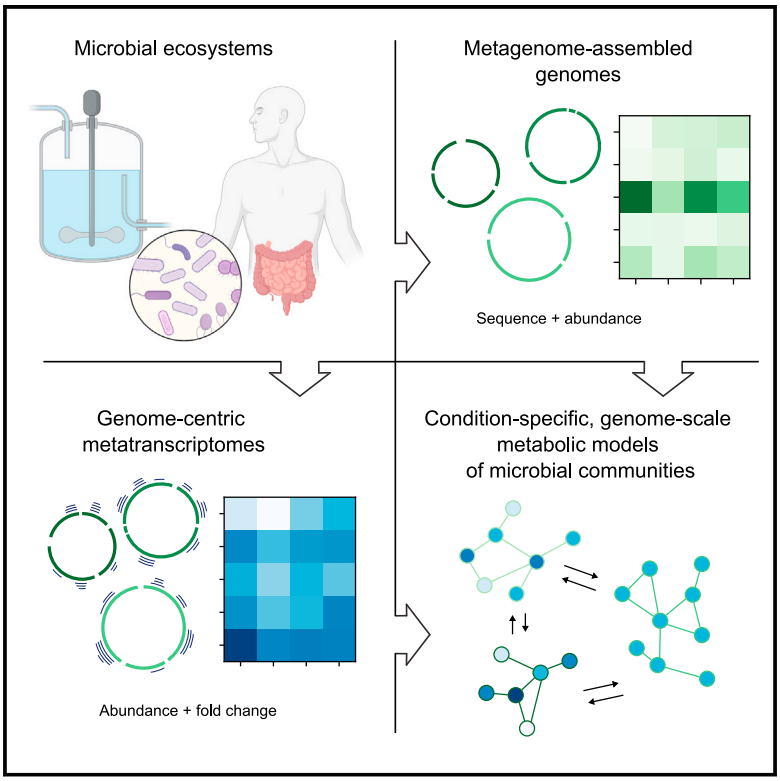
Cell Reports
Methods

Article

Metatranscriptomics-guided genome-scale metabolic modeling of microbial communities

Graphical abstract



Authors

Guido Zampieri, Stefano Campanaro, Claudio Angione, Laura Treu

Correspondence

stefano.campanaro@unipd.it

In brief

Zampieri et al. present a genome-scale modeling approach for integrating transcriptomics profiles collected from the environment, also called metatranscriptomics data, into genome-scale models of microbial community metabolism, showing that such an approach can better capture metabolic activity in anaerobic digestion consortia and in the human gut microbiota.

Highlights

- Genome-centric metatranscriptomics are integrated in genome-scale metabolic models
- Metabolic activity of complex microbial communities is better captured upon integration
- Validations are performed on anaerobic digestion and human gut microbiomes
- The proposed approach is culture independent

Zampieri et al., 2023, Cell Reports Methods 3, 100383
January 23, 2023
https://doi.org/10.1016/j.crmeth.2022.100383

CellPress

Cell Reports Methods

CellPress
OPEN ACCESS

Article

Metatranscriptomics-guided genome-scale metabolic modeling of microbial communities

Guido Zampieri,[1] Stefano Campanaro,[1,2,6,*] Claudio Angione,[3,4,5] and Laura Treu[1]
[1]Department of Biology, University of Padova, Padova 35121, Italy
[2]CRIBI Biotechnology Center, University of Padova, Padova 35121, Italy
[3]School of Computing, Engineering and Digital Technologies, Teesside University, Middlesbrough TS1 3BX, UK
[4]National Horizons Centre, Teesside University, Darlington DL1 1HG, UK
[5]Centre for Digital Innovation, Teesside University, Middlesbrough TS1 3BX, UK
[6]Lead contact
*Correspondence: stefano.campanaro@unipd.it
https://doi.org/10.1016/j.crmeth.2022.100383

MOTIVATION Genome-scale models of metabolism are effective tools for exploring the behavior of microbial ecosystems and are increasingly used for this purpose. However, current modeling approaches for microbial communities are limited to considering the functional potential encoded in microbial genomes, which can often be scarcely indicative of real functional activity. To better guide modeling efforts, we designed an approach for the incorporation of metatranscriptomics data in microbial community models and implemented it in two alternative pipelines that can be applied in culture-independent settings.

SUMMARY

Multi-omics data integration via mechanistic models of metabolism is a scalable and flexible framework for exploring biological hypotheses in microbial systems. However, although most microorganisms are unculturable, such multi-omics modeling is limited to isolate microbes or simple synthetic communities. Here, we developed an approach for modeling microbial activity and interactions that leverages the reconstruction of metagenome-assembled genomes and associated genome-centric metatranscriptomes. At its core, we designed a method for condition-specific metabolic modeling of microbial communities through the integration of metatranscriptomic data. Using this approach, we explored the behavior of anaerobic digestion consortia driven by hydrogen availability and human gut microbiota dysbiosis associated with Crohn's disease, identifying condition-dependent amino acid requirements in archaeal species and a reduced short-chain fatty acid exchange network associated with disease, respectively. Our approach can be applied to complex microbial communities, allowing a mechanistic contextualization of multi-omics data on a metagenome scale.

INTRODUCTION

The vast majority of living microorganisms cannot be easily isolated and cultured in the laboratory. Despite recent notable efforts,[1] culture-independent approaches are thus essential to understand the organization and evolution of microbial ecosystems.[2,3] These approaches make possible the investigation of microbial communities within their natural environment, often starting from *omics* information characterizing DNA, RNA, proteins, and metabolites. In particular, *de novo* assembly and binning of bulk DNA sequences allows the definition of distinctive functional roles within a community, thus dissecting the biological potential at the single-species level. This process involves the reconstruction of metagenome-assembled genomes (MAGs) for individual microbes, whose diversity is starting to emerge at an unprecedented level.[4,5] However, while new data-generating technologies in this area have progressively been adopted and are now widely used, *in silico* techniques for contextualizing and interpreting metagenomics data still lack adequate depth for pinpointing key biological and ecological processes.

