Jahnes et al. 2019), are warranted to address these hypotheses, ascertain the degree to which the host relies upon specific lineages for normal growth and development, and determine the amount of functional redundancy that exists between commensals comprising the gut community.

Proposed Names

Aequitasia blattaphilus gen. nov., sp. nov. (Ae.qui.tas'ia L. n. *aequitas*, equity; bla.tta'phi.lus L. n. blatta, light-shunning insect; L. adj. philus friend, loving). The type strain is PAL113.

Ohessyouella blattaphilus gen. nov., sp. nov. (Oh.ess'. you.el.la. L. dim. n. "Ohessyou" is a written pronunciation of the acronym of Ohio State University, where several substantive contributions to the natural sciences have been

made and the strain was isolated; bla.tta'phi.lus L. n. blatta, light-shunning insect; L. adj. philus friend, loving). The type strain is PAL227.

Materials and Methods

PAL113 and PAL227 were isolated from gut tissues of adult *P. americana* using procedures described previously (supplemental text, Supplementary Material online; Vera-Ponce de León et al. 2020). Genomic DNA was extracted from isolate cultures using the DNeasy blood and tissue kit (Qiagen) following the manufacturer's instructions for Gram-positive bacteria. Bacterial DNA was used as a template to amplify near-full-length 16S rRNA genes with 27F (5′-AGA GTT TGA TCCTGG CTC AG-3′) and 1492R (5′-GGT TAC CTT GTT ACG ACT T-3′) primers. Amplicons were produced using the following thermocycler parameters: 1 cycle of 97 °C for 5 min; 40 cycles of 97 °C for 30 s, 40–55 °C for 30 s, and 68 °C for 2 min; and 1 cycle of 68 °C for 15 min. The amplification of 1.5-kbp products was confirmed by gel electrophoresis, and the products were purified using the DNA Clean and Concentrator kit according to the manufacturer's instructions (Zymo Research, Irvine, CA, USA). 16S rRNA gene amplicons were sequenced by Sanger capillary chemistry and manually checked to ensure a single read per base position in roughly 90% of the sequence; sequences meeting this criterion were deemed pure. These sequences were matched to their closest bacterial relative by BLASTN (Camacho et al. 2009) search for classification. To refine classification, 16S sequences were searched in the SILVA bacterial rRNA database with SINA (v.1.2.11) (Pruesse et al. 2012).

Isolate genomes were sequenced on an Illumina HiSeq platform using the Illumina Nextera XT Library kit. Before assembly, read quality trimming and adaptor removal were performed with TrimGalore v. 0.4.1 (parameters: -q 30 –phred 33 –illumina –paired; https://github.com/FelixKrueger/TrimGalore). Trimmed, high-quality reads were then assembled using SPAdes v. 3.7. Scaffolds were retrieved by SSPACE 1.3 with BWA as an aligner (parameters: -x 0 -z 0 -k 0 -g 0 -a 0.70 -n 15 -T 30 -p 0). Prodigal v. 2.6 was used with default parameters for the prediction of ORFs and CDS in the assembled genomes.

Genome completeness was evaluated with Benchmarking Universal Single-Copy Orthologs (BUSCO; v. 4.1.4) (Simão et al. 2015) in protein mode with the clostridiales_odb10 lineage data set (number of species: 347, number of BUSCOs: 264).

Phylogenetic Analysis

Close to full-length 16S rDNA sequences were obtained for 83 Lachnospiraceae taxa and aligned with PAL113 and PAL227 16S rDNA sequences (supplementary table S1, Supplemental Material online) using MAFFT (parameters: G-INS-I; version 7.475) (Katoh and Standley 2013). A concatenated protein-based phylogeny was constructed using 219 orthologous proteins selected from 44 Lachnospiraceae taxa (supplementary table S2, Supplementary Material online) that were concatenated and aligned using the multigenome2blocks pipeline developed for this work (https://github.com/avera1988/multigenome2blocks). RAxML (parameters: GTRGAMMA model of substitution; version 8.2.12) was used for phylogenetic reconstructions and visualized using the Interactive Tree of Life (https://itol.embl.de/; version 4) (Letunic and Bork 2021).

