Title

The phenotype and genotype of fermentative prokaryotes.

Journal

Science Advances, 9(39)

Authors

Zhang, Bo

Hackmann, Timothy

Publication Date

2023-09-29

DOI

10.1126/sciadv.adg8687

Peer reviewed

MICROBIOLOGY

The phenotype and genotype of fermentative prokaryotes

Timothy J. Hackmann[1]* and Bo Zhang[2]

Fermentation is a type of metabolism pervasive in oxygen-deprived environments. Despite its importance, we know little about the range and traits of organisms that carry out this metabolism. Our study addresses this gap with a comprehensive analysis of the phenotype and genotype of fermentative prokaryotes. We assembled a dataset with phenotypic records of 8350 organisms plus 4355 genomes and 13.6 million genes. Our analysis reveals fermentation is both widespread (in ~30% of prokaryotes) and complex (forming ~300 combinations of metabolites). Furthermore, it points to previously uncharacterized proteins involved in this metabolism. Previous studies suggest that metabolic pathways for fermentation are well understood, but metabolic models built in our study show gaps in our knowledge. This study demonstrates the complexity of fermentation while showing that there is still much to learn about this metabolism. All resources in our study can be explored by the scientific community with an online, interactive tool.

INTRODUCTION

Fermentation is a major type of metabolism carried out in the absence of oxygen. During fermentation, organic molecules (e.g., glucose) are catabolized and donate electrons to other organic molecules. In the process, adenosine 5′-triphosphate (ATP) and organic end products (e.g., lactate) are formed. Because fermentation forms ATP without using O_2, prokaryotes in the environment use it as one alternative to aerobic respiration. The gut (*1*, *2*), sediments (*3*), and anaerobic bioreactors (*4*) are just some environments where oxygen is scarce and fermentative microbes are common. Fermentation is also important because of the end products it forms. These products are metabolized by human and animal hosts (*1*, *5*, *6*), present in food (*7*), and valuable as biofuels or other chemicals (*8*, *9*). Fermentation is thus important in many contexts.

Despite the importance of fermentation, we know little about the range of prokaryotes that carry it out and their traits. There are excellent reviews of fermentation (*10*–*15*), but their focus is on a few model organisms and their biochemical pathways. Information on more organisms is available in journal articles or book chapters (*16*, *17*), but each covers just a few (usually related) organisms. With no central resource for information, it is hard to answer even simple questions about the prokaryotes carrying out fermentation.

Our laboratory and others have started to collect information on fermentative prokaryotes. Our laboratory has cataloged prokaryotes that carry out fermentation and which form one end product (acetate) (*18*, *19*). Another group has cataloged end products of fermentation, but the scope is limited to bacteria of the human gut (*20*). Other groups have collected information on fermentation when building databases on microbial phenotypes (*21*, *22*). Given the scope of these databases, information specific to fermentation tends to be less complete. Although limited in scope or completeness, these resources provide a good starting point and, if expanded, could give a full picture of fermentative prokaryotes.

Here, we assemble a large dataset on fermentative prokaryotes and use it to explore the phenotype and genotype of such organisms. This dataset includes phenotypic records on n = 8350 organisms (prokaryotes) as well as n = 4355 genomes and n = 13.6 million genes. With it, we have answered simple but important questions about fermentation. We also built an interactive tool for the microbiology community to explore our dataset and make predictions about organisms of interest.

RESULTS

Our dataset contains information on more than 8300 prokaryotes

To examine fermentative prokaryotes, we assembled a dataset from multiple sources (Fig. 1 and data S1 and S2). Our starting point was n = 1828 articles from *Bergey's Manual of Systematics of Archaea and Bacteria* (*16*) and the primary literature. From these articles, we obtained names, written descriptions, and phenotypic information for n = 8350 organisms (type strains). The phenotypic information we obtained was for traits related to fermentation (fermentative ability, fermentation end products, and fermentation substrates). We used computational approaches to automate some steps (e.g., obtaining names of organisms). Other steps (e.g., obtaining phenotypic information) were done by manually reading articles.

We obtained genomic and more phenotypic information from additional sources. We used Genomes OnLine Database (GOLD) (*23*) and Integrated Microbial Genomes and Microbiomes (IMG/M) (*24*) database to obtain information on genomes and genes from these organisms. We also used HydDB (*25*) for information on genes for hydrogenases. Using BacDive database (*21*), we obtained phenotypic information for n = 14 traits. The traits (e.g., cell length and habitat) were those not directly related to fermentation. Data from BacDive comes from other databases (*26*), the primary literature, and other sources. We considered other sources of information. For example, we considered one source with data for n = 701 prokaryotes of the human gut (*20*). However, only n = 279 of these were type strains, and n = 234 (84%) were already in our dataset. Furthermore, the source used

[1]Department of Animal Science, University of California, Davis, CA, USA. [2]Department of Chemical Engineering, University of California, Santa Barbara, CA, USA.
*Corresponding author. Email: tjhackmann@ucdavis.edu

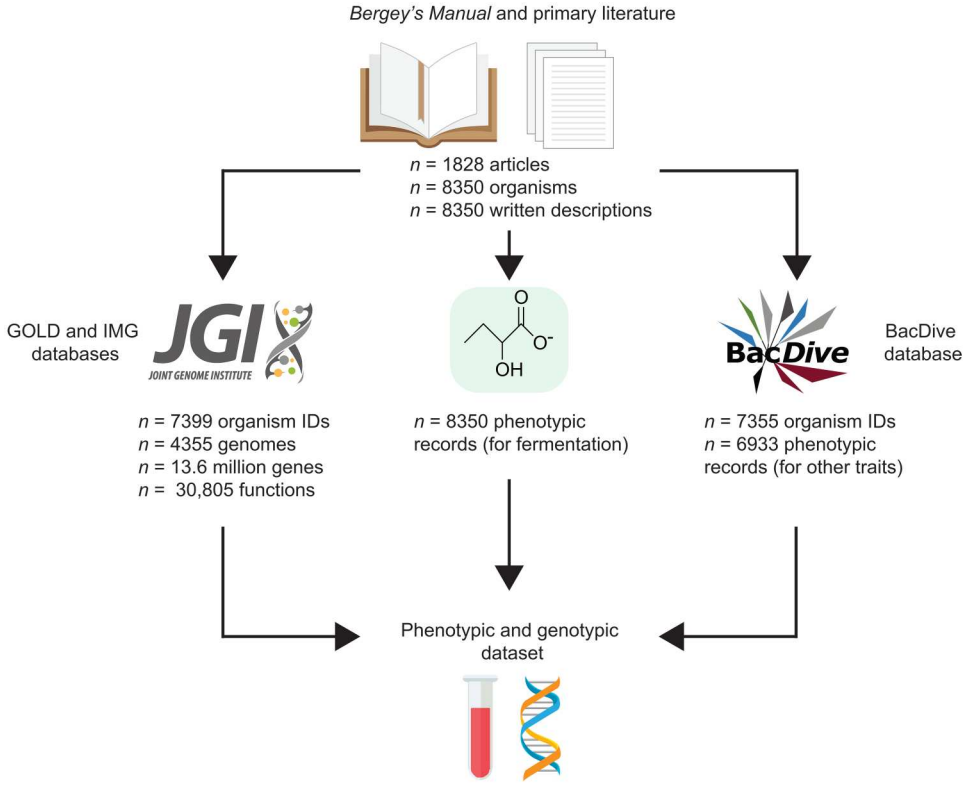

Fig. 1. To study fermentative prokaryotes, we assembled a phenotypic and genotypic dataset. The dataset includes *n* = 8350 organisms (type strains).

genus-level data when information for specific organisms was not available. Our dataset included only strain-level data, and we chose not to mix these two different types of information. In summary, our dataset was comprehensive. It allowed us to probe both the phenotype and genotype of fermentative prokaryotes.

Fermentative prokaryotes are diverse

With our dataset, we first wanted to answer first how widespread fermentation is across prokaryotes. In total, we found that *n* = 2357 of the 8350 organisms were capable of fermentation or 28% of the total (data file S1). In earlier work with a smaller dataset (*19*), we found a comparable value (33%).

We then built a phylogenetic tree for the *n* = 3822 organisms with enough information (ribosomal protein sequences) available (Fig. 2A). Fermentative prokaryotes were found all over the tree, although they appeared more abundant on some branches than others. The phylum *Bacillota* was one group in which they were abundant. Nonfermentative prokaryotes were abundant in *Actinomycetota*.

We quantified the difference in abundance across phyla using a statistic known as risk difference (see fig. S1A for definition). This analysis was done with *n* = 7394 organisms. We found the risk difference was highest for *Bacillota* and lowest for *Actinomycetota* (Fig. 2B), in agreement with the tree (Fig. 2A). It identified *Thermodesulfobacteriota*, *Spirochaetota*, and *Mycoplasmatota* as other phyla where fermentative prokaryotes were abundant.

Risk difference is only one example of an effect size statistic. Another commonly used statistic is the log-odds ratio. When we calculate the log-odds ratio, the results are generally similar to those with the risk difference (fig. S2). However, many values were undefined (negative or positive infinity), which makes interpretation more difficult. Together, these results show the ability to ferment is found all over the tree of life, being particularly common for members of *Bacillota* and certain other phyla.

Fermentative prokaryotes have distinct phenotypic traits

After identifying which prokaryotes were fermentative, we wanted to answer whether their phenotype differed from that of nonfermentative prokaryotes. We examined *n* = 14 types of phenotypic traits, such as habitat, that are not immediately related to fermentation. We calculated the risk difference to determine whether traits were more common in fermentative versus nonfermentative prokaryotes. For continuous traits, we instead calculated a statistic we call the probability difference. The probability difference is simply the difference in the probability density functions between the two groups (see fig. S1B for definition).

We found several traits were more common in fermentative versus nonfermentative prokaryotes (Fig. 3 and figs. S3 to S5). Some differences in traits, such as oxygen tolerance, were expected. Fermentative prokaryotes were more likely to be oxygen intolerant (anaerobes or microaerophiles) (Fig. 3A), which reflects that fermentation does not use O_2. They were also more likely to be isolated from host-associated habitats (Fig. 3A and figs. S3 and S4). Many host-associated habitats, such as the gut, are well known to harbor fermentative prokaryotes (*1*, *2*).

Some differences in traits were more unexpected. For example, fermentative prokaryotes were also more likely to grow faster (with shorter incubation period) (Fig. 3B) and have longer cell length

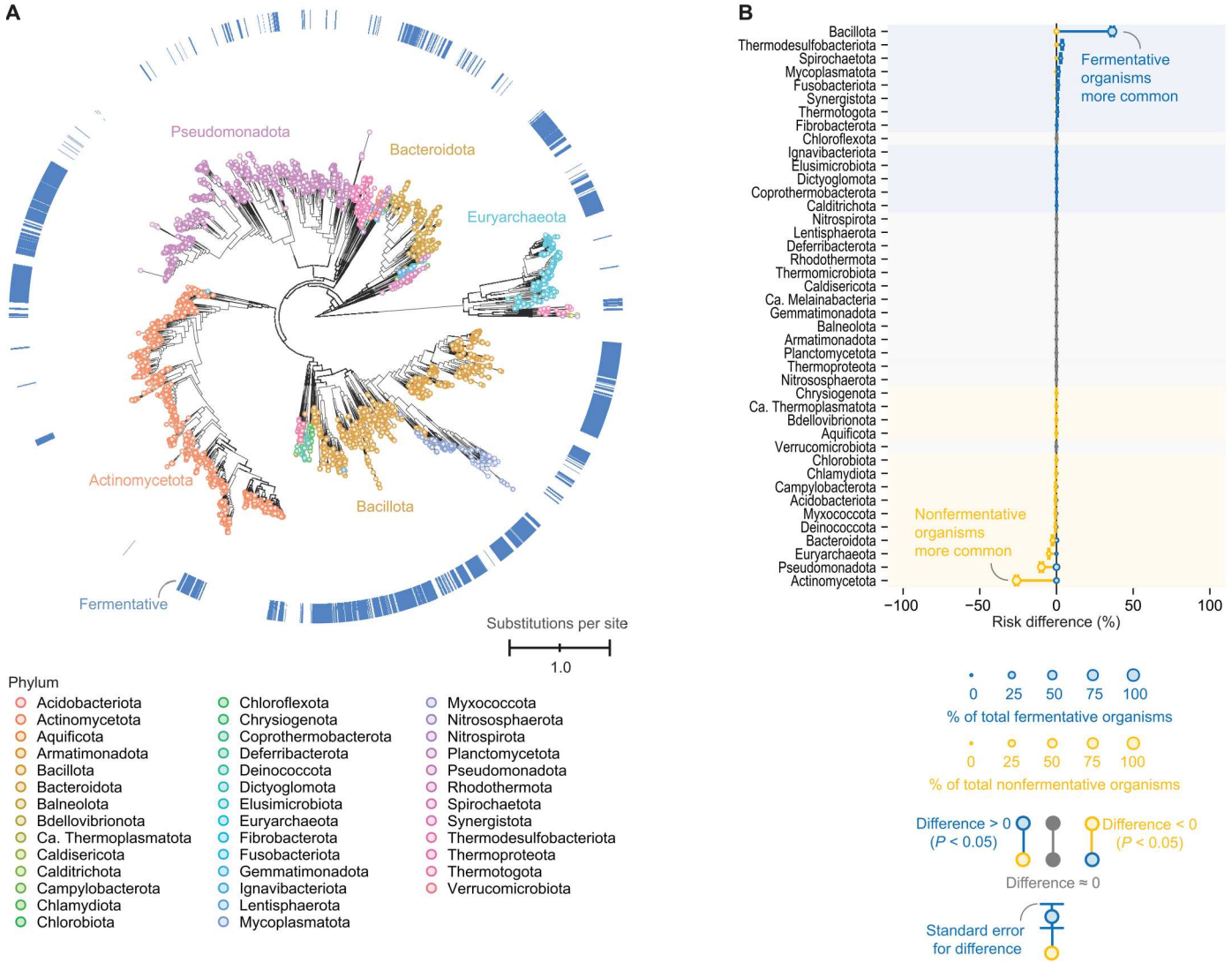

Fig. 2. Fermentation is widespread, although it is more common in some groups of prokaryotes than others. (**A**) Phylogenetic tree of n = 3822 prokaryotes and their fermentative ability. (**B**) Phyla of n = 7394 prokaryotes and their fermentative ability. Differences between fermentative and nonfermentative prokaryotes were determined using a two-tailed z test. Names of phyla correspond to the NCBI taxonomy of organisms (see data file S1). Ca., *Candidatus*.

(Fig. 3B). Examining the raw data confirms a clear difference in cell length (fig. S6). Although this was unexpected, it has been observed that growing cells under microaerophilic or anaerobic conditions increases length (*27*, *28*).