Among the emerging tools, genome-scale metabolic models (GEMs) represent a scalable formulation for simulating biological activity over whole cells and cellular communities.[6–11] Activity is here quantified by the flux through metabolic reaction networks based on the constraints provided by physico-chemical laws, network structure, and potentially additional biological data and knowledge. For instance, genome-scale modeling approaches for microbial communities can often account for varying microbial abundances.[12–15] While GEM application in this context is a relatively young field, latest progress demonstrated

that these models are becoming mature for exploring complex microbiomes,[12,16] with recent studies beginning to apply GEM reconstruction and analysis starting from MAGs.[17,18] However, large-scale community modeling remains centered on the functional potential encoded in genomic sequences and lacks effective guidance over how this translates into biological activity in different conditions.

To complement metagenomic profiling, other *meta*-omics data are necessary. For instance, transcriptomics applied on metagenomic samples can give insights into microbial community activity when mapping transcript abundance profiles onto MAGs.[19–21] Such a genome-centric metatranscriptomics (GCM) could thus be exploited to inform and contextualize community GEMs much like traditional transcriptomics is used in the modeling of single organisms.[22–24] Yet, although methods for transcriptomics data integration in GEMs exist and are widely adopted, they were developed for single-species models and are not suited to the specific challenges of GCM. Indeed, some previous studies combined GEMs with metatranscriptomic data, but either they avoided direct integration[25,26] or they focused on simple synthetic consortia of known species.[26,27] Particularly, different members of a community often possess highly heterogeneous distributions in abundance and activity. GCM profile heterogeneity reflects the underlying variable gene regulation associated with a continuous adaptation and the inter-species interaction network. This leads to highly unbalanced representation in terms of functional activity, which is only partly correlated with MAG abundance. Moreover, inter-sample RNA abundance heterogeneity often translates into a varying ability in quantifying real gene expression changes. Ultimately, a valid approach for the integration of GCM into community-scale GEMs is currently missing, preventing the exploration of complex microbiomes through multi-omics mechanistic models.

Here, we present an approach for the integration of GCM profiles in community GEMs that explicitly accounts for a wide range of gene expression across MAGs. Our approach thus produces condition-specific community GEMs (CoCo-GEMs), hence we refer to it as CoCo. To demonstrate the use of the proposed approach, we applied it to investigate the reconstruction and analysis of genome-scale microbial community metabolism in two widely studied ecosystems: anaerobic digestion reactors and human gut microbiota. We show that CoCo can improve community-level phenotypic predictions and detect key metabolic adaptation responses in such complex communities. In anaerobic digestion consortia, our results suggest syntrophic mechanisms linked to hydrogen availability and identify specific amino acid requirements for species using alternative trophic pathways. In the gut microbiota of Crohn's disease patients, our findings are consistent with an impaired short-chain fatty acid cross-feeding network.

Our work thus expands the investigation of complex microbial communities by means of omics data, with a 3-fold contribution: (1) the design of a computational methodology – CoCo – for condition-specific genome-scale modeling over microbial communities; (2) the introduction of end-to-end workflows for probing metabolic interactions in culture-independent microbial systems based both on their functional potential and transcriptional activity; and (3) the genome-scale model verification of crucial cross-feedings within real complex communities. The developed tools can be applied in genomics studies of microbial consortia and provide a step forward in culture-independent ecosystem modeling.

RESULTS

Modeling transcriptional regulation over microbial communities

The first considered biological system was constituted by anaerobic reactors operating in a continuous mode, whose microbial communities originate from a biogas plant and have been previously characterized.[28] The core biological process for this type of community is the anaerobic digestion of organic substrates, ultimately converging on methanogenesis.[29] Known methanogenic organisms mainly belong to the Euryarchaeota phylum and largely exploit acetate or hydrogen and carbon dioxide as fueling compounds via acetoclastic or hydrogenotrophic pathways, respectively. These organisms thrive also thanks to syntrophic interactions with bacterial species that benefit from the disposal of these compounds.[30] In this work, the microbes were subject to two feeding regimens: simple medium with acetate as the sole carbon source, temporally followed by the same medium with the addition of exogenous hydrogen (Figure 1A). The low environmental complexity allowed the simplification of the microbiome while retaining the most critical species, yet complex interactions remained.

Such ecosystem served as a primary test-bed for the proposed multi-omics workflow, which was conceived to be applicable in any scenario where culture-independent strategies are required. Briefly, the first step involves shotgun metagenomics followed by metagenome assembly and MAG reconstruction for comprehensively characterizing the functional potential of the microbiome. To increase MAG quality, here we performed long-read DNA sequencing and devised a multi-assembly-and-binning strategy combining long and short reads with multiple tools at each analysis stage (for details consult STAR Methods). The second step is GCM analysis to profile the condition-specific functional activity of individual species by mapping RNA reads on the MAGs. MAG sequences are then used for the reconstruction of GEMs, which are assembled within a shared virtual environment, thereby generating a community GEM (Co-GEM) for each individual sample. Finally, to explicitly model a varying activity in the communities, we introduced an approach, named CoCo, to convert Co-GEMs into CoCo-GEMs. Specifically, CoCo accounts for variation in both gene transcript abundance and whole-transcriptome abundance of individual MAGs. The former is used to shape the biochemical reaction bounds like traditional condition-specific GEM building methods,[23,24] and in particular analogously to Metabolic and Transcriptomics Adaptation Estimator (METRADE), which uses a logarithmic map to connect enzyme abundance and flux constraints.[31] The use of the latter is instead 2-fold: on one hand, the base bounds of individual GEMs are re-scaled depending on the associated transcriptome abundance to model the global activity heterogeneity in different community members. Second, lowly active species suffer from sparse expression count

