# UC San Diego
## UC San Diego Previously Published Works

**Title**
Children with Autism and Their Typically Developing Siblings Differ in Amplicon Sequence Variants and Predicted Functions of Stool-Associated Microbes

**Authors**
David, Maude M
Tataru, Christine
Daniels, Jena
et al.

# Children with Autism and Their Typically Developing Siblings Differ in Amplicon Sequence Variants and Predicted Functions of Stool-Associated Microbes

Maude M. David,[a,b,c] Christine Tataru,[a,b] Jena Daniels,[a] Jessey Schwartz,[a] Jessica Keating,[a] Jarrad Hampton-Marcell,[d,e] Neil Gottel,[d,e] Jack A. Gilbert,[d,e] Dennis P. Wall[a,f]

aDepartment of Pediatrics, Division of Systems Medicine, Stanford University, Stanford, California, USA
bDepartment of Microbiology, Oregon State University, Corvallis, Oregon, USA
cDepartment of Pharmaceutical Sciences, Oregon State University, Corvallis, Oregon, USA
dBioscience Division, The Microbiome Center, Argonne National Laboratory, Argonne, Illinois, USA
eDepartment of Pediatrics, University of California San Diego, La Jolla, California, USA
fDepartment of Biomedical Data Science, Stanford University, Stanford, California, USA

**ABSTRACT** The existence of a link between the gut microbiome and autism spectrum disorder (ASD) is well established in mice, but in human populations, efforts to identify microbial biomarkers have been limited due to a lack of appropriately matched controls, stratification of participants within the autism spectrum, and sample size. To overcome these limitations, we crowdsourced the recruitment of families with age-matched sibling pairs between 2 and 7 years old (within 2 years of each other), where one child had a diagnosis of ASD and the other did not. Parents collected stool samples, provided a home video of their ASD child's natural social behavior, and responded online to diet and behavioral questionnaires. 16S rRNA V4 amplicon sequencing of 117 samples (60 ASD and 57 controls) identified 21 amplicon sequence variants (ASVs) that differed significantly between the two cohorts: 11 were found to be enriched in neurotypical children (six ASVs belonging to the *Lachnospiraceae* family), while 10 were enriched in children with ASD (including *Ruminococcaceae* and *Bacteroidaceae* families). Summarizing the expected KEGG orthologs of each predicted genome, the taxonomic biomarkers associated with children with ASD can use amino acids as precursors for butyragenic pathways, potentially altering the availability of neurotransmitters like glutamate and gamma aminobutyric acid (GABA).

**IMPORTANCE** Autism spectrum disorder (ASD), which now affects 1 in 54 children in the United States, is known to have comorbidity with gut disorders of a variety of types; however, the link to the microbiome remains poorly characterized. Recent work has provided compelling evidence to link the gut microbiome to the autism phenotype in mouse models, but identification of specific taxa associated with autism has suffered replicability issues in humans. This has been due in part to sample size that sufficiently covers the spectrum of phenotypes known to autism (which range from subtle to severe) and a lack of appropriately matched controls. Our original study proposes to overcome these limitations by collecting stool-associated microbiome on 60 sibling pairs of children, one with autism and one neurotypically developing, both 2 to 7 years old and no more than 2 years apart in age. We use exact sequence variant analysis and both permutation and differential abundance procedures to identify 21 taxa with significant enrichment or depletion in the autism cohort compared to their matched sibling controls. Several of these 21 biomarkers have been identified in previous smaller studies; however, some are new to autism and known to be important in gut-brain interactions and/or are associated with specific fatty acid biosynthesis pathways.

Autism spectrum disorder (ASD) is a heterogeneous developmental disorder affecting social and behavioral functioning in 1 out of 59 children in the United States (1). Recent studies have identified several environmental factors associated with ASD etiology and susceptibility, including prenatal infection (2), zinc deficiency (3), maternal diabetes (4), toxins and pesticides (5), and advanced parental age (6). Individuals with ASD have demonstrated a high prevalence of gastrointestinal (GI) and immunologic abnormalities pertaining to GI motility and intestinal permeability (7, 8). The legitimacy of the proposed microbiome-ASD connection is supported by research on ASD phenotype mouse models and the microbiota compositions of human individuals with ASD (9, 11, 15–17). Hsiao et al. found that administrating *Bacteroides fragilis* to ASD mouse models improved ASD-typified behavioral traits by reducing anxiety, restoring communicative behaviors, and improving sensorimotor gating (9). Bacterial taxa, such as members of *Lactobacillus* and a genus of *Bifidobacterium*, have demonstrated microbially induced behavioral modulation in both rats and humans (10, 12–14). Moreover, several studies have identified microbial trends among the ASD population such as an increased abundance of *Clostridium* (15, 16). More recently, a study involving fecal microbiome transplant between neurotypical controls and children with ASD demonstrated a significant improvement in both GI and neurobehavioral symptoms following the treatment (17). This study particularly demonstrates a potential causative relationship between the gut microbiome and ASD symptoms. While the data on the microbiome-ASD symptom link is compelling, studies attempting to identify the specific microbes responsible have maintained small sample sizes, lack of appropriately matched controls, limited phenotype scoring, and a lack of bacterial phylogenetic resolution, factors that impact the reproducibility of findings.

The present study aims to determine the specific intestinal microbiota that associate with behavioral traits in children with ASD. We recruited families with age-matched neurotypical and ASD siblings via crowdsourcing to reach a sufficient sample size (18–20). Recruited families had a child clinically diagnosed with ASD and a neurotypical sibling who were both between the ages of 2 to 7 years old and no more than 2 years apart in age. We confirmed the self-reported autism diagnosis of each child by leveraging validated machine-learning classification tools that assess ASD-typified features obtained from parent reports and home video showcasing social interactions (21–26). Lack of ASD diagnosis relied on parent reporting. Each family completed an extensive dietary questionnaire online and collected a stool sample from each child at home.

## RESULTS

**Crowdsourcing recruitment and participant demographics.** Between March 2015 and September 2017, 20,478 unique users visited our study website, 1,953 were electronically screened for eligibility by survey, and 297 of them met our study inclusion criteria. A total of 194 users electronically consented to participate, and 164 began responding to the online surveys. Of 164 participants, 100 completed the online component and were mailed sampling kits. Seventy-one families, or parents of 142 sibling pairs, completed the online and at-home sampling procedures for the study, and 117 child subjects (60 ASD and 57 neurotypical [NT]) met all eligibility criteria, including the required confirmation of diagnosis obtained from the mobile autism risk assessment (MARA) and video classifier, when submitted. Of the 117 child subjects, there were 55 sibling pairs, two sibling pairs were accompanied by a third sibling with autism, and 5 were singleton samples.

The ASD cohort comprised 72% male participants ($n = 43$) compared to 55% of the NT cohort ($n = 27$). Dietary and lifestyle questionnaires were completed, in entirety, for 106 of the 117 participants. Among the 106 child subjects, 66% ($n = 79$) identified as

Caucasian, 7.5% ($n = 8$) identified as Asian or Pacific Islander, 3.8% ($n = 4$) identified as African American, and 7.5% ($n = 8$) identified as Hispanic (participants were also given the option to select more than one identifying ethnicity, not reported here). Participant age was not significantly different between ASD and NT cohorts. Samples were collected from 24 states: California, Colorado, Florida, Georgia, Hawaii, Illinois, Indiana, Massachusetts, Maryland, Michigan, Minnesota, Missouri, North Carolina, Nebraska, New Jersey, Nevada, New York, Ohio, Pennsylvania, Tennessee, Texas, Utah, Washington, and Wisconsin. Additional demographic and lifestyle data are in Table S1 in the supplemental material.

**ASD diagnosis confirmation using the mobile autism risk assessment (MARA) and video classifier.** For all child subjects with ASD, the MARA was completed, and scorable video was provided for 29 of these child subjects (including one family with two siblings with ASD and one NT sibling meeting the age criteria). There was a 100% agreement in class assignment between the MARA and the video classifier in all 37 cases (Table S2). In 12 instances, the output from either or both classifiers (2 supported by the video classifier) did not confirm the parent-reported ASD diagnosis. These participants were therefore excluded from analysis. The results reported hereafter include only the remaining 60 child subjects with confirmed ASD.

**Diet differences between children with ASD and neurotypical siblings.** We found three categorical factors (supplements, dairy intolerance, and dietary restrictions) to be significantly different between the two cohorts according to a chi-square test (Table 1). Nutritional/herbal supplements showed significant differences between the two cohorts, with 63.6% ($n = 35$) ASD child subjects taking an herbal supplement compared to 35.3% ($n = 18$) NT child subjects ($q$ value [qval], 2.3E−2). Dairy intolerance was also more prevalent in the ASD cohort (1 NT child subject versus 16 ASD child subjects) (qval, 2.6E−3), which correlates with a statistically significant deviation in the frequency of consumption of both milk/cheese and milk substitutes: only 33.3% ($n = 18$) of ASD child subjects consumed milk/cheese on a "regular" or "daily" basis compared to 72% ($n = 36$) of NT child subjects (qval, 2.3E−3). Finally, gluten intolerance was found to be more prevalent in the ASD cohort (qval, 3.5E−4). Additionally, 20 ASD child subjects had other special dietary restrictions apart from dairy and gluten constraints compared to only 6 NT child subjects (qval, 2.3e−3). Refer to Table 1 and Table S1 for a summary of the remaining reported data.

**Dietary and lifestyle habits influencing the microbial community.** Figure S1 in the supplemental material detailed the five variables that seem to significantly influence the microbial community: probiotics, multivitamin, sugary sweets, olive oil, and sequencing batch. Constraint principal coordinate analysis (PCoA) were used to identify the amplicon sequence variants (ASVs) for which the abundance was the most influenced by these variables.

Because none of the factors that may influence the microbial structure (probiotics, multivitamin, sugary sweets, olive oil, and sequencing batch) was significantly different between the two cohorts, we did not expect confounding effects when identifying ASVs specific to each cohort. Note that we did not detect any significant differences in the microbial community of our samples significantly associated with the reported binary gender of the children.

**Similarity between sibling lifestyles.** Sibling lifestyles, as measured by our online questionnaire were significantly more similar to each other than to other participants ($P \ll 0.01$) (Fig. S2). We observed significant differences in gluten and dairy intolerances. We did not, however, observe any significant differences between our cohorts regarding the reported gastrointestinal motility (Fig. S2C and S2D) or the frequency distribution of bowel movements (two-sample Wilcoxon signed-rank test, $P$ value = 0.8313, Fig. S2E, S2A, and S2B). We did not observe statistically significant results either when samples were agglomerated into two categories, typical bowel movement (one per day) and abnormal bowel movement (less or more than one per day) (chi-square $P = 0.17$), despite the percentage of "tend to have normally formed stool" higher in the neurotypical

**TABLE 1** Clinical characteristics for ASD and NT participants with significant difference between the cohorts

| Clinical characteristic | ASD participants ($n = 60$) | | | NT participants ($n = 57$) | | | Adjusted $P$ value |
|---|---|---|---|---|---|---|---|
| | No. of participants/ total no. of participants | % | | No. of participants/ total no. of participants | % | | |
| Gender[a] | | | | | | | |
| Male | 43/60 | 71.7 | | 27/49 | 55.1 | | 2.3e−2 |
| Female | 17/60 | 28.3 | | 22/49 | 44.9 | | |
| Gluten intolerance[b] | | | | | | | |
| Yes | 22/59 | 34.4 | | 1/56 | 1.8 | | 3.5E−4 |
| No | 42/59 | 71.2 | | 55/56 | 98.2 | | |
| Nutritional/herbal supplement[c] | | | | | | | |
| Yes | 35/55 | 63.6 | | 18/51 | 35.3 | | 2.3E−2 |
| No | 20/55 | 36.4 | | 33/51 | 64.7 | | |
| Special diet restrictions[d] | | | | | | | |
| Yes | 20/54 | 37.0 | | 6/51 | 11.8 | | 2.3E−2 |
| No | 34/54 | 63.0 | | 45/51 | 88.2 | | |
| Dairy intolerance | | | | | | | |
| Yes | 16/60 | 26.7 | | 1/57 | 1.8 | | 2.6E−6 |
| No | 44/60 | 73.3 | | 56/57 | 98.2 | | |
| Consumption of at least 2 servings of milk or cheese a day[e] | | | | | | | |
| Never | 23/54 | 42.6 | | 6/50 | 12.0 | | |
| Rarely (a few times a month) | 8/54 | 14.8 | | 2/50 | 4.0 | | |
| Occasionally (1 or 2/week) | 5/54 | 9.3 | | 6/50 | 12.0 | | 2.3E−3 |
| Regularly (3 to 5/week) | 7/54 | 12.9 | | 18/50 | 36.0 | | |
| Daily | 11/54 | 20.4 | | 18/50 | 36.0 | | |

[a]Missing 8 NT gender responses.
[b]Missing 1 NT and 1 ASD responses.
[c]Missing 6 NT and 5 ASD responses.
[d]Missing 6 NT and 6 ASD responses.
[e]Missing 7 NT and 6 ASD responses.

cohort (63% compared to 52%). Finally, we performed a Levene test and determine that the variance was not significantly different between the cohorts as well.