Although the differences were large, they were not as stark as they might be imagined to be. Fermentative prokaryotes were less likely than nonfermentative prokaryotes to be aerobic, but 18% of them still were (see Fig. 3A). The aerobic, fermentative prokaryotes belonged mostly to *Paenibacillus*, *Corynebacterium*, *Staphylococcus*, *Vibrio*, *Enterococcus*, and *Streptococcus*. End products and substrates for fermentation have been reported for many of these aerobic prokaryotes (see data file S1), confirming that they are fermentative.

Our analysis above considered all prokaryotes, and we next examined anaerobic prokaryotes only. We examined a total of n = 1496 anaerobic organisms, n = 1029 of which were fermentative. Even when looking at this subset of data, we still found that many traits were more common among fermentative prokaryotes (fig. S7). As before, fermentative prokaryotes were more likely to be isolated from host-associated habitats, grow with shorter incubation time, and have longer cells. Thus, their long cells may owe partly to growth under anaerobic conditions (*27*, *28*), but it is not the full explanation. These results show that fermentative prokaryotes differ from anaerobic prokaryotes at large.

The analyses above examined phenotypic traits at a broad level. Our next analysis focused on metabolic traits specifically. To help carry out this analysis, we used FAPROTAX, a tool that predicts traits from organism names (taxonomy). After inputting names of the n = 8350 organisms in our dataset, this tool outputted n = 72 metabolic traits of n = 6065 organisms. We found several of these traits differed between fermentative versus nonfermentative prokaryotes (fig. S8A). For example, fermentative prokaryotes were more likely to carry out respiration with sulfur, iron, and arsenate. Similar to fermentation, these types of respiration do not use

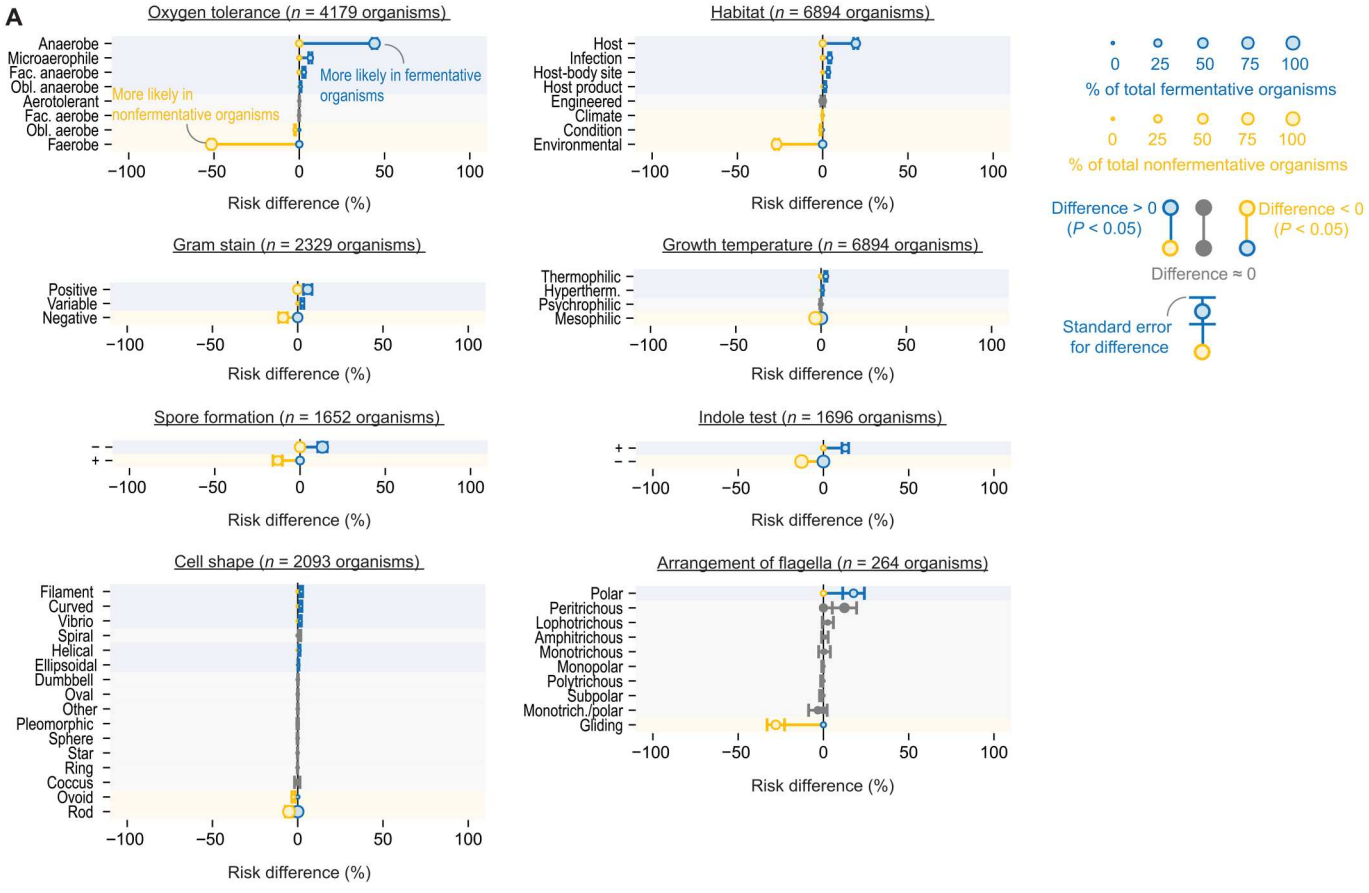

Fig. 3. Examining phenotypic records of *n* = 6933 prokaryotes shows that fermentative prokaryotes have distinct traits. (**A**) Discrete traits. The number of organisms analyzed per trait is shown in the figure. Differences between fermentative and nonfermentative prokaryotes were determined using a two-tailed *z* test. Habitat refers to the category 1 isolation source. See figs. S3 and S4 for category 2 and category 3 isolation sources. Fac., facultative; hypertherm., hyperthermophilic; monotrich., monotrichous; obl., obligate. (**B**) Continuous traits. All traits differed between fermentative and nonfermentative prokaryotes according to a two-sample, two-sided Kolmogorov-Smirnov test ($P < 0.001$ for cell length, $P < 0.001$ for incubation period, and $P = 0.002$ for optimum salt for growth). See fig. S5 for traits that did not differ according to this test. (**C**) Summary of the most common phenotypic traits of fermentative prokaryotes.

oxygen and are alternatives to aerobic respiration. This shows that fermentative prokaryotes are versatile and can carry out multiple types of metabolism independent of oxygen.

While traits outputted by FAPROTAX are illuminating, they are based on an organism's taxonomy and are thus ultimately predictions. To determine their reliability, we compared prokaryotes predicted as fermentative by FAPROTAX with those observed as fermentative in our dataset (fig. S8B). We found good agreement, although our dataset identified 54% more organisms as fermentative than did FAPROTAX. This shows value of tools such as FAPROTAX and the need for more focused datasets such as ours.

These results show that fermentative prokaryotes have a distinct phenotype (Fig. 3C). Some aspects of this phenotype, such as oxygen tolerance, are easy to explain. Others, such as cell length, are unexpected but clear. Most fermentative prokaryotes are anaerobic, but many are not, and they still have a phenotype distinct from anaerobic prokaryotes.

Fermentative prokaryotes have a distinct genotype

Next, we wanted to see whether fermentative and nonfermentative prokaryotes differed in genotype. We looked at their genes and whether their predicted functions differed. Our analysis included n = 13.6 million genes and n = 30805 predicted functions. These were from n = 4355 organisms, n = 1490 of which are fermentative. The predicted functions correspond to Kyoto Encyclopedia of Genes and Genome Orthology (KO) (*29*), Clusters of Orthologous Genes (COG) (*30*), pfam (*31*), and TIGRFAM (*32*) IDs. We also included functions corresponding to HydDB names (*25*).

We looked for general differences in gene functions by using *t*-distributed stochastic neighbor embedding (t-SNE) (*33*). This technique projected all n = 30,805 gene functions onto a two-dimensional plot (Fig. 4A). In that plot, fermentative and nonfermentative prokaryotes generally fell into two different groups, with some overlap. This suggests that fermentative prokaryotes have different gene functions from nonfermentative ones, although it does not identify which specific functions differ.

To identify specific functions that differed, we calculated the risk difference for each function. With this approach, we found n = 9450 predicted gene functions that were more common in fermentative versus nonfermentative prokaryotes (Fig. 4B and data S3) ($P < 0.05$). We examined gene functions corresponding to KO IDs first (Fig. 4B). Of the top 15 KO IDs, several are involved production of fermentation end products (e.g., K00656) or their utilization (e.g., K21636). We next looked at COG, pfam, and TIGRFAM IDs. Most predicted functions were similar to those of KO IDs, although COG and pfam had more unknown functions (data file S3 and below).

Most gene functions were clear and well defined, but many were not. In total, n = 1229 gene functions more common in fermentative prokaryotes were "uncharacterized" or of "unknown function" (data file S3). We found or predicted protein structures for several of these (Fig. 4C). Despite having no established function, they appear to play an important role in fermentation. In one case, this has been verified; COG3610 is still reported as uncharacterized in its database, but one member was recently characterized as a transporter of a fermentation product (succinate) (*34*).

We also found n = 13,945 gene functions that were more common in nonfermentative prokaryotes (Fig. 4B and data file S3) ($P < 0.05$). Many corresponded to enzymes used in oxidative phosphorylation (e.g., K02275).

Besides looking at the level of individual genes and functions, we examined whether certain metabolic pathways were more common. To do this, we examined the abundance of all n = 3445 metabolic pathways in MetaCyc (data file S4) (*35*). The prokaryotes in our dataset encoded n = 1579 of these pathways, and n = 337 were more abundant in fermentative ones ($P < 0.05$). The top 15 pathways in fermentative prokaryotes included those for production of fermentation products (e.g., PWY-5481) or their utilization (e.g., PWY-1722). More expectedly, there were pathways involved in quorum sensing (e.g., PWY-6154) and nucleotide metabolism. Thus, pathways were as expected, although a few surprises were also present.

Our analysis above considered all prokaryotes, and we next examined anaerobic prokaryotes only. We examined a total of n = 1039 anaerobic organisms, n = 745 of which were fermentative. Again, we found gene functions and pathways more common in fermentative prokaryotes (data files S5 and S6). Furthermore, they were similar to functions and pathways identified above. Of the top 15 KO IDs in Fig. 4B, 13 were still more common in fermentative versus nonfermentative prokaryotes (see data file S5). The situation was similar for MetaCyc pathways, with 11 of the 15 top pathways still being more common (see data S6). These results show the genotype of fermentative prokaryotes differs from anaerobic prokaryotes at large.

These results together show that fermentative prokaryotes have a distinct genotype. Intriguingly, several genes appear to be important in fermentation but have no known function. These are targets for further study in understanding fermentation and protein function.

Fermentation forms many end products

Another important trait of fermentative prokaryotes is which end products of fermentation they form. We collected information on end products for n = 1455 organisms. On the basis of the text of the written description, we divided these into major and minor end products. We also recorded the n = 100 substrates reported to form these end products (information available for n = 1260 organisms).

In total, we found prokaryotes formed n = 55 fermentation products (Fig. 5, A and B). Acetate and lactate were the most common products, with at least one being formed by 97% of organisms. Most (83%) organisms formed multiple organic products (Fig. 5C). Of the organisms that formed only one organic product, nearly all formed lactate (56%) or acetate (34%). There were many (n = 289) unique combinations of products altogether.

We also examined end products formed by substrate (Fig. 5, A and B). We focused on the n = 46 substrates that are single, chemically defined molecules (e.g., glucose). This represented information for n = 805 organisms. Glucose was the most common substrate reported, and acetate was the most common product (Fig. 5A). Other common products were as before, although the order of their abundance differed somewhat (Fig. 5, A and B).

The number of end products in our dataset is higher than in other sources. It is >3-fold higher than reported for prokaryotes of the human gut (*20*) (fig. S9). The large size of our dataset makes it a useful resource.

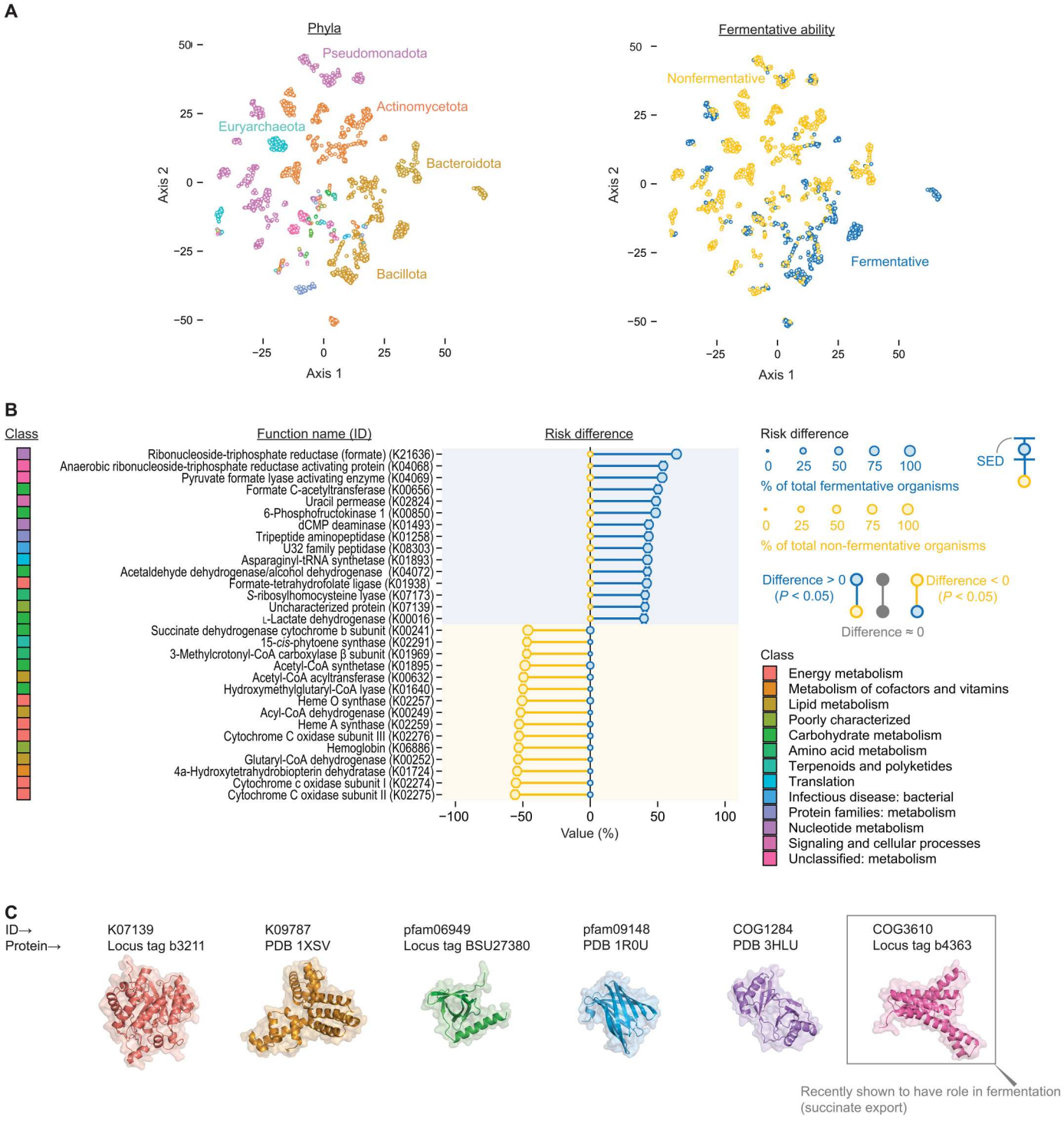

Fig. 4. Examining *n* = 13.6 million genes reveals that fermentative prokaryotes have a distinct genotype. (**A**) Two-dimensional plot of gene functions after performing t-SNE. A total of *n* = 4301 organisms was analyzed. Colors of phyla are same as in Fig. 2. (**B**) Top 15 KO IDs for gene functions. See data S3 for full list of gene functions. The number of organisms analyzed was as above. Differences between fermentative and nonfermentative prokaryotes were determined using a two-tailed *z* test. The class of KO IDs refers to level B of the BRITE hierarchy. Some class names have been abbreviated for display. SED, standard error for the difference. (**C**) Structures of uncharacterized proteins common in fermentative prokaryotes.