CellPress
OPEN ACCESS

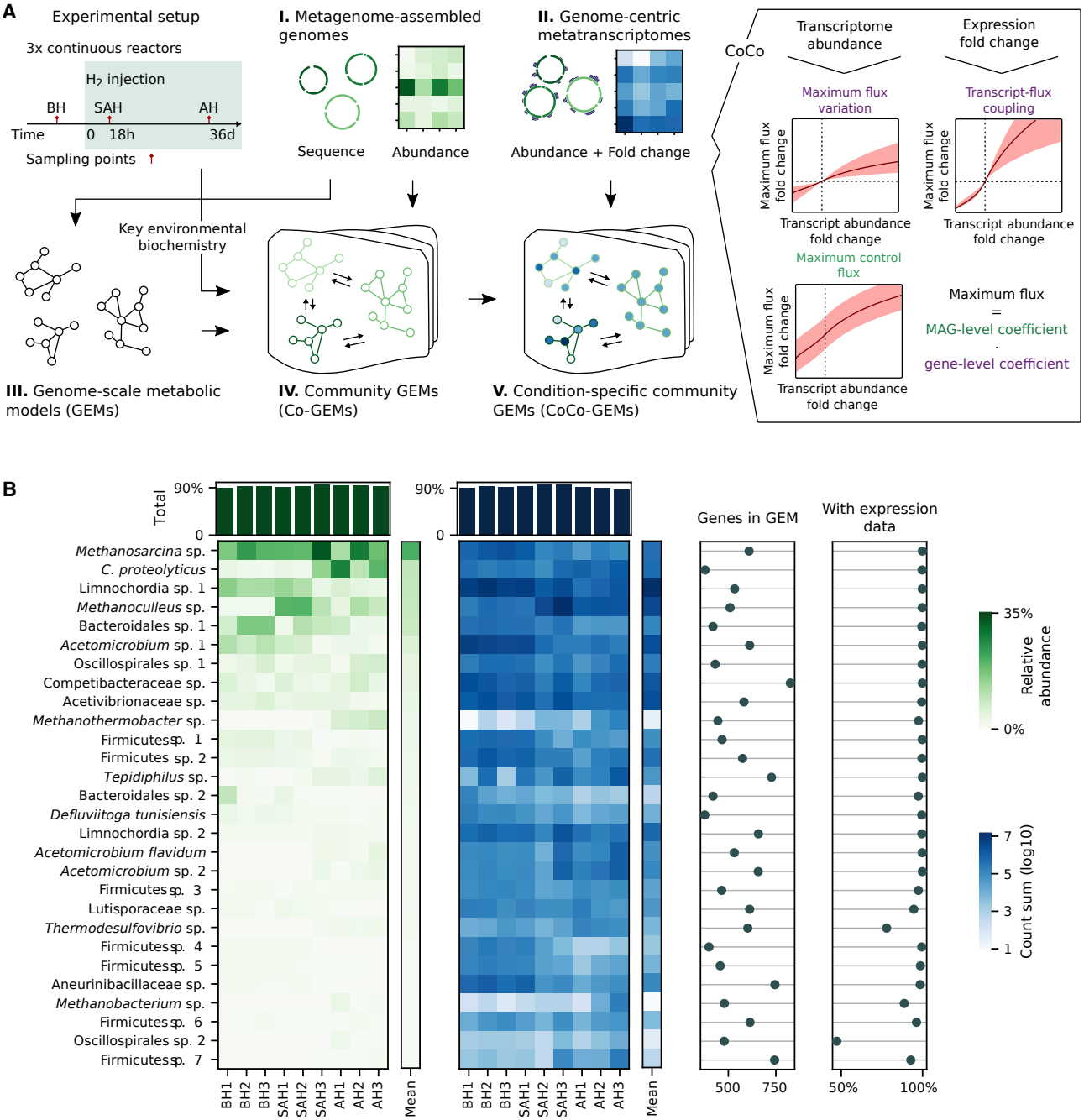


Figure 1. Overview of the proposed approach and the considered anaerobic digestion ecosystem

(A) Schematic representation of the proposed workflow. The system was composed of three continuous bioreactors fed on basal anaerobic medium with acetate as the sole carbon source, subject to exogenous hydrogen injection. Sampling of biological material was performed before, shortly after, and after the start of hydrogen injection (BH, SAH, and AH, respectively). MAG sequences were used to reconstruct GEMs of microbial species, and their abundance was used to define community GEMs (Co-GEMs) for each sample together with biochemical measurements for key compounds. The newly introduced approach CoCo was then applied to build condition-specific community GEMs (CoCo-GEMs) starting from GCM profiles.

(B) Relative abundance and total GCM counts for the considered MAGs over the samples, which comprise 24 bacteria and four archaea. Overall, these organisms account for over 90% of functional potential and activity. On the right, number of genes for each reconstructed GEM and their fraction having expression information. See also Figure S1.

distributions, which can affect the reliability of detected activity changes. Thus, the magnitude of gene-level bound modulation is tuned depending on transcriptome abundance, attenuating gene expression changes supported by less reliable count distributions. CoCo therefore generalizes METRADE to microbial communities. The whole workflow is depicted in Figure 1A, and details are described in the STAR Methods.

Continuous bioreactors operate at a constant feeding inflow and microbial matter dilution, similarly to the human gut. Thus, we chose the cooperative trade-off approach as a framework for our simulations.[12] In its original form, this method uses microbial abundances as an input to estimate their growth rate while balancing between the growth of the entire community and that of individual species. Cooperation and competition can thus be explicitly taken into account. The resulting metabolic fluxes are tailored to specific conditions solely based on microbial abundance profiles. Through CoCo, we thus introduced gene expression to better estimate condition-specific metabolic activity. Moreover, in our case study, external model boundaries were defined based on the medium composition and on biochemical measurements describing the net consumption and production of key compounds.

As a result of the hybrid short- and long-read shotgun sequencing analysis, we obtained a total of 44 high-quality and 25 medium-quality MAGs, as defined by the Genomic Standards Consortium.[32] Compared with a previous analysis of the same metagenomes, the improvement was a 25% increase in high-quality MAGs and the complete elimination of low-quality ones (Figure S1A and Data S1).[28] Throughout the following steps, we considered all the MAGs having comprehensive "multi-omic abundance" above 1% in at least one condition, for a total of 28 MAGs (for details see STAR Methods). Figure 1B shows the relative abundance of these MAGs and their corresponding fraction of RNA counts detected, whose total in each sample was around 90% for both data types. Archaeal members of the communities comprise three hydrogenotrophic organisms (*Methanothermobacter*, *Methanobacterium,* and *Methanoculleus*) and a generalistic species capable of using both the acetotrophic, hydrogenotrophic, and methylotrophic pathways (*Methanosarcina*). While *Methanosarcina* is initially the most active member, *Methanoculleus* has a strong increase in both abundance and activity in the second phase. Among bacteria, Bacteroidales, *Limnochordia*, and *Acetomicrobium* species characterize the first phase of the process, with a shift in favor of *Coprothermobacter proteolyticus* following exogenous hydrogen addition. Hypotheses have previously been advanced over the microbial adaptations and interactions underlying these shifts,[21,28] but were lacking a quantitative framework to test them. In the present study, CoCo was applied to gain quantitative insights on the metabolic shift occurring upon hydrogen addition.