Pan and core genome analyses were conducted using GETHOMOLOGUES 2.0 (Contreras-Moreira and Vinuesa 2013) and OrthoMCL (Li 2003) was used for ortholog clustering (parameters: -A -c -t 0 -M -n 35). Core genes corresponding to each genome were (supplemental table S2, Supplementary Material online) parsed from the "GETHOMOLOGUES" pangenome matrix results using custom Bash and Perl scripts (https://github.com/avera1988/Comparative_genomics), and they were used to calculate the average AAI between Lachnospiraceae genomes using the AAI calculator from the "enveomics" collection tools (Rodriguez-R and Konstantinidis 2016). AAI distance matrices were calculated and visualized by custom Perl and R scripts (deposited in https://github.com/avera1988/Comparative_genomics). BLASTN search of PAL113 and PAL227 16S sequences in the nt/nr database was performed with default parameters on May 12, 2021 (Camacho et al. 2009).

Functional Analysis and Comparative Genomics

PROKKA pipeline version 1.12 (Seemann 2014) was used for open reading frame discovery and gene annotation of PAL227 and PAL113 genomes, with SignalP (Almagro Armenteros et al. 2019) for Gram-positive signal peptides and Infernal (Nawrocki and Eddy 2013) for noncoding RNA annotation (–gram -/neg –rfam parameters, respectively). Protein function and pathways were predicted for PAL227 and PAL113, as well as 38 unique publicly available genomes in the Lachnospiraceae family, the names and accessions of which can be found in Supplemental Materials. Functional annotation of proteins for all genomes was performed by running RPS-BLAST+ against the NCBI conserved domain database, then assigning categories to proteins from the COG database using the cdd2cog pipeline (https://github.com/aleimba/bac-genomics-scripts). Predicted metabolic pathways of the proteins in all genomes were also annotated with the KEGG database using the KEGG GhostKoala tool (Kanehisa et al. 2016) and the KEGGREST bioconductor package for R. After completing COG and KEGG annotations, further specialized analysis of isolates PAL113 and PAL227 was performed for categories and pathways of interest. Carbohydrate active enzymes

(CAZy) coding genes and major SCFA anabolic genes were annotated using the Distilled and Refined Annotation of Metabolism (DRAM) tool version 1.1.1 (Shaffer et al. 2020). Proteins from PAL113, PAL227, and public genomes were compared with all HMM profiles in the TXSSscan database (Abby et al. 2016) using the "hmmscan" (parameters: hmmscan –cpu 40 –domtblout) tool from HMMER 3.1b2 (Eddy 2011) to identify genes encoding domains that comprised potential components of flagellar and bacterial secretion systems.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

Data Availability

The isolate genome assembly data have been deposited under NCBI BioProject accession number PRJNA842349 in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/). All relevant data are also available from the authors upon request.

Literature Cited

Abby SS, et al. 2016. Identification of protein secretion systems in bacterial genomes. Sci Rep. 6(1):23080.

Almagro Armenteros JJ et al. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol. 37(4):420–423.

Bell W, Adiyodi K. 1982, editors. The American cockroach. London: Chapman and Hall.

Bertino-Grimaldi D, et al. 2013. Bacterial community composition shifts in the gut of *Periplaneta americana* fed on different lignocellulosic materials. Springerplus 2(1):609.

Biddle A, Stewart L, Blanchard J, Leschine S. 2013. Untangling the genetic basis of fibrolytic specialization by Lachnospiraceae and Ruminococcaceae in diverse gut communities. Diversity 5(3): 627–640.

Camacho C, et al. 2009. BLAST+: Architecture and applications. BMC Bioinformatics 10(1):421.

Contreras-Moreira B, Vinuesa P. 2013. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl Environ Microbiol. 79(24):7696–7701.

Eddy SR. 2011. Accelerated profile HMM searches. PLOS Comput Biol. 7(10):e1002195.

Erlandson MA, Toprak U, Hegedus DD. 2019. Role of the peritrophic matrix in insect-pathogen interactions. J Insect Physiol. 117:103894.

Giraud MF, Naismith JH. 2000. The rhamnose pathway. Curr Opin Struct Biol. 10(6):687–696.

Groussin M, Mazel F, Alm EJ. 2020. Co-evolution and co-speciation of host-gut bacteria systems. Cell Host Microbe 28(1):12–22.

Huttenhower C, et al. 2012. Structure, function and diversity of the healthy human microbiome. Nature 486:207–214.