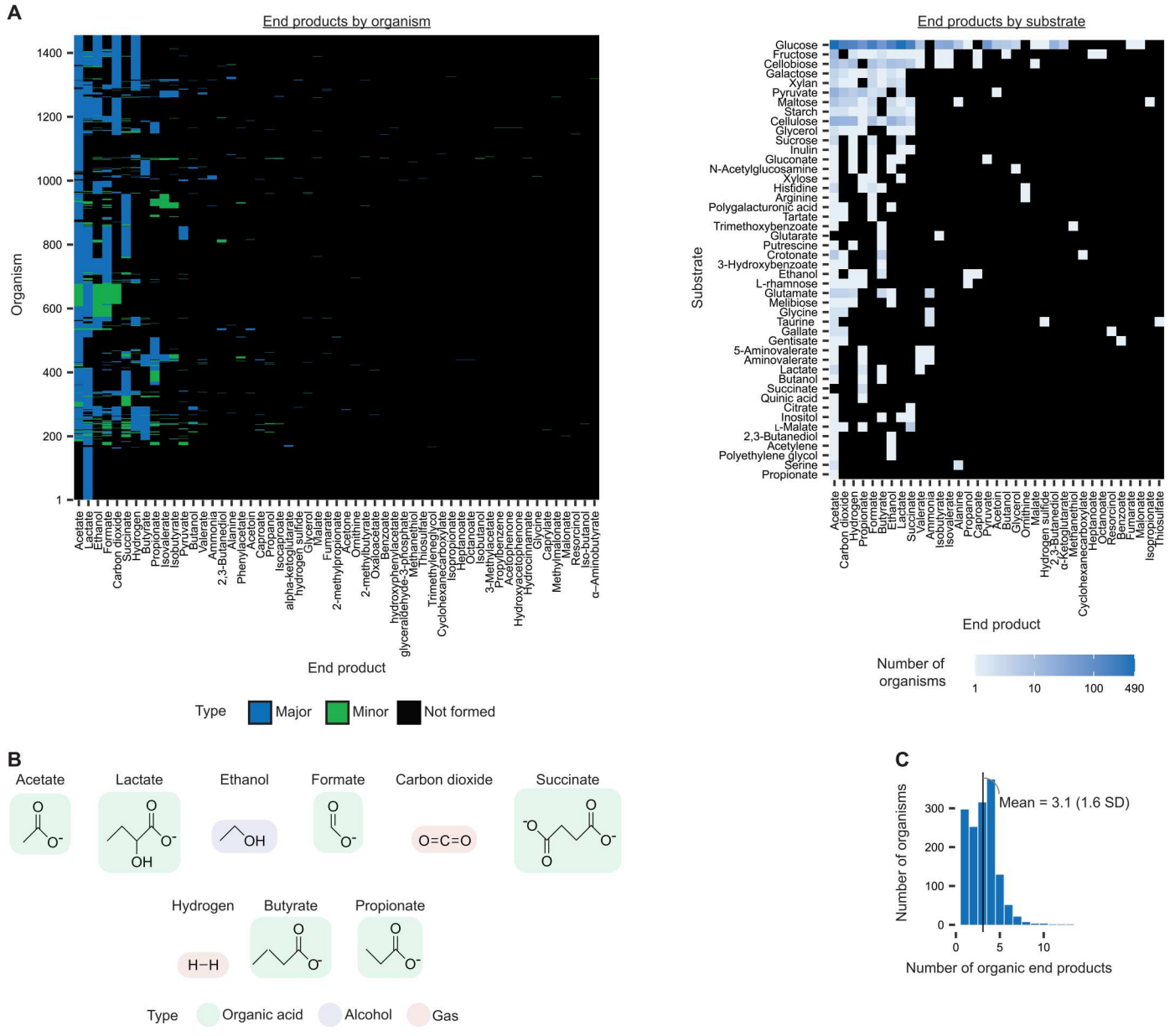

Fig. 5. Examining end products of *n* = 1455 prokaryotes shows that a wide number and range are formed. (**A**) Range of products across organisms and substrates. Lactate represents two potentially different products [(S) lactate and (R) lactate], but articles did not always distinguish which was formed. (**B**) Most common fermentation products. (**C**) Number of products formed. SD, standard deviation.

Our dataset gives simple but important insights about end products of fermentation. It shows which products are most common, and it shows nearly all fermentations form multiple products and in many combinations. It shows the relationship between substrate, organism, and end product with a dataset of unprecedented size.

End products can be predicted from genes using metabolic models

Many studies use an organism's genes to predict their metabolic pathways and fermentation products they form (*20*, *36*–*38*). We wanted to test how reliable is this practice with our dataset. We built metabolic models of our prokaryotes and then used flux balance analysis (*39*) to predict end products. We then compared predicted end products to those observed in our dataset. Because glucose metabolism is well studied, we examined prokaryotes known to ferment glucose (n = 406) and focused on the n = 9 most common fermentation end products (see Fig. 5B).

The metabolic models we built were networks of up to n = 98 metabolites connected by n = 110 biochemical reactions and n = 210 enzymes (Fig. 6A). Of the total metabolites, n = 64 were carbon-carrying metabolites (e.g., glucose), and an additional n = 34 were cofactors [e.g., reduced form of NAD^+ (NADH)]. For each organism, we built a model containing only those enzymes and reactions corresponding to the KO IDs of its genes (see

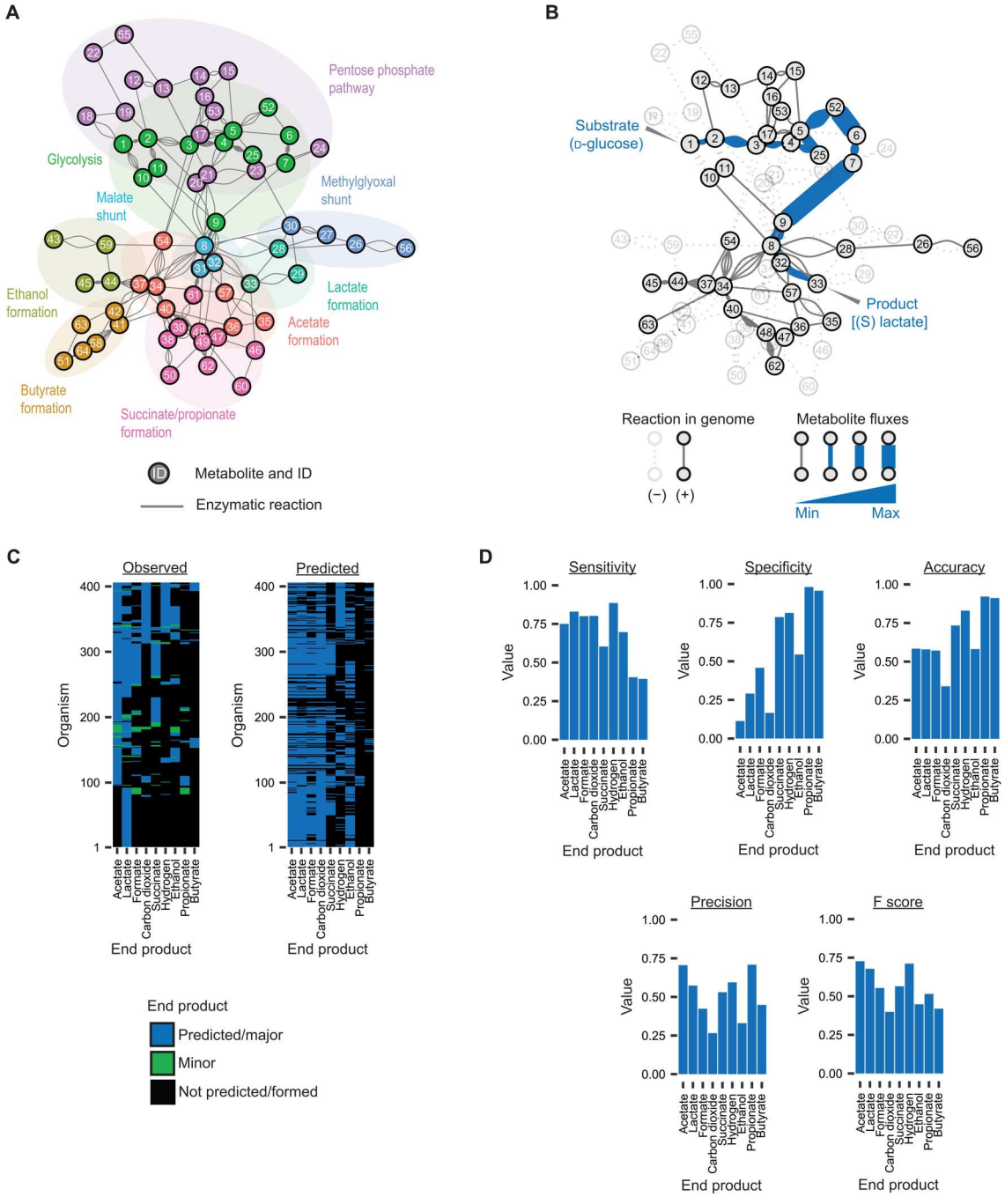

Fig. 6. Metabolic models of *n* = 406 organisms predict end products of fermentation. (**A**) The reference model containing and *n* = 210 enzymes, *n* = 110 biochemical reactions, and *n* = 64 metabolites (cofactors excluded). (**B**) An example model for one organism (*Lactococcus plantarum*), showing the predicted flux from D-glucose to (S) lactate. (**C**) End products predicted by metabolic models compared to those observed in our dataset. (**D**) Performance of predictions. In (A) and (B), the metabolites are (1) D-glucose, (2) D-glucose 6-phosphate, (3) D-fructose 6-phosphate, (4) D-fructose 1,6-bisphosphate, (5) D-glyceraldehyde 3-phosphate, (6) 3-phospho-D-glycerate, (7) 2-phospho-D-glycerate, (8) pyruvate, (9) phosphoenolpyruvate, (10) protein histidine, (11) protein n (pi)-phospho-*l*-histidine, (12) D-glucono-1,5-lactone 6-phosphate, (13) 6-phospho-D-gluconate, (14) D-ribulose 5-phosphate, (15) D-ribose 5-phosphate, (16) sedoheptulose 7-phosphate, (17) D-xylulose 5-phosphate, (18) D-glucono-1,5-lactone, (19) D-gluconic acid, (20) 2-dehydro-3-deoxy-D-gluconate, (21) 2-dehydro-3-deoxy-6-phospho-D-gluconate, (22) 2-keto-D-gluconic acid, (23) D-glyceraldehyde, (24) D-glycerate, (25) glycerone phosphate, (26) (R)-*S*-lactoylglutathione, (27) (R)-lactaldehyde, (28) (R)-lactate, (29) (S)-lactaldehyde, (30) methylglyoxal, (31) (S)-malate, (32) oxaloacetate, (33) (S)-lactate, (34) acetyl-CoA, (35) enzyme N6- (lipoyl)lysine, (36) enzyme N6- (dihydrolipoyl)lysine, (37) acetate, (38) succinyl-CoA, (39) succinate, (40) CoA, (41) butanoyl-CoA, (42) butanoic acid, (43) thiamin diphosphate, (44) acetaldehyde, (45) ethanol, (46) lactoyl-CoA, (47) propanoate, (48) propanoyl-CoA, (49) (S)-methylmalonyl-CoA, (50) (R)-methylmalonyl-CoA, (51) (S)-3-hydroxybutanoyl-CoA, (52) 3-phospho-D-glyceroyl phosphate, (53) D-erythrose 4-phosphate, (54) acetyl phosphate, (55) 6-phospho-2-dehydro-D-gluconate, (56) glutathione, (57) [dihydrolipoyllysine-residue acetyltransferase] *s*-acetyldihydrolipoyllysine, (58) butanoylphosphate, (59) 2-(alpha-hydroxyethyl)thiamine diphosphate, (60) propenoyl-CoA, (61) fumarate, (62) propanoyl phosphate, (63) acetoacetyl-CoA, and (64) crotonoyl-CoA.

Fig. 6B). For hydrogen formation, we used reactions corresponding to the HydDB name instead. With each model, we calculated the flux (flow) of metabolites through the network. If there was a positive flux from the substrate (glucose) to the end product (e.g., lactate), that end product was predicted to be formed (see Fig. 6B).

We compared products predicted by models with the observed ones (Fig. 6C). Some products (acetate, lactate, formate, and CO_2) were predicted in nearly all organisms, whether they were observed to be formed or not. This meant that the predictions had high sensitivity but low specificity (Fig. 6D). Other products (propionate, butyrate) were predicted in far fewer organisms than were observed to form them (Fig. 6C). The predictions had high specificity but low sensitivity (Fig. 6D). Three products (succinate, ethanol, and hydrogen) were predicted with high sensitivity and specificity (Fig. 6D).