CoCo improves community-scale metabolic phenotype predictions

Starting from the MAG sequences, we reconstructed GEMs for the 28 species under consideration. In this stage, we sought to maximize the detail of physiological capabilities for each organism while utilizing the least strict assumptions. We thus used two different strategies for gap-filling bacterial and archaeal GEMs. Given the key role of archaea in anaerobic digestion, we specifically required them to be able to carry out methanogenesis starting from appropriate substrates, which include carbon dioxide and hydrogen for the hydrogenotrophic organisms and alternatively acetate or methanol for the generalist *Methanosarcina*. Bacteria cover instead more diversified roles, therefore we only imposed the ability to grow anaerobically on a medium representing the environment of origin (for details see STAR Methods). The quality of all the GEMs was assessed, verifying their general soundness and the absence of structural issues.[33] Validation reports are available as Supplemental Material (see also Figure S1B). As a result of the high metagenome and metatranscriptome relative abundance associated with the selected communities, the percentage of genes captured by the GEMs having successfully detected GCM data was above 90% for most species (Figure 1B).

We then simulated microbial growth in each of the bioreactors at every time point, where 13 to 17 species were simultaneously present over a 1% abundance threshold, comparing the results of Co-GEMs and CoCo-GEMs. In this validation stage, both model types were constrained by biochemical data regarding acetate and hydrogen consumption rates and volatile fatty acids (VFA) non-accumulation in the medium and without the biogas production data (STAR Methods). CoCo parameters were selected based on the Pearson's correlation with independently estimated microbial proliferation rates via analysis of coverage differences across genomic regions with an approach called CoPTR (STAR Methods).[34] Moreover, different cooperative trade-off values α were explored for both Co-GEMs and CoCo-GEMs, examining the degree of cooperation among community members. While this can generally vary depending on the environment,[16] anaerobic digestion communities are postulated to be governed by syntrophies between bacterial acetate oxidizers and archaeal methanogens.[28,29]

Figure 2A shows individual species growth rates obtained through cooperative trade-off with and without gene expression integration in the models. Compared with Co-GEMs, CoCo-GEMs gained a marked and statistically significant improvement in the Pearson correlation between predicted growth rates and independently estimated replication rates (Steiger test $p = 1.8 \cdot 10^{-4}$). These results were obtained with $\alpha = 0.5$ for base models and $\alpha = 0.9$ for CoCo-GEMs, supporting the idea that such communities thrive on key cross-feedings and showing that transcriptional information can affect our ability to determine the cooperation degree of a community. To test whether CoCo's MAG-level bound modulation contributed to identifying more meaningful solutions, we performed the same analysis by using METRADE. In this case, flux bounds were affected to the same extent across all the MAGs. The growth rate predictions show indeed a partial redistribution, but without an overall correlation with replication rates. Correcting for inter-species RNA abundance heterogeneity and uncertainty in expression fold changes is thus beneficial to the identification of biologically meaningful growth rates.

Having optimized CoCo-GEMs according to the growth rate, we next sought to test whether they could predict trends in metabolic activity that were not specifically targeted during model optimization. We thus verified the models' ability to predict the community-level production rates of the main anaerobic

