Construction and Characterization of Synthetic Bacterial Community for Experimental Ecology and Evolution

Johannes Cairns[1†], Roosa Jokela[1‡], Jenni Hultman[1§], Manu Tamminen[2||], Marko Virta[1¶] and Teppo Hiltunen[1,2*]

[1] Department of Microbiology, University of Helsinki, Helsinki, Finland, [2] Department of Biology, University of Turku, Turku, Finland

Experimental microbial ecology and evolution have yielded foundational insights into ecological and evolutionary processes using simple microcosm setups and phenotypic assays with one- or two-species model systems. The fields are now increasingly incorporating more complex systems and exploration of the molecular basis of observations. For this purpose, simplified, manageable and well-defined multispecies model systems are required that can be easily investigated using culturing and high-throughput sequencing approaches, bridging the gap between simpler and more complex synthetic or natural systems. Here we address this need by constructing a completely synthetic 33 bacterial strain community that can be cultured in simple laboratory conditions. We provide whole-genome data for all the strains as well as metadata about genomic features and phenotypic traits that allow resolving individual strains by amplicon sequencing and facilitate a variety of envisioned mechanistic studies. We further show that a large proportion of the strains exhibit coexistence in co-culture over serial transfer for 48 days in the absence of any experimental manipulation to maintain diversity. The constructed bacterial community can be a valuable resource in future experimental work.

Keywords: microbial community, model system, synthetic ecology, experimental evolution, whole-genome sequencing

INTRODUCTION

Testing ecological and evolutionary theory in a highly controlled manner using simple laboratory setups with one or two microbial species (Fraser and Keddy, 1997; Buckling et al., 2009) has produced important insights into ecological interactions–e.g., competition, cooperation, and cross-feeding interactions (Helling et al., 1987; Treves et al., 1998; Rozen and Lenski, 2000; Shou et al., 2007; Harcombe, 2010); the role of cheaters (MacLean and Gudelj, 2006); predator–prey interactions (Shertzer et al., 2002); and host–parasite interactions (Morgan et al., 2005)–and evolutionary processes–e.g., the evolution of coexistence (Good et al., 2017), coevolution between species (Hall et al., 2011; Brockhurst and Koskella, 2013), and eco-evolutionary feedback dynamics (Yoshida et al., 2003; Hiltunen and Becks, 2014). However, there is an increasing awareness that ecological and evolutionary processes can be fundamentally altered in more complex multispecies communities owing to several features such as altered competitive interactions and multiple selection pressures (Dunham, 2007). Recent empirical findings show, for example, that pairwise

interactions can be strongly altered in the presence of other species (Kastman et al., 2016) and the rate of adaptation of species can differ between monocultures and communities (Lawrence et al., 2012). Even a basic understanding of certain characteristics of microbial life such as horizontal gene transfer (Smillie et al., 2011), metabolic interactions and spatial heterogeneity (Elias and Banin, 2012; van Gestel et al., 2014) requires investigation of multispecies settings integral to them. Furthermore, several key ecological features are specific to multispecies communities, such as diversity, stability, succession and high-order (e.g., four-way) species interactions (Bairey et al., 2016). There is therefore a profound need to expand the biotic complexity of study systems used in the fields of experimental microbial ecology and evolution.

The design of multispecies model communities in experimental ecology and evolution is part of the emerging field of synthetic ecology where synthetic communities are used for a plethora of basic and applied purposes. Several research attempts have sought mechanistic understanding of specific natural systems, such as methane consuming communities (Yu et al., 2017), plant root colonizing bacteria (Lebeis et al., 2015), the human gut microbiota (Goodman et al., 2011), and cheese rind communities (Wolfe et al., 2014), by complementing observational findings with findings from controlled *in vitro* or *in vivo* studies using synthetic communities. These studies focus on designing synthetic communities that capture the essential features of the natural system being investigated. A typical approach is to determine the prevalent taxa in the natural system or in the core microbiome (Shade and Handelsman, 2012) common to similar systems, to construct a synthetic community of taxonomically representative strains isolated from the natural system, and to culture the community in conditions mimicking the natural system. The limitations of this approach include the potentially important role in ecological functions or evolutionary processes of low-abundance taxa (Liu et al., 2017), microdiversity (Chase et al., 2017), or interactions between bacteria and members of other taxonomic groups such as viruses, unicellular eukaryotic predators or fungi, as well as technical difficulties in mimicking natural conditions in the laboratory (Wolfe, 2018). Likely owing to such limitations, among these studies, cases have been observed both where simple synthetic communities representing predominant taxa in natural systems recapitulate the dynamics in natural systems (Goodman et al., 2011; Wolfe et al., 2014; Lebeis et al., 2015) and where major differences are observed between synthetic and natural systems (Yu et al., 2016).

Compared with studies employing synthetic communities to understand specific natural systems, more applied studies focusing, among others, on medical therapeutics (Petrof et al., 2013; Sheth et al., 2016), bioremediation (Dejonghe et al., 2003; Zomorrodi and Segre, 2016), or biofuel production (Wang et al., 2015), rely even more heavily on a detailed understanding of the characteristics and functions of specific bacterial taxa in natural systems to engineer communities that can successfully perform desired functions. In contrast, studies attempting to investigate highly general ecological or evolutionary processes, similar to traditional experimental microbial ecology and evolution

using one- or two-species model systems, do not necessarily seek to, or prefer simple culture conditions over the ability to, accurately represent a particular natural community. For instance, a synthetic community of 72 bacterial strains isolated from tree-hole bacterial communities–but limited to aerobic heterotrophs cultivable in simple laboratory conditions–has been used to investigate several key ecological questions, including the relationship between diversity and ecosystem productivity (Bell et al., 2005) and the success of multispecies invasions during different stages of ecological succession (Rivett et al., 2018). Celiker and Gore (2014), in turn, used a completely synthetic model community comprising six apparently random soil bacterial strains from culture collections to examine the repeatability of change in community composition over time.

In simplified synthetic communities, verisimilitude is sacrificed to obtain relative ease of analysis and modeling, and control of species interactions, non-linear effects from added traits and strains, and evolutionary change (Widder et al., 2016). There is ongoing debate about the utility of simple microcosm setups to understand ecological and evolutionary phenomena (Carpenter, 1996, 1999; Fraser and Keddy, 1997; Drenner and Mazumder, 1999; Benton et al., 2007; Buckling et al., 2009), yet the approach continues to produce major scientific discoveries (van Houte et al., 2016; Good et al., 2017; Betts et al., 2018; Frickel et al., 2018). Similarly, the definition of, need for, and necessary level of representativeness of synthetic communities remain matters of debate (Dolinsek et al., 2016; Widder et al., 2016; Zomorrodi and Segre, 2016; Wolfe, 2018), and are likely strongly dependent on the research questions. Although debated, completely synthetic communities composed of strains isolated from different habitats can also be used to study general questions as well as having special use in studying questions such as community assembly and the evolution of coexistence in newly formed communities.

In this context, the detailed, mechanistic understanding of simpler, less representative synthetic communities can be thought to inform, or even be a prerequisite to, understanding more complex synthetic systems, and ultimately, natural systems. It has been argued that such efforts should focus on understanding a limited set of well-defined model synthetic communities, which would make results comparable between studies and allow collaborative efforts toward mechanistic understanding (Widder et al., 2016). However, not many such systems exist to our knowledge despite the general boom in synthetic ecology. Furthermore, the highly general level studies that exist primarily focus on simple phenotypic analyses (Bell et al., 2005; Foster and Bell, 2012; Celiker and Gore, 2014; Rivett et al., 2016, 2018; Rivett and Bell, 2018), although high-throughput molecular methods, such as amplicon sequencing, (meta)genomics, (meta) transcriptomics, proteomics and metabolomics, which have been promisingly utilized in studies focusing on synthetic systems mimicking natural systems (Goodman et al., 2011; Wolfe et al., 2014; Lebeis et al., 2015; Stopnisek et al., 2016; Yu et al., 2017), could provide valuable insights into the mechanisms behind observed phenotypic and community level features.

To address these needs, we here constructed a simplified experimental system comprising a completely synthetic

community of 33 bacterial strains representing two phyla and six classes that can be cultured individually and in co-culture in highly simple laboratory conditions, and that can be individually resolved based on amplicon sequencing. Furthermore, we present draft-level whole genome sequence data as well as information regarding genomic features and phenotypic traits for all strains in the community, facilitating further mechanistic studies. We also present proof of concept for coexistence of a large proportion (14/33) of the strains in co-culture after serial passage of cultures for 48 days. Such a community can be a highly useful resource for experimental microbial ecology and evolution. For instance, we recently used a closely related model system to track the mobility of antibiotic resistance genes in a complex bacterial community (Cairns et al., 2018). For future studies, we envision, for example, exploration of the trajectories of ecosystem composition and genetic structure in response to environmental perturbations or variability in functional trait space. We are also considering using the community as an internal control for improving high-throughput microbial single cell genome sequencing techniques such as epicPCR (Spencer et al., 2016) or metagenome assemblies.