We explored why certain products were predicted with low sensitivity (i.e., had many false-negative predictions). For butyrate, we were able to identify a missing reaction in 45% of cases (fig. S10). The most common reaction missing was butyryl–coenzyme A (CoA):acetate-CoA transferase [enzyme commission EC) 2.8.3.8], and adding it back to an organism′s model restored its ability to produce butyrate. This reaction may be missing because of poor gene annotation—databases such as Kyoto Encyclopedia of Genes and Genomes (KEGG) are missing annotations for 86% of enzymes known to carry out this reaction (*40*). In 20% of cases, more than one reaction could restore the ability to produce butyrate, making it unclear which reaction was really missing. In the remaining cases, the missing reaction(s) could not be identified. For propionate, another product predicted with low sensitivity, the situation was similar (fig. S10).

Our study shows that some, but not all, end products could be reliably predicted from metabolic models. Some products were predicted with low specificity, indicating that more organisms have the ability to produce them than observed (or reported). Others were predicted with low sensitivity instead. Furthermore, our study shows a number of reactions apparently missing from organisms. Poor gene annotation is likely to blame in some cases, although the reason is not clear in others. Our study shows there are still gaps in our knowledge of metabolic pathways and their prediction, even for glucose.

Our work is available in an interactive tool

To maximize the usefulness of our work, we built an interactive tool we call Fermentation Explorer (Fig. 7). This tool allows users to explore our dataset (Fig. 7B) and make predictions about prokaryotes in their samples (Fig. 7, C to E).

With this tool, users can build metabolic models and predict fermentation end products for their organisms (Fig. 7C). This allows users to predict traits from an organism′s genome. Users first upload KO IDs for genes for an organism. After this, they specify substrates and end products to check. The tool builds a metabolic model specific to the organism and predicts end products formed. Users can also visualize the model with predicted fluxes. To provide KO IDs, users can use the KEGG Automatic Annotation Server (KAAS) (*41*)—the output can be uploaded into our tool. Users can also use our library of KO IDs for $n = 987$ organisms (preloaded into the tool).

Users can also predict traits for prokaryotes from their taxonomy (Fig. 7D). After the user uploads the taxonomy (names) of their organisms, the tool finds matching organisms from our dataset. If the matching organisms share a trait, the trait is predicted for the user′s organisms. By default, 50% of the matching organisms must share the trait, but this threshold is adjustable by the user. The method of prediction is similar to FAPROTAX (*22*), but with the added flexibility of an adjustable threshold. The tool accepts partial taxonomy (e.g., resolved at genus only), making it useful for prokaryotes identified with ribosomal DNA sequencing. The user can also choose different systems for taxonomy [that from *Bergey's Manual* or the National Center for Biotechnology Information (NCBI)].

To demonstrate potential applications of our tool, we have used it to predict traits of prokaryotes in three studies. First, we used it to predict traits of $n = 410$ prokaryotes in the Hungate 1000 Project (fig. S11) (*42*). These are cultured prokaryotes from the rumen, and the predictions demonstrate that the tool can be used with large datasets. Second, we used it with $n = 733$ metagenome-assembled genomes (MAGs) from Stewart *et al.* (*43*) (fig. S12). These belong to uncultured prokaryotes from the rumen, and our tool gives insight into possible functions. We show that quality of MAGs is crucial; genomes with low completeness also have low numbers of predicted end products. Third, we used it with $n = 1053$ operational taxonomic units (OTUs) found in the gut of human infants over the first year of life (fig. S13) (*44*). Our analysis shows that fermentative prokaryotes rise and dominate the gut within weeks after birth. This rise mirrors the increase in fermentation acids in feces observed over this same time (*45*).

As a final exercise, we tested our tool with previously uncharacterized prokaryotes (Fig. 7E). We chose $n = 5$ bacteria from the rumen with no fermentation products reported in the primary literature (see Materials and Methods). This includes $n = 2$ bacteria isolated by our laboratory for this study. All belong to previously uncharacterized species or genera. We grew these organisms and measured $n = 13$ end products. For the $n = 2$ bacteria isolated by our laboratory, we also performed de novo genome sequencing. We found general agreement between end products observed and those predicted by the tool (Fig. 7E). Agreement was closest when combining predictions from genomes with those from taxonomy; accuracy of 90% was achieved (fig. S14). In summary, our tool works with large datasets of cultured and uncultured prokaryotes, and it is useful in making predictions for previously uncharacterized prokaryotes.

Our tool is available for use online and for download (https://github.com/thackmann/FermentationExplorer). It is built as an R Shiny app, and the online version is hosted by shinyapps.io. The app can be downloaded for use within R or as a Docker container image. Being easy to access and use, the tool enables the microbial community to apply the resources from our work in their own research.

DISCUSSION

Fermentation is a major type of metabolism, and it is important in microbial ecology, host health, food production, and industry. Despite this importance, there has been no systematic study of fermentative prokaryotes and their properties. Reviews of fermentation have been based on information from model organisms (*10*–*15*), which may not capture the full diversity of this metabolism. Some work, including our own, had started to accumulate information on more prokaryotes (*18*, *19*, *21*, *22*), but a full picture has still been lacking.

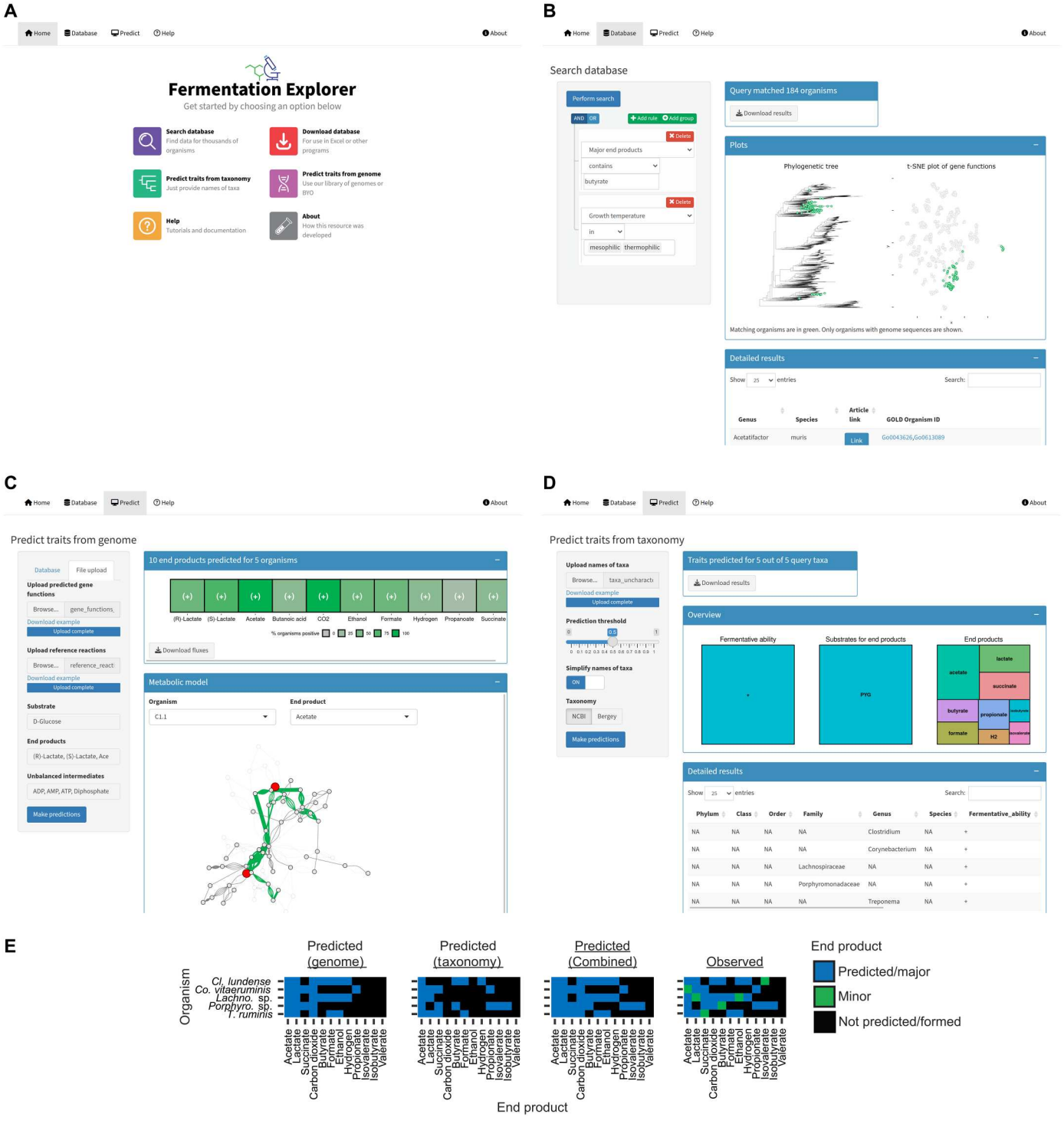

Fig. 7. Screenshot of Fermentation Explorer, an interactive tool for exploring our dataset and making predictions about organisms. (**A**) Home page. (**B**) Database search. (**C**) Predicting end products from genomes using metabolic models. (**D**) Predicting fermentative traits from taxonomy. (**E**) End products predicted by tool versus those observed in $n = 5$ previously uncharacterized bacteria. Lactate represents two different products [(S) lactate and (R) lactate]. *Cl.*, *Clostridium*; *Co.*, *Corynebacterium*; *Lachno.*, *Lachnospiraceae*; *Porphyro.*, *Porphyromonadaceae*; *T.*, *Treponema* (*81*–*161*).

By using a dataset of n = 8350 organisms (28% of which are fermentative), the current study paints a fuller picture. It shows that fermentative prokaryotes are both abundant and widespread. It reveals key insights into their genotype and phenotype and thus the concept of fermentation as a whole.

Some of the insights we reach make sense in context of earlier work, while others are unexpected. Some are both expected and unexpected at the same time. Given that fermentation does not use oxygen, it is expected that many organisms were reported as anaerobes. Fermentation was first described as "la vie sans air" (life without air) (*46*), and it is still regarded as a principally anaerobic metabolism (*10*–*15*). It is thus all the more unexpected that 18% of fermentative prokaryotes were reported as aerobes. These organisms may be more oxygen tolerant than previously realized.

One important insight is fermentation is complex and forms many end products. Historically, fermentations have been named by the major end product they form—e.g., alcohol fermentation forms ethanol (*10*–*12*, *14*, *15*). This is a long-standing practice (*47*, *48*) and remains useful for teaching, but our dataset shows that it does not match reality. Nearly all fermentations formed multiple products and in nearly 300 combinations. Our work reveals a complexity to fermentation not fully apparent before.

With the rise of genome sequencing, it is common to use an organism's genome to predict its metabolic pathways and fermentation products (*20*, *36*–*38*). This practice is especially important when an organism is uncultured and known only by a genome sequence (*37*, *38*). Our study shows that this practice is useful but has limits. We used genomes of n = 406 organisms to build metabolic models of fermentation. Some end products were predicted with low specificity or low sensitivity. Low specificity means that more organisms are predicted to form a product than observed. It could be due to many factors, such as a product being formed only under certain growth conditions [often the case for lactate (*49*)]. Alternatively, the product could be used for anabolism [often the case for formate (*50*)] and not leave the cell. Low sensitivity means that fewer organisms are predicted to form a product than observed, which represents a different issue.

One approach for raising sensitivity of predictions is gap filling or adding enzymatic reactions apparently missing from models (*51*). This approach is common and used by a study predicting end products for bacteria from the human gut (*20*). Our study identified reactions that are often missing, and adding these would have raised sensitivity. However, it was not always clear which reaction was missing, much less the reason. If anything else, missing reactions show areas of metabolism needing further study; such study has led to previously unknown pathways for fermentation being discovered (*18*).

To maximize the value of our work to the microbial community, we built an interactive tool called Fermentation Explorer. This tool has many applications. One is for identifying organisms for producing end products for biotechnological purposes (e.g., biofuel production). With data on 55 fermentation end products formed in nearly 300 combinations, our tool can pinpoint the best organism for an application. Such an organism could be used outright, or it could be used as a source of genes to genetically engineer other organisms (*52*–*54*). Another application is predicting fermentative traits of prokaryotes present in a user's sample. Users can predict traits in two different ways—either from an organism's genome or its taxonomy. We showcase this ability with four datasets, including one with uncharacterized prokaryotes that was generated for this study. Because fermentation is widespread, the ability to make these predictions will be useful to microbiologists working in many systems.

Fermentation has been studied for 185 years [since the time of Theodor Schwann (*55*) and Louis Pasteur (*47*, *48*)], and our study fills key gaps in our knowledge of this metabolism. It also shows that there is much to learn. It shows common genes in fermentative prokaryotes and that some genes have no defined function. The latter are targets for further study, and their abundance in fermentative prokaryotes may help narrow down possible roles. In one case, a gene with no defined function was confirmed to have a role in fermentation (transport of succinate) (*34*). Our study shows a wealth of information exists on fermentative prokaryotes, but it speaks little on eukaryotes. Both unicellular and multicellular eukaryotes carry out fermentation (*13*), and they merit further study. The next 185 years of study will be illuminating.

MATERIALS AND METHODS

Collection of information from *Bergey's Manual* and the primary literature

We collected information from *Bergey's Manual of Systematic of Archaea and Bacteria* using an approach similar to *(18*, *19*, *56)*. We downloaded all n = 1751 articles for genera. We then extracted name, strain ID (s), and written description of each of the n = 8331 organisms using R scripts.

To obtain phenotypic traits related to fermentation, we read all n = 5465 written descriptions containing the keyword "ferment." The traits we recorded were fermentative ability, fermentation end products, and the substrate used to form those end products. We also recorded the original text reporting the end products, and then, we used this text to divide fermentation end products into major and minor types. Minor products were those produced in only small quantities or only under certain conditions.

Many articles used the term fermentation without defining it. To be consistent, we defined it as catabolism where organic compounds are both the electron donor and acceptor. Protons can be another electron acceptor, forming hydrogen (H_2). H_2 can be another electron donor but only if an organic electron donor (e.g., glucose) is also used. This definition draws a solid line between fermentation and other types of metabolism, and it is consistent with most of the literature. For example, butyrate fermentation (where protons are an electron acceptor) is indeed a fermentation according to our definition and most others (*10*–*12*, *15*). Nitrate respiration (where nitrate is the electron acceptor); however, is excluded [see (*10*–*13*, *15*)]. Other examples are shown in table S1.

To check information in *Bergey's Manual*, we found n = 77 articles in the primary literature on prokaryotes of the rumen, an environment that our laboratory studies. These articles cover all n = 88 type strains from this environment. We found these articles using an approach similar to that of (*19*, *36*). We recorded the types of information above, plus yield of fermentation end products (mol/mol substrate fermented) in the n = 50 cases it was available. For polymers of hexose (e.g., cellulose), yield was expressed as mol/mol hexose equivalents. The moles of substrate fermented was calculated from moles of carbon in products. If CO_2 was not reported, its production was calculated using stoichiometry of other end

products (*36*). When yields were available, minor products were defined as those with yield <0.05 mol/mol.