**Microbial alpha-diversity.** We calculated the phylogenetic diversity (PD) and Shannon diversity metric for each sample (27). Performing a rank sum test using each metric, we found no significant difference between cohorts. However, the variance of diversity (distribution of scores) in the ASD cohort was significantly greater than the NT cohort (bootstrap $P < 0.001$; Fig. 1) with both diversity estimators (Fig. 1C and F). Shannon diversity was not associated with ASD, but rather with bowel movement quality (exact Fisher $P = 0.02$) with low diversity associated with diarrhea (Fig. 1A and B), but not significantly related to bowel movement frequency (exact Fisher $P = 0.17$). Shannon diversity was also significantly related to bowel movement quality (exact Fisher $P = 0.02$), with low diversity associated with diarrhea, but not significantly related to bowel movement frequency (exact Fisher $P = 0.17$).

**Permutation test on sibling pairs to determine ASVs that differentiate between ASD and NT.** Ten ASVs were determined to be differentially abundant between sibling pairs (ASD versus NT) as determined by a permutation test with false discovery rate (FDR) correction (Table 2; also see Text S1A in the supplemental material). The mean differential abundance drawn from the null distribution was never more extreme than the actual differential abundance mean, and all $P$ values were 0 and increased to $9.36 \times 10^{-3}$ upon correction (see distribution plot in Fig. S4). The genera *Aggregatibacter*, *Anaerococcus*, and *Oscillospira* were significantly enriched in the ASD cohort, while *Porphyromonas*, *Slackia*, *Desulfovibrio*, *Clostridium colinum*, and *Acinetobacter johnsonii* were enriched in the NT cohort.
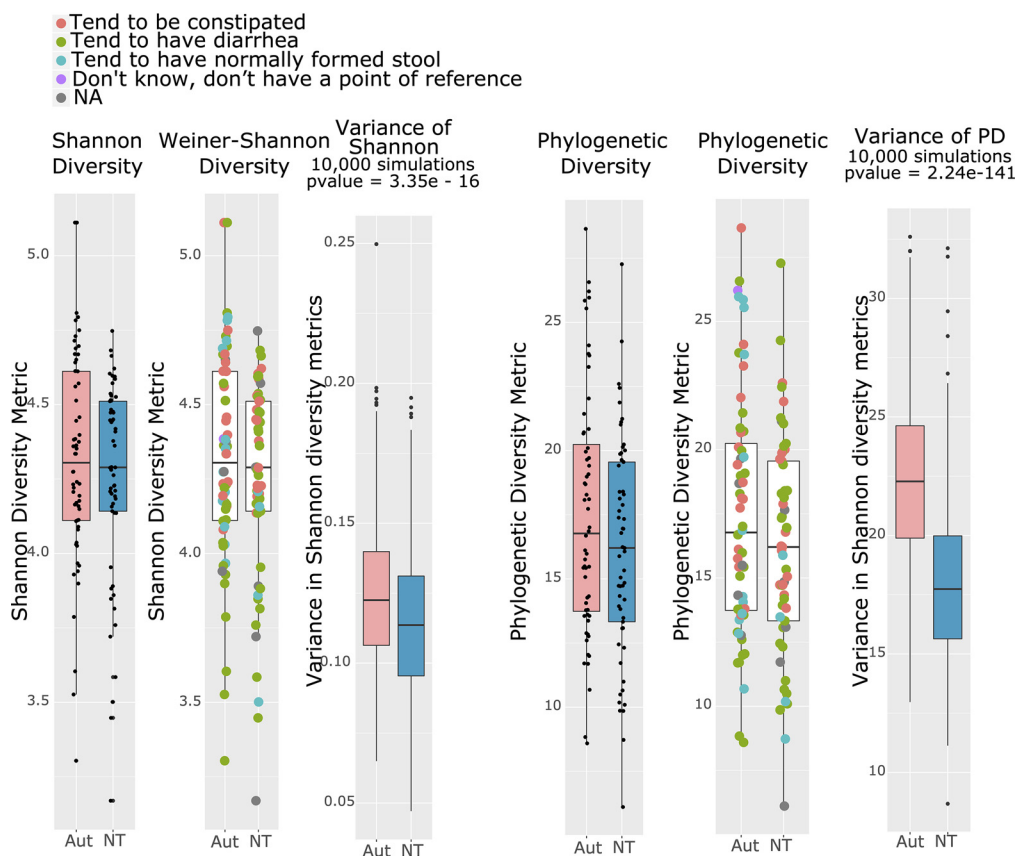
FIG 1 Phylogenetic diversity and Shannon diversity used as estimators of microbial alpha-diversity. The variance of diversity (distribution of scores) in the ASD cohort was significantly greater than that in the NT cohort (bootstrap $P < 0.001$) with both diversity estimators. Shannon diversity was also significantly related to bowel movement quality (exact Fisher $P = 0.02$), with low diversity associated with diarrhea, but not significantly related to bowel movement frequency (exact Fisher $P = 0.17$). NA, not available; PD, phylogenetic diversity.

**Models to maximize the likelihood of detecting rare ASVs.** We implemented a mixture model using a zero-inflated Gaussian (ZIG) distribution of mean group abundance for each ASV in metagenomeSeq (28) in order to quantify the fold change in taxa between the ASD cohort and the NT cohort. This analysis revealed 10 ASVs differentially present in the two cohorts: four were enriched in the ASD cohort (Table 2), and six were enriched in the NT cohort. The NT cohort was enriched in the *Lachnospiraceae* (five of six ASVs), including *Coprococcus catus*, *Clostridium colinum*, and the genera *Bifidobacterium*. In comparison, the ASD cohort was enriched in *Ruminococcus* and *Holdemania*, as well as the species *Bacteroides uniformis* and *Clostridium celatum*.

We also used a negative binomial distribution analysis to identify ASVs between ASD and NT cohorts, through which we identified a single ASV, from the genus *Bacteroides* (ASV1), enriched in the ASD cohort. Among the aforementioned statistical analyses, the microbial genus types identified in more than one statistical abundance test in the ASD cohort included the family *Ruminococcaceae* (by three ASVs including the genera *Oscillospira* and *Ruminococcus*) and the genus *Bacteroides* (by two ASVs). The NT cohort presents six ASVs belonging to the family *Lachnospiraceae*.

**Correlation between ASVs and MARA scores.** Four taxa were correlated with the MARA score in the ASD cohort, *Bacteroidales* genus *Prevotella*, *Campylobacterales* genus *Campylobacter*, and two ASVs from the *Clostridiales* order, and classified as genera *Peptoniphilus* and WAL_1855D.

**Functional profile prediction.** The software Piphillan predicted ~6,900 active KEGG orthologs (KOs) that were part of ~170 metabolic pathways as defined by KEGG Brite. Overall, we were able to associate 105 ASVs with full genome annotations. From the

predicted KOs that were present in these genomes, we observed 17 predicted metabolic pathways with significantly differential abundance between ASD and NT cohorts. Two pathways were significantly enriched in the ASD cohort: flagellar assembly (ko02040) and aminoacyl-tRNA biosynthesis (ko00970) (Fig. 2). Fifteen pathways were significantly enriched in the NT cohort, including butanoate metabolism (ko00650), propanoate metabolism (ko00640), sulfur metabolism (ko00920), phosphotransferase system (ko02060), and microbial metabolism in diverse environments (ko01120). A full list of significantly differential pathways is shown in Fig. 2 and Fig. S6.

## DISCUSSION

**Crowdsourcing recruitment, lifestyle, dietary practices, and GI symptoms.** By targeting the Internet-active autism community, we were able to crowdsource study subject recruitment from diverse geographical areas and reach our targeted sample size for each cohort in a short amount of time. This study design highlights ASD biomarker candidates while minimizing the impact of confounding environmental factors. By recruiting only sibling pairs who are within 2 years of age of one another, living in similar home environments, and eating similar diets (see Fig. S2F in the supplemental material), we successfully controlled for diet and lifestyle among our two cohorts. Using this approach of working with sibling cohorts, other studies have also shown very similar microbial structure between the two cohorts (29) to the point of not being able to identify taxa specific to a cohort.

We observed no overlap between factors that heavily influence the microbial structure (Fig. S1) and the factors that were significantly different between the ASD and the NT cohorts (Table 1). Therefore, it is unlikely that the ASVs identified as ASD or NT biomarkers were differentially abundant due to diet or lifestyle as confounding influences.

We did not observe significant differences in the GI motility or stool quality between cohorts; however, there was an increased prevalence of dairy and gluten sensitivities among ASD child subjects which may imply a propensity for GI distress and have been previously found to be associated with children with ASD (30–32). We also observed an increase in special dietary restrictions among our ASD child subjects, which could imply that many parents had already implemented limitations to their child's diet to alleviate any potential or previously identified gastrointestinal issues. Perhaps due to high parent involvement, we did not observe the expected differences in GI motility or gastrointestinal abnormalities, which is unusual for a study, especially with this sample size. Failure to detect systematic GI distress in ASD could also be attributed to differences in the populations recruited in this study, where the families involved were very interested by the potential impact of the diet on autism, and therefore potentially very aware and proactive in relieving GI discomforts. It can be noted that when focusing on the quality of the gastrointestinal motility (rather than quantitative), we observe a trend with a chi-square of 0.12, suggesting that maybe the frequency of bowel movements are a sensitive enough measure.

**Microbial community diversity.** There is much debate in the literature as to whether microbial diversity is significantly different in children with ASD versus children with typical neurodevelopment. While we found no significant rank sum relationship between alpha-diversity of the microbiota and ASD diagnosis, we observed a significant relationship when considering samples "high," "medium," or "low" diversity. The variance in diversity scores was significantly greater in the ASD cohort than in the NT cohort, which may explain some of the discrepancies seen in smaller cohort studies such as those conducted by Finegold et al. and Kang et al. (33, 34), which report increased and decreased microbial diversity, respectively, in the ASD cohort. It is worth noting that young children (between 2 and 7 years old) also experience some drastic changes in their gut microbial diversity (35), which may also contribute to the high variation in the diversity of the samples.