MATERIALS AND METHODS

Constructing Model Community

Strains from the University of Helsinki culture collection (HAMBI) representing diverse taxa were initially screened for the ability to grow individually at 28°C in two complex liquid media: the nutrient-rich proteose peptone yeast extract (PPY: 20 g proteose peptone and 2.5 g yeast extract in 1 l deionized H_2O) medium or a custom lower-nutrient-level medium containing M9 salt solution and King's B (KB) nutrients at a 1% concentration compared with full-strength medium (concentrations used: 0.2 g peptone number 3 and 0.1 ml of 85% glycerol in 1 l of dH_2O), and 0.2 g l^{-1} protozoan pellets (Carolina Biological Supply Co., Burlington, United States). Protozoan pellets were prepared by dissolving in dH_2O, bringing to boil and filtering through 40 μm to remove particulate matter. Notably, the strains were not selected based on being representative of any particular natural system but rather as a collection of strains, each representing a different taxonomic group, that can grow in simple, uniform laboratory conditions and therefore be easily used to test general ecological and evolutionary theory or techniques (Wolfe, 2018).

Previously, we performed a serial transfer experiment with an initial version of the model community consisting of 62 strains, including the *Escherichia coli* K-12 strain JE2571(RP4) harboring the multidrug resistance plasmid RP4 (Bradley, 1980; Cairns et al., 2018). The study demonstrates the ability of dozens of the strains to coexist in culture over 40 days, as well as a method for manipulating the level of spatial heterogeneity (biofilm mass) in community cultures. For tracking strain abundance over time in the study, we Sanger sequenced the near-full-length 16S rRNA gene sequences of the strains for use as a reference database for mapping high-throughput sequencing amplicons from experimental samples (Cairns et al., 2018). Based on these
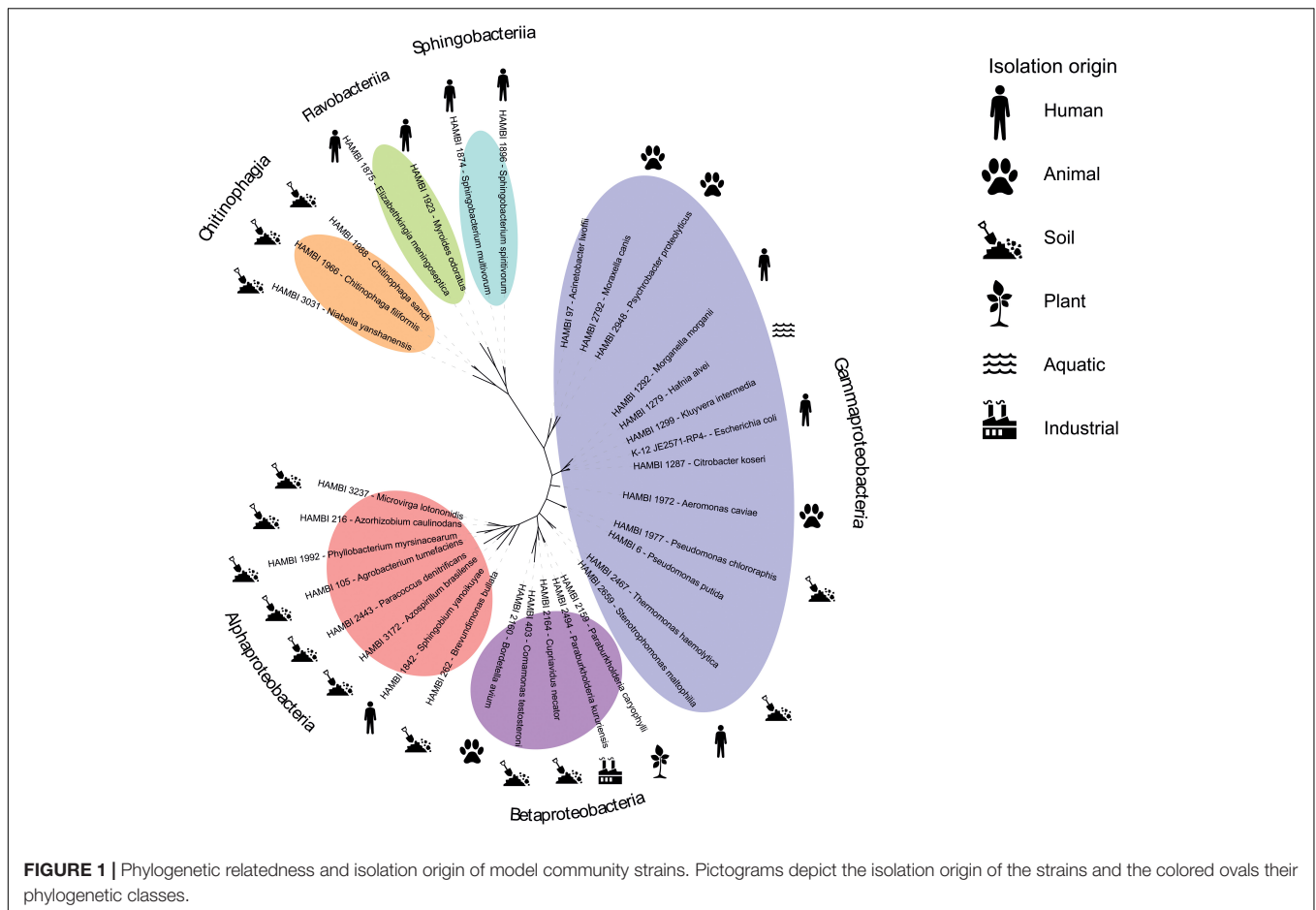
previous data, we refined the model community, removing strains with > 97% identity of 16S rRNA gene sequence with another strain in the community or uncertain strain identity. This resulted in a set of 33 gram-negative strains with confirmed culture collection identity that can be tracked at strain-level resolution through amplicon sequencing (**Supplementary Tables S1, S2**). The community was given the acronym HMC 33.1 (standing for HAMBI Mock Community containing 33 strains, version 1). The strains represent three classes in the phylum Proteobacteria (Alpha-, Beta- and Gammaproteobacteria) and three classes in the phylum Bacteroidetes (Chitinophagia, Flavobacteriia and Sphingobacteriia) (**Figure 1**). Each class contains a minimum of two strains, with the most representatives (13 strains) for the class Gammaproteobacteria. The strains have been isolated from diverse environments across the globe (**Supplementary Table S2**). Most of the strains originate from the soil, and the second highest group is human associated bacteria. Other sources include animals, aquatic environments, plants and the industry.

We further refined the co-culture medium to contain 1 g l^{-1} R2A solution and 0.5 g l^{-1} of cereal grass medium (Ward's Science, St. Catharines, ON, Canada) in M9 salt solution. Cereal grass medium stock was prepared by autoclaving in dH_2O and removing particulate matter by filtering through 5 μm.

To provide proof of concept for the coexistence of the strains in the medium, we performed a 48-day serial transfer experiment. The strains were grown separately for 4 days at 28°C with constant rotation at 50 rpm in the new medium and pooled together at equal volumes for the serial transfer experiment. The growth curves of strains when growth separately for 4 days are shown in **Supplementary Figure S1**. The stock community was stored in 30% glycerol at –80°C. The experiment was started by adding 10 μl of 100-fold diluted freeze-stored stock community to wells in a ABgene[TM] 96 Well 2.2 ml Polypropylene Deepwell Storage Plate (Thermo Fisher Scientific, Waltham, MA, United States) containing 550 μl of medium. The community was grown at 28°C with constant rotation at 100 rpm. The experiment was maintained every 96 h by transferring 10% to fresh medium. To determine bacterial density, optical density values at 600 nm were obtained for undiluted grown cultures after each growth cycle (Tecan Infinite M200 well-plate reader), and samples were frozen in glycerol for further analysis.

Phenotypic Analyses

Minimum inhibitory concentration (MIC) was determined for each strain for six antibiotics representing different antibiotic classes: ampicillin (class: penam), erythromycin (macrolide), kanamycin (aminoglycoside), nalidixic acid (fluoroquinolone), rifampicin (rifamycin), and streptomycin (aminoglycoside). For MIC determination, bacterial inoculum from an individual colony was suspended in M9 salt solution to 0.5 McFarland standard turbidity. Subsequently, 100 μl was spread-plated on 50% PPY agar medium containing a Liofilchem® MIC test strip (Liofilchem, Italy) for a specific antibiotic. The MIC was interpreted according to manufacturer's instructions after culturing at 28°C for 2 days, or for slow-growing strains, 6 days. Therefore, MICs were estimated in customized standard

FIGURE 1 | Phylogenetic relatedness and isolation origin of model community strains. Pictograms depict the isolation origin of the strains and the colored ovals their phylogenetic classes.

conditions rather than strain-specific optimal growth conditions or standard conditions recommended by the European Committee on Antimicrobial Susceptibility Testing (EUCAST) for clinical bacterial isolates (Kahlmeter et al., 2006), and the data might therefore not be comparable with clinical MIC data.