CellPress
OPEN ACCESS

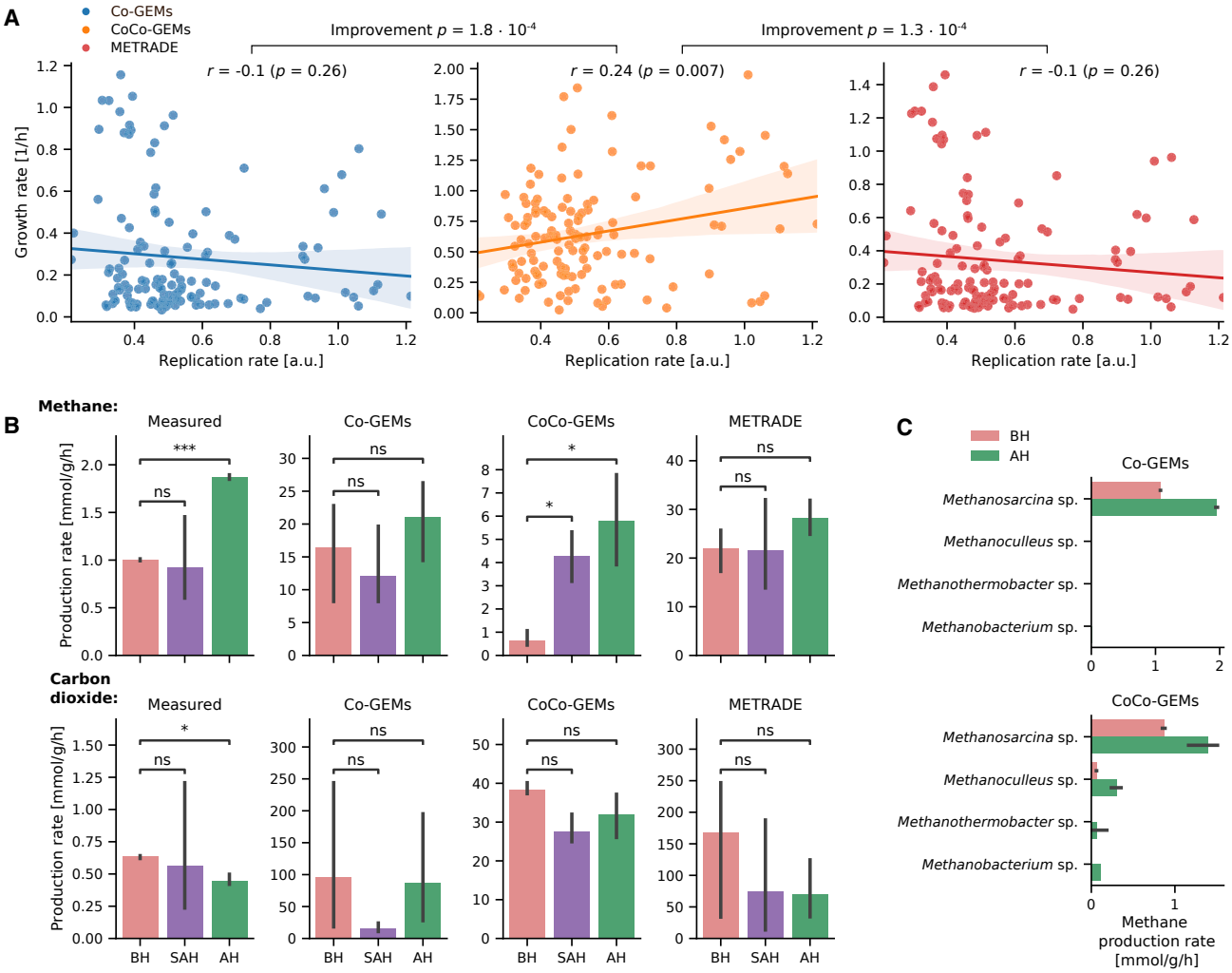


Figure 2. Validation of CoCo in an anaerobic digestion ecosystem

(A) Relationship between the microbial replication rate estimated by CoPTR and the growth rate predicted via the cooperative trade-off approach using Co-GEMs (left) CoCo-GEMs (center), and Co-GEMs constrained by GCM through METRADE. On the top of each panel, the Pearson's correlation coefficient and associated p value are indicated. Improvement p values were obtained by the Steiger test for dependent groups.

(B) Comparison between experimentally measured and predicted gas production rates in the three process stages. Statistical comparison between the conditions was performed by two-sided t tests for dependent samples. CoCo-GEMs are able to more precisely estimate methane production rates both in terms of scale and trends, compared with other models. Data are represented as mean, and error bars represent 95% confidence intervals.

(C) Methane production rates predicted for individual archaeal species when constraining the models with ecosystem-level gas measurements. Data are represented as mean and error bars represent 95% confidence intervals.

digestion products, namely methane and carbon dioxide, as shown in Figure 2B. Experimental data display a consistent increase in methane production upon hydrogen injection, with a concomitant decrease in net carbon dioxide levels. In Co-GEMs, these trends were not clearly identifiable, with predicted production rates one or two orders of magnitude larger than observed. In spite of a large variation among the reactors, CoCo-GEMs instead precisely predicted methane production in the initial feeding regimen, followed by a significant rise in the second phase. Although the production of carbon dioxide results over-estimated by CoCo-GEMs, its decrease along process stages was correctly predicted. METRADE also reproduced the observed trends, yet at a non-significant level and retaining a large absolute error in export fluxes. Further, when integrating the measured community-level gas production rates in the models, Co-GEMs predicted a methanogenic activity essentially driven by *Methanosarcina*, at odds with the increasing abundance of the other archaea. On the contrary, CoCo-GEMs correctly captured the emergence of hydrogenotrophic species (Figure 2C). Overall, these results demonstrate the benefit of GCM integration for identifying key metabolic activities.

CoCo enables probing metabolic niche flexibility

Upon verifying the predictive ability of CoCo-GEMs, we integrated within them the biogas production data that were

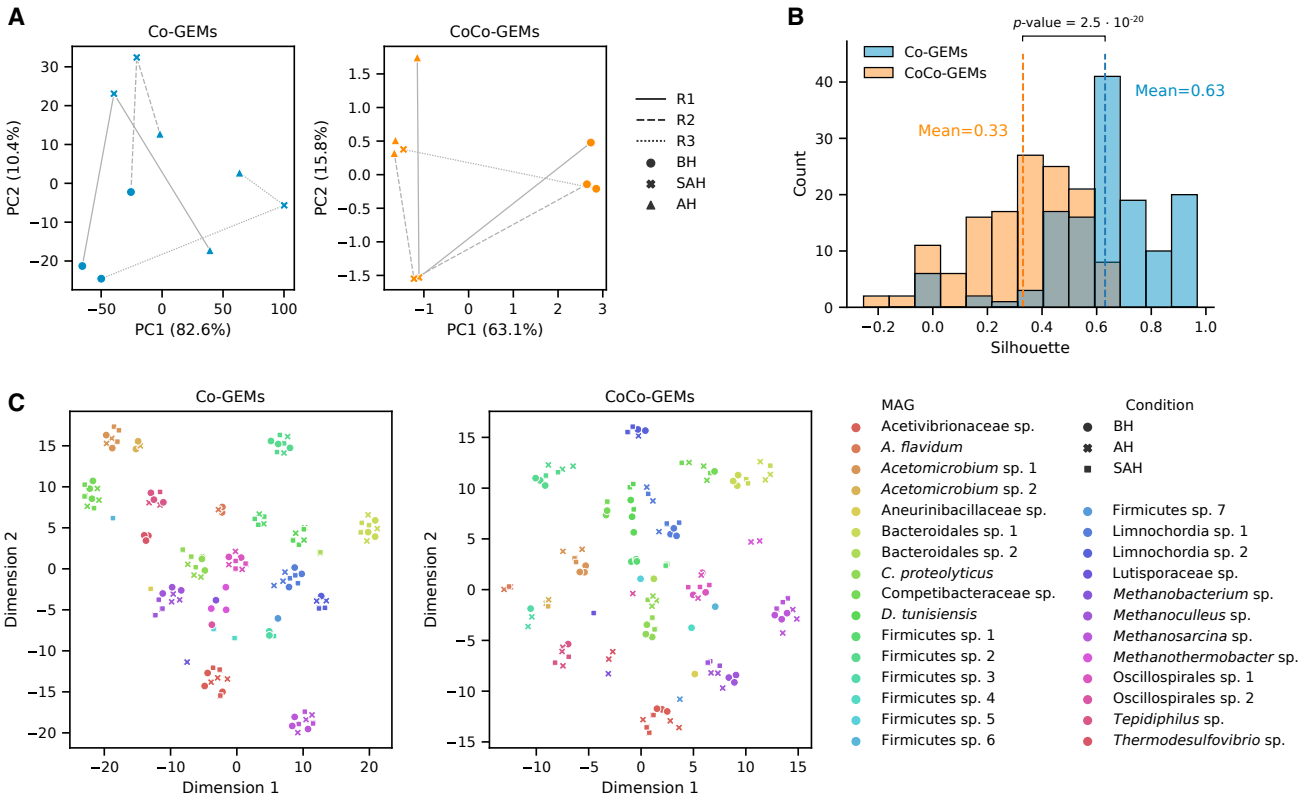


Figure 3. Multi-omics integration reveals metabolic niche flexibility across taxa

(A) Principal component representation of ecosystem metabolic input/outputs expressed by community-level exchange fluxes. Lines depict the trajectory of individual reactors in such metabolic rate space.