**Microbial taxon analysis.** The control cohort chosen in this study, typically developing siblings of ages within 2 years of the children with autism, has the advantage of closely matching the environment and diet within each sibling pair. The literature

**TABLE 2** Candidate 16S biomarkers enriched and depleted in the autism cohort and their annotation

| ASV | Phylum | Class | Order | Family | Genus | Species | Analysis | qvals | IMG genome best alignment (number of hits) |
|---|---|---|---|---|---|---|---|---|---|
| ASV1 | Bacteroidetes | Bacteroidia | Bacteroidales | *Bacteroidaceae* | *Bacteroides* | | DEseq2 | 2.34E-14 | *Bacteroides vulgatus* |
| ASV2 | Firmicutes | Clostridia | Clostridiales | | | | Pair | 8.92E-03 | - |
| ASV3 | Firmicutes | ASV10 | Clostridiales | *Ruminococcaceae* | | | Pair | 8.92E-03 | - |
| ASV4 | Firmicutes | Clostridia | Clostridiales | *Ruminococcaceae* | *Oscillospira* | | Pair | 8.92E-03 | - |
| ASV5 | Firmicutes | Clostridia | Clostridiales | *Tissierellaceae* | *Anaerococcus* | | Pair | 8.92E-03 | *Anaerococcus senegalensis* |
| ASV6 | Proteobacteria | Gammaproteobacteria | Pasteurellales | *Pasteurellaceae* | *Aggregatibacter* | | Pair | 8.92E-03 | *Haemophilus pittmaniae, Aggregatibacter* |
| ASV7 | Firmicutes | Clostridia | Clostridiales | *Ruminococcaceae* | *Ruminococcus* | | ZIG | 1.09E-05 | - |
| ASV8 | Bacteroidetes | Bacteroidia | Bacteroidales | *Bacteroidaceae* | *Bacteroides* | *uniformis* | ZIG | 4.88E-04 | *Bacteroides uniformis* (6), *Bacteroides sp. 4_1_36* (1) |
| ASV9 | Firmicutes | Erysipelotrichi | *Erysipelotrichales* | *Erysipelotrichaceae* | *Holdemania* | | ZIG | 2.20E-02 | *Holdemania filiformis* |
| ASV10 | Firmicutes | Clostridia | Clostridiales | *Clostridiaceae* | *Clostridium* | *celatum* | ZIG | 2.70E-02 | *Clostridium saudiense* [1*] |
| ASV11 | Firmicutes | Clostridia | Clostridiales | *Lachnospiraceae* | *[Clostridium]* | *colinum* | Pair | 8.92E-03 | - |
| ASV12 | Bacteroidetes | Bacteroidia | Bacteroidales | *Porphyromonadaceae* | *Porphyromonas* | | Pair | 8.92E-03 | - |
| ASV13 | Actinobacteria | Coriobacteriia | *Coriobacteriales* | *Coriobacteriaceae* | *Slackia* | | Pair | 8.92E-03 | *Slackia piriformis* |
| ASV14 | Proteobacteria | Deltaproteobacteria | *Desulfovibrionales* | *Desulfovibrionaceae* | *Desulfovibrio* | | Pair | 8.92E-03 | *Desulfovibrio fairfieldensis* (3), *Desulfovibrio sp. 3_1_syn3* (3), *Desulfovibrio sp. 6_1_46AFAA* (1) |
| ASV15 | Proteobacteria | Gammaproteobacteria | *Pseudomonadales* | *Moraxellaceae* | *Acinetobacter* | *johnsonii* | Pair | 8.92E-03 | *Acinetobacter johnsonii* (18), *Acinetobacter johnsonii* (4), *Acinetobacter schindleri* |
| ASV16 | Firmicutes | Clostridia | Clostridiales | *Lachnospiraceae* | *Lachnospira* | | ZIG | 2.12E-03 | *Eubacterium eligens* [2*] |
| ASV17 | Firmicutes | Clostridia | Clostridiales | *Lachnospiraceae* | | | ZIG | 9.73E-03 | - |
| ASV18 | Firmicutes | Clostridia | Clostridiales | *Lachnospiraceae* | | | ZIG | 3.52E-03 | - |
| ASV19 | Firmicutes | Clostridia | Clostridiales | *Lachnospiraceae* | *[Ruminococcus]* | | ZIG | 1.64E-02 | *Ruminococcaceae bacterium GD1* (2), *Ruminococcus sp. DSM 100440* (1), *Clostridiales bacterium VE202-14* (1), *Sellimonas intestinalis BR72* (1) |
| ASV20 | Firmicutes | Clostridia | Clostridiales | *Lachnospiraceae* | *Coprococcus* | *catus* | ZIG | 3.00E-02 | *Coprococcus catus* |
| ASV21 | Actinobacteria | Actinobacteria | *Bifidobacteriales* | *Bifidobacteriaceae* | *Bifidobacterium* | | ZIG | 4.87E-02 | *Bifidobacterium pseudocatenulatum* (9), *Bifidobacterium catenulatum* (6), *Bifidobacterium gallicum* (4), *Bifidobacterium kashiwanohense* (2) |

*ASVs with taxa specific to the ASD cohort* (ASV1–ASV10, ASD CANDIDATE BIOMARKERS)

*ASVs with taxa specific to the NT cohort* (ASV11–ASV21, NEUROTYPICAL CANDIDATE BIOMARKERS)

[1*] *Clostridium celatum* being the second hit with 232/233 pb [2*] *Lachnospira* first hit 221/233

indicates that interpersonal distances of the microbial community as a whole using beta-diversity analysis of children within family is significantly smaller than between families (35, 36). Because they share the same environment and because of the possible horizontal microbial transmission, our study design may hinder some 16S amplicon biomarkers that would be exchanged between the two siblings.

This study relied exclusively on ASVs, which means that the taxonomic comparisons were performed without any clustering of 16S rRNA amplicon sequences (Table 2) (37). This approach produces single sequence variants which can be reproducibly detected between studies and across sequencing runs and improving taxonomic resolution (38). Among those variants, one ASV was also correlated with the MARA score within the ASD cohort.

We provided in Table 3 and Table 4 a comprehensive list of biological information and information related to other reports on each bacterial association with ASD and other pertinent phenotypes. Our findings agree overall with the literature in that the *Bacteroides* genus and families such as *Erysipelotrichaceae* and *Clostridiaceae* (member of clostridial cluster I) have already been widely reported as enriched in ASD (Table 3). We also observed that members of clostridial cluster IV (genus *Ruminococcus*) and family *Pasteurellaceae* (ASV5) were both enriched in the ASD cohort. This entire family was previously reported as being depleted in ASD participants (34); this discrepancy could be explained by the aggregation of all *Pasteurellaceae*, while we implicate a single member of the family. *Pasteurellaceae* was also detected as one of the most abundant
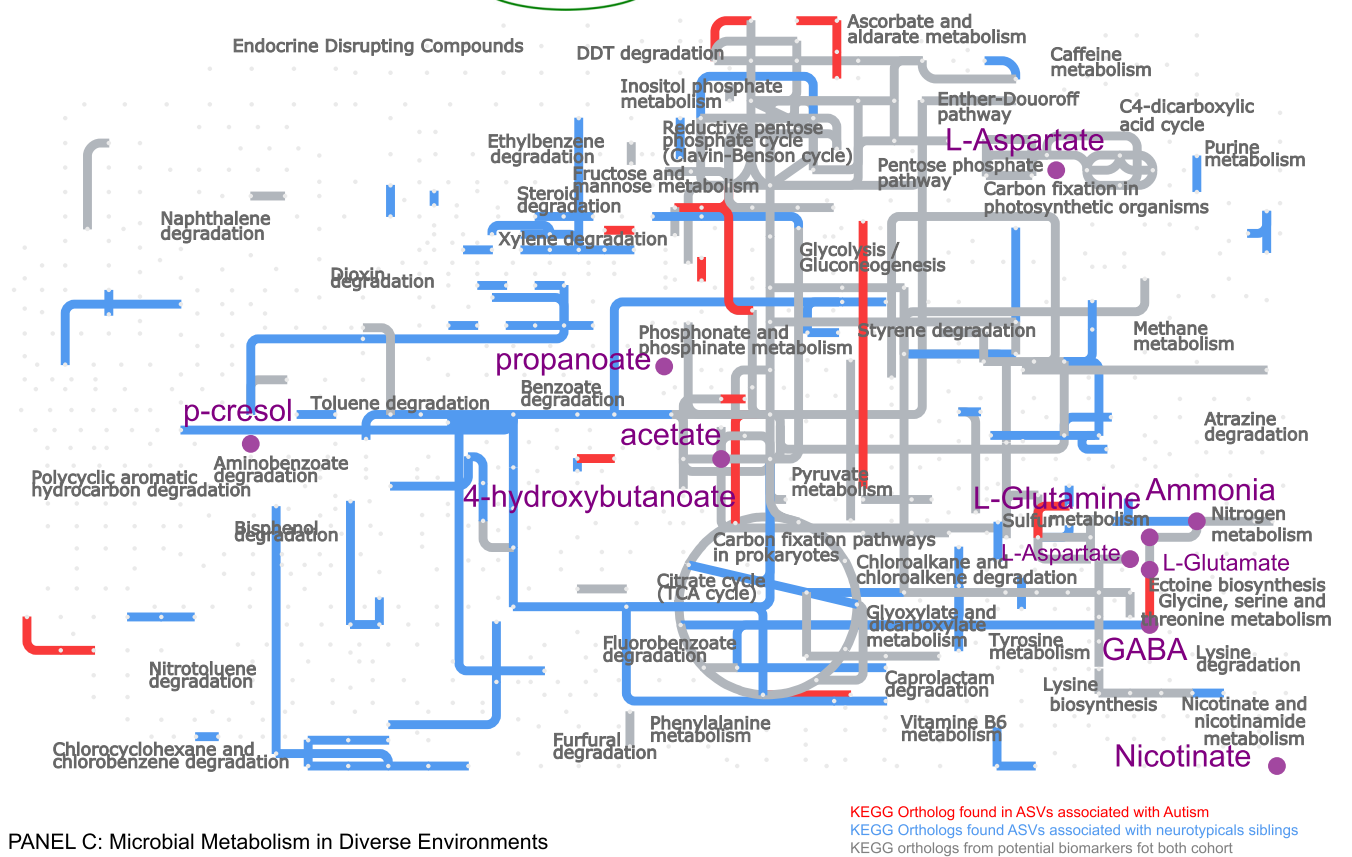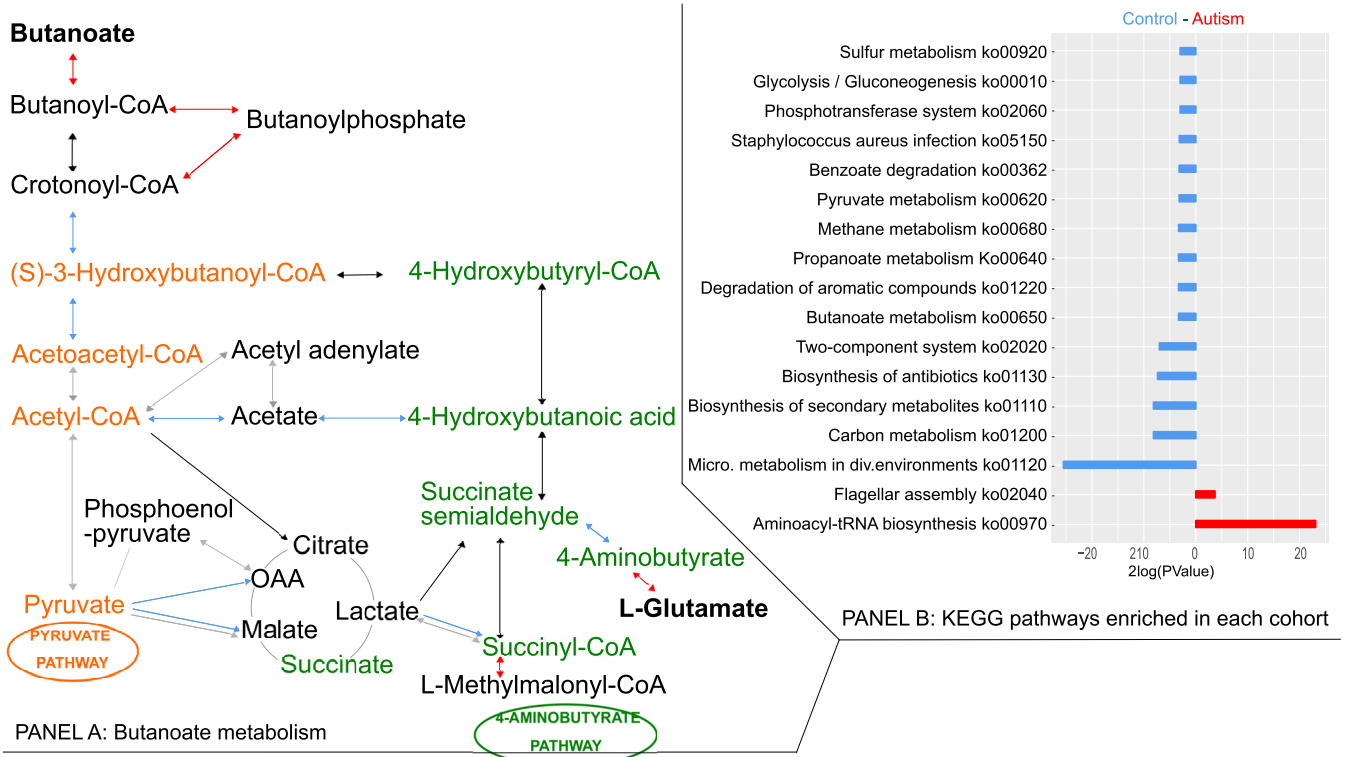
FIG 2 Pathway analyses derived from amplicon sequence variant analysis. (A) Butanoate metabolism. Detailed analysis of the butanoate pathway is shown. The color of the arrow reflects the cohort in which the ASV carrying the KEGG ortholog was detected. CoA, coenzyme A; OAA, oxaloacetate. (B) KEGG pathways enriched in each cohort. A list of the 17 pathways enriched in the gene set enrichment analysis using genomes and abundances estimated from the ribosomal analysis is shown. (C) Microbial metabolism in diverse environments. A detailed analysis of the pathway microbial metabolism in diverse environments and metabolites of interest for the gut-brain axis interaction is shown. TCA cycle, trichloroacetic cycle; DDT, 1,1,1-trichloro-2,2-bis(p-chlorophenyl)ethane.