The ability of the strains to utilize 31 different carbon sources was assessed using the Biolog EcoPlate system (Biolog, Inc., Hayward, CA, United States), with the exception of HAMBI 2948 which grew poorly in experimental conditions. The experimental procedure was modified from MacLean et al. (2004). Strains were precultured in liquid PPY medium for 48 or 96 h, depending on strain growth ability, at 28°C with constant rotation. Subsequently, cells were spinned down, resuspended in M9 salt solution and nutrient-starved for 48–72 h. Starved cells were spinned down to remove any carryover nutrients in the supernatant, and resuspended in fresh M9 salt solution. Following 100-fold dilution into M9 salt solution, 150 μl of the culture was pipetted to wells in an EcoPlate (one EcoPlate per strain) containing three technical replicate wells for each carbon source and a negative control. The ability of a strain to utilize a particular carbon source was interpreted as a significantly higher optical density, based on a *t*-test, at 590 nm (measured with Tecan Infinite M200 well-plate reader) compared to the negative control after culturing for 7 days at 28°C.

DNA Extraction and Sequencing for 16S rRNA Amplicon Analysis

DNA from three technical replicates from the original pooled bacterial community and three replicate communities from days 16, 32, and 48 in the serial transfer experiment was extracted using DNeasy 96 Blood & Tissue Kit (Qiagen, Hilden, Germany). DNA extraction was performed according to the manufacturer's instructions using 400–600 μl of sample. DNA concentrations were measured using the Qubit® 3.0 fluorometer (Thermo Fisher Scientific, Waltham, MA, United States).

Paired-end sequencing was performed using the Illumina MiSeq platform at the Institute for Molecular Medicine Finland (FIMM) amplifying the V3 and V4 regions of ribosomal RNA with Phusion High Fidelity PCR Master Mix (Thermo Fisher Scientific, Waltham, MA, United States). Reactions were done with a 2-step PCR method with primers carrying Illumina adapter tails (PCR1: forward 5′-ACACTCTTTCCCT ACACGACGCTCTTCCGATCTCTCCTACGGGAGGCAGCAG-3′, reverse 5-′AGACGTGTGCTCTTCCGATCTTCTRCGMATT YCACYKCTACAC-3′; PCR2: forward 5′-ACACTCTTTCCCT ACACGACGCTCTTCCGATCTGACTACHVGGGTATCTAATC C-3′, reverse 5′-AGACGTGTGCTCTTCCGATCTCCTACGGGN GGCWGCAG-3′). For PCR2, the primers are the same but the locus-specific part and adapter part have been paired vice versa

to create more diversity in the final sequencing library. Products from PCR1 and PCR2 were pooled together and indexed with two Illumina P5/P7 index primers (every sample had their own unique combination).

PCR amplification was performed in a volume of 20 µl containing approx. 20 ng of sample DNA, 1 µl (5 µM) of each locus-specific primer (final concentration 0.25 µM), and 10 µl of 2 × Phusion High-Fidelity PCR Master Mix, and the reaction mix was brought to a final volume with laboratory grade water. The cycling conditions were as follows: 98°C for 30 s, 27 cycles of 98°C for 10 s, 62°C for 30 s, and 72°C for 15 s, with a final extension at 72°C for 10 min, followed by hold at 10°C.

Sample indexing was performed in a volume of 20 µl containing 1.5 µl (5 µM) of each index primer (final concentration 0.375 µM), 10 µl of 2 × Phusion High-Fidelity PCR Master Mix (Thermo Fisher Scientific, Waltham, MA, United States) and 1 µl of pooled PCR product. The reaction mix was brought to a final volume with laboratory grade water. The cycling conditions were as follows: 98°C for 30 s, 8 cycles of 98°C for 10 s, 65°C for 30 s, and 72°C for 20 s, with a final extension at 72°C for 5 min, followed by hold at 10°C.

After PCR, random samples were measured with LabChip GX Touch HT DNA High Sensitivity Reagent Kit (Perkin Elmer, Waltham, MA, United States) to check that the PCR was successful with the correct product size. The same run was repeated after indexing. Samples were pooled together in equal volumes and purified with Agencourt® AMPure® XP beads (Beckman Coulter, Brea, CA, United States) twice using 0.8 × volume of beads compared to the sample pool volume (40 µl). The ready amplicon library was diluted to 1:10 and 1:20 and quantified with the Agilent 2100 Bioanalyzer High Sensitivity DNA Analysis Kit (Agilent Genomics, Santa Clara, CA, United States). The 16S rRNA gene amplicon pool was sequenced with Illumina MiSeq System using the Illumina MiSeq Reagent Kit v3 600 cycles kit (Illumina, San Diego, CA, United States). The read length for the paired-end run was 2 × 300 bp.

DNA Extraction and Sequencing for Whole-Genome Analysis

For this study, 16 of the 33 model community strains were whole-genome sequenced (**Table 1**). For the remaining 17 strains, 16 assembled genomes were obtained from the NCBI Reference Sequence Database (RefSeq) and 1 raw sequence dataset was obtained from the Joint Genome Institute Genomes OnLine Database (JGI GOLD; **Supplementary Table S1**). For the strains sequenced, DNA was extracted from 1 ml of liquid culture using the DNeasy 96 Blood & Tissue Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. DNA concentrations were measured using the Qubit® 3.0 fluorometer (Thermo Fisher Scientific, Waltham, MA, United States). High-throughput sequencing was performed by the Next Generation Sequencing Services, Institute for Molecular Medicine Finland (FIMM). For this, 2.5 ng of dsDNA was prepared according to the Nextera XT DNA Library Prep Kit Reference Guide (Illumina, San Diego, GA, United States) with the following modifications: All reactions

were performed in half of the normal volume, and normalization was done according to the concentration measured on LabChip GX Touch HT (PerkinElmer, United States). 530–670 bp fragments were size selected from the pool using BluePippin (Sage Science, United States). Sequencing was performed with the Illumina HiSeq2500 system in the HiSeq high output mode using v4 kits (Illumina, San Diego, CA, United States). The read length for the paired-end run was 2 × 101 bp.

16S rRNA Amplicon Analysis

Sequencing adapters were removed from unpaired sequence data using Cutadapt 1.12 (Martin, 2011), with the parameter –minimum-length 100. Sequence pairing was done with Pear 0.9.11 (Zhang et al., 2014). Quality was assessed before and after Cutadapt and Pear with FASTQC[1], and further trimming was done with PRINSEQ (Schmieder and Edwards, 2011) using the parameters -trim_left 5 and -trim_right 40 to obtain better quality. Quality filtering was done with the USEARCH 10 (Edgar, 2013) command –fastq_filter with fastq_maxee setting 1.0 (to discard all reads with > 1.0 total expected errors). The command –usearch_global was used to align the sequences to the 16S rRNA fragment database of the strains with > 97% identity. The Shannon diversity index was computed using the USEARCH 10 command -alpha_div. Difference in community composition between days 16–48 (disregarding initial stock at time point 0 that might not represent viable cell counts for each strain) was assessed in R 3.4.0 (R Core Team, 2017) using permutational multivariate analysis of variance (PERMANOVA) as implemented in the adonis function in the vegan package (Anderson, 2001; Oksanen et al., 2016).

Genome Assembly and Annotation

Prior to genome assembly, Cutadapt 1.12 (Martin, 2011) was used to remove sequencing adapters and quality trim sequence data, with the parameters -O 10 (minimum overlap for an adapter match) and -q 28 (quality cutoff for the 3′ end of each read). Sequence data quality before and after Cutadapt was assessed using FASTQC[1] and multiQC (Ewels et al., 2016). Genome assembly was performed using SPAdes 3.11.1 (Bankevich et al., 2012) with default settings (i.e., no specified parameters since SPAdes performs estimation of e.g., k-mer sizes, coverage cutoff value and PHRED quality offset for input reads). Following assembly, contigs were filtered by minimum coverage 30 and minimum length 1000 bp. Genome assembly quality was assessed using QUAST 4.0 (Gurevich et al., 2013), and by mapping reads back to the assembly using Bowtie 2 with default settings (Langmead and Salzberg, 2012) combined with the Picard[2] command CollectWgsMetrics performed for alignment files after sorting with SAMtools (Li et al., 2009) and marking duplicates with Picard (**Table 1**). Subsequently, all assemblies, including those obtained from databases, were annotated using Prokka 1.12 (Seemann, 2014), providing genus and species names, with the parameters –centre X, –compliant, and –usegenus.

---

[1]http://www.bioinformatics.babraham.ac.uk/projects/fastqc
[2]http://broadinstitute.github.io/picard

TABLE 1 | Genome sequencing and assembly metrics.