We compared information in *Bergey's Manual* and the primary literature, and we found good agreement between the sources (fig. S15 and data file S2). In total, 78% of organisms in the primary literature were present in *Bergey's Manual*, and those absent tended to be recently described. Fermentative ability and fermentation end products also closely agreed. For organisms described both in *Bergey's Manual* and the primary literature, we used information from the latter in subsequent analysis.

Collection of information from other sources

We obtained information on genomes and genes from GOLD (*23*) and IMG/M (*24*) databases. Following *(18*, *19*, *56)*, we searched GOLD database for a GOLD organism ID, GOLD project ID, and IMG genome ID for each organism. Organisms were matched to GOLD organism ID by their strain ID plus genus or species name. If the strain ID was assigned by a large culture collection [Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ), American Type Culture Collection, Japan Collection of Microorganisms (JCM), or Collection de l'Institut Pasteur (CIP)], only the strain ID had to match. We then searched IMG/M database for all protein-coding genes belonging to each genome (IMG genome ID). If an organism had multiple genomes, we chose the one with the most genes. For each gene, we downloaded the locus tag as well as gene, KO, COG, pfam, and TIGRFAM IDs.

We obtained phenotypic traits (those not related to fermentation) from BacDive (*21*). We searched for BacDive IDs for organisms, matching them by strain ID plus the genus or species name (see above). We downloaded information for traits using the "Advanced search" and "Isolation sources" features. We formatted data as needed, such as by averaging data given as ranges (e.g., pH 7 to 9). Unformatted data are presented in data file S1.

We classified putative genes for hydrogenases using HydDB (*25*). We obtained sequences of genes using IMG/M and database IDs (pfam02906, pfam00346, pfam00374, and TIGR03295) covering hydrogenases of interest. We then classified sequences with HydDB (*25*). HydDB classified the sequences as [NiFe] Group 3B, [NiFe] Group 3d, [NiFe] Group 4A, [FeFe] Group A, or other. We further classified [FeFe] Group A hydrogenases as group A1, A2, A3, or A4 according to which accessory proteins were adjacent (*25*). Members of group A2 had GltA (COG0493) adjacent, group A3 had HydB (pfam01512) adjacent, group A4 had HytE1 (K05796) adjacent, and group A1 had no accessory protein adjacent.

We obtained definitions (short descriptions) of KO IDs from KEGG (*29*), COG IDs from COG (*30*), and TIGRFAM IDs (*32*) and pfam IDs from InterPro (*31*). We obtained NCBI taxonomy ID and names from GOLD database.

We obtained information about organisms in (*20*) by downloading it at https://vmh.life/#microbes/fermcarb. We determined which organisms were type strains by using BacDive. We found BacDive IDs for the organisms in that source following methods above. An updated version of this study has been released (*57*), but the data are not yet available at the link above.

FAPROTAX

We used FAPROTAX v. 1.2.6 to predict metabolic traits from the taxonomy of our organisms. The tool outputted predictions for n = 88 traits. We removed n = 16 traits related to parasitism, disease, and habitat, leaving n = 72 traits in our final analysis.

MetaCyc pathways

We predicted which of the n = 3445 pathways on MetaCyc (*35*) were encoded by genomes. We navigated to the MetaCyc webpage for each pathway, and then, we downloaded EC numbers for enzymes of each pathway using "Download Genes." We then found a KO ID corresponding to each EC number on KEGG (*29*).

We defined a genome as encoding a MetaCyc pathway if it had KO IDs for each enzyme. If an EC number had multiple KO IDs, only one had to match. If an enzyme had no EC number or no KO ID, it was ignored.

Construction of metabolic models

We constructed metabolic models for organisms using R and the fbar package. We first constructed a reference model that contained all n = 110 biochemical reactions we found relevant to fermentation (see data file S7). The reactions were identified mostly from (*36*) and are catalyzed by n = 210 enzymes. We obtained information on each reaction (including equation, EC number, and KO ID) from KEGG (*29*) and other sources. For reactions catalyzed by hydrogenases, we list the HydDB name in place of the KO ID. Many reactions could be catalyzed by multiple enzymes, and many enzymes had multiple KO IDs. Our model preserves the relationship between reaction, enzyme, and KO ID (see data file S7). After defining the main reactions, we added exchange reactions representing entry of cofactors, substrates, and products.

We predicted which reactions are catalyzed by each organism. We predicted that a reaction is catalyzed if an organism has a gene with the appropriate KO ID (or HydDB name). If a reaction is catalyzed by an enzyme with multiple KO IDs, genes for all must be present for it to be predicted.

Last, we set constraints for fluxes and solved the model. Fluxes for most reactions were constrained between −1000 and 1000 (arbitrary units). Fluxes for reactions that usually proceed in the forward direction (e.g., EC 2.7.1.1) were constrained between 0 and 1000. Similarly, fluxes for reactions proceeding in reverse were constrained between −1000 and 0. Fluxes of exchange reactions were constrained to be between -10^6 and 10^6 (for cofactors), −1000 and 0 (for substrates), or 0 and 10^6 (for products). For reactions not predicted to be catalyzed, fluxes were constrained to 0. The model was solved once for each product; this was done by setting the objective function of its exchange reaction to be 1. Products with flux >1 were considered to be produced.

Cofactors included molecules such as NADH and ATP. They also included any metabolite, such as H_2O, that we wished to be unbalanced. They did not include molecules, such as CoA, that are balanced by reactions in close proximity. With cofactors (unbalanced metabolites) in the model, the structure could be much simpler and did not have to include reactions for anabolism.

For propionate and butyrate, we explored how adding reactions would affect predictions. We built models of organisms observed but not predicted to form propionate or butyrate. We then added one-by-one all n = 110 biochemical reactions in the reference model. We recorded the flux and determined whether adding a reaction would change it from a negative prediction (flux of <1) to a positive one (flux of >1). Models were plotted using R and the igraph package (*58*).

Construction of phylogenetic trees

Phylogenetic trees were constructed as in *(18*, *19)*. The construction used sequences of 14 ribosomal proteins (*59*). We downloaded sequences from IMG/M (*24*) then aligned and concatenated them in R. We used aligned and concatenated sequences to create a phylogenetic tree with Randomized Axelerated Maximum Likelihood (RAxML) (*60*). Branch lengths of the consensus tree were calculated using R and the phytools package (*61*). The consensus tree was visualized using R and the ggtree package (*62*).

Protein structures

Structures were downloaded from Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) (*63*) or predicted with ColabFold (*64*). Structures on PDB were found by using COG (*30*) and InterPro (*31*) as references. If none existed for a given database ID, we predicted structures with ColabFold with protein sequences from *Escherichia coli* or *Bacillus subtilis*. The protein sequence used for prediction was downloaded from IMG/M. Structures were visualized using PyMOL (v2.0, Schrödinger LLC) following (*40*).

Other bioinformatics

We constructed heatmaps, lollipop charts, and bar charts using R and the ggplot2 package (*65*). Our interactive tool was constructed using R and the Shiny package.

Source and growth of organisms

We obtained *Corynebacterium vitaeruminis* DSM 20294, *Clostridium lundense* R1, and *Treponema ruminis* Ru1 from the DSMZ. These organisms were isolated by other laboratories, but their fermentation end products had not been reported in the primary literature (see data file S2).

Lachnospiraceae sp. C1.1 and *Porphyromonadaceae* sp. W3.11 were isolated by our laboratory from the rumen of a Holstein heifer. All procedures with animals have been approved by University of California Davis's Institutional Animal Care and Use Committee. Rumen contents were collected through a rumen fistula and strained through two layers of cheesecloth into a bottle. The bottle was sealed to exclude air and maintained at 39°C. Contents were brought to the laboratory and bubbled under O_2-free CO_2 within 15 min. At the laboratory, serial dilutions were made with anaerobic dilution solution for *Lachnospiraceae* sp. C1.1 and propionibacterium diluent for *Porphyromonadaceae* sp. W3.11 (table S2). Aliquots (0.1 ml) of each dilution were injected into anaerobic bottle plates (*66*) containing 9 ml of LH medium (table S2). After incubation at 37°C for 7 days, isolated colonies were picked. *Lachnospiraceae* sp. C1.1 was picked from a bottle inoculated with a 10^4 dilution of rumen contents, and *Porphyromonadaceae* sp. W3.11 was picked from a bottle inoculated with a 10^3 dilution. After initial isolation, these organisms were purified by growing on anaerobic roll tubes (*67*) and picking isolated colonies.

In subsequent experiments, organisms were grown anaerobically under O_2-free CO_2 in 9 ml of liquid medium at 39°C. Balch tubes with sealed rubber stoppers were used to exclude air, and syringes flushed with O_2-free CO_2 were used for all liquid transfers. The medium for most organisms was PC-VFA, which contains glucose as the only carbohydrate (table S2). One organism (*T. ruminis* Ru1) was grown on PC + VFA medium, which also contains short-chain fatty acids and stimulated growth. Organisms were grown on media without glucose (PC-VFA-glucose or PC + VFA-glucose) as controls.

Tubes were inoculated with a 0.1-ml volume of a stationary phase culture, and then, growth was monitored by measuring optical density at 600 nm on a spectrophotometer (Genesys 20, Thermo Fisher Scientific, Waltham, MA). Growth on media with glucose was higher than without glucose, showing glucose was used as a substrate (fig. S16).

Measurement of fermentation products

Fermentation products were measured after cultures reached early stationary phase. Aliquots (2 ml) of liquid culture were collected by syringe, and cells were removed by centrifugation (21,100*g* for 10 min at 4°C). The resulting supernatant was stored at −20°C until analysis. Separate cultures were grown for analysis of gas headspace. Cultures were grown at least three times and represent biological replicates.

Supernatant was analyzed for concentration of fermentation acids and alcohols (millimole per liter culture). Acetate, propionate, butyrate, isobutyrate, isovalerate, and valerate were measured by gas chromatography (*18*). (R) lactate and (S) lactate were analyzed using an enzymatic kit from R-Biopharm (product code 11112821035). Formate, succinate, and ethanol were measured using kits from Megazyme (product codes K-FORM, K-SUCC, and K-ETOH). The kit for ethanol cannot distinguish ethanol from propanol or butanol. When kits were used, supernatant was heated at 100°C for 10 min to inactivate interfering enzymes.

Headspace from cultures was analyzed for H_2. Total gas production (milliliter per culture) was measured with a syringe. Aliquots of gas (1 ml) were then collected and analyzed for H_2 by gas chromatography (*56*). Production of H_2 (millimole per liter of culture) was then calculated.

Production of CO_2 (millimole per liter culture) was determined with two methods. The first method was by measuring the difference between total gas production and H_2 production. This method cannot distinguish between CO_2 produced from glucose versus buffer, and it gave high values (see fig. S17). A second method, using stoichiometry of other end products, was also used (*36*). This method gave lower values.

Products were measured for cultures grown on media with glucose (e.g., PC-VFA) and without glucose (e.g., PC-VFA-glucose). Concentrations (millimole per liter culture) were higher for cultures grown on media with glucose, showing that glucose was used to form fermentation end products (fig. S17).

Final yield of products was expressed as millimole formed per millimole substrate fermented (fig. S18). We calculated millimole formed as millimole per liter culture with glucose minus millimole per liter culture without glucose. We assumed 6 mmol C in products/mmol substrate. For these calculations, we used CO_2 calculated using stoichiometry.

Sequencing and analysis of genomes

We performed de novo sequencing of *Lachnospiraceae* sp. C1.1 and *Porphyromonadaceae* sp. W3.11. Aliquots of liquid culture (9 and 1.5 ml, respectively) were collected by syringe and centrifuged (21,000*g* for 10 min at 4°C). Cell pellets were submitted to Molecular Research LP for DNA extraction, library preparation, and sequencing. After resuspending pellets in 180 ul of ATL buffer (Qiagen), DNA was extracted using the MagAttract HMW DNA

Kit (Qiagen). DNA was eluted in 100 ul of AE buffer (Qiagen) and then cleaned using the DNEasy PowerClean Pro Cleanup Kit (Qiagen). DNA was then sheared using the Covaris g-TUBE (Covaris). Sequencing libraries were prepared using the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences) and 1500 ng of the sheared and purified DNA. The SMRTbell libraries were size-selected (>6 Kb) using a BluePippin instrument (Sage Science) and 0.75% agarose gel. Libraries were then sequenced using the PacBio Sequel II (Pacific Biosciences) platform and a 30-hour movie time.

After obtaining raw sequence reads, we assembled them into contigs. We performed these steps using apps in the Department of Energy Systems Biology Knowledgebase (KBase) (*68*). We filtered low-quality reads using Trimmomatic (v0.36) (*69*), assembled filtered reads with SPAdes (v3.15.3) (*70*), and then checked completeness and contamination of the assembled genomes with CheckM (v1.0.18) (*71*). Statistics for sequencing and assembly are in table S3.

Using the assembled contigs (genomes), we called genes and annotated them. Protein-coding genes were called using Prodigal (v2.6.3) (*72*) locally or using KBase [via RASTtk (v1.073) (*73*)], with identical results. Genes were annotated with KO IDs using KAAS (*41*). They were further annotated with pfam and TIGRFAM IDs using KBase and the Annotate Domains in a Genome app. We classified putative genes for hydrogenases using HydDB. Genes for 16*S* ribosomal RNA (rRNA) were called using RASTtk (v1.073) in KBase.

The contigs (genomes) were analyzed to determine whether they belonged to new species. Taxonomy was assigned using GTDB-Tk (v1.7.0) (*74*) in KBase. The identity of 16*S* rRNA genes to other organisms was found using EzBioCloud (*75*). Values of digital DNA-DNA hybridization (dDDH) were found with Type (Strain) Genome Server (*76*). These analyses suggest that *Lachnospiraceae* sp. C1.1 and *Porphyromonadaceae* sp. W3.11 represent novel species or genera. GTDB-Tk assigned *Lachnospiracae* sp. C1.1 to family *Lachnospiraceae* and genus NK4A144, which contains no type strains. It assigned *Porphyromonadaceae* sp. W3.11 to *Porphyromonadaceae* and genus *Porphyromonas*_A. Values of 16*S* rRNA identity and dDDH with respect to type strains were low (table S4). Although more phenotypic data are needed, available evidence supports assignment of genomes to new species or genera.

Prediction of traits with our interactive tool

We used our interactive tool to predict traits of organisms grown in our study and three external datasets (*42*–*44*). Taxonomy and KO IDs were downloaded from IMG. Sequences of putative genes for hydrogenases were obtained similarly and then classified with HydDB as above. Information for *Lachnospiraceae* sp. C1.1 and *Porphyromonadaceae* sp. W3.11 was obtained as above. Completeness of genomes in (*43*) was that reported in the original paper. A newer version (*43*) has been published [see (*77*)], but data are not yet available on IMG. For (*44*), taxonomy and abundances of OTUs were taken from the paper.

To predict traits from taxonomy, we ran the tool using a prediction threshold of 0.5 and with NCBI taxonomy. To simulate typical use, we used partial taxonomies that included names of genera only. If names of genera were not available, names of the next highest rank (e.g., family) were used instead. For *Lachnospiraceae* sp. C1.1 and *Porphyromonadaceae* sp. W3.11, we used names of their families. To predict traits from genomes, we used the reference model for glucose metabolism described above. To combine predictions, we added traits predicted from genomes to traits predicted from taxonomy. All datasets are available as example datasets in our interactive tool.

Statistical analysis

We calculated risk difference for discrete variables as

$$P_1 - P_2$$

where P_1 is the percentage of fermentative prokaryotes positive for a trait and P_2 is the corresponding value for nonfermentative prokaryotes (see fig. S1A). The corresponding SEM is

$$\mathrm{SEM} = [P_1(100 - P_1)/n_1 + P_2(100 - P_2)/n_2]^{0.5}$$

where n_1 is the total number of fermentative prokaryotes and n_2 is the corresponding value for nonfermentative prokaryotes. We tested whether the statistic was different from 0 using a two-tailed z test (*78*). Values of P_1 and P_2 were arcsine-transformed (*78*) before the z test, and they were untransformed for presentation in the figures. P values from the z test were corrected for multiple comparisons using the Benjamini-Hochberg procedure (*79*).

For continuous variables, we calculated what we call the probability difference (see fig. S1B). It is

$$P_1(x) - P_2(x)$$

where $P_1(x)$ is the probability density function for fermentative prokaryotes. $P_2(x)$ is the corresponding function for nonfermentative prokaryotes. We fitted probability density functions to data using the density function of R (setting the bandwidth to triple the default value). We calculated 95% confidence limits using bootstrapping with 10,000 replicates.

For continuous variables, we also performed a two-sample, two-sided Kolmogorov-Smirnov test using R and the stats package. This is a nonparametric test for comparing empirical cumulative distribution functions, and it complements the probability difference calculated above. We performed it using the ks.test function of the stats package of R. We performed t-SNE (*33*) using R and the Rtsne package. The value of perplexity was set to 45.

We calculated log-odds ratios according to (*80*). As with risk difference, we tested whether the statistic was different from 0 using a two-tailed z test and corrected P values using the Benjamini-Hochberg procedure (*79*). We regressed number of end products on genome completeness using the loess function of R.

We analyzed data from (*44*) using a linear model. We first calculated the response (% fermentative prokaryotes) using predictions from our tool (see fig. S13A) and abundance of OTUs in (*44*). We then analyzed the data with the nlme package of R, using the gls function and model Response ~ Day. To reflect the repeated measures design, infant was the grouping factor. An unstructured variance-covariance matrix was chosen, as it gave the lowest value of Akaike information criterion. SEM was estimated using the emmeans package of R. We tested whether means differed from each other using a two-tailed t test, and we corrected P values for multiple comparisons with the Tukey procedure. Other statistical tests are as described in the figure legends.

Supplementary Materials

This PDF file includes:
Figs. S1 to S18
Tables S1 to S4
Legends for data S1 to S7

Other Supplementary Material for this manuscript includes the following:
Data file S1 to S7

REFERENCES AND NOTES

1. I. Mizrahi, R. J. Wallace, S. Morais, The rumen microbiome: Balancing food security and environmental impacts. *Nat. Rev. Microbiol.* **19**, 553–566 (2021).
2. E. T. Hillman, H. Lu, T. Yao, C. H. Nakatsu, Microbial ecology along the gastrointestinal tract. *Microbes Environ.* **32**, 300–313 (2017).
3. A. J. Kessler, Y. J. Chen, D. W. Waite, T. Hutchinson, S. Koh, M. E. Popa, J. Beardall, P. Hugenholtz, P. L. M. Cook, C. Greening, Bacterial fermentation and respiration processes are uncoupled in anoxic permeable sediments. *Nat. Microbiol.* **4**, 1014–1023 (2019).
4. S. K. Khanal, *Anaerobic Biotechnology for Bioenergy Production: Principles and Applications* (Wiley-Blackwell, 2008).
5. E. N. Bergman, Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol. Rev.* **70**, 567–590 (1990).
6. A. Koh, F. De Vadder, P. Kovatcheva-Datchary, F. Backhed, From dietary fiber to host physiology: Short-chain fatty acids as key bacterial metabolites. *Cell* **165**, 1332–1345 (2016).
7. R. W. Hutkins, *Microbiology and Technology of Fermented Foods* (John Wiley & Sons, ed. 2, 2019).
8. V. K. Varghese, B. J. Poddar, M. P. Shah, H. J. Purohit, A. A. Khardenavis, A comprehensive review on current status and future perspectives of microbial volatile fatty acids production as platform chemicals. *Sci. Total Environ.* **815**, 152500 (2022).
9. J. Keasling, H. Garcia Martin, T. S. Lee, A. Mukhopadhyay, S. W. Singer, E. Sundstrom, Microbial production of advanced biofuels. *Nat. Rev. Microbiol.* **19**, 701–715 (2021).
10. G. Gottschalk, *Bacterial Metabolism* (Springer Verlag, 1986).
11. V. Müller, in *Encyclopedia of Life Sciences* (Wiley, 2008).
12. R. Meganathan, Y. Ranganathan, C. A. Reddy, in *Methods for General and Molecular Microbiology*, C. A. Reddy *et al.*, Eds. (ASM Press, 2007), pp. 558–585.
13. M. Müller, M. Mentel, J. J. van Hellemond, K. Henze, C. Woehle, S. B. Gould, R. Y. Yu, M. van der Giezen, A. G. M. Tielens, W. F. Martin, Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol. Mol. Biol. Rev.* **76**, 444–495 (2012).
14. R. A. Schmitz, R. Daniel, U. Deppenmeier, G. Gottschalk, in *The Prokaryotes: Prokaryotic Communities and Ecophysiology* (Springer, 2013), vol. 2, pp. 259–273.
15. D. White, J. Drummond, C. Fuqua, *The Physiology and Biochemistry of Prokaryotes* (Oxford University Press, 2012).
16. W. B. Whitman, *Bergey's Manual of Systematics of Archaea and Bacteria* (Wiley, 2023).
17. E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, F. Thompson, *The Prokaryotes: Prokaryotic Biology and Symbiotic Associations* (Springer, ed. 4, 2013).
18. B. Zhang, C. Lingga, C. Bowman, T. J. Hackmann, A new pathway for forming acetate and synthesizing ATP during fermentation in bacteria. *Appl. Environ. Microbiol.* **87**, e0295920 (2021).
19. T. J. Hackmann, B. Zhang, Using neural networks to mine text and predict metabolic traits for thousands of microbes. *PLoS Comput. Biol.* **2**, e1008757 (2021).
20. S. Magnusdottir, A. Heinken, L. Kutt, D. A. Ravcheev, E. Bauer, A. Noronha, K. Greenhalgh, C. Jäger, J. Baginska, P. Wilmes, R. M. T. Fleming, I. Thiele, Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* **35**, 81–89 (2017).
21. L. C. Reimer, J. Sardà Carbasse, J. Koblitz, C. Ebeling, A. Podstawka, J. Overmann, BacDive in 2022: The knowledge base for standardized bacterial and archaeal data. *Nucleic Acids Res.* **50**, D741–D746 (2022).
22. S. Louca, L. W. Parfrey, M. Doebeli, Decoupling function and taxonomy in the global ocean microbiome. *Science* **353**, 1272–1277 (2016).
23. S. Mukherjee, D. Stamatis, C. T. Li, G. Ovchinnikova, J. Bertsch, J. C. Sundaramurthi, M. Kandimalla, P. A. Nicolopoulos, A. Favognano, I.-M. A. Chen, N. C. Kyrpides, T. B. K. Reddy, Twenty-five years of Genomes OnLine Database (GOLD): Data updates and new features in v.9. *Nucleic Acids Res.* **51**, D957–D963 (2023).
24. I. A. Chen, K. Chu, K. Palaniappan, A. Ratner, J. Huang, M. Huntemann, P. Hajek, S. Ritter, N. Varghese, R. Seshadri, S. Roux, T. Woyke, E. A. Eloe-Fadrosh, N. N. Ivanova, N. C. Kyrpides, The IMG/M data management and analysis system v.6.0: New tools and advanced capabilities. *Nucleic Acids Res.* **49**, D751–D763 (2021).
25. D. Sondergaard, C. N. Pedersen, C. Greening, HydDB: A web tool for hydrogenase classification and analysis. *Sci. Rep.* **6**, 34212 (2016).
26. A. Barberan, H. Caceres Velazquez, S. Jones, N. Fierer, Hiding in plain sight: Mining bacterial species records for phenotypic trait information. *mSphere* **2**, e00237-17 (2017).
27. O. N. Murashko, S. Lin-Chao, *Escherichia coli* responds to environmental changes using enolasic degradosomes and stabilized DicF sRNA to alter cellular morphology. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E8025–E8034 (2017).
28. R. Holman, Development of the long forms of bacteria. *Nature* **178**, 424–425 (1956).
29. M. Kanehisa, M. Furumichi, Y. Sato, M. Ishiguro-Watanabe, M. Tanabe, KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551 (2021).
30. M. Y. Galperin, Y. I. Wolf, K. S. Makarova, R. Vera Alvarez, D. Landsman, E. V. Koonin, COG database update: Focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* **49**, D274–D281 (2021).
31. M. Blum, H. Y. Chang, S. Chuguransky, T. Grego, S. Kandasaamy, A. Mitchell, G. Nuka, T. Paysan-Lafosse, M. Qureshi, S. Raj, L. Richardson, G. A. Salazar, L. Williams, P. Bork, A. Bridge, J. Gough, D. H. Haft, I. Letunic, A. Marchler-Bauer, H. Mi, D. A. Natale, M. Necci, C. A. Orengo, A. P. Pandurangan, C. Rivoire, C. J. A. Sigrist, I. Sillitoe, N. Thanki, P. D. Thomas, S. C. E. Tosatto, C. H. Wu, A. Bateman, R. D. Finn, The InterPro protein families and domains database: 20 Years on. *Nucleic Acids Res.* **49**, D344–D354 (2021).
32. D. H. Haft, J. D. Selengut, R. A. Richter, D. Harkins, M. K. Basu, E. Beck, TIGRFAMs and genome properties in 2013. *Nucleic Acids Res.* **41**, D387–D395 (2013).
33. L. Van der Maaten, G. Hinton, Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
34. K. Fukui, K. Nanatani, Y. Hara, S. Yamakami, D. Yahagi, A. Chinen, M. Tokura, K. Abe, *Escherichia coli* yjjPB genes encode a succinate transporter important for succinate production. *Biosci. Biotechnol. Biochem.* **81**, 1837–1844 (2017).
35. R. Caspi, R. Billington, I. M. Keseler, A. Kothari, M. Krummenacker, P. E. Midford, W. K. Ong, S. Paley, P. Subhraveti, P. D. Karp, The MetaCyc database of metabolic pathways and enzymes–A 2019 update. *Nucleic Acids Res.* **48**, D445–D453 (2020).
36. T. J. Hackmann, D. K. Ngugi, J. L. Firkins, J. Tao, Genomes of rumen bacteria encode atypical pathways for fermenting hexoses to short-chain fatty acids. *Environ. Microbiol.* **19**, 4670–4683 (2017).
37. D. K. Ngugi, S. Miyake, M. Cahill, M. Vinu, T. J. Hackmann, J. Blom, M. D. Tietbohl, M. L. Berumen, U. Stingl, Genomic diversification of giant enteric symbionts reflects host dietary lifestyles. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7592–E7601 (2017).
38. K. C. Wrighton, B. C. Thomas, I. Sharon, C. S. Miller, C. J. Castelle, N. C. VerBerkmoes, M. J. Wilkins, R. L. Hettich, M. S. Lipton, K. H. Williams, P. E. Long, J. F. Banfield, Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* **337**, 1661–1665 (2012).
39. J. D. Orth, I. Thiele, B. O. Palsson, What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
40. T. J. Hackmann, Redefining the coenzyme A transferase superfamily with a large set of manually annotated proteins. *Protein Sci.* **31**, 864–881 (2022).
41. Y. Moriya, M. Itoh, S. Okuda, A. C. Yoshizawa, M. Kanehisa, KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**, W182–W185 (2007).
42. R. Seshadri, S. C. Leahy, G. T. Attwood, K. H. Teh, S. C. Lambie, A. L. Cookson, E. A. Eloe-Fadrosh, G. A. Pavlopoulos, M. Hadjithomas, N. J. Varghese, D. Paez-Espino; Hungate Project Collaborators, R. Perry, G. Henderson, C. J. Creevey, N. Terrapon, P. Lapebie, E. Drula, V. Lombard, E. Rubin, N. C. Kyrpides, B. Henrissat, T. Woyke, N. N. Ivanova, W. J. Kelly, Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nat. Biotechnol.* **36**, 359–367 (2018).
43. R. D. Stewart, M. D. Auffret, A. Warr, A. H. Wiser, M. O. Press, K. W. Langford, I. Liachko, T. J. Snelling, R. J. Dewhurst, A. W. Walker, R. Roehe, M. Watson, Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nat. Commun.* **9**, 870 (2018).
44. L. Wampach, A. Heintz-Buschart, A. Hogan, E. E. L. Muller, S. Narayanasamy, C. C. Laczny, L. W. Hugerth, L. Bindl, J. Bottu, A. F. Andersson, C. de Beaufort, P. Wilmes, Colonization and succession within the human gut microbiome by archaea, bacteria, and microeukaryotes during the first year of life. *Front. Microbiol.* **8**, 738 (2017).
45. A. C. Midtvedt, T. Midtvedt, Production of short chain fatty acids by the intestinal microflora during the first 2 years of human life. *J. Pediatr. Gastroenterol. Nutr.* **15**, 395–403 (1992).
46. L. Pasteur, *Études sur la bière, ses maladies, causes qui les provoquent, procédé pour la rendre inaltérable: Avec une théorie nouvelle de la fermentation* (Gauthier-Villars, 1876).
47. L. Pasteur, Mémoire sur la fermentation alcoolique. *C. R. Acad. Sci.* **45**, 1032–1036 (1857).
48. L. Pasteur, Memaire sur la fermentation appelee lactique. *C. R. Acad. Sci.* **45**, 913–916 (1857).

49. T. D. Thomas, D. C. Ellwood, V. M. Longyear, Change from homo- to heterolactic fermentation by *Streptococcus lactis* resulting from glucose limitation in anaerobic chemostat cultures. *J. Bacteriol.* **138**, 109–117 (1979).
50. J. Zhou, D. G. Olson, A. A. Lanahan, L. Tian, S. J. L. Murphy, J. Lo, L. R. Lynd, Physiological roles of pyruvate ferredoxin oxidoreductase and pyruvate formate-lyase in *Thermoanaerobacterium saccharolyticum* JW/SL-YS485. *Biotechnol. Biofuels* **8**, 138 (2015).
51. S. Pan, J. L. Reed, Advances in gap-filling genome-scale metabolic models and model-driven experiments lead to novel metabolic discoveries. *Curr. Opin. Biotechnol.* **51**, 103–108 (2018).
52. S. Atsumi, A. F. Cann, M. R. Connor, C. R. Shen, K. M. Smith, M. P. Brynildsen, K. J. Y. Chou, T. Hanai, J. C. Liao, Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metab. Eng.* **10**, 305–311 (2008).
53. V. Kandasamy, H. Vaidyanathan, I. Djurdjevic, E. Jayamani, K. B. Ramachandran, W. Buckel, G. Jayaraman, S. Ramalingam, Engineering *Escherichia coli* with acrylate pathway genes for propionic acid synthesis and its impact on mixed-acid fermentation. *Appl. Microbiol. Biotechnol.* **97**, 1191–1200 (2013).
54. H. Park, B. S. Jeon, B.-I. J. B. Sang, B. Engineering, Efficient, simple production of corresponding alcohols from supplemented C2-C8 carboxylic acids in *Escherichia coli* using acyl-CoA transferase from *Megasphaera hexanoica*. *Biotechnol. Bioprocess. Eng.* **25**, 599–606 (2020).
55. T. Schwann, Vorläufige mittheilung, betreffend versuche über die weingährung und fäulniss. *Ann. Phys.* **117**, 184–193 (1837).
56. B. Zhang, C. Lingga, H. De Groot, T. J. Hackmann, The enzyme Rnf completes the pathway for forming propionate during fermentation in *Prevotella*. *bioRxiv*, 10.1101/2022.1111.1108.515646 (2022).
57. A. Heinken, J. Hertel, G. Acharya, D. A. Ravcheev, M. Nyga, O. E. Okpala, M. Hogan, S. Magnúsdóttir, F. Martinelli, B. Nap, G. Preciat, J. N. Edirisinghe, C. S. Henry, R. M. T. Fleming, I. Thiele, Genome-scale metabolic reconstruction of 7,302 human microorganisms for personalized medicine. *Nat. Biotechnol.*, 10.1038/s41587-022-01628-0 (2023).
58. G. Csardi, T. Nepusz, The igraph software package for complex network research. *Int. J. Complex Syst.* **1695**, 1–9 (2006).
59. C. J. Castelle, J. F. Banfield, Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell* **172**, 1181–1197 (2018).
60. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
61. L. Revell, phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
62. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T. Y. Lam, ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *J. Method Ecol. Evol.* **8**, 28–36 (2017).
63. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
64. M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, ColabFold: Making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
65. H. Wickham, in *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, ed. 2, 2016), pp. 260.
66. K. D. Olson, Modified bottle plate for the cultivation of strict anaerobes. *J. Microbiol. Methods* **14**, 267–269 (1992).
67. R. Hungate, Chapter IV A roll tube method for cultivation of strict anaerobes. *Methods Microbiol.* **3**, 117–132 (1969).
68. A. P. Arkin, R. W. Cottingham, C. S. Henry, N. L. Harris, R. L. Stevens, S. Maslov, P. Dehal, D. Ware, F. Perez, S. Canon, M. W. Sneddon, M. L. Henderson, W. J. Riehl, D. Murphy-Olson, S. Y. Chan, R. T. Kamimura, S. Kumari, M. M. Drake, T. S. Brettin, E. M. Glass, D. Chivian, D. Gunter, D. J. Weston, B. H. Allen, J. Baumohl, A. A. Best, B. Bowen, S. E. Brenner, C. C. Bun, J. M. Chandonia, J. M. Chia, R. Colasanti, N. Conrad, J. J. Davis, B. H. Davison, M. DeJongh, S. Devoid, E. Dietrich, I. Dubchak, J. N. Edirisinghe, G. Fang, J. P. Faria, P. M. Frybarger, W. Gerlach, M. Gerstein, A. Greiner, J. Gurtowski, H. L. Haun, F. He, R. Jain, M. P. Joachimiak, K. P. Keegan, S. Kondo, V. Kumar, M. L. Land, F. Meyer, M. Mills, P. S. Novichkov, T. Oh, G. J. Olsen, R. Olson, B. Parrello, S. Pasternak, E. Pearson, S. S. Poon, G. A. Price, S. Ramakrishnan, P. Ranjan, P. C. Ronald, M. C. Schatz, S. M. D. Seaver, M. Shukla, R. A. Sutormin, M. H. Syed, J. Thomason, N. L. Tintle, D. Wang, F. Xia, H. Yoo, S. Yoo, D. Yu, KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat. Biotechnol.* **36**, 566–569 (2018).
69. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
70. A. Bankevich, S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, P. A. Pevzner, SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
71. D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
72. D. Hyatt, G. L. Chen, P. F. LoCascio, M. L. Land, F. W. Larimer, L. J. Hauser, Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
73. T. Brettin, J. J. Davis, T. Disz, R. A. Edwards, S. Gerdes, G. J. Olsen, R. Olson, R. Overbeek, B. Parrello, G. D. Pusch, M. Shukla, J. A. Thomason III, R. Stevens, V. Vonstein, A. R. Wattam, F. Xia, RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* **5**, 8365 (2015).
74. P. A. Chaumeil, A. J. Mussig, P. Hugenholtz, D. H. Parks, GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2019).
75. S. H. Yoon, S. M. Ha, S. Kwon, J. Lim, Y. Kim, H. Seo, J. Chun, Introducing EzBioCloud: A taxonomically united database of 16*S* rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* **67**, 1613–1617 (2017).
76. J. P. Meier-Kolthoff, J. S. Carbasse, R. L. Peinado-Olarte, M. Goker, TYGS and LPSN: A database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res.* **50**, D801–D807 (2022).
77. R. D. Stewart, M. D. Auffret, A. Warr, A. W. Walker, R. Roehe, M. Watson, Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat. Biotechnol.* **37**, 953–961 (2019).
78. J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Routledge, 2013).
79. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B. Stat. Methodol.* **57**, 289–300 (1995).
80. J. M. Bland, D. G. Altman, Statistics notes. The odds ratio. *BMJ* **320**, 1468 (2000).
81. J. M. Chow, J. A. Van Kessel, J. B. Russell, Binding of radiolabeled monensin and lasalocid to ruminal microorganisms and feed. *J. Anim. Sci.* **72**, 1630–1635 (1994).
82. R. I. Mackie, S. Heath, Enumeration and isolation of lactate-utilizing bacteria from the rumen of sheep. *Appl. Environ. Microbiol.* **38**, 416–421 (1979).
83. B. A. Dehority, *Rumen Microbiology* (Nottingham University Press, 2003).
84. M. Kishishita, T. Ushijima, Y. Ozaki, Y. Ito, New medium for isolating propionibacteria and its application to assay of normal flora of human facial skin. *Appl. Environ. Microbiol.* **40**, 1100–1105 (1980).
85. R. C. Greening, J. A. Leedle, Enrichment and isolation of *Acetitomaculum ruminis*, gen. nov., sp. nov.: Acetogenic bacteria from the bovine rumen. *Arch. Microbiol.* **151**, 399–406 (1989).
86. M. V. Guettler, D. Rumler, M. K. Jain, *Actinobacillus succinogenes* sp. nov., a novel succinic-acid-producing strain from the bovine rumen. *Int. J. Syst. Bacteriol.* **49**Pt 1, 207–216 (1999).
87. D. An, S. Cai, X. Dong, *Actinomyces ruminicola* sp. nov., isolated from cattle rumen. *Int. J. Syst. Evol. Microbiol.* **56**, 2043–2048 (2006).
88. J. A. Rosero, J. Killer, H. Sechovcová, J. Mrázek, O. Benada, K. Fliegerová, J. Havlík, J. Kopečný, Reclassification of *Eubacterium rectale* (Prévot et al., 1967) in a new genus *Agathobacter* gen. nov., as *Agathobacter rectalis* comb. nov., within the family *Lachnospiraceae*, and description of *Agathobacter ruminis* sp. nov., from the rumen. *Int. J. Syst. Evol. Microbiol.* **66**, 768–773 (2015).
89. I. M. Robinson, M. J. Allison, P. A. Hartman, *Anaeroplasma abactoclasticum* gen. nov., sp. nov.: An obligately anaerobic mycoplasma from the rumen. *Int. J. Syst. Evol. Microbiol.* **25**, 173–181 (1975).
90. J. P. Robinson, R. Hungate, *Acholeplasma bactoclasticum* sp. n., an anaerobic mycoplasma from the bovine rumen. *Int. J. Syst. Evol. Microbiol.* **23**, 171–181 (1973).
91. I. M. Robinson, M. J. Allison, Transfer of *Acholeplasma bactoclasticum* Robinson and Hungate to the genus *Anaeroplasma* (*Anaeroplasma bactoclasticum* [Robinson and Hungate] comb. nov.): Emended description of the species. *Int. J. Syst. Evol. Microbiol.* **25**, 182–186 (1975).
92. P. N. Hobson, Continuous culture of some anaerobic and facultatively anaerobic rumen bacteria. *J. Gen. Microbiol.* **38**, 167–180 (1965).
93. P. Kuhnert, E. Scholten, S. Haefner, D. Mayor, J. Frey, *Basfia succiniciproducens* gen. nov., sp. nov., a new member of the family *Pasteurellaceae* isolated from bovine rumen. *Int. J. Syst. Evol. Microbiol.* **60**, 44–50 (2010).
94. V. Scardovi, L. Trovatelli, B. Biavati, G. Zani, *Bifidobacterium cuniculi*, *Bifidobacterium choerinum*, *Bifidobacterium boum*, and *Bifidobacterium pseudocatenulatum*. *Int. J. Syst. Evol. Microbiol.* **29**, 291–311 (1979).
95. B. Biavati, P. Mattarelli, *Bifidobacterium ruminantium* sp. nov. and *Bifidobacterium merycicum* sp. nov. from the rumens of cattle. *Int. J. Syst. Bacteriol.* **41**, 163–168 (1991).