bacterial families in children with developmental disabilities in Japan (39), supporting its potential association with atypical behavioral phenotypes. Finally, we also found that the genus *Anaerococcus* (ASV5) was enriched in the ASD cohort which to our knowledge has not yet been reported in previous research literature.

Our analysis also identified six ASVs belonging to the *Lachnospiraceae* family that are depleted in the ASD cohort and enriched in the neurotypical cohort. This family overall has already been reported as associated with autism phenotype (Table 3). Of these ASVs, the Ribosomal Database Project (RDP) classifier was able to assign only two species names, and we were only able to identify three genomes carrying similar ribosomal sequences (Table 2), indicating that we may have identified novel variants from our analyses (40, 41). These homogeneous phylogenetic groups of ASVs seem especially interesting as they cluster near each other on the 16S rRNA phylogenic tree (Fig. 3). Additionally, members from clostridial cluster IV were associated with ASD, while members from *Clostridium* cluster XIVa were associated with the control cohort. We could hypothesize that these two families exhibit metabolic pathways that are distinct among their functional redundancies. As microbes from this genus are some of human gut-associated microbiomes main butyrate producers, we examine the differences in butyrate production pathways of these two clusters in our pathway analysis below. The genera *Desulfovibrio* and *Bifidobacterium* were also depleted in the ASD cohort, consistent with results from at least three other studies (Table 4). *Bifidobacterium* has been characterized for its ability to normalize gut permeability (14). Finally, we identified two more families depleted in the ASD cohort, *Porphyromonadaceae* and *Moraxellaceae*, which produce butyrate or use it as the sole source of carbon, respectively.

**Pathway analysis.** We identified 17 predicted pathways associated with either cohort (Fig. 2). It should be noted that connecting enriched predicted pathways to the potential biomarker ASVs reported can be tenuous, as Piphillin was able to match only 6 of our 21 markers to full genomes and extract their associated KOs. Notable differentially predicted pathways include the following.

**(i) Butyrate production pathway.** The potential role of short-chain fatty acids (SCFAs) in autism has been discussed in multiple studies. Wang et al. reported elevated SCFA concentration in children with ASD (42), while two other studies reported the opposite trend when looking at total SCFAs (43, 44). Macfabe et al. found that intravenous administration of the SCFA propionate induced ASD typified behavior in mouse models (45). Butyrate, in particular, has been proposed as a potential major mediator of the gut-brain axis either through modulation of the density of cholinergic enteric neurons through epigenetic mechanisms or through direct modulation of the vagus nerve and hypothalamus (46). Our gene set enrichment analysis (GSEA) revealed that the stool microbiome of NT participants carry microbial genomes capable of butyrate metabolism, implying that butyrate production and consumption pathways may be enriched in the NT cohort, in Fig. 2. Butyrate production pathways in commensal microbial species and pathogens are thought to have evolved divergently; there are four pathways for butyrate production, each branching from a different initial substrate: pyruvate, 4-aminobutyrate, glutarate, and lysine (47). The by-products and influences of these major butyrogenic pathways could be relevant to host physiology. In our cohort, the predicted bacterial genetic potential in NT samples showed an enrichment for KOs associated with butyrate production from pyruvate, while in the ASD samples, the predicted functional potential was enriched for butyrate production via the 4-aminobutanoate (4Ab) pathway (Fig. 2A and Fig. S6. 4Ab (or gamma aminobutyric acid [GABA]) is a major inhibitory neurotransmitter and its biosynthesis can directly interfere with the amount of available glutamate (47). GABA has also been found in a few case reports in high levels in the urine and blood of young children with autism (48). This pathway can also potentially release harmful by-products such as ammonia (47), which were found to be elevated in feces of children with ASD (42). Genes identified as part of the pyruvate biosynthesis pathway were either found in both cohorts or only in predicted genomes associated with biomarkers from the NT cohort (Fig. 2 and Fig. S6B).

**TABLE 3** Taxa enriched in the ASD cohort

| Phylogeny | Commensal activity and/or potential relevance to ASD |
| --- | --- |
| Family: **Clostridiaceae** | Belongs to clostridial cluster IV |
| ASV 10: *Clostridium celatum* | Enrichment in ASD cohort: *Clostridium* genus (19, 33), *Clostridium histolyticum* (72), *Clostridium bolteae* in ASD (73) |
| | *Clostridia* and *Bacteroides* classes drive differences between ASD and NT in mouse guts (9) |
| | Depletion of *Clostridium leptum* in ASD (33) |
| | Correlated with high-fat diets and subsequently cognitive inflexibility (74) |
| | Produce *m*-tyrosine which has been shown to decrease neural catecholamine concentration levels and induce ASD typified behavioral abnormalities in animal models (75) |
| | Antibiotic vancomycin can be used to target the *Clostridium* species to provide short-term alleviation of ASD symptoms (76) |
| | Increased *Clostridiales* correlated with shorter gap between GI symptoms and time of onset of ASD symptoms (32) |
| | *Clostridia* produce both an enterotoxin and a neurotoxin and are generally very active metabolically. They may produce toxic substances like phenols, *p*-cresol, and various indole derivatives (16) |
| Genus: **Bacteroides** | Largest portion of the gut microbiome and helps digest vegetables and whole grain; produce butyrate and ferment glycans |
| ASV1, ASV8 | Responsible for biotransformation of bile acids, which in turn are associated with GI dysfunction in mouse (77) |
| | *Bacteroides vulgatus* found enriched ASD (33) |
| | Increased levels of the bacterium *Bacteroides vulgatus* lead to increased brain levels of propionic acid, known to cause symptoms characteristic of autism when injected into the brains of rats (78) |
| | *In situ* hybridization targeting *Bacteroides* found no association with ASD (72) |
| Family: **Ruminococcaceae** | Aids in digestion of resistant starches, belongs to clostridial cluster IV, and produces butyrate |
| ASV3 and ASV4: *Ruminococcus* | Helps reverse infectious diarrhea by slowly digesting resistant starches (79) |
| | *Ruminococcus torques* associated with increased severity of irritable bowel syndrome (80) and enriched in ASD (81) |
| | *Ruminococcus* potentially predictive fecal biomarkers for dysregulation of central brain neurometabolite *N*-acetylaspartate mediated through serum cortisol in young pigs (82). Same neurometabolite was reported altered in ASD (83) |
| ASV7: *Oscillospira* | *Oscillospira* aids in the breakdown of complex carbohydrates by fermenting resistant starches and produces butyrate |
| Family: **Erysipelotrichaceae** | Family commonly found in gut microbiome of mice on high-fat diet (84), which itself is often associated with neurobehavioral change (85) |
| ASV9: *Holdemania* | *Erysipelotrichaceae Turicibacter sanguinis* found enriched in ASD (86) |
| | Entire *Erysipelotrichaceae* family also found depleted in ASD individuals (87) |
| Family: **Pasteurellaceae** | Periodontopathic species. Demonstrate opportunistic pathogenicity, including brain abscess (88) |
| ASV6: *Aggregatibacter/ Haemophilus pittmaniae*[a] | *Aggregatibacter* was detected as one of the most abundant bacterial genera in children with developmental disabilities in Japan (39) |
| | *Aggregatibacter* was isolated from urine of children with ASD (17) |
| Family: **Tissierellaceae** | Isolated for infectious sites and bacterial abscesses, especially ovarian (89) |
| ASV5: *Anaerococcus* | |

[a]Discrepancy in annotation. The RDP classifier using GreenGenes 13.8 (last updated 2013) assigned *Aggregatibacter* Unknown, while the IMG database (updated continuously) assigned *Haemophilus pittmaniae*. Both genera are strongly related to each other, and the new genus *Aggregatibacter* was created to accommodate some former *Haemophilus* and *Actinobacillus* species.

**(ii) Propionate pathway.** The predicted pathways for the synthesis of propionate, another SCFA, appears depleted in the ASD cohort (Fig. 2C). This pathway generates isopropanol as a by-product, which has recently been found at a greater concentration in the feces of children with autism (34).