Strain	Raw sequence yield (Gb)	Length (bp)	Coverage $\pm$ SD	% reads mapped	Largest contig (bp)	No. contigs	GC%	N50
HAMBI 6	0.76	6,425,774	88.4 ± 32.6	98.8	459,923	99	61.80	167,218
HAMBI 97	1.36	3,256,685	213 ± 63.1	97.4	201,770	162	41.41	32,660
HAMBI 105	0.79	5,387,408	110 ± 31.4	99.7	1,656,016	35	59.42	689,457
HAMBI 262	0.70	3,407,715	147 ± 49.7	98.6	203,726	88	67.02	103,424
HAMBI 1287	0.91	4,685,430	142 ± 40.2	99.7	540,344	31	53.81	453,729
HAMBI 1874	—*	5,845,984	219 ± 36.4	94.9	337,658	94	39.90	117,696
HAMBI 1972	1.02	4,463,252	160 ± 54.3	99.2	404,185	61	61.77	179,321
HAMBI 1977	0.74	6,621,816	83.0 ± 31.5	99.4	675,539	57	63.08	355,467
HAMBI 1992	0.86	5,412,359	119 ± 35.6	99.6	544,523	57	59.33	262,180
HAMBI 2159	0.86	6,527,402	96.1 ± 30.2	99.0	590,902	59	64.73	222,449
HAMBI 2160	0.89	3,679,287	172 ± 49.3	99.2	544,789	48	61.65	278,194
HAMBI 2467	0.77	2,584,354	200 ± 48.3	98.5	519,133	29	70.03	174,888
HAMBI 2494	0.90	6,648,385	102 ± 32.1	98.9	781,864	68	65.45	259,448
HAMBI 2792	1.28	2,116,559	245 ± 22.5	95.7	118,108	103	45.03	51,649
HAMBI 2948	1.23	3,031,855	218 ± 55.3	100	1,099,313	22	42.82	657,030
HAMBI 3031	1.09	5,519,621	147 ± 55.8	99.9	1,089,344	28	42.70	633,609
HAMBI 3172	0.79	7,139,720	78.4 ± 29.8	97.7	230,491	186	68.98	71,229

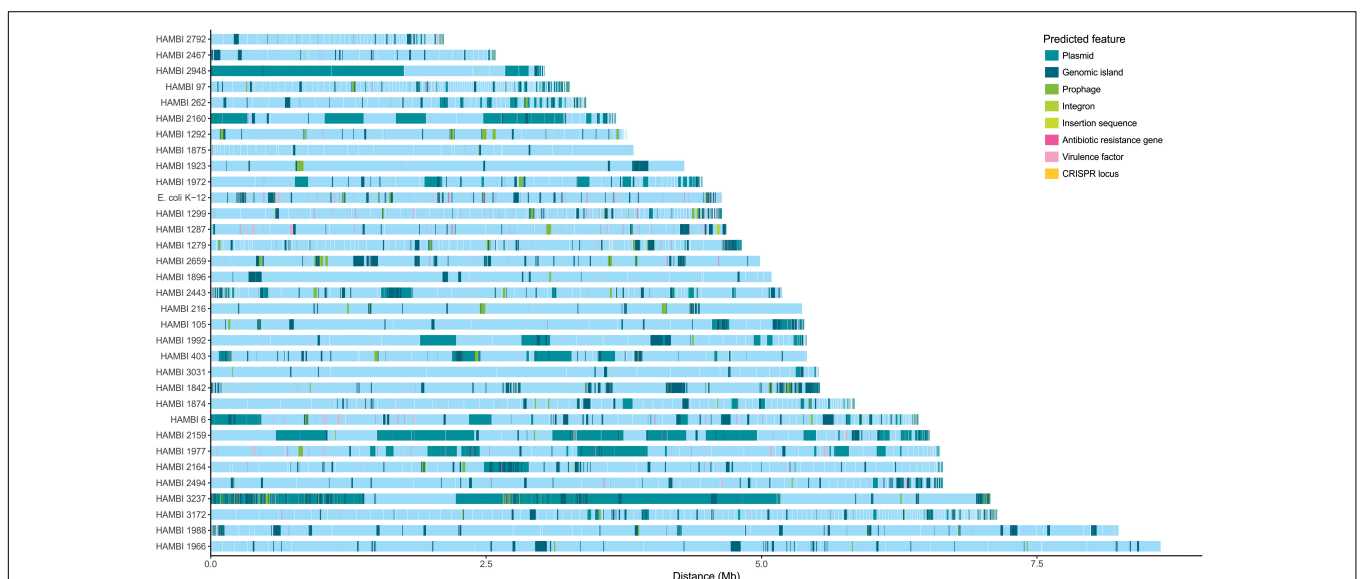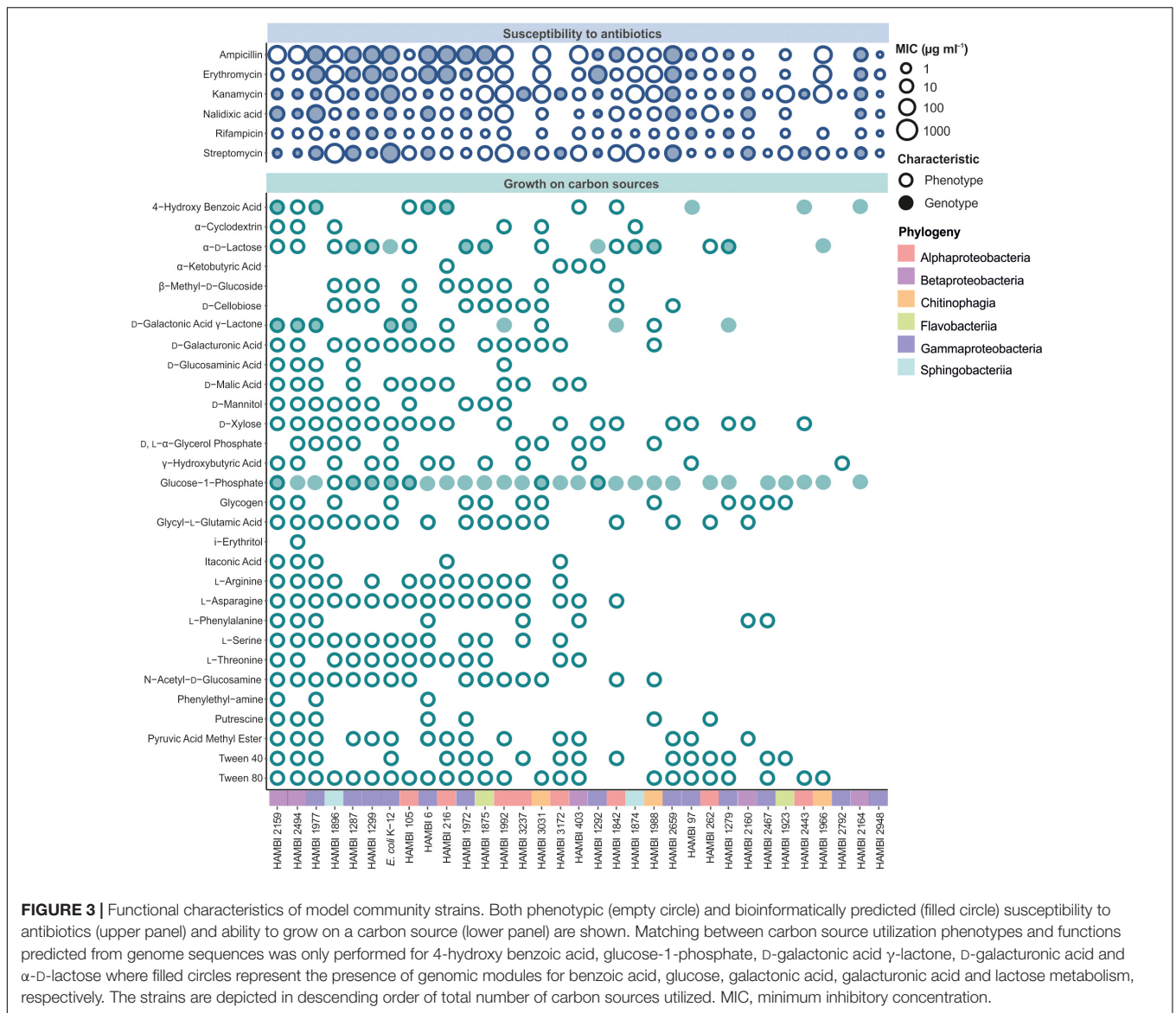*HAMBI 1874 was assembled but not sequenced in this study.*

FIGURE 2 | Genomic features of model community strains. *E. coli* K-12, HAMBI 216, HAMBI 1923 and HAMBI 2659 are closed genomes composed of a single chromosome. Other genomes are composed of blocks of adjacent contigs whose order may not reflect genomic position. Contigs are ordered by size for genomes assembled in this study and by order of appearance in assembly file for genomes obtained from databases. Predicted features are depicted by colored blocks along the genome. These include plasmid-derived sequences, genomic islands, prophages, integrons, insertion sequences, antibiotic resistance genes, virulence factors, and CRISPR loci.

Phylogenetic Analysis

We identified and aligned genes conserved among all community members using Roary 3.8.0-0 (Page et al., 2015) with minimum BLASTp identity set to 65% and protein group limit set to 200,000 to account for high diversity in the community. The core gene alignment, consisting of four ribosomal proteins (*rplN*, *rpmA*, *rpsL*, and *rpsS*) and the ATP synthase subunit beta (*atpD*), was used to create a phylogenetic tree with PhyML (Guindon et al., 2010), visualized in iTol (Letunic and Bork, 2016).