96. V. Scardovi, L. D. Trovatelli, F. Crociani, B. Sgorbati, Bifid bacteria in bovine rumen. New species of the genus *Bifidobacterium*: *B. globosum* n.sp. and *B. ruminale* n.sp. *Arch. Mikrobiol.* **68**, 278–294 (1969).
97. F. Rieu-Lesme, B. Morvan, M. Collins, G. Fonty, A. Willems, A new H_2/CO_2-using acetogenic bacterium from the rumen: Description of *Ruminococcus schinkii* sp. nov. *FEMS Microbiol. Lett.* **140**, 281–286 (1996).
98. F. Diez-Gonzalez, D. R. Bond, E. Jennings, J. B. Russell, Alternative schemes of butyrate production in *Butyrivibrio fibrisolvens* and their relationship to acetate utilization, lactate production, and phylogeny. *Arch. Microbiol.* **171**, 324–330 (1999).
99. J. Kopecný, M. Zorec, J. Mrázek, Y. Kobayashi, R. Marinsek-Logar, *Butyrivibrio hungatei* sp. nov. and *Pseudobutyrivibrio xylanivorans* sp. nov., butyrate-producing bacteria from the rumen. *Int. J. Syst. Evol. Microbiol.* **53**, 201–209 (2003).
100. C. D. Moon, D. M. Pacheco, W. J. Kelly, S. C. Leahy, D. Li, J. Kopecny, G. T. Attwood, Reclassification of *Clostridium proteoclasticum* as *Butyrivibrio proteoclasticus* comb. nov., a butyrate-producing ruminal bacterium. *Int. J. Syst. Evol. Microbiol.* **58**, 2041–2045 (2008).
101. S. Cai, X. Dong, *Cellulosilyticum ruminicola* gen. nov., sp. nov., isolated from the rumen of yak, and reclassification of *Clostridium lentocellum* as *Cellulosilyticum lentocellum* comb. nov. *Int. J. Syst. Evol. Microbiol.* **60**, 845–849 (2010).
102. N. Van Gylswyk, J. Van Der Toorn, *Clostridium aerotolerans* sp. nov., a xylanolytic bacterium from corn stover and from the rumina of sheep fed corn stover. *Int. J. Syst. Evol. Microbiol.* **37**, 102–105 (1987).
103. G. Chen, J. B. Russell, More monensin-sensitive, ammonia-producing bacteria from the rumen. *Appl. Environ. Microbiol.* **55**, 1052–1057 (1989).
104. R. E. Hungate, Studies on cellulose fermentation: I. The culture and physiology of an anaerobic cellulose-digesting bacterium. *J. Bacteriol.* **48**, 499–513 (1944).
105. W. J. Kelly, R. V. Asmundson, D. H. Hopcroft, Isolation and characterization of a strictly anaerobic, cellulolytic spore former: *Clostridium chartatabidum* sp. nov. *Arch. Microbiol.* **147**, 169–173 (1987).
106. M. Bryant, N. Small, C. Bouma, I. Robinson, Studies on the composition of the ruminal flora and fauna of young calves. *J. Dairy Sci.* **41**, 1747–1767 (1958).
107. N. O. van Gylswyk, *Fusobacterium polysaccharolyticum* sp.nov., a Gram-negative rod from the rumen that produces butyrate and ferments cellulose and starch. *J. Gen. Microbiol.* **116**, 157–163 (1980).
108. N. O. van Gylswyk, J. P. Hoffman, Characteristics of cellulolytic cillobacteria from the rumens of sheep fed teff (*Eragrostis tef*) hay diets. *J. Gen. Microbiol.* **60**, 381–386 (1970).
109. L. R. Krumholz, M. Bryant, *Eubacterium oxidoreducens* sp. nov. requiring H_2 or formate to degrade gallate, pyrogallol, phloroglucinol and quercetin. *Arch. Microbiol.* **144**, 8–14 (1986).
110. R. J. Wallace, N. McKain, N. R. McEwan, E. Miyagawa, L. C. Chaudhary, T. P. King, N. D. Walker, J. H. A. Apajalahti, C. J. Newbold, *Eubacterium pyruvativorans* sp. nov., a novel non-saccharolytic anaerobe from the rumen that ferments pyruvate and amino acids, forms caproate and utilizes acetate and propionate. *Int. J. Syst. Evol. Microbiol.* **53**, 965–970 (2003).
111. M. P. Bryant, Bacterial species of the rumen. *Bacteriol. Rev.* **23**, 125–153 (1959).
112. N. Van Gylswyk, J. Van der Toorn, *Eubacterium uniforme* sp. nov. and *Eubacterium xylanophilum* sp. nov., fiber-digesting bacteria from the rumina of sheep fed corn stover. *Int. J. Syst. Evol. Microbiol.* **35**, 323–326 (1985).
113. C. Béra-Maillet, Y. Ribot, E. Forano, Fiber-degrading systems of different strains of the genus *Fibrobacter*. *Appl. Environ. Microbiol.* **70**, 2172–2179 (2004).
114. M. Bryant, R. Doetsch, A study of actively cellulolytic rod-shaped bacteria of the bovine rumen. *J. Dairy Sci.* **37**, 1176–1183 (1954).
115. M. E. Sharpe, M. J. Latham, E. I. Garvie, J. Zirngibl, O. Kandler, Two new species of *Lactobacillus* isolated from the bovine rumen, *Lactobacillus ruminis* sp.nov. and *Lactobacillus vitulinus* sp.nov. *J. Gen. Microbiol.* **77**, 37–49 (1973).
116. M. F. Whitford, L. J. Yanke, R. J. Forster, R. M. Teather, *Lachnobacterium bovis* gen. nov., sp. nov., a novel bacterium isolated from the rumen and faeces of cattle. *Int. J. Syst. Evol. Microbiol.* **51**, 1977–1981 (2001).
117. M. P. Bryant, N. Small, Characteristics of two new genera of anaerobic curved rods isolated from the rumen of cattle. *J. Bacteriol.* **72**, 22–26 (1956).
118. M. Marounek, K. Fliegrova, S. Bartos, Metabolism and some characteristics of ruminal strains of *Megasphaera elsdenii*. *Appl. Environ. Microbiol.* **55**, 1570–1573 (1989).
119. G. Q. Lan, Y. W. Ho, N. Abdullah, *Mitsuokella jalaludinii* sp. nov., from the rumens of cattle in Malaysia. *Int. J. Syst. Evol. Microbiol.* **52**, 713–718 (2002).
120. G. H. Lee, M. S. Rhee, D. H. Chang, J. Lee, S. Kim, M. H. Yoon, B. C. Kim, *Oscillibacter ruminantium* sp. nov., isolated from the rumen of Korean native cattle. *Int. J. Syst. Evol. Microbiol.* **63**, 1942–1946 (2013).
121. L. Krumholz, M. Bryant, *Clostridium pfennigii* sp. nov. uses methoxyl groups of monobenzenoids and produces butyrate. *Int. J. Syst. Evol. Microbiol.* **35**, 454–456 (1985).
122. J. L. Johnson, L. V. Holdeman, *Bacteroides intermedius* comb. nov. and descriptions of *Bacteroides corporis* sp. nov. and *Bacteroides levii* sp. nov. *Int. J. Syst. Evol. Microbiol.* **33**, 15–25 (1983).
123. M. P. Bryant, N. Small, C. Bouma, H. Chu, *Bacteroides ruminicola* n. sp. and *Succinimonas amylolytica*; the new genus and species; species of succinic acid-producing anaerobic bacteria of the bovine rumen. *J. Bacteriol.* **76**, 15–23 (1958).
124. M. R. Howlett, D. O. Mountfort, K. W. Turner, A. M. Roberton, Metabolism and growth yields in *Bacteroides ruminicola* strain $B_1$4. *Appl. Environ. Microbiol.* **32**, 274–283 (1976).
125. J. D. Vaidya, B. V. H. Hornung, H. Smidt, J. E. Edwards, C. M. Plugge, *Propionibacterium ruminifibrarum* sp. nov., isolated from cow rumen fibrous content. *Int. J. Syst. Evol. Microbiol.* **69**, 2584–2590 (2019).
126. K. Zhang, L. Song, X. Dong, *Proteiniclasticum ruminis* gen. nov., sp. nov., a strictly anaerobic proteolytic bacterium isolated from yak rumen. *Int. J. Syst. Evol. Microbiol.* **60**, 2221–2225 (2010).
127. N. Van Gylswyk, H. Hippe, F. Rainey, *Pseudobutyrivibrio ruminis* gen. nov., sp. nov., a butyrate-producing bacterium from the rumen that closely resembles *Butyrivibrio fibrisolvens* in phenotype. *Int. J. Syst. Evol. Microbiol.* **46**, 559–563 (1996).
128. E. P. Cato, W. Moore, M. P. Bryant, Designation of neotype strains for *Bacteroides amylophilus* Hamlin and Hungate 1956 and *Bacteroides succinogenes* Hungate 1950. *Int. J. Syst. Evol. Microbiol.* **28**, 491–495 (1978).
129. M. Bryant, N. Small, C. Bouma, I. Robinson, Characteristics of ruminal anaerobic cellulolytic cocci and *Cillobacterium cellulosolvens* n. sp. *J. Bacteriol.* **76**, 529–537 (1958).
130. N. O. van Gylswyk, H. Hippe, F. A. Rainey, *Schwartzia succinivorans* gen. nov., sp. nov., another ruminal bacterium utilizing succinate as the sole energy source. *Int. J. Syst. Bacteriol.* **47**, 155–159 (1997).
131. K. Zhang, X. Dong, *Selenomonas bovis* sp. nov., isolated from yak rumen contents. *Int. J. Syst. Evol. Microbiol.* **59**, 2080–2083 (2009).
132. M. P. Bryant, The characteristics of strains of *Selenomonas* isolated from bovine rumen contents. *J. Bacteriol.* **72**, 162–167 (1956).
133. J. Brooker, L. A. O'Donovan, I. Skene, K. Clarke, L. Blackall, P. Muslera, *Streptococcus caprinus* sp. nov., a tannin-resistant ruminal bacterium from feral goats. *Lett. Appl. Microbiol.* **18**, 313–318 (1994).
134. N. O. van Gylswyk, *Succiniclasticum ruminis* gen. nov., sp. nov., a ruminal bacterium converting succinate to propionate as the sole energy-yielding mechanism. *Int. J. Syst. Bacteriol.* **45**, 297–300 (1995).
135. C. S. McSweeny, M. J. Allison, R. I. Mackie, Amino acid utilization by the ruminal bacterium *Synergistes jonesii* strain 78-1. *Arch. Microbiol.* **159**, 131–135 (1993).
136. L. Krumholz, M. Bryant, *Syntrophococcus sucromutans* sp. nov. gen. nov. uses carbohydrates as electron donors and formate, methoxymonobenzenoids or *Methanobrevibacter* as electron acceptor systems. *Arch. Microbiol.* **143**, 313–318 (1986).
137. T. B. Stanton, E. Canale-Parola, *Treponema bryantii* sp. nov., a rumen spirochete that interacts with cellulolytic bacteria. *Arch. Microbiol.* **127**, 145–156 (1980).
138. B. J. Paster, E. Canale-Parola, *Treponema saccharophilum* sp. nov., a large pectinolytic spirochete from the bovine rumen. *Appl. Environ. Microbiol.* **50**, 212–219 (1985).
139. E. L. Emerson, P. J. Weimer, Fermentation of model hemicelluloses by *Prevotella* strains and *Butyrivibrio fibrisolvens* in pure culture and in ruminal enrichment cultures. *Appl. Microbiol. Biotechnol.* **101**, 4269–4278 (2017).
140. I. Robinson, E. Freundt, Proposal for an amended classification of anaerobic mollicutes. *Int. J. Syst. Evol. Microbiol.* **37**, 78–81 (1987).
141. M. Lanéelle, J. Asselineau, M. Welby, M. V. Norgard, T. Imaeda, M. C. Pollice, L. Barksdale, Biological and chemical bases for the reclassification of *Brevibacterium vitarumen* (Bechdel *et al.*) Breed (Approved Lists, 1980) as *Corynebacterium vitarumen* (Bechdel *et al.*) comb. nov. and *Brevibacterium liquefaciens* Okabayashi and Masuo (Approved Lists, 1980) as *Corynebacterium liquefaciens* (Okabayashi and Masuo) comb. nov. *Int. J. Syst. Evol. Microbiol.* **30**, 539–546 (1980).
142. O. Angen, R. Mutters, D. A. Caugant, J. E. Olsen, M. Bisgaard, Taxonomic relationships of the [*Pasteurella*] *haemolytica* complex as evaluated by DNA-DNA hybridizations and 16*S* rRNA sequencing with proposal of *Mannheimia haemolytica* gen. nov., comb. nov., *Mannheimia granulomatis* comb. nov., *Mannheimia glucosida* sp. nov., *Mannheimia ruminalis* sp. nov. and *Mannheimia varigena* sp. nov. *Int. J. Syst. Bacteriol.* **49**, 67–86 (1999).
143. M. R. Garner, J. F. Flint, J. B. Russell, *Allisonella histaminiformans* gen. nov., sp. nov. *Allisonella histaminiformans* gen. nov., sp. nov. *Syst. Appl. Microbiol.* **25**, 498–506 (2002).
144. D. G. Cirne, O. D. Delgado, S. Marichamy, B. Mattiasson, *Clostridium lundense* sp. nov., a novel anaerobic lipolytic bacterium isolated from bovine rumen. *Int. J. Syst. Evol. Microbiol.* **56**, 625–628 (2006).
145. R. C. Anderson, M. A. Rasmussen, N. S. Jensen, M. J. Allison, *Denitrobacterium detoxificans* gen. nov., sp. nov., a ruminal bacterium that respires on nitrocompounds. *Int. J. Syst. Evol. Microbiol.* **50** Pt 2, 633–638 (2000).

146. G. S. Coleman, A sulphate-reducing bacterium from the sheep rumen. *J. Gen. Microbiol.* **22**, 423–436 (1960).
147. A. R. Cook, P. W. Riley, H. Murdoch, P. N. Evans, I. R. McDonald, *Howardella ureilytica* gen. nov., sp. nov., a Gram-positive, coccoid-shaped bacterium from a sheep rumen. *Int. J. Syst. Evol. Microbiol.* **57**, 2940–2945 (2007).
148. M. J. Allison, K. A. Dawson, W. R. Mayberry, J. G. Foss, *Oxalobacter formigenes* gen. nov., sp. nov.: Oxalate-degrading anaerobes that inhabit the gastrointestinal tract. *Arch. Microbiol.* **141**, 1–7 (1985).
149. G. W. Lanigan, *Peptococcus heliotrinreducans*, sp. nov., a cytochrome-producing anaerobe which metabolizes pyrrolizidine alkaloids. *J. Gen. Microbiol.* **94**, 1–10 (1976).
150. M. J. Wolin, E. A. Wolin, N. J. Jacobs, Cytochrome-producing anaerobic *Vibrio succinogenes*, sp. n. *J. Bacteriol.* **81**, 911–917 (1961).
151. P. H. Smith, R. E. Hungate, Isolation and characterization of *Methanobacterium ruminantium* n. sp. *J. Bacteriol.* **75**, 713–718 (1958).
152. J. H. Lee, S. Kumar, G. H. Lee, D. H. Chang, M. S. Rhee, M. H. Yoon, B. C. Kim, *Methanobrevibacter boviskoreani* sp. nov., isolated from the rumen of Korean native cattle. *Int. J. Syst. Evol. Microbiol.* **63**, 4196–4201 (2013).
153. J. Gaffney, J. Embree, S. Gilmore, M. Embree, *Ruminococcus bovis* sp. nov., a novel species of amylolytic *Ruminococcus* isolated from the rumen of a dairy cow. *Int. J. Syst. Evol. Microbiol.* **71**, 004924 (2021).
154. S. W. Na, B. H. Chun, S. H. Beak, S. A. Khan, M. N. Haque, J. S. Lee, C. O. Jeon, S. S. Lee, M. Baik, *Pseudoprevotella muciniphila* gen. nov., sp. nov., a mucin-degrading bacterium attached to the bovine rumen epithelium. *PLOS ONE* **16**, e0251791 (2021).
155. T. Shinkai et al., *Prevotella lacticifex* sp. nov., isolated from the rumen of cows. *Int. J. Syst. Evol. Microbiol.* **72**, 005278 (2022).
156. M. J. Paynter, R. E. Hungate, Characterization of *Methanobacterium mobilis*, sp. n., isolated from the bovine rumen. *J. Bacteriol.* **95**, 1943–1951 (1968).
157. S. Rea, J. P. Bowman, S. Popovski, C. Pimm, A. G. Wright, *Methanobrevibacter millerae* sp. nov. and *Methanobrevibacter olleyae* sp. nov., methanogens from the ovine and bovine rumen that can utilize formate for growth. *Int. J. Syst. Evol. Microbiol.* **57**, 450–456 (2007).
158. S. B. Kim, M. Goodfellow, Reclassification of *Amycolatopsis rugosa* Lechevalier *et al*. 1986 as *Prauserella rugosa* gen. nov., comb. nov. *Int. J. Syst. Bacteriol.* **49**Pt 2, 507–512 (1999).
159. B. H. Chun, H. J. Lee, S. E. Jeong, P. Schumann, C. O. Jeon, *Leucobacter ruminantium* sp. nov., isolated from the bovine rumen. *Int. J. Syst. Evol. Microbiol.* **67**, 2634–2639 (2017).
160. B. S. Jeon, S. Kim, B. I. Sang, *Megasphaera hexanoica* sp. nov., a medium-chain carboxylic acid-producing bacterium isolated from a cow rumen. *Int. J. Syst. Evol. Microbiol.* **67**, 2114–2120 (2017).
161. K. Newbrook, G. J. Staton, S. R. Clegg, R. J. Birtles, S. D. Carter, N. J. Evans, *Treponema ruminis* sp. nov., a spirochaete isolated from the bovine rumen. *Int. J. Syst. Evol. Microbiol.* **67**, 1349–1354 (2017).

Acknowledgments: We thank H. De Groot for assistance in analysis of fermentation products. We also thank J. Parris, M. Yonathan, and A. Sen for assistance in culturing organisms. All those acknowledged are from the University of California, Davis. **Funding:** This work was supported by the United States Department of Agriculture National Institute of Food and Agriculture grant 2018-67015-27495, the United States Department of Agriculture National Institute of Food and Agriculture grant 1019985, and the University of California, Davis Small Grant in Aid of Research.
Author contributions: Investigation: T.J.H. Writing—original draft: T.J.H. Writing—review and editing: T.J.H. and B.Z. Software: T.J.H. and B.Z. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All strains generated for this study are available from T.J.H. pending scientific review and a completed material transfer agreement. Sequence data are available at the NCBI database under BioProject accession numbers PRJNA990468 and PRJNA990469. Code for the interactive tool is available at GitHub https://github.com/thackmann/FermentationExplorer) and Zenodo (https://zenodo.org/record/8111166).

Submitted 8 February 2023
Accepted 25 August 2023
Published 27 September 2023
10.1126/sciadv.adg8687