(B) Silhouette distributions for fluxome clusters defined by the MAG of origin. CoCo-GEMs capture significantly lower silhouette values, reflecting a higher metabolic flexibility in individual MAGs. The p value was obtained by a two-tailed Wilcoxon signed-rank test.

(C) Bidimensional t-SNE representation of individual MAGs' fluxomes, representing their genome-scale metabolic activity. CoCo-GEMs fluxomes for individual MAGs cluster less evidently than Co-GEMs'. See also Figure S2.

previously excluded for an in-full evaluation of microbial metabolic activity. Figure 3A shows the dynamics of community-level metabolism over the three time points, which was quantified in terms of the total imports and exports into and from the ecosystem. As expected, both in Co-GEMs and CoCo-GEMs, hydrogen injection is the main driver of the global metabolic state of the microbiome; however, only in the latter models the final states are clearly distinguishable from the initial ones, in agreement with the permanent modification of the environmental conditions. In the metabolic exchange space, the system trajectory follows the first principal component upon perturbation, while it adjusts on the final steady state by moving along the second principal component. At the same time, differences distinguishing the individual reactors emerge. Reactor R1 displays a markedly different final state from the other two, which was observed experimentally by a diverging gas production. Reactor R3 has instead a short-term perturbed state more similar to its final state, anticipating the end-point activity and microbial representation (Figure 1B). CoCo-GEMs thus better identify the new metabolic steady states associated with a high hydrogen availability and which are biologically distinct from the initial states both in terms of microbial composition and activity. Microbial growth is in general sustained by a net uptake of sulfite and iron(-III), with corresponding production of hydrogen sulfate and iron(II). Sulfite is consumed by a few organisms including Bacteroidales and *Methanosarcina*, the latter of which exploits it for converting NADH into NAD. By considering the overall trends of global metabolic inputs/outputs, we could thus show that CoCo-GEMs correctly recapitulate ecosystem-level experimental evidence.

We next focused on the intracellular activity of the individual MAGs, exploring how the metabolic niche of different taxonomic groups is affected by accounting for transcriptional information. To this end, we quantified niche flexibility through the silhouette coefficient, a measure used in clustering analysis to estimate the separation between clusters (for details see STAR Methods). In our case, clusters identify the MAGs and the higher their silhouette and the more they belong in restricted specialist niches. Silhouette distributions for Co-GEMs and CoCo-GEMs are shown in Figure 3B, where it is possible to see that the latter have significantly lower mean silhouette (Wilcoxon signed-rank test $p = 2.5 \cdot 10^{-20}$), mirroring an elevated niche flexibility. This indicates that modeling transcript abundance correctly captures metabolic adaptation over variable environmental and

CellPress
OPEN ACCESS

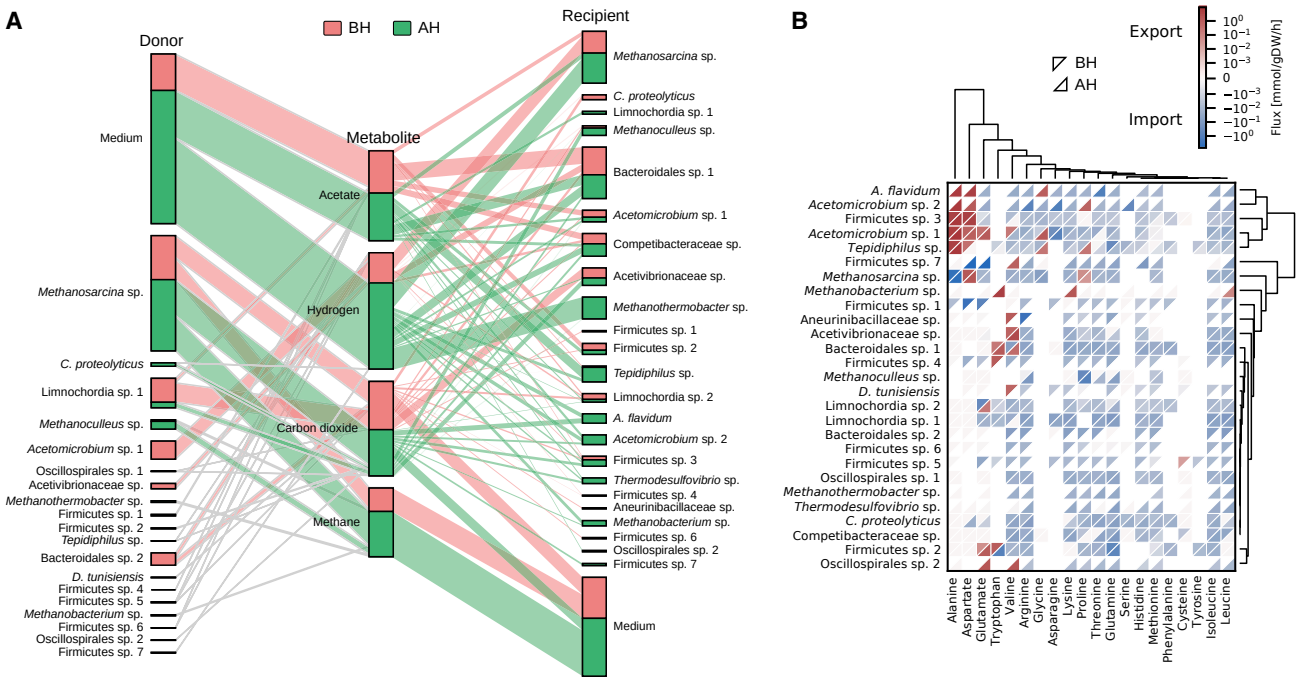


Figure 4. Decomposing metabolic cross-feeding in the methanogenic consortia

(A) Cross-species exchange of the main metabolites driving methanogenesis in the absence (BH) and presence (AH) of external hydrogen supplementation.

(B) Amino acid exchange distribution in the absence (BH) and presence (AH) of external hydrogen supplementation. Blue and red triangles denote consumption and production, respectively. See also Figure S3.

compositional conditions. The higher niche specialization predicted by Co-GEMs is displayed in Figures 3C and S2, where individual MAGs form well-distinguished bulks. In contrast, CoCo-GEM metabolic phenotypes present more variable distributions according to the species to which they belong, better reflecting metabolic adaptations throughout the environmental conditions and the corresponding microbial interaction variations.