Characterization of Genomic Features

To generate comprehensive genomic metadata for model community strains, the genome assemblies were scanned for genomic elements using several tools, with default parameters unless otherwise specified. The Metabolic And Physiological potentiaL Evaluator (MAPLE) 2.3.0 (Takami et al., 2016) was used to map genes in assemblies to functional modules as defined by the Kyoto Encyclopedia of Genes and Genomes (KEGG) and for calculating module completion ratios (MCR).

FIGURE 3 | Functional characteristics of model community strains. Both phenotypic (empty circle) and bioinformatically predicted (filled circle) susceptibility to antibiotics (upper panel) and ability to grow on a carbon source (lower panel) are shown. Matching between carbon source utilization phenotypes and functions predicted from genome sequences was only performed for 4-hydroxy benzoic acid, glucose-1-phosphate, D-galactonic acid γ-lactone, D-galacturonic acid and α-D-lactose where filled circles represent the presence of genomic modules for benzoic acid, glucose, galactonic acid, galacturonic acid and lactose metabolism, respectively. The strains are depicted in descending order of total number of carbon sources utilized. MIC, minimum inhibitory concentration.
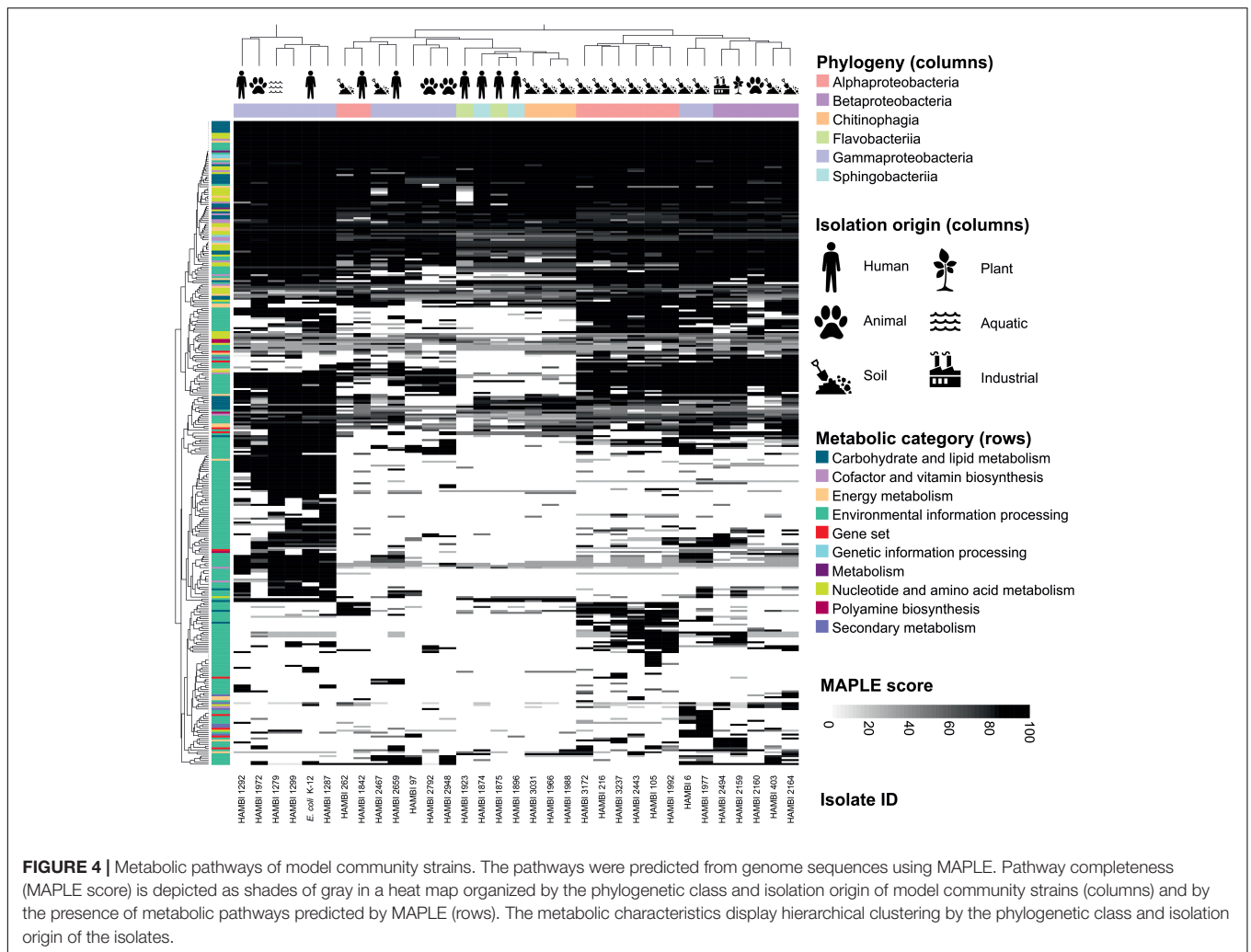
Genomic islands, i.e., large genomic regions assumed to have horizontal origins, were predicted using IslandViewer 4 (Bertelli et al., 2017). For strains other than *E. coli* K-12, HAMBI 216, HAMBI 1923 and HAMBI 2659 which have closed genomes, genomic island predictions were obtained for incomplete genomes with contigs ordered against closely related reference genomes (**Supplementary Table S3**) using the Mauve contig orderer (Rissman et al., 2009; Bertelli et al., 2017).

Plasmid-derived sequences were predicted using cBar 1.2 (Zhou and Xu, 2010) and PlasmidFinder 1.3 (Carattoli et al., 2014), which was implemented through ABRicate[3] using databases dated 2018-01-02. In addition, for genomes assembled for this study, plasmidSPAdes (Antipov et al., 2016) implemented in SPAdes 3.11.1 (Bankevich et al., 2012) was used, with

assignment of plasmid annotation to > 1000 bp regions in genome assembly contigs displaying 100% BLASTn (Altschul et al., 1990) identity with plasmid assemblies generated by plasmidSPAdes.

Antibiotic resistance and virulence genes were predicted with ABRicate (https://github.com/tseemann/abricate, databases dated 2018-01-02) using ResFinder 3.0 (Zankari et al., 2012), the comprehensive antibiotic resistance database (CARD) (Jia et al., 2017), and the virulence factors database (VFDB) (Chen et al., 2016), discarding hits with proportion of gene covered < 50%. Prophages, integrons, insertion sequences, and clustered regularly interspaced short palindromic repeats (CRISPR) were predicted using the PHAge Search Tool – Enhanced Release (PHASTER; accessed 2018-01-08) (Arndt et al., 2016), Integron Finder 1.5.1 (Cury et al., 2016), ISEScan 1.5.4.3 (Xie and Tang, 2017) and CRISPRFinder (accessed 2018-02-01) (Grissa et al., 2007), respectively.

[3]https://github.com/tseemann/abricate

FIGURE 4 | Metabolic pathways of model community strains. The pathways were predicted from genome sequences using MAPLE. Pathway completeness (MAPLE score) is depicted as shades of gray in a heat map organized by the phylogenetic class and isolation origin of model community strains (columns) and by the presence of metabolic pathways predicted by MAPLE (rows). The metabolic characteristics display hierarchical clustering by the phylogenetic class and isolation origin of the isolates.