**TABLE 4** Taxa enriched in the neurotypical cohort[a]

| Phylogeny | Commensal activity and/or potential relevance to ASD |
|---|---|
| Family: *Lachnospiraceae* | Member of *Clostridium* cluster XIV |
| ASV17 and ASV18 | Help digest fiber and produces butyrate (40) |
| ASV16: *Lachnospira* | Family shown depleted in inflammatory bowel disorder (90) |
| ASV20: *Coprococcus catus* | Genus *Coprococcus* found depleted in ASD (11) |
| ASV19: *Ruminococcus* | *Lachnospiraceae* depletion in ASD suggests lack of bacterial taxa important for carbohydrate degradation. |
| | |
| Family: *Desulfovibrionaceae* | *Desulfovibrio* genus potentially key influential organisms in ASD (91) as children with ASD commonly have low blood levels of sulfur and high urinary excretion |
| ASV14: *Desulfovibrio* | *D. pigers*, *D. desulfuricans*, and *D. intestinalis* found enriched in severe ASD, but multiple hypothesis testing correction was not performed (91) |
| | Increased abundance correlated with improvement of GI and improvement of behavioral ASD symptoms after microbial transfer therapy (17) |
| | |
| Family: *Bifidobacteriaceae* | Provides protection from pathogenic infections (92) |
| ASV21: *Bifidobacterium* | Assists in normalization of gut permeability and inhibits inflammatory cytokine interleukin-10 |
| | Aid in prevention of diarrhea, reduce food allergies, and help digest lactose (93) |
| | Found depleted in ASD (33) |
| | Found increased in ASD after microbiota transfer therapy that correlated with improvement of GI and behavioral symptoms (17) |
| | |
| Family: *Porphyromonadaceae* | Produces butyrate (94) |
| ASV12 | Genus *Porphyromonas* harbors several species known to be pathogens for oral cavity, namely *P. gingivalis* (95) |
| | *P. gingivalis* produces a unique capsular polysaccharide that triggers Toll-like receptor 2- dependent anti-inflammatory mechanisms in autoimmune encephalomyelitis mice (model for multiple sclerosis) (96, 97) |
| | |
| Family: *Coriobacteriaceae* | Relatively little physiological information available about the *Slackia* genus |
| ASV13: *Slackia* | |
| | |
| Family: *Moraxellaceae* | Has been characterized in soil and water for its catabolic property in degrading aromatic compounds (98) |
| ASV15: *Acinetobacter johnsonii* | Multiple strains of this species have also been reported to develop multiantibiotic resistance, most likely through horizontal transfer reshuffling (99) |
| | Has capacity to proliferate on butyrate as the sole carbon source (100) |

[a]We will not discuss the results related to ASV2, as it was not possible to assign any taxonomy beyond the order level.

**(iii) Sulfur pathway.** The association between the predicted sulfur pathway and the NT cohort, though somewhat surprising, parallels our finding of *Acinetobacter* and *Desulfovibrio* enrichment in the NT cohort and has already been hypothesized as possible route modulating the gut-brain interaction in autism (49) (Fig. S6D).

**(iv) Phosphotransferase system.** ASVs associated with ASD seemed to show a much greater variety of carbohydrate transporters (Fig. S6G).

**(v) Microbial metabolism in diverse environments.** This high level category comprises many different pathways, and it is interesting to note that this KEGG category is associated with several metabolites already known in the literature as enriched or depleted in subjects with ASD (Fig. 2C). Among these metabolites were *p*-cresol (34) and ammonia (43) that have been found in greater abundance in the feces of children with autism; SCFAs (propanoate and acetate), which have mixed reports associating them with either children with ASD or controls (43, 44, 50); and neurotransmitters such as L-glutamate and GABA, which tend to be greater and lower, respectively, in feces of children with ASD (34). Glutamine, found in greater levels in the feces of ASD participants (34), also belongs to this KEGG pathway, as do several metabolites such as nicotinate and aspartate, another neurotransmitter (34, 43, 51).

**Limitations.** While these results show promising microbial differences between autism and typically developing children, potential limitations included reliance on
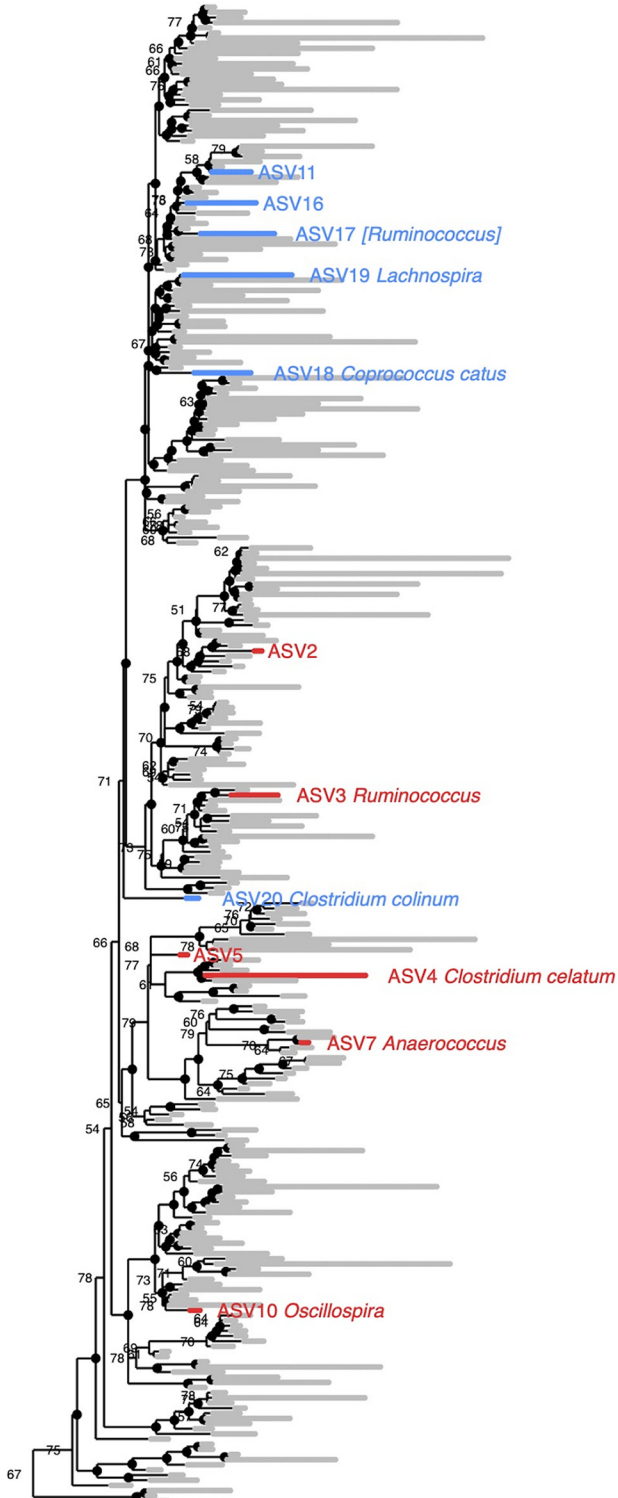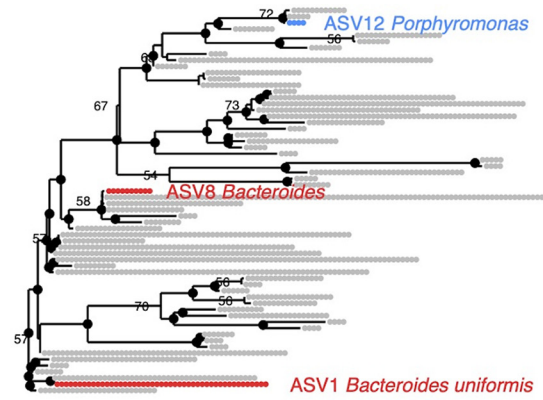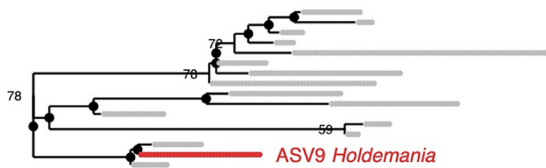
**FIG 3** Selected order from V4 ribosome-derived phylogenetic tree. Ribosome-derived phylogenetic tree (V4 alignments) of the order carrying amplicon sequence variants (ASVs) of interest. The ASVs in blue are enriched in the neurotypical cohort and the ones in blue in the autism cohort.

self-reported information, limited identification of species or strain level variants, limited single time point sampling, and lack of consideration of host genetic variation.

While we safeguarded against self-report bias through two validated machine-learning algorithms that adapt well to mobile testing and we gathered numerous lifestyle information, there remains bias in self-report of diagnosis. In particular, because we required MARA only for the child with ASD, we could not confirm the typical development of their siblings. In addition, the compliance with the optional request for video was slightly under <50% of the cohort studied. There is also a higher probability for a second child to have autistic symptoms (even subphenotypes of autism); and in these scenarios, the parents have a heightened sensitivity to and are more likely to watch for autistic symptoms and to request formal screening, following the CDC recommendations that doctors screen all siblings of autistic children for developmental delays starting at 9 months old (52). While it was encouraging to see perfect alignment between the MARA and the video classifier outcomes, bolstering confidence in the confirmation of self-reporting, it would be better to require this dual check for all participants in future work.

Although widely used (17), self-reported GI symptoms can also suffer some discrepancies compared to data reported by a pediatric gastroenterologist (45). Furthermore, while we observed physiological distinctions between the microbiomes of the cohorts on the level of amplicon sequence variants, it was often not possible to assign a taxonomic annotation or full genome to these sequences because of incomplete coverage in public databases. As the predicted pathways discussed were highly dependent on availabilities of full genome information, further metagenome and multi-omics analyses in this space will be needed to confirm the metabolic hypotheses presented.

## MATERIALS AND METHODS

**Crowdsourcing recruitment and data collection.** Data were collected from March 2015 to September 2017 under an approved Stanford University Institutional Review Board protocol (eProtocol 30205). We crowdsourced study subject recruitment by an online survey (https://microbiome.stanford.edu) via popular social media networking platforms and engaged with nonprofit and for-profit companies. Each sampling kit included two sets of collection tubes containing swabs to collect stool samples, instructions, and a detailed, 53-question dietary questionnaire for each child (see Table S3 and Table S1 in the supplemental material).

**ASD diagnosis confirmation.** The study enrolled only children that the parents indicated as previously diagnosed by a clinician with ASD, using standard of care with behavioral instruments such as the autism diagnostic observation schedule (ADOS) and the autism diagnostic interview-revised. To confirm this parent-provided ASD diagnosis of a child subject, we applied two machine-learning classifiers, one based on a parent-directed questionnaire (22, 25) and one based on a home video of the child with ASD (21, 22, 26). The parent-directed questionnaire is described below as "mobile autism risk assessment" or MARA, and the video-based classifier is referred to as "video classifier."

All participants electronically completed the clinically validated mobile autism risk assessment (MARA) (22, 25). Briefly, this system uses a set of seven behavioral features developed through machine learning for rapid screening for autism, focusing on the child's language ability, make-believe play, social activity, restricted and repetitive behaviors, general signs of developmental delays by or before age 3, and eye contact; the responses generate a score that classifies the child as either "ASD" (positive score) or "no ASD" (negative score).

We also requested a home video of their child with ASD. Due to both privacy and technical barriers to video upload, we made the video upload optional but required that all subjects complete the MARA as a strict inclusion criterion. The specific details of scoring and the validation of the classifier are described in previous publications (21, 22, 26). For the purposes of this study, we had at least three video raters who were blind to diagnosis independently tag the specific behavioral features that our video classifier requires to produce a risk score. To safeguard against ascertainment bias due to increasing familiarity, we required our video analysts to score an unlabeled mixture of ASD participant videos and similar home videos of neurotypical children aged 2 to 7 years mined from YouTube's publicly available videos. We took the majority consensus diagnosis as the outcome for comparison with the caregiver's self-reported diagnosis. Finally, we worked individually with the families that enrolled in the study to minimize the potential of false-negative controls by asking them to confirm that their second child was not on the autism spectrum (having never received a formal diagnosis, despite the other child receiving one).

**Lifestyle, antibiotics, and dietary practices.** Participants electronically completed questionnaires on behalf of their children, which included both dietary, supplement (including antibiotics) lifestyle

using a five-point frequency-based Likert scale and categorical answers (Table S1 and Table S3), including categorical data items were assigned either 1 or 0, and Likert scale items were assigned a value from 1 to 5 (1 = "Never," to 5 = "Always"). Differences of qualities with ordinal values were investigated using a linear-by-linear association test, and qualities with categorical values were tested using a chi-squared test. To verify that family relation was a practical criterion to ensure similarity between case and control lifestyle and dietary habits, we performed a permutation test (999 permutations) on the Euclidean distances between participants' numerical responses. Data were standardized to mean 0 and variance 1 to account for the differences in scale between categorical and Likert scale numerical values (see Fig. S2F in the supplemental material).