CoCo reveals the plasticity of key cross-feedings underlying methanogenesis

Upon verifying the soundness of CoCo-GEMs, we sought to use these models for studying which metabolic processes sustain microbial growth and biogas production in the ecosystem. Figure 4A shows the redistribution of acetate and hydrogen within the communities in the two steady states along with the associated methane and carbon dioxide output. While major acetate consumers are *Methanosarcina*, Bacteroidales sp. 1, *Acetomicrobium* sp. 1, Firmicutes sp. 2, and *Tepidiphilus*, acetate is also shared among numerous other community members in variable amounts. This confirms that, despite the simplicity of the growth medium, this carbon source can sustain the vast majority of this relatively complex community. Concomitantly, carbon dioxide is exchanged in consistent amounts within the ecosystem, with the experimentally measured accumulation in the medium being only a fraction of the total amount produced and consumed. Carbon dioxide is largely produced by *Methanosarcina* and *Limnochordia* in the lack of exogenous hydrogen, although the abundance decrease of *Limnochordia*, consequent to the external introduction of this gas, determined a concomitant decrease in carbon dioxide export by this organism. Just like for acetate, carbon dioxide is also absorbed by numerous taxa and contributes to shaping community composition. Further, *Acetomicrobium* sp. 1 was found to produce hydrogen when external supplementation was not active. As expected, in the first steady state endogenous hydrogen production feeds hydrogenotrophic activity, yet various bacterial taxa appear to exploit it as well. For example, hydrogen fuels ferredoxin NADPH-linked hydrogenase in Bacteroidales sp. 1.

In contrast, metabolic cross-feeding estimated without the use of GCM predicted a disproportionately large activity within the community compared with the net exchanges with the environment (Figure S3A). In this case, *Methanosarcina* resulted as a major acetate producer, which is unlikely due to the dominance of this species in terms of abundance. Moreover, *Methanosarcina* and *Acetomicrobium* spp. Release hydrogen even during the external supplementation. Overall, Co-GEMs predictions display evident differences with CoCo-GEMs', some of which can hardly be reconciled with biological knowledge.

Among the most abundant bacteria, multiple metabolic patterns reminiscent of syntrophic behavior with archaea were observed (Figure 4A). Before exogenous hydrogen addition, *Limnochordia* sp. 1 emerges as the main carbon dioxide producer, while *Acetomicrobium* sp. 1 displays the largest export of hydrogen, which is predominantly used by *Methanosarcina*. Besides archaea, *Bacteroidales* sp. 1 appears to thrive on the availability of these gases as well as of acetate in both steady states. During hydrogen injection, *C. proteolyticus* becomes the most abundant bacterium displaying an activity shift from being a carbon dioxide consumer to a

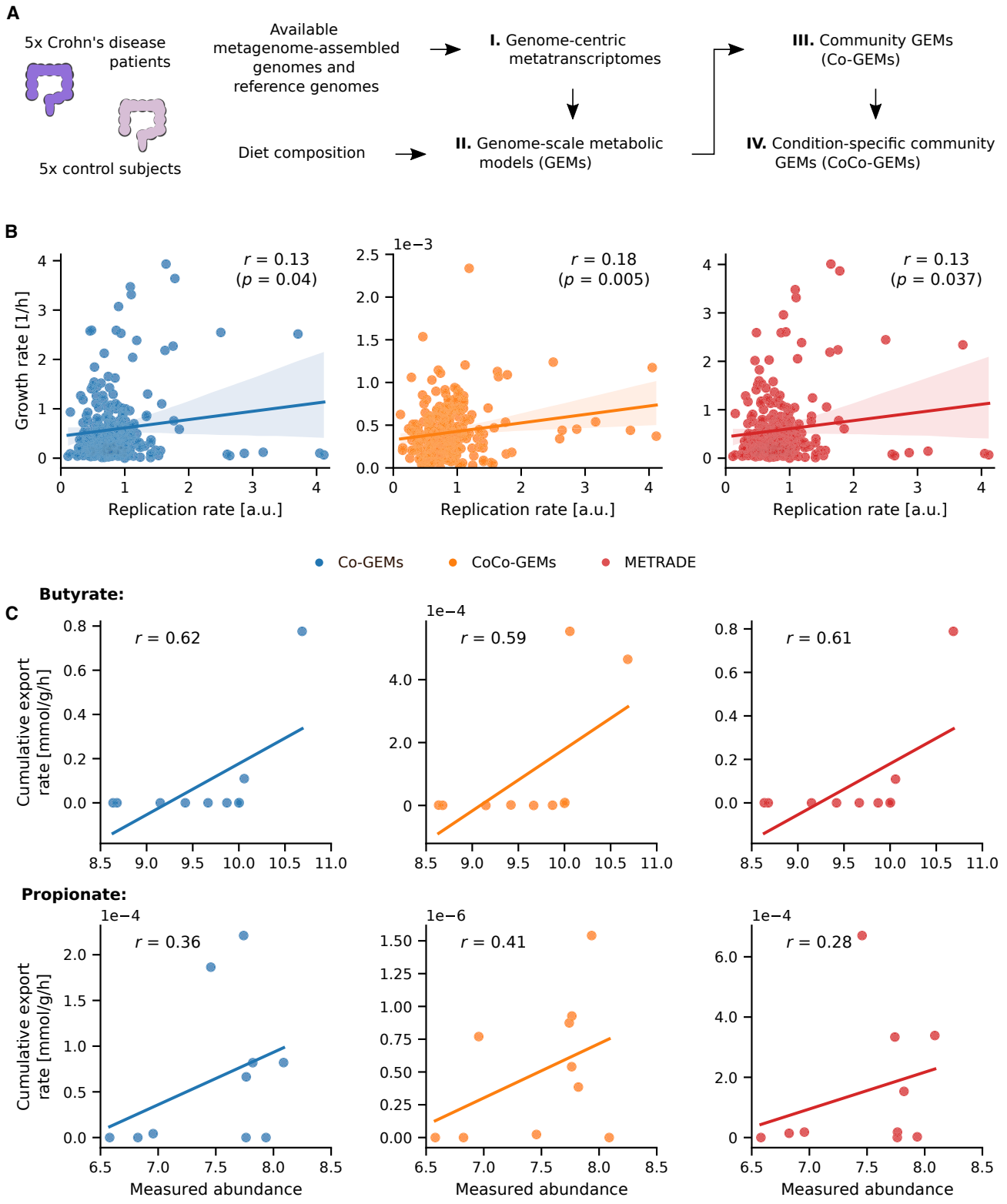


Figure 5. Validation of CoCo in human gut microbiota

(A) Schematic representation of the workflow applied to the human microbiota, for which high-quality MAGs and reference genomes are publicly available. See also Figure S4.