Notably, in the absence of further validation steps, plasmid predictions obtained with cBar or plasmidSPAdes have low accuracy (Arredondo-Alonso et al., 2017), and genomic island predictions for draft level genomes may contain errors (Bertelli et al., 2017), and the predicted loci must therefore be considered as regions of interest rather than high-confidence predictions.
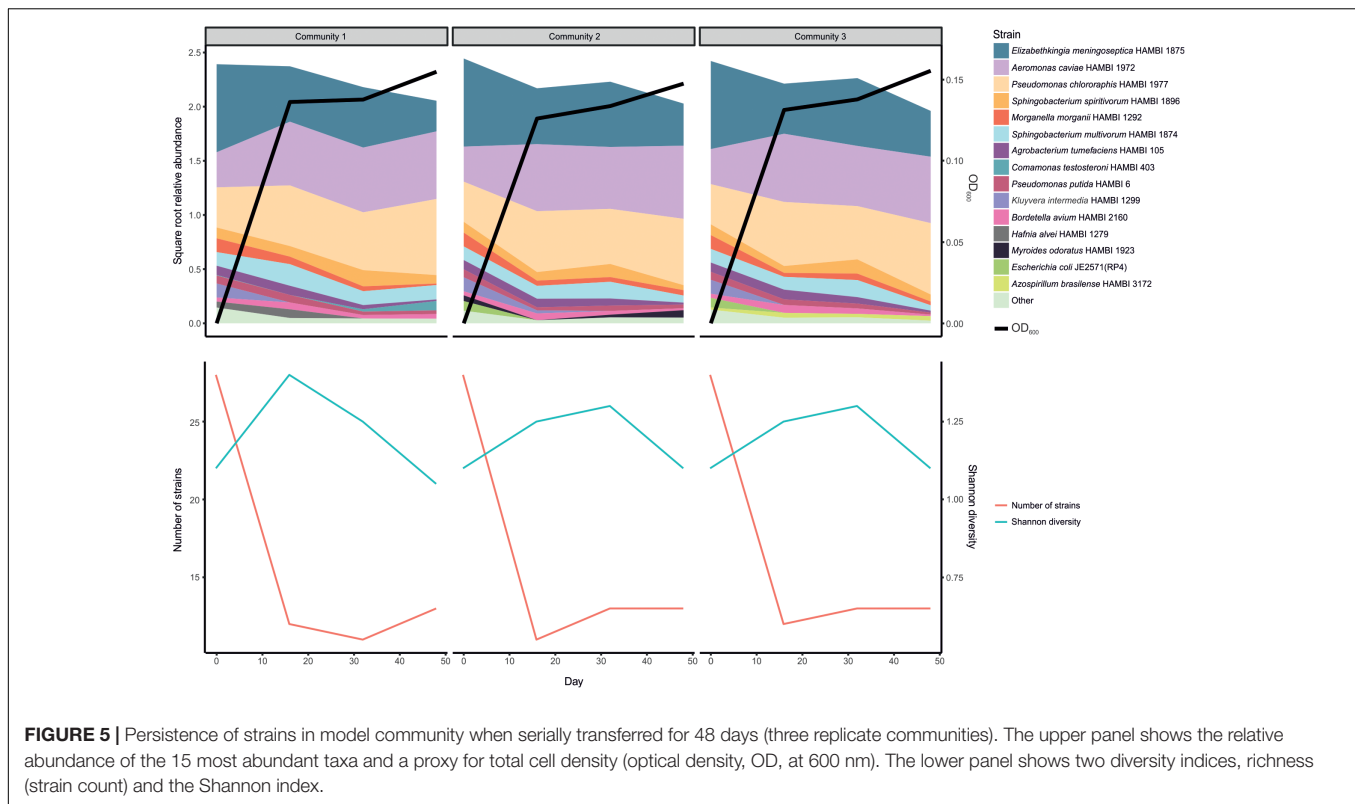
RESULTS

Genomic Composition

The strains display high variability in genome size and genomic content, including the resistome, mobilome and functionome (**Figure 2**). Antibiotic resistance and virulence genes are particularly prevalent among strains in the Gammaproteobacteria class (**Figures 1**, **2**). Genomic elements associated with horizontal gene transfer, including plasmids, genomic islands, prophages, integrons, and insertion sequences, occur frequently among the strains, as do CRISPR arrays that might be associated with CRISPR/Cas

systems conferring adaptive immunity against mobile genetic elements.

Metabolic and Physiological Characteristics

The strains display diverse antibiotic susceptibility and carbon source utilization phenotypes (**Figure 3**). Antibiotic resistance phenotypes are variably associated with the presence of bioinformatically predicted resistance factors to the antibiotic class. The number of carbon sources on which the strains can grow displays high variability, with a maximum of 25 out of 31 carbon sources (HAMBI 2159 and 2494, which both belong to the genus *Paraburkholderia*), indicating large differences between strains in metabolic capacity and fastidiousness. Furthermore, the presence of genomic functional modules for benzoic acid, galactonic acid, galacturonic acid or lactose metabolism is frequently associated with the ability of a strain to grow with a corresponding compound as the sole carbon source (**Figure 3**).

The metabolic characteristics predicted from the genome sequences by MAPLE display hierarchical clustering by

FIGURE 5 | Persistence of strains in model community when serially transferred for 48 days (three replicate communities). The upper panel shows the relative abundance of the 15 most abundant taxa and a proxy for total cell density (optical density, OD, at 600 nm). The lower panel shows two diversity indices, richness (strain count) and the Shannon index.

phylogenetic classes of the isolates as well as by their isolation origin (**Figure 4**). Soil isolates typically cluster separately from isolates from humans and animals. Most of this clustering is driven by metabolic functions related to environmental information processing. Core metabolic pathways such as the tricarboxylic acid cycle are present across all isolates while specialized metabolic features, such as utilization of less common carbohydrates, exhibit differences in their distribution across the isolates.

Coexistence of Strains in Co-culture

Serially transferring the 33 strain community for 48 days seemingly led to a loss of over half of the strains within the first 16 days, although the initial number of viable strains is uncertain owing to different growth dynamics of individual strains prior to initiating the experiment (**Supplementary Figure S1** and **Figure 5**). After these transient dynamics, there was only a marginally significant change in community composition over time between days 16 and 48 (PERMANOVA: $F_{1,7} = 3.018$, $r^2 = 0.301$, $p = 0.074$). Furthermore, the apparent initial loss in richness (strain count) was accompanied by an increase in evenness, seen as a relatively small temporal change in Shannon diversity, which accounts for both factors. There were three dominant strains: *Elizabethkingia meningoseptica* HAMBI 1875, *Aeromonas caviae* HAMBI 1972 and *Pseudomonas chlororaphis* HAMBI 1977. However, the degree to which the dominance of these strains is caused by their relative abundance or viability in the input community (strain mix) or competitive ability during co-culture is uncertain.

DISCUSSION AND CONCLUSION

We developed a multispecies synthetic bacterial community, which was characterized at the genomic and phenotypic levels, revealing high diversity in the resistome, mobilome and functionome of the community. Furthermore, we demonstrated the utility of the community by showing coexistence of a large proportion (14/33) of the strains in co-culture over serial transfer for 48 days. These observations indicate that the community is suitable for use as a cultivated model community to ask a wide range of biological questions.

Along with characterizing the community, we produced comprehensive genomic and phenotypic metadata for the community members. This facilitates future mechanistic work with the community. For instance, our pipeline allows the use of amplicon sequencing to track community composition over time at strain-level resolution. Moreover, the genome assemblies provide a reference database for (meta)genomic and (meta)transcriptomic studies. Isolation of individual colonies combined with colony PCR, in turn, allows the rapid identification of substrains possessing mutations or horizontal gene transfer events of interest.

Our model community is composed of diverse bacterial strains isolated from different environments and hence does not mimic any specific natural community. Therefore, the community is more suited to study general questions about ecology and evolution such as community assembly and response to environmental perturbations rather than to explain patterns in any particular natural microbial community (Wolfe, 2018).

The ability of the community to answer general questions may be limited by the potential absence of focal taxonomic groups or high-order interactions that occur in more complex communities. The community is composed of bacteria alone, while interactions between bacteria and fungi, protozoa or bacteriophages often play a key role in natural microbial communities. In a separate study, however, we have introduced a method for incorporating protozoa in the community (Cairns et al., 2018).

The utility of the 33-strain model community and its predecessor is demonstrated in the current and previous work by us (Cairns et al., 2018). Here we present the community and a collection of methods and metadata to the scientific community. A previous version of the community has already been successfully used for tracking the mobility of antibiotic resistance genes (Cairns et al., 2018), and as prominent cases of future use we envision, for instance, replicated ecosystem microcosms to explore the trajectories of ecosystem composition and genetic structure in response to various environmental perturbations. We also envision the community as an efficient spike-in control to increase the statistical rigor of high-throughput microbial single cell assays such as epicPCR (Spencer et al., 2016) or as a test system for validating metagenome assemblers.

DATA AVAILABILITY

Raw fastq files and genome assemblies for the genomes sequenced or assembled in this study ($N = 17$) have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) and GenBank, respectively, under the BioProject Accession Number PRJNA476209. Accession numbers for raw fastq files or genome assemblies obtained for this study from databases are listed in **Supplementary Table S1**. Previously Sanger sequenced near-full-length 16S rRNA gene sequences for all strains are available in the European Nucleotide Archive (ENA) under the Accession Number PRJEB21728 (https://www.ebi.ac.uk/ena/data/view/PRJEB21728). The complete sequence for the plasmid RP4 harbored by *E. coli* K-12 JE2571(RP4) is available from GenBank under the accession BN000925 (Pansegrau et al., 1994). Genome annotations, including those obtained in this study for previously sequenced or assembled genomes, are available upon request.

The following community metadata is available via Dryad (doi: 10.5061/dryad.53b6n5f): Genomic predictions for functional modules obtained using MAPLE 2.3.0 (Takami et al.,

2016); genomic feature data collated into a single data frame, using custom R scripts (R Core Team, 2017), displaying strain, contig, locus, and genomic feature information; antibiotic MICs; and carbon source utilization phenotypes. Raw output files for each strain from all genomic feature prediction tools are available upon request.

The model community strains, except for *E. coli* K-12 strain JE2571(RP4), are available individually from the HAMBI culture collection at the University of Helsinki, Finland (http://www.helsinki.fi/hambi/). For further requests concerning the use of the model community, please contact the corresponding author.