**DNA extraction, amplification, and sequencing.** Microbiome samples were processed according to the procedures outlined by the American Gut Project Protocol (53–56). Samples were shipped the day of collection by express mail (prestamped envelopes were provided to the participants) (56). Samples were stored at −80°C upon arrival (within 48 h of sampling) (9). DNA was extracted using the 96-well Powersoil DNA isolation kit (MO BIO, Carlsbad, CA). We utilized the manufacturer's protocol with the following modification: after the addition of the sample and solution C1, we partially submerged the sealed extraction plates in a water bath for 10 min at 65°C. We amplified the extracted DNA using the 5PRIME MasterMix (5 PRIME, Inc., Gaithersburg, MD) and the 515F/806R (targeting the V4 region) primers for a final concentration of $0.2\,\mu M$ per primer. The thermocycler settings for generating amplicons were as follows: (i) 3 min at 94°C; (ii) 35 cycles, with 1 cycle consisting of 45 s at 94°C, 1 min at 50°C, and 1.5 min at 72°C; (iii) a final extension step of 10 min at 72°C. After PCR, we quantified the DNA concentration of each sample using the Quant-iT PicoGreen double-stranded DNA (dsDNA) assay kit and then pooled to 70 ng DNA per sample. We generated clean pools using the QIAquick PCR purification kit (Qiagen, Hilden, Germany). The clean pools were then submitted to the Environmental Sample Preparation and Sequencing Facility at Argonne National Laboratory to be sequenced on an Illumina MiSeq using V4 chemistry.

**Sequence filtering, chimera removal, taxonomic assignment, and phylogenic tree.** We obtained a total of 3,419,972 reads. Raw sequences were processed using the workflow available in the software package DADA2 (31), which models and corrects amplicon errors. Reads were trimmed to include base pairs 10 through 140 and truncated at the first instance of a Phred quality score less than 20. Reads with more than two expected errors were filtered out. Reads were then dereplicated and denoised, and samples with less than 10,000 reads were discarded. Forward and reverse reads were merged, and chimeras were removed. Taxonomy was assigned to each amplicon sequence variant (ASV) generated by this pipeline by running the Ribosomal Database Project's (RDP) naive Bayesian classifier (57), implemented in DADA2, against the GreenGenes data set maintained in DADA2 package (58). The phylogenetic tree was rooted using an *Archaea* sequence from *Halorhabdus rudnickae* as an outgroup (see available GitHub code): all the sequences were aligned using the phangorn package and a neighbor-joining tree was built (59) using ape. The tree was bootstrapped 100 times with phangorn (60).

**Statistical analyses.** We performed statistical analyses with R version 3.4.2 (28 September 2017) using R Studio Integrated development environment for R v1.0.136 (open source software) (Boston, MA). All code used for this work is publicly available: https://github.com/walllab/Microbiome_16S_mbio.

**Analysis of alpha-diversity differences.** We calculated alpha-diversity for each sample using Shannon-Weiner diversity, a traditional metric that takes into account richness and evenness of taxonomic species, and phylogenetic diversity, a metric that measures the total length of phylogenetic branches necessary to span the set of taxa in a sample (60). We then used a Wilcoxon rank sum test (61) to quantify the significance of differences observed between the two cohorts. Next, we performed 1,000 bootstraps to calculate the variance of diversity metrics observed in each cohort and again used a rank sum test to quantify the significance of the difference in variances.

**Identification of dietary and lifestyle habits influencing the microbial community and identification of confounding factors.** To determine whether or not the parent-reported dietary and lifestyle questionnaires contained influential data and insight into the microbial communities observed in our samples, we used a permutational multivariate analysis of variance (PERMANOVA) test (ADONIS function in vegan package) on Bray-Curtis distances agglomerating all the samples (without stratification). We also performed a test to measure the homogeneity of the dispersion (PERMDISP2 procedure) of each cluster in order to be more confident that the cluster-specific centroids were robustly different rather than due to disparities in cohort dispersions.

Once we identify which dietary and lifestyle habits are significantly associated with a microbial community structure across all samples, we used the intersection of covariate significantly different between the two cohorts (see "Lifestyle, antibiotics, and dietary practices" above) to identify which of these covariates could be confounding factors.

**Permutation test on sibling pair differentials.** In addition to community level trends, we investigated differential abundances of specific taxa. We ran a permutation test on mean taxon abundance differences between sibling pairs to determine whether any taxa were systematically enriched or depleted in our ASD samples compared to our NT controls. Details can be found in Text S1A in the supplemental material.

**Models to maximize the likelihood of detecting rare ASVs.** Given the sparsity of 16S rRNA gene sequencing due to sequencing depth limitations, we used differential ribosomal analysis based on the negative binomial distribution and zero-inflated Gaussian analysis, to estimate log fold changes of taxa abundances between our ASD and NT groups (62).

**Differential analysis of 16S amplicon. (i) Negative binomial distribution.** We performed differential analysis of taxa counts between groups by modeling taxon abundances under a negative binomial

model using the DESeq2 framework (63). This method performs variance stabilization on taxon counts and then fits a generalized linear model with a log link on normalized count data.

**(ii) Zero-inflated Gaussian analysis.** To account for sparsity due to undersampling, we used the method developed by Paulson et al. (28), a mixture model that uses a zero-inflated Gaussian distribution to account for various depths of coverage. This method uses cumulative sum scaling (CSS) for normalization to account for undersampling and increase the sensitivity and specificity of identifiable taxa.

**(iii) Spearman correlation.** Finally, we also leverage MARA classifier scores to determine whether any ASVs were correlated with phenotype scoring within the ASD cohort only.

**Functional profile prediction using Piphillan.** We inferred metabolic activity using Piphillan (64). Piphillan aligns 16S sequences to sequences in the GreenGenes database (58) and assigns functional profiles based on 16S sequences, 16S sequences that do not match a database entry are assigned the functional profile of their nearest neighbor. Piphillan was run on DESeq2-normalized data (see Fig. S5 for normalization justification) to produce estimates of KEGG ortholog abundances (65–67). We then performed a modified gene enrichment analysis: a set was considered a metabolic pathway as defined by the KEGG Brite database, and an element was considered a KEGG ortholog (68).

**Gene set enrichment analysis using KEGG orthologs.** Using KO prediction provided by Piphillan, we assigned to each ASV the functional pathways using the KEGG database (67). As some KOs contribute to multiple pathways, we first divided the abundance of each KO in a sample by the number of pathways the KO participates in, so that a single functional unit's activity or output may contribute to only one pathway at a time (as a rule of thumb). Then we used the Gage implementation of gene set enrichment analysis (GSEA) (69) to perform a modified gene enrichment analysis: a set was considered a metabolic pathway as defined by the KEGG Brite database, and an element was considered a KEGG ortholog (68).

**Power calculation.** We modeled microbial abundances as a Dirichlet-Multinomial, a model which has been proven to successfully reflect the abundances seen in naturally occurring microbial communities (70). Under this model, we estimated Method-of-Moments (MoM) parameters for each taxon and determined the stability of those estimates by comparing likelihood ratio test statistics over 1,000 Monte-Carlo simulations (71). From this simulation, we can determine the number of samples necessary to reach a given level of power when estimating parameter values. At a rejection threshold value of 0.05, 70 ASD child subjects and 70 NT subjects were required to obtain power above 0.99, and 45 ASD child subjects and 45 NT child subjects were sufficient to provide a power greater than 0.95. Though we do not explicitly use the Dirichlet-Multinomial model in further analyses, a high power in this context implies a high power in more complex downstream analyses that are not able to be simulated.

**Data accessibility.** All fastq files can be downloaded at http://files.cgrb.oregonstate.edu/David_Lab/ASD_study1/. Code and analysis are available on GitHub at https://github.com/walllab/Microbiome_16S_mbio.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**TEXT S1**, PDF file, 0.1 MB.
**FIG S1**, PDF file, 0.9 MB.
**FIG S2**, PDF file, 0.2 MB.
**FIG S3**, PDF file, 0.4 MB.
**FIG S4**, PDF file, 0.2 MB.
**FIG S5**, PDF file, 0.1 MB.
**FIG S6**, PDF file, 0.8 MB.
**TABLE S1**, XLSX file, 0.03 MB.
**TABLE S2**, PDF file, 0.02 MB.
**TABLE S3**, XLSX file, 0.02 MB.

and Cancer. D.P.W. is cofounder of Cognoa, a company focused on digital methods for healthy child development.

## REFERENCES

1. Maenner MJ, Shaw KA, Baio J, et al. 2020. Prevalence of autism spectrum disorder among children aged 8 years—Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2016. MMWR Surveill Summ 69:1–12. https://doi.org/10.15585/mmwr.ss6904a1.

2. Hall D, Huerta MF, McAuliffe MJ, Farber GK. 2012. Sharing heterogeneous data: the national database for autism research. Neuroinformatics 10:331–339. https://doi.org/10.1007/s12021-012-9151-4.

3. Lakshmi Priya MD, Geetha A. 2011. Level of trace elements (copper, zinc, magnesium and selenium) and toxic elements (lead and mercury) in the hair and nail of children with autism. Biol Trace Elem Res 142:148–158. https://doi.org/10.1007/s12011-010-8766-2.

4. Gardener H, Spiegelman D, Buka SL. 2009. Prenatal risk factors for autism: comprehensive meta-analysis. Br J Psychiatry 195:7–14. https://doi.org/10.1192/bjp.bp.108.051672.

5. Moore SJ, Turnpenny P, Quinn A, Glover S, Lloyd DJ, Montgomery T, Dean JC. 2000. A clinical study of 57 children with fetal anticonvulsant syndromes. J Med Genet 37:489−497. https://doi.org/10.1136/jmg.37.7.489.

6. Parner ET, Baron-Cohen S, Lauritsen MB, Jørgensen M, Schieve LA, Yeargin-Allsopp M, Obel C. 2012. Parental age and autism spectrum disorders. Ann Epidemiol 22:143–150. https://doi.org/10.1016/j.annepidem.2011.12.006.

7. de Magistris L, Familiari V, Pascotto A, Sapone A, Frolli A, Iardino P, Carteni M, De Rosa M, Francavilla R, Riegler G, Militerni R, Bravaccio C. 2010. Alterations of the intestinal barrier in patients with autism spectrum disorders and in their first-degree relatives. J Pediatr Gastroenterol Nutr 51:418−424. https://doi.org/10.1097/MPG.0b013e3181dcc4a5.

8. Boukthir S, Matoussi N, Belhadj A, Mammou S, Dlala SB, Helayem M, Rocchiccioli F, Bouzaidi S, Abdennebi M. 2010. Abnormal intestinal permeability in children with autism. Tunis Med 88:685–686.

9. Hsiao EY, McBride SW, Hsien S, Sharon G, Hyde ER, McCue T, Codelli JA, Chow J, Reisman SE, Petrosino JF, Patterson PH, Mazmanian SK. 2013. Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. Cell 155:1451–1463. https://doi.org/10.1016/j.cell.2013.11.024.

10. Bravo JA, Forsythe P, Chew MV, Escaravage E, Savignac HM, Dinan TG, Bienenstock J, Cryan JF. 2011. Ingestion of Lactobacillus strain regulates emotional behavior and central GABA receptor expression in a mouse via the vagus nerve. Proc Natl Acad Sci U S A 108:16050–16055. https://doi.org/10.1073/pnas.1102999108.