(legend continued on next page)

producer. This trend may support a niche swap between *Limnochordia* sp. 1 and *C. proteolyticus* as the main archaeal syntrophic partner. A syntrophic relationship between *C. proteolyticus* and the hydrogenotrophic *Methanothermobacter wolfeii* has been previously hypothesized based on GCM patterns of microbial communities in two-phase reactors,[35] and has been tested *in silico* in pairwise community models.[17] Yet, these computational results accounted for the interaction between the two species in isolation, overlooking the global metabolic context. Our results support the establishment of positive interactions between *C. proteolyticus* and hydrogenotrophic archaea showing that this bacterium can provide substrates fueling hydrogenotrophic methanogenesis and also revealing that such substrates can change depending on the environmental conditions.

Next, given the relevance of amino acid auxotrophies in anaerobic digestion communities,[17] we inspected the associated exchange fluxes obtained by CoCo-GEMs. Figure 4B shows the amino acid exchange distribution in the absence and presence of external hydrogen supplementation. Here, major amino acid producers were identified to populate a cluster of taxa belonging to the *Acetomicrobium* genus, *Tepidiphilus* sp., and Firmicutes sp. 3. Alanine and aspartate are consistently secreted by all the taxa in the cluster. At the same time, *Methanosarcina* displays a strong uptake of alanine in both nutritional conditions. Members of Synergistaceae, such as *Acetomicrobium*, are known for their ability to degrade amino acids to VFAs. Here the reverse process is suggested to support a beneficial relationships between them and archaea. This finding is consistent with a syntrophic cross-feeding previously identified between *Desulfovibrio vulgaris* and *Methanococcus maripaludis*.[36] In such syntrophy, alanine was hypothesized to be used by *M. maripaludis* as a nitrogen source and to reduce the energy cost in the autotrophic synthesis of pyruvate, while the advantage for *D. vulgaris* is less clear. Further, the emergence of *C. proteolyticus* in the community and its shift to carbon dioxide production is associated with an increased amino acid demand, which expands to including aspartate, glutamate, leucine, and lysine.

Also on the level of amino acid exchanges, GCM integration provides a markedly different picture compared with base community models (Figure S3B). While *Acetomicrobium* spp. Are still among the major amino acid producers, the importance of alanine for *Methanosarcina* is less clear and *Methanoculleus* results as a proline producer. Although specific metabolic exchanges are challenging to be experimentally validated, such differences combined with the above CoCo evaluation results suggest that GCM integration in a genome-scale modeling framework can substantially improve the detection of key activity and cross-feeding patterns.

CoCo supports the characterization of personal gut microbiota metabolism

As a second validation setting, we considered the human gut microbiome of subjects monitored over multiple omic levels. The selected cohort consists of a group of five patients diagnosed with Crohn's disease (CD) displaying an altered microbiota and five control subjects.[37] Such a microbial ecosystem represents a more problematic scenario than the one considered above, as human microbiota is highly personal and often presents numerous lowly abundant species, with a globally high diversity.[3,37,38] In addition, diet is heterogeneous across individuals and represents a source of uncertainty for metabolic modeling, requiring some approximations (details in STAR Methods).

Given the wide availability of high-quality MAGs and genomes for such microbial ecosystems, it can be advantageous to use these directly rather than reconstructing MAGs *de novo*. As an alternative workflow, we thus started from one such genome collection[38] and followed the rest of the process described in the section "Modeling transcriptional regulation over microbial communities," as depicted in Figure 5A, thus generating CoCo-GEMs for each individual's microbiota. Using a 0.1% abundance threshold, considered microbial communities comprised from 22 up to 91 taxa, depending on the individual, for a total relative abundance between 90% and 97% and 289 species. In this case, MAGs captured a fraction of GCM counts between 15% and 47%, with an extremely large variation in model genes coverage, reflecting the dynamics of the system and the large number of lowly abundant microorganisms (Figure S4). Despite that, CoCo could be applied, as it was designed with these difficulties in mind (details in STAR Methods).

Again, we explored the relation between model-predicted growth rates and corresponding species replication rates. In this case, base community models achieved a positive correlation, a sign that diet was approximated at an acceptable level. The removal of dietary constraints in fact almost completely eliminated such trend (Figure S5A). Also here, improvements were found upon introducing transcriptional information, albeit milder (Figure 5B). Of note is that even without dietary constraints, CoCo is able to comparably recover significant microbial growth trends (Figure S5A). The best cooperation trade-off decreased from 0.9 to 0.6 when comparing Co-GEMs with CoCo-GEMs, indicative of a microbiota more balanced between cooperation and competition.

Short-chain fatty acids (SCFA) are among the main fermentation products in the gut, resulting from the catabolism of carbohydrates and proteins.[39] Specifically, butyrate and propionate were identified and tracked in the considered cohort via untargeted metabolomics.[38] For these reasons, we analyzed genome-scale model predictions over these metabolites to verify whether they match experimental observations. We assumed that detectable metabolite concentration is directly proportional to global microbial secretion flux and determined the correlation between cumulative production over all community members and metabolomics abundance. Figure 5B shows that positive Pearson's correlations are found both for Co-GEMs and CoCo-GEMs, yet it can be noticed that Co-GEMs

(B) Relationship between the microbial replication rate estimated by CoPTR and the growth rate predicted via the cooperative trade-off approach using Co-GEMs (left) CoCo-GEMs (center), and Co-GEMs constrained by GCM through METRADE. On the top of each panel, the Pearson's correlation coefficient and associated p value are indicated. See also Figure S5.

(C) Relationship between the sum of individual microbes' export rate of butyrate and propionate and their abundance detected by metabolomics. While positive correlations are found for all model types, CoCo-GEMs and METRADE return a lower number of null total exports, consistent with the experimental data.