AUTHOR CONTRIBUTIONS

JC and TH designed the experiments. RJ contributed to phenotypic measurements, proof of concept experiments, and performed sequence analyses for 16S rRNA amplicon data. TH, JH, and MV provided sequencing resources. JC and JH designed the sequence analyses. JC performed sequence analyses for whole-genome data and drafted the manuscript. JC, MT, and RJ contributed to data visualization. RJ, JH, MT, MV, and TH edited and commented on the manuscript. All authors gave final approval for publication and accept accountability for the content and work performed.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2018.00312/full#supplementary-material

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 26, 32–46. doi: 10.1111/j.1442-9993.2001.01070.pp.x

Antipov, D., Hartwick, N., Shen, M., Raiko, M., Lapidus, A., and Pevzner, P. A. (2016). plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics* 32, 3380–3387. doi: 10.1093/bioinformatics/btw493

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y. J., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387

Arredondo-Alonso, S., Willems, R. J., van Schaik, W., and Schürch, A. C. (2017). On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. *Microb. Genom.* 3:e000128. doi: 10.1099/mgen.0.000128

Bairey, E., Kelsic, E. D., and Kishony, R. (2016). High-order species interactions shape ecosystem diversity. *Nat. Commun.* 7:12285. doi: 10.1038/ncomms12285

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Bell, T., Newman, J. A., Silverman, B. W., Turner, S. L., and Lilley, A. K. (2005). The contribution of species richness and composition to bacterial services. *Nature* 436, 1157–1160. doi: 10.1038/nature03891

Benton, T. G., Solan, M., Travis, J. M. J., and Sait, S. M. (2007). Microcosm experiments can inform global ecological problems. *Trends Ecol. Evol.* 22, 516–521. doi: 10.1016/j.tree.2007.08.003

Bertelli, C., Laird, M. R., Williams, K. P., Lau, B. Y., Hoad, G., Winsor, G. L., et al. (2017). IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* 45, W30–W35. doi: 10.1093/nar/gkx343

Betts, A., Gray, C., Zelek, M., MacLean, R. C., and King, K. C. (2018). High parasite diversity accelerates host adaptation and diversification. *Science* 360, 907–911. doi: 10.1126/science.aam9974

Bradley, D. E. (1980). Morphological and serological relationships of conjugative pili. *Plasmid* 4, 155–169. doi: 10.1016/0147-619X(80)90005-0

Brockhurst, M. A., and Koskella, B. (2013). Experimental coevolution of species interactions. *Trends Ecol. Evol.* 28, 367–375. doi: 10.1016/j.tree.2013.02.009

Buckling, A., MacLean, R. C., Brockhurst, M. A., and Colegrave, N. (2009). The Beagle in a bottle. *Nature* 457, 824–829. doi: 10.1038/nature07892

Cairns, J., Ruokolainen, L., Hultman, J., Tamminen, M., Virta, M., and Hiltunen, T. (2018). Ecology determines how low antibiotic concentration impacts community composition and horizontal transfer of resistance genes. *Commun. Biol.* 1:35. doi: 10.1038/s42003-018-0041-7

Carattoli, A., Zankari, E., Garcia-Fernandez, A., Larsen, M. V., Lund, O., Villa, L., et al. (2014). *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* 58, 3895–3903. doi: 10.1128/AAC.02412-14

Carpenter, S. R. (1996). Microcosm experiments have limited relevance for community and ecosystem ecology. *Ecology* 77, 677–680. doi: 10.2307/2265490

Carpenter, S. R. (1999). Microcosm experiments have limited relevance for community and ecosystem ecology: reply. *Ecology* 80, 1085–1088. doi: 10.1890/0012-9658(1999)080[1085:MEHLRF]2.0.CO;2

Celiker, H., and Gore, J. (2014). Clustering in community structure across replicate ecosystems following a long-term bacterial evolution experiment. *Nat. Commun.* 5:4643. doi: 10.1038/ncomms5643

Chase, A. B., Karaoz, U., Brodie, E. L., Gomez-Lunar, Z., Martiny, A. C., and Martiny, J. B. H. (2017). Microdiversity of an abundant terrestrial bacterium encompasses extensive variation in ecologically relevant traits. *mBio* 8:e1809-17. doi: 10.1128/mBio.01809-17

Chen, L. H., Zheng, D. D., Liu, B., Yang, J., and Jin, Q. (2016). VFDB 2016: hierarchical and refined dataset for big data analysis-10 years on. *Nucleic Acids Res.* 44, D694–D697. doi: 10.1093/nar/gkv1239

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Cury, J., Jove, T., Touchon, M., Neron, B., and Rocha, E. P. C. (2016). Identification and analysis of integrons and cassette arrays in bacterial genomes. *Nucleic Acids Res.* 44, 4539–4550. doi: 10.1093/nar/gkw319

Dejonghe, W., Berteloot, E., Goris, J., Boon, N., Crul, K., Maertens, S., et al. (2003). Synergistic degradation of linuron by a bacterial consortium and isolation of a single linuron-degrading *Variovorax* strain. *Appl. Environ. Microbiol.* 69, 1532–1541. doi: 10.1128/AEM.69.3.1532-1541.2003

Dolinsek, J., Goldschmidt, F., and Johnson, D. R. (2016). Synthetic microbial ecology and the dynamic interplay between microbial genotypes. *FEMS Microbiol. Rev.* 40, 961–979. doi: 10.1093/femsre/fuw024

Drenner, R. W., and Mazumder, A. (1999). Microcosm experiments have limited relevance for community and ecosystem ecology: comment. *Ecology* 80, 1081–1085. doi: 10.1890/0012-9658(1999)080[1081:MEHLRF]2.0.CO;2

Dunham, M. J. (2007). Synthetic ecology: a model system for cooperation. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1741–1742. doi: 10.1073/pnas.0611067104

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 10, 996–998. doi: 10.1038/nmeth.2604

Elias, S., and Banin, E. (2012). Multi-species biofilms: living with friendly neighbors. *FEMS Microbiol. Rev.* 36, 990–1004. doi: 10.1111/j.1574-6976.2012.00325.x

Ewels, P., Magnusson, M., Lundin, S., and Kaller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048. doi: 10.1093/bioinformatics/btw354

Foster, K. R., and Bell, T. (2012). Competition, not cooperation, dominates interactions among culturable microbial species. *Curr. Biol.* 22, 1845–1850. doi: 10.1016/j.cub.2012.08.005

Fraser, L. H., and Keddy, P. (1997). The role of experimental microcosms in ecological research. *Trends Ecol. Evol.* 12, 478–481. doi: 10.1016/S0169-5347(97)01220-2

Frickel, J., Feulner, P. G. D., Karakoc, E., and Becks, L. (2018). Population size changes and selection drive patterns of parallel evolution in a host-virus system. *Nat. Commun.* 9:1706. doi: 10.1038/s41467-018-03990-7

Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E., and Desai, M. M. (2017). The dynamics of molecular evolution over 60,000 generations. *Nature* 551, 45–50. doi: 10.1038/nature24287

Goodman, A. L., Kallstrom, G., Faith, J. J., Reyes, A., Moore, A., Dantas, G., et al. (2011). Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6252–6257. doi: 10.1073/pnas.1102938108

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35, W52–W57. doi: 10.1093/nar/gkm360

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Hall, A. R., Scanlan, P. D., Morgan, A. D., and Buckling, A. (2011). Host-parasite coevolutionary arms races give way to fluctuating selection. *Ecol. Lett.* 14, 635–642. doi: 10.1111/j.1461-0248.2011.01624.x

Harcombe, W. (2010). Novel cooperation experimentally evolved between species. *Evolution* 64, 2166–2172. doi: 10.1111/j.1558-5646.2010.00959.x

Helling, R. B., Vargas, C. N., and Adams, J. (1987). Evolution of *Escherichia coli* during growth in a constant environment. *Genetics* 116, 349–358.