Jahnes BC, Poudel K, Staats AM, Sabree ZL. 2021. Microbial colonization promotes model cockroach gut tissue growth and development. J Insect Physiol. 133:104274.

Jahnes BC, Herrmann M, Sabree ZL. 2019. Conspecific coprophagy stimulates normal development in a germ-free model invertebrate. PeerJ. 7(5):e6914.

Jiang N, Dillon FM, Silva A, Gomez-Cano L, Grotewold E. 2021. Rhamnose in plants – from biosynthesis to diverse functions. Plant Sci. 302:110687.

Kajikawa A, et al. 2016. Characterization of flagellins isolated from a highly motile strain of *Lactobacillus agilis*. BMC Microbiol. 16:49.

Kajikawa A, Suzuki S, Igimi S. 2018. The impact of motility on the localization of *Lactobacillus agilis* in the murine gastrointestinal tract. BMC Microbiol. 18:68.

Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 428:726–731.

Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.

Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 49(W1):W293–W296.

Li L. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13(9):2178–2189.

Liu D, et al. 2019. The N-glycan profile of the peritrophic membrane in the Colorado potato beetle larva (*Leptinotarsa decemlineata*). J Insect Physiol. 115:27–32.

Luis AS, et al. 2018. Dietary pectic glycans are degraded by coordinated enzyme pathways inhuman colonic Bacteroides. Nat Microbiol. 3:210.

Meehan CJ, Beiko RG. 2014. A phylogenomic view of ecological specialization in the Lachnospiraceae, a family of digestive tract-associated bacteria. Genome Biol Evol. 6(3):703–713.

Naas AE, et al. 2018. 'Candidatus Paraporphyromonas polyenzymogenes' encodes multi-modular cellulases linked to the type IX secretion system. Microbiome 6:44.

Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29(22):2933–2935.

Pérez-Cobas AE, et al. 2015. Diet shapes the gut microbiota of the omnivorous cockroach *Blatella germanica*. FEMS Microbiol Ecol. 91(4):fiv022.

Pruesse E, Peplies J, Glöckner FO. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 28(14):1823–1829.

Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Prepr. 4:e1900v1.

Sabree ZL, Moran NA. 2014. Host-specific assemblages typify gut microbial communities of related insect species. Springerplus. 3:138.

Salyers AA, West SEH, Vercellotti JR, Wilkins TD. 1977. Fermentation of mucins and plant polysaccharides by anaerobic bacteria from the human colon. Appl Environ Microbiol. 34(5): 529–533.

Schauer C, Thompson CL, Brune A. 2012. The bacterial community in the gut of the cockroach *Shelfordella lateralis* reflects the close evolutionary relatedness of cockroaches and termites. Appl Environ Microbiol. 78(8):2758–2767.

Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30(14):2068–2069.

Shaffer M, et al. 2020. DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Res. 48(16):8883–8900.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19):3210–3212.

Tellam RL, Wijffels G, Willadsen P. 1999. Peritrophic matrix proteins. Insect Biochem Mol Biol. 29(2):87–101.

Terra WR. 2001. The origin and functions of the insect peritrophic membrane and peritrophic gel. Arch Insect Biochem Physiol. 47(2):47–61.

Tinker KA, Ottesen EA. 2016. The core gut microbiome of the American cockroach, *Periplaneta americana*, is stable and resilient to dietary shifts. Appl Environ Microbiol. 82(22): 6603–6610.

Turnbaugh PJ, et al. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 444:1027–31.

Vacca M, et al. 2020. The controversial role of human gut Lachnospiraceae. Microorganisms. 8(4):573.

Vera-Ponce de León A, Jahnes BC, Duan J, Camuy-Vélez LA, Sabree ZL. 2020. Cultivable, host-specific bacteroidetes symbionts exhibit diverse polysaccharolytic strategies. Appl Environ Microbiol. 86(8): e00091-20.

Vera-Ponce de León A, Jahnes BC, Otero-Bravo A, Sabree ZL. 2021. Microbiota perturbation or elimination can inhibit normal development and elicit a starvation-like response in an omnivorous model invertebrate. mSystems. 6(4):e0080221.

Wada-Katsumata A, et al. 2015. Gut bacteria mediate aggregation in the German cockroach. Proc Natl Acad Sci U S A. 112(51): 15678–15683.

Associate editor: Emmanuelle Lerat