11. Kang D-W, Park JG, Ilhan ZE, Wallstrom G, Labaer J, Adams JB, Krajmalnik-Brown R. 2013. Reduced incidence of Prevotella and other fermenters in intestinal microflora of autistic children. PLoS One 8:e68322. https://doi.org/10.1371/journal.pone.0068322.

12. Messaoudi M, Violle N, Bisson J-F, Desor D, Javelot H, Rougeot C. 2011. Beneficial psychological effects of a probiotic formulation (Lactobacillus helveticus R0052 and Bifidobacterium longum R0175) in healthy human volunteers. Gut Microbes 2:256–261. https://doi.org/10.4161/gmic.2.4.16108.

13. Bercik P, Park AJ, Sinclair D, Khoshdel A, Lu J, Huang X, Deng Y, Blennerhassett PA, Fahnestock M, Moine D, Berger B, Huizinga JD, Kunze W, McLean PG, Bergonzelli GE, Collins SM, Verdu EF. 2011. The anxiolytic effect of Bifidobacterium longum NCC3001 involves vagal pathways for gut-brain communication. Neurogastroenterol Motil 23:1132–1139. https://doi.org/10.1111/j.1365-2982.2011.01796.x.

14. Desbonnet L, Garrett L, Clarke G, Bienenstock J, Dinan TG. 2008. The probiotic Bifidobacteria infantis: an assessment of potential antidepressant properties in the rat. J Psychiatr Res 43:164–174. https://doi.org/10.1016/j.jpsychires.2008.03.009.

15. Song Y, Liu C, Finegold SM. 2004. Real-time PCR quantitation of clostridia in feces of autistic children. Appl Environ Microbiol 70:6459–6465. https://doi.org/10.1128/AEM.70.11.6459-6465.2004.

16. Finegold SM, Molitoris D, Song Y, Liu C, Vaisanen M, Bolte E, McTeague M, Sandler R, Wexler H, Marlowe EM, Collins MD, Lawson PA, Summanen P, Baysallar M, Tomzynski TJ, Read E, Johnson E, Rolfe R, Nasir P, Shah H, Haake DA, Manning P, Kaul A. 2002. Gastrointestinal microflora studies

in late-onset autism. Clin Infect Dis 35:S6–S16. https://doi.org/10.1086/341914.

17. Kang D-W, Adams JB, Gregory AC, Borody T, Chittick L, Fasano A, Khoruts A, Geis E, Maldonado J, McDonough-Means S, Pollard EL, Roux S, Sadowsky MJ, Lipson KS, Sullivan MB, Caporaso JG, Krajmalnik-Brown R. 2017. Microbiota transfer therapy alters gut ecosystem and improves gastrointestinal and autism symptoms: an open-label study. Microbiome 5:10. https://doi.org/10.1186/s40168-016-0225-7.

18. Swan M. 2012. Crowdsourced health research studies: an important emerging complement to clinical trials in the public health research ecosystem. J Med Internet Res 14:e46. https://doi.org/10.2196/jmir.1988.

19. Comber A, Brunsdon C, See L, Fritz S, McCallum I. 2013. Comparing expert and non-expert conceptualisations of the land: an analysis of crowdsourced land cover data, p 243–260. In Spatial information theory. Springer International Publishing, Cham, Switzerland.

20. David MM, Babineau BA, Wall DP. 2016. Can we accelerate autism discoveries through crowdsourcing? Res Autism Spectr Disord 32:80–83. https://doi.org/10.1016/j.rasd.2016.09.001.

21. Duda M, Kosmicki JA, Wall DP. 2014. Testing the accuracy of an observation-based classifier for rapid detection of autism risk. Transl Psychiatry 4:e424. https://doi.org/10.1038/tp.2014.65.

22. Wall DP, Kosmicki J, DeLuca TF, Harstad E, Fusaro VA. 2012. Use of machine learning to shorten observation-based screening and diagnosis of autism. Transl Psychiatry 2:e100. https://doi.org/10.1038/tp.2012.10.

23. Kosmicki JA, Sochat V, Duda M, Wall DP. 2015. Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. Transl Psychiatry 5:e514. https://doi.org/10.1038/tp.2015.7.

24. Fusaro VA, Daniels J, Duda M, DeLuca TF, D'Angelo O, Tamburello J, Maniscalco J, Wall DP. 2014. The potential of accelerating early detection of autism through content analysis of YouTube videos. PLoS One 9:e93533. https://doi.org/10.1371/journal.pone.0093533.

25. Duda M, Daniels J, Wall DP. 2016. Clinical evaluation of a novel and mobile autism risk assessment. J Autism Dev Disord 46:1953–1961. https://doi.org/10.1007/s10803-016-2718-4.

26. Levy S, Duda M, Haber N, Wall DP. 2017. Sparsifying machine learning models identify stable subsets of predictive features for behavioral detection of autism. Mol Autism 8:65. https://doi.org/10.1186/s13229-017-0180-6.

27. Cadotte MW, Davies TJ, Regetz J, Kembel SW, Cleland E, Oakley TH. 2010. Phylogenetic diversity metrics for ecological communities: integrating species richness, abundance and evolutionary history. Ecol Lett 13:96–105. https://doi.org/10.1111/j.1461-0248.2009.01405.x.

28. Paulson JN, Stine OC, Bravo HC, Pop M. 2013. Differential abundance analysis for microbial marker-gene surveys. Nat Methods 10:1200–1202. https://doi.org/10.1038/nmeth.2658.

29. Gondalia SV, Palombo EA, Knowles SR, Cox SB, Meyer D, Austin DW. 2012. Molecular characterisation of gastrointestinal microbiota of children with autism (with and without gastrointestinal dysfunction) and their neurotypical siblings. Autism Res 5:419–427. https://doi.org/10.1002/aur.1253.

30. Vuong HE, Hsiao EY. 2017. Emerging roles for the gut microbiome in autism spectrum disorder. Biol Psychiatry 81:411–423. https://doi.org/10.1016/j.biopsych.2016.08.024.

31. Lau NM, Green PHR, Taylor AK, Hellberg D, Ajamian M, Tan CZ, Kosofsky BE, Higgins JJ, Rajadhyaksha AM, Alaedini A. 2013. Markers of celiac disease and gluten sensitivity in children with autism. PLoS One 8:e66155. https://doi.org/10.1371/journal.pone.0066155.

32. Williams BL, Hornig M, Buie T, Bauman ML, Cho Paik M, Wick I, Bennett A, Jabado O, Hirschberg DL, Lipkin WI. 2011. Impaired carbohydrate digestion and transport and mucosal dysbiosis in the intestines of children with autism and gastrointestinal disturbances. PLoS One 6:e24585. https://doi.org/10.1371/journal.pone.0024585.

33. Finegold SM, Dowd SE, Gontcharova V, Liu C, Henley KE, Wolcott RD, Youn E, Summanen PH, Granpeesheh D, Dixon D, Liu M, Molitoris DR, Green JA, III. 2010. Pyrosequencing study of fecal microflora of autistic and control children. Anaerobe 16:444–453. https://doi.org/10.1016/j.anaerobe.2010.06.008.

34. Kang D-W, Ilhan ZE, Isern NG, Hoyt DW, Howsmon DP, Shaffer M, Lozupone CA, Hahn J, Adams JB, Krajmalnik-Brown R. 2018. Differences

in fecal microbial metabolites and microbiota of children with autism spectrum disorders. Anaerobe 49:121–131. https://doi.org/10.1016/j.anaerobe.2017.12.007.

35. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI. 2012. Human gut microbiome viewed across age and geography. Nature 486:222–227. https://doi.org/10.1038/nature11053.

36. Hedin CR, van der Gast CJ, Stagg AJ, Lindsay JO, Whelan K. 2017. The gut microbiota of siblings offers insights into microbial pathogenesis of inflammatory bowel disease. Gut Microbes 8:359–365. https://doi.org/10.1080/19490976.2017.1284733.

37. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583. https://doi.org/10.1038/nmeth.3869.

38. Callahan BJ, McMurdie PJ, Holmes SP. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J 11:2639–2643. https://doi.org/10.1038/ismej.2017.119.

39. Naka S, Yamana A, Nakano K, Okawa R, Fujita K, Kojima A, Nemoto H, Nomura R, Matsumoto M, Ooshima T. 2009. Distribution of periodontopathic bacterial species in Japanese children with developmental disabilities. BMC Oral Health 9:24. https://doi.org/10.1186/1472-6831-9-24.

40. Yoon S-H, Ha S-M, Kwon S, Lim J, Kim Y, Seo H, Chun J. 2017. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol 67:1613–1617. https://doi.org/10.1099/ijsem.0.001755.

41. Microbiology Society. 2010. List of new names and new combinations previously effectively, but not validly, published. Int J Syst Evol Microbiol 60:1009–1010. https://doi.org/10.1099/ijs.0.024562-0.

42. Wang L, Christophersen CT, Sorich MJ, Gerber JP, Angley MT, Conlon MA. 2012. Elevated fecal short chain fatty acid and ammonia concentrations in children with autism spectrum disorder. Dig Dis Sci 57:2096–2102. https://doi.org/10.1007/s10620-012-2167-7.

43. De Angelis M, Piccolo M, Vannini L, Siragusa S, De Giacomo A, Serrazzanetti DI, Cristofori F, Guerzoni ME, Gobbetti M, Francavilla R. 2013. Fecal microbiota and metabolome of children with autism and pervasive developmental disorder not otherwise specified. PLoS One 8:e76993. https://doi.org/10.1371/journal.pone.0076993.

44. Adams JB, Johansen LJ, Powell LD, Quig D, Rubin RA. 2011. Gastrointestinal flora and gastrointestinal status in children with autism – comparisons to typical children and correlation with autism severity. BMC Gastroenterol 11:22. https://doi.org/10.1186/1471-230X-11-22.

45. Macfabe D. 2013. Autism: metabolism, mitochondria, and the microbiome. Glob Adv Health Med 2:52–66. https://doi.org/10.7453/gahmj.2013.089.

46. Liu H, Wang J, He T, Becker S, Zhang G, Li D, Ma X. 2018. Butyrate: a double-edged sword for health? Adv Nutr 9:21–29. https://doi.org/10.1093/advances/nmx009.

47. Anand S, Kaur H, Mande SS. 2016. Comparative in silico analysis of butyrate production pathways in gut commensals and pathogens. Front Microbiol 7:1945. https://doi.org/10.3389/fmicb.2016.01945.

48. Dhossche D, Applegate H, Abraham A, Maertens P, Bland L, Bencsath A, Martinez J. 2002. Elevated plasma gamma-aminobutyric acid (GABA) levels in autistic youngsters: stimulus for a GABA hypothesis of autism. Med Sci Monit 8:PR1–PR6.

49. Midtvedt T. 2012. The gut: a triggering place for autism – possibilities and challenges. Microb Ecol Health Dis 2012:23. https://doi.org/10.3402/mehd.v23i0.18982.

50. Wang L, Christophersen CT, Sorich MJ, Gerber JP, Angley MT, Conlon MA. 2011. Low relative abundances of the mucolytic bacterium Akkermansia muciniphila and Bifidobacterium spp. in feces of children with autism. Appl Environ Microbiol 77:6718–6721. https://doi.org/10.1128/AEM.05212-11.

51. Yap IKS, Angley M, Veselkov KA, Holmes E, Lindon JC, Nicholson JK. 2010. Urinary metabolic phenotyping differentiates children with autism from their unaffected siblings and age-matched controls. J Proteome Res 9:2996–3004. https://doi.org/10.1021/pr901188e.

52. Centers for Disease Control and Prevention. 2020. Screening and diagnosis of autism spectrum disorder for healthcare providers. Centers for Disease Control and Prevention, Atlanta, GA.

53. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci U S A 108(Suppl 1):4516–4522. https://doi.org/10.1073/pnas.1000080107.

54. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6:1621–1624. https://doi.org/10.1038/ismej.2012.8.

55. Parada AE, Needham DM, Fuhrman JA. 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ Microbiol 18:1403–1414. https://doi.org/10.1111/1462-2920.13023.

56. Song SJ, Amir A, Metcalf JL, Amato KR, Xu ZZ, Humphrey G, Knight R. 2016. Preservation methods differ in fecal microbiome stability, affecting suitability for field studies. mSystems 1:e00021-16. https://doi.org/10.1128/mSystems.00021-16.

57. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73:5261–5267. https://doi.org/10.1128/AEM.00062-07.

58. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069–5072. https://doi.org/10.1128/AEM.03006-05.

59. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454.

60. Schliep KP. 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27:592–593. https://doi.org/10.1093/bioinformatics/btq706.

61. Finegold SM. 2011. State of the art; microbiology in health and disease. Intestinal bacterial flora in autism. Anaerobe 17:367–368. https://doi.org/10.1016/j.anaerobe.2011.03.007.

62. Dixon P. 2003. VEGAN, a package of R functions for community ecology. J Veg Sci 14:927–930. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x.

63. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8.

64. Iwai S, Weinmaier T, Schmidt BL, Albertson DG, Poloso NJ, Dabbagh K, DeSantis TZ. 2016. Piphillin: improved prediction of metagenomic content by direct inference from human microbiomes. PLoS One 11:e0166104. https://doi.org/10.1371/journal.pone.0166104.

65. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45:D353–D361. https://doi.org/10.1093/nar/gkw1092.

66. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44:D457–D462. https://doi.org/10.1093/nar/gkv1070.

67. Kanehisa M. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28:27–30. https://doi.org/10.1093/nar/28.1.27.

68. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102:15545–15550. https://doi.org/10.1073/pnas.0506580102.

69. Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. 2009. GAGE: generally applicable gene set enrichment for pathway analysis. BMC Bioinformatics 10:161. https://doi.org/10.1186/1471-2105-10-161.

70. La Rosa PS, Brooks JP, Deych E, Boone EL, Edwards DJ, Wang Q, Sodergren E, Weinstock G, Shannon WD. 2012. Hypothesis testing and power calculations for taxonomic-based human microbiome data. PLoS One 7:e52078. https://doi.org/10.1371/journal.pone.0052078.

71. Mooney CZ. 1997. Monte Carlo simulation. SAGE Publications, Thousand Oaks, CA.

72. Parracho HM, Bingham MO, Gibson GR, McCartney AL. 2005. Differences between the gut microflora of children with autistic spectrum disorders and that of healthy children. J Med Microbiol 54:987–991. https://doi.org/10.1099/jmm.0.46101-0.

73. Kandeel WA, Meguid NA, Bjørklund G, Eid EM, Farid M, Mohamed SK, Wakeel KE, Chirumbolo S, Elsaeid A, Hammad DY. 2020. Impact of Clostridium bacteria in children with autism spectrum disorder and their anthropometric measurements. J Mol Neurosci 70:897–907. https://doi.org/10.1007/s12031-020-01482-2.

74. Magnusson KR, Hauck L, Jeffrey BM, Elias V, Humphrey A, Nath R, Perrone A, Bermudez LE. 2015. Relationships between diet-related changes in the gut microbiome and cognitive flexibility. Neuroscience 300:128–140. https://doi.org/10.1016/j.neuroscience.2015.05.016.

75. Shaw W. 2010. Increased urinary excretion of a 3-(3-hydroxyphenyl)-3-hydroxypropionic acid (HPHPA), an abnormal phenylalanine metabolite of Clostridia spp. in the gastrointestinal tract, in urine samples from patients with autism and schizophrenia. Nutr Neurosci 13:135–143. https://doi.org/10.1179/147683010X12611460763968.

76. Sandler RH, Finegold SM, Bolte ER, Buchanan CP, Maxwell AP, Väisänen ML, Nelson MN, Wexler HM. 2000. Short-term benefit from oral vancomycin treatment of regressive-onset autism. J Child Neurol 15:429–435. https://doi.org/10.1177/088307380001500701.

77. Golubeva AV, Joyce SA, Moloney G, Burokas A, Sherwin E, Arboleya S, Flynn I, Khochanskiy D, Moya-Pérez A, Peterson V, Rea K, Murphy K, Makarova O, Buravkov S, Hyland NP, Stanton C, Clarke G, Gahan CGM, Dinan TG, Cryan JF. 2017. Microbiota-related changes in bile acid & tryptophan metabolism are associated with gastrointestinal dysfunction in a mouse model of autism. EBioMedicine 24:166–178. https://doi.org/10.1016/j.ebiom.2017.09.020.

78. MacFabe DF, Cain DP, Rodriguez-Capote K, Franklin AE, Hoffman JE, Boon F, Taylor AR, Kavaliers M, Ossenkopp KP. 2007. Neurobiological effects of intraventricular propionic acid in rats: possible role of short chain fatty acids on the pathogenesis and characteristics of autism spectrum disorders. Behav Brain Res 176:149–169. https://doi.org/10.1016/j.bbr.2006.07.025.

79. Ze X, Duncan SH, Louis P, Flint HJ. 2012. Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. ISME J 6:1535–1543. https://doi.org/10.1038/ismej.2012.4.

80. Malinen E, Krogius-Kurikka L, Lyra A, Nikkilä J, Jääskeläinen A, Rinttilä T, Vilpponen-Salmela T, von Wright AJ, Palva A. 2010. Association of symptoms with gastrointestinal microbiota in irritable bowel syndrome. World J Gastroenterol 16: 4532–4540. https://doi.org/10.3748/wjg.v16.i36.4532.

81. Wang L, Christophersen CT, Sorich MJ, Gerber JP, Angley MT, Conlon MA. 2013. Increased abundance of Sutterella spp. and Ruminococcus torques in feces of children with autism spectrum disorder. Mol Autism 4:42. https://doi.org/10.1186/2040-2392-4-42.

82. Mudd AT, Berding K, Wang M, Donovan SM, Dilger RN. Serum cortisol mediates the relationship between fecal Ruminococcus and brain N-acetylaspartate in the young pig. Gut Microbes 8:589–600. https://doi.org/10.1080/19490976.2017.1353849.

83. Friedman SD, Shaw DW, Artru AA, Richards TL, Gardner J, Dawson G, Posse S, Dager SR. 2003. Regional brain chemical alterations in young children with autism spectrum disorder. Neurology 60:100–107. https://doi.org/10.1212/WNL.60.1.100.

84. Kaakoush NO. 2015. Insights into the role of Erysipelotrichaceae in the human host. Front Cell Infect Microbiol 5:84. https://doi.org/10.3389/fcimb.2015.00084.

85. Bruce-Keller AJ, Salbaum JM, Luo M, Blanchard E, IV, Taylor CM, Welsh DA, Berthoud HR. 2015. Obese-type gut microbiota induce neurobehavioral changes in the absence of obesity. Biol Psychiat 77:607–615. https://doi.org/10.1016/j.biopsych.2014.07.012.

86. DeAngelis KM, D'Haeseleer P, Chivian D, Simmons B, Arkin AP, Mavromatis K, Malfatti S, Tringe S, Hazen TC. 2013. Metagenomes of tropical soil-derived anaerobic switchgrass-adapted consortia with and without iron. Stand Genomic Sci 7:382–398. https://doi.org/10.4056/sigs.3377516.

87. Gilbert JA, Krajmalnik-Brown R, Porazinska DL, Weiss SJ, Knight R. 2013. Toward effective probiotics for autism and other neurodevelopmental disorders. Cell 155:1446–1448. https://doi.org/10.1016/j.cell.2013.11.035.

88. Nørskov-Lauritsen N. 2014. Classification, identification, and clinical significance of Haemophilus and Aggregatibacter species with host specificity for humans. Clin Microbiol Rev 27:214–240. https://doi.org/10.1128/CMR.00103-13.

89. Ezaki T, Kawamura Y, Li N, Li ZY, Zhao L, Shu S. 2001. Proposal of the genera Anaerococcus gen. nov., Peptoniphilus gen. nov. and Gallicola gen. nov. for members of the genus Peptostreptococcus. Int J Syst Evol Microbiol 51:1521–1528. https://doi.org/10.1099/00207713-51-4-1521.

90. Haas KN, Blanchard JL. 2017. Kineothrix alysoides, gen. nov., sp. nov., a saccharolytic butyrate-producer within the family Lachnospiraceae. Int J Syst Evol Microbiol 67:402–410. https://doi.org/10.1099/ijsem.0.001643.

91. Frye RE, Rose S, Slattery J, MacFabe DF. 2015. Gastrointestinal dysfunction in autism spectrum disorder: the role of the mitochondria and the enteric microbiome. Microb Ecol Health Dis 26:27458. https://doi.org/10.3402/mehd.v26.27458.

92. Fukuda S, Toh H, Hase K, Oshima K, Nakanishi Y, Yoshimura K, Tobe T, Clarke JM, Topping DL, Suzuki T, Taylor TD, Itoh K, Kikuchi J, Morita H, Hattori M, Ohno H. 2011. Bifidobacteria can protect from enteropathogenic infection through production of acetate. 469:543–547. https://doi.org/10.1038/nature09646.

93. Di Gioia D, Aloisio I, Mazzola G, Biavati B. 2014. Bifidobacteria: their impact on gut microbiota composition and their applications as probiotics in infants. Appl Microbiol Biotechnol 98:563–577. https://doi.org/10.1007/s00253-013-5405-9.

94. Vital M, Karch A, Pieper DH. 2017. Colonic butyrate-producing communities in humans: an overview using omics data. mSystems 2:e00130-17. https://doi.org/10.1128/mSystems.00130-17.

95. Spooner R, Weigel KM, Harrison PL, Lee K, Cangelosi GA, Yilmaz Ö. 2016. In situ anabolic activity of periodontal pathogens Porphyromonas gingivalis and Filifactor alocis in chronic periodontitis. Sci Rep 6:33638. https://doi.org/10.1038/srep33638.

96. Nichols FC, Housley WJ, O'Conor CA, Manning T, Wu S, Clark RB. 2009. Unique lipids from a common human bacterium represent a new class of Toll-like receptor 2 ligands capable of enhancing autoimmunity. Am J Pathol 175:2430–2438. https://doi.org/10.2353/ajpath.2009.090544.

97. Wang Y, Kasper LH. 2014. The role of microbiome in central nervous system disorders. Brain Behav Immun 38:1–12. https://doi.org/10.1016/j.bbi.2013.12.015.

98. Jiang Y, Qi H, Zhang XM. 2018. Co-biodegradation of anthracene and naphthalene by the bacterium Acinetobacter johnsonii. J Environ Sci Health A Tox Hazard Subst Environ Eng 53:448–456. https://doi.org/10.1080/10934529.2017.1409579.

99. Montaña S, Schramm ST, Traglia GM, Chiem K, Parmeciano Di Noto G, Almuzara M, Barberis C, Vay C, Quiroga C, Tolmasky ME, Iriarte A, Ramírez MS. 2016. The genetic analysis of an Acinetobacter johnsonii clinical strain evidenced the presence of horizontal genetic transfer. PLoS One 11:e0161528. https://doi.org/10.1371/journal.pone.0161528.

100. Van Veen HW, Abee T, Kleefsman AW, Melgers B, Kortstee GJ, Konings WN, Zehnder AJ. 1994. Energetics of alanine, lysine, and proline transport in cytoplasmic membranes of the polyphosphate-accumulating Acinetobacter johnsonii strain 210A. J Bacteriol 176:2670–2676. https://doi.org/10.1128/JB.176.9.2670-2676.1994.