Hiltunen, T., and Becks, L. (2014). Consumer co-evolution as an important component of the eco-evolutionary feedback. *Nat. Commun.* 5:5226. doi: 10.1038/ncomms6226

Jia, B. F., Raphenya, A. R., Alcock, B., Waglechner, N., Guo, P. Y., Tsang, K. K., et al. (2017). CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 45, D566–D573. doi: 10.1093/nar/gkw1004

Kahlmeter, G., Brown, D. F. J., Goldstein, F. W., MacGowan, A. P., Mouton, J. W., Odenholt, I., et al. (2006). European Committee on Antimicrobial Susceptibility Testing (EUCAST) Technical Notes on antimicrobial susceptibility testing. *Clin. Microbiol. Infect.* 12, 501–503. doi: 10.1111/j.1469-0691.2006.01454.x

Kastman, E. K., Kamelamela, N., Norville, J. W., Cosetta, C. M., Dutton, R. J., and Wolfe, B. E. (2016). Biotic interactions shape the ecological distributions of *Staphylococcus* species. *mBio* 7:e1157-16. doi: 10.1128/mBio.01157-16

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lawrence, D., Fiegna, F., Behrends, V., Bundy, J. G., Phillimore, A. B., Bell, T., et al. (2012). Species interactions alter evolutionary responses to a novel environment. *PLoS Biol.* 10:e1001330. doi: 10.1371/journal.pbio.1001330

Lebeis, S. L., Paredes, S. H., Lundberg, D. S., Breakfield, N., Gehring, J., McDonald, M., et al. (2015). Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science* 349, 860–864. doi: 10.1126/science.aaa8764

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245. doi: 10.1093/nar/gkw290

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Liu, W. Z., Russel, J., Roder, H. L., Madsen, J. S., Burmolle, M., and Sorensen, S. J. (2017). Low-abundant species facilitates specific spatial organization that promotes multispecies biofilm formation. *Environ. Microbiol.* 19, 2893–2905. doi: 10.1111/1462-2920.13816

MacLean, R. C., Bell, G., and Rainey, P. B. (2004). The evolution of a pleiotropic fitness tradeoff in *Pseudomonas fluorescens*. *Proc. Natl. Acad. Sci. U.S.A.* 101, 8072–8077. doi: 10.1073/pnas.0307195101

MacLean, R. C., and Gudelj, I. (2006). Resource competition and social conflict in experimental populations of yeast. *Nature* 441, 498–501. doi: 10.1038/nature04624

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12. doi: 10.14806/ej.17.1.200

Morgan, A. D., Gandon, S., and Buckling, A. (2005). The effect of migration on local adaptation in a coevolving host-parasite system. *Nature* 437, 253–256. doi: 10.1038/nature03913

Oksanen, J., Blanchet, G. F., Kindt, R., et al. (2016). *vegan: Community Ecology package. R Package version 2.3-5*. Available at: http://CRAN.R-project.org/package=vegan

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421

Pansegrau, W., Lanka, E., Barth, P. T., Figurski, D. H., Guiney, D. G., Haas, D., et al. (1994). Complete nucleotide sequence of Birmingham IncPα plasmids: compilation and comparative analysis. *J. Mol. Biol.* 239, 623–663. doi: 10.1006/jmbi.1994.1404

Petrof, E. O., Gloor, G. B., Vanner, S. J., Weese, S. J., Carter, D., Daigneault, M. C., et al. (2013). Stool substitute transplant therapy for the eradication of *Clostridium difficile* infection: 'RePOOPulating' the gut. *Microbiome* 1:3. doi: 10.1186/2049-2618-1-3

Rissman, A. I., Mau, B., Biehl, B. S., Darling, A. E., Glasner, J. D., and Perna, N. T. (2009). Reordering contigs of draft genomes using the Mauve Aligner. *Bioinformatics* 25, 2071–2073. doi: 10.1093/bioinformatics/btp356

Rivett, D. W., and Bell, T. (2018). Abundance determines the functional role of bacterial phylotypes in complex communities. *Nat. Microbiol.* 3, 767–772. doi: 10.1038/s41564-018-0180-0

Rivett, D. W., Jones, M. L., Ramoneda, J., Mombrikotb, S. B., Ransome, E., and Bell, T. (2018). Elevated success of multispecies bacterial invasions impacts community composition during ecological succession. *Ecol. Lett.* 21, 516–524. doi: 10.1111/ele.12916

Rivett, D. W., Scheuerl, T., Culbert, C. T., Mombrikotb, S. B., Johnstone, E., Barraclough, T. G., et al. (2016). Resource-dependent attenuation of species interactions during bacterial succession. *ISME J.* 10, 2259–2268. doi: 10.1038/ismej.2016.11

Rozen, D. E., and Lenski, R. E. (2000). Long-term experimental evolution in *Escherichia coli*. VIII. Dynamics of a balanced polymorphism. *Am. Nat.* 155, 24–35. doi: 10.1086/303299

Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–864. doi: 10.1093/bioinformatics/btr026

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Shade, A., and Handelsman, J. (2012). Beyond the Venn diagram: the hunt for a core microbiome. *Environ. Microbiol.* 14, 4–12. doi: 10.1111/j.1462-2920.2011.02585.x

Shertzer, K. W., Ellner, S. P., Fussmann, G. F., and Hairston, N. G. (2002). Predator-prey cycles in an aquatic microcosm: testing hypotheses of mechanism. *J. Anim. Ecol.* 71, 802–815. doi: 10.1046/j.1365-2656.2002.00645.x

Sheth, R. U., Cabral, V., Chen, S. P., and Wang, H. H. (2016). Manipulating bacterial communities by in situ microbiome engineering. *Trends Genet.* 32, 189–200. doi: 10.1016/j.tig.2016.01.005

Shou, W. Y., Ram, S., and Vilar, J. M. G. (2007). Synthetic cooperation in engineered yeast populations. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1877–1882. doi: 10.1073/pnas.0610575104

Smillie, C. S., Smith, M. B., Friedman, J., Cordero, O. X., David, L. A., and Alm, E. J. (2011). Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* 480, 241–244. doi: 10.1038/nature10571

Spencer, S. J., Tamminen, M. V., Preheim, S. P., Guo, M. T., Briggs, A. W., Brito, I. L., et al. (2016). Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *ISME J.* 10, 427–436. doi: 10.1038/ismej.2015.124

Stopnisek, N., Zuhlke, D., Carlier, A., Barberan, A., Fierer, N., Becher, D., et al. (2016). Molecular mechanisms underlying the close association between soil *Burkholderia* and fungi. *ISME J.* 10, 253–264. doi: 10.1038/ismej.2015.73

Takami, H., Taniguchi, T., Arai, W., Takemoto, K., Moriya, Y., and Goto, S. (2016). An automated system for evaluation of the potential functionome: MAPLE version 2.1.0. *DNA Res.* 23, 467–475. doi: 10.1093/dnares/dsw030

Treves, D. S., Manning, S., and Adams, J. (1998). Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*. *Mol. Biol. Evol.* 15, 789–797. doi: 10.1093/oxfordjournals.molbev.a025984

van Gestel, J., Weissing, F. J., Kuipers, O. P., and Kovacs, A. T. (2014). Density of founder cells affects spatial pattern formation and cooperation in *Bacillus subtilis* biofilms. *ISME J.* 8, 2069–2079. doi: 10.1038/ismej.2014.52

van Houte, S., Ekroth, A. K. E., Broniewski, J. M., Chabas, H., Ashby, B., Bondy-Denomy, J., et al. (2016). The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* 532, 385–388. doi: 10.1038/nature17436

Wang, H., Luo, H., Fallgren, P. H., Jin, S., and Ren, Z. J. (2015). Bioelectrochemical system platform for sustainable environmental remediation and energy generation. *Biotechnol. Adv.* 33, 317–334. doi: 10.1016/j.biotechadv.2015.04.003

Widder, S., Allen, R. J., Pfeiffer, T., Curtis, T. P., Wiuf, C., Sloan, W. T., et al. (2016). Challenges in microbial ecology: building predictive understanding of community function and dynamics. *ISME J.* 10, 2557–2568. doi: 10.1038/ismej.2016.45

Wolfe, B. E. (2018). Using cultivated microbial communities to dissect microbiome assembly: challenges, limitations, and the path ahead. *mSystems* 3:e161-17. doi: 10.1128/mSystems.00161-17

Wolfe, B. E., Button, J. E., Santarelli, M., and Dutton, R. J. (2014). Cheese rind communities provide tractable systems for in situ and in vitro studies of microbial diversity. *Cell* 158, 422–433. doi: 10.1016/j.cell.2014.05.041

Xie, Z. Q., and Tang, H. X. (2017). ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* 33, 3340–3347. doi: 10.1093/bioinformatics/btx433

Yoshida, T., Jones, L. E., Ellner, S. P., Fussmann, G. F., and Hairston, N. G. (2003). Rapid evolution drives ecological dynamics in a predator-prey system. *Nature* 424, 303–306. doi: 10.1038/nature01767

Yu, Z., Beck, D. A. C., and Chistoserdova, L. (2017). Natural selection in synthetic communities highlights the roles of *Methylococcaceae* and *Methylophilaceae* and suggests differential roles for alternative methanol dehydrogenases in methane consumption. *Front. Microbiol.* 8:2392. doi: 10.3389/fmicb.2017.02392

Yu, Z., Krause, S. M. B., Beck, D. A. C., and Chistoserdova, L. (2016). A synthetic ecology perspective: how well does behavior of model organisms in the laboratory predict microbial activities in natural habitats? *Front. Microbiol.* 7:946. doi: 10.3389/fmicb.2016.00946

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., et al. (2012). Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67, 2640–2644. doi: 10.1093/jac/dks261

Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30, 614–620. doi: 10.1093/bioinformatics/btt593

Zhou, F. F., and Xu, Y. (2010). cBar: a computer program to distinguish plasmid-derived from chromosome-derived sequence fragments in metagenomics data. *Bioinformatics* 26, 2051–2052. doi: 10.1093/bioinformatics/btq299

Zomorrodi, A. R., and Segre, D. (2016). Synthetic ecology of microbes: mathematical models and applications. *J. Mol. Biol.* 428, 837–861. doi: 10.1016/j.jmb.2015.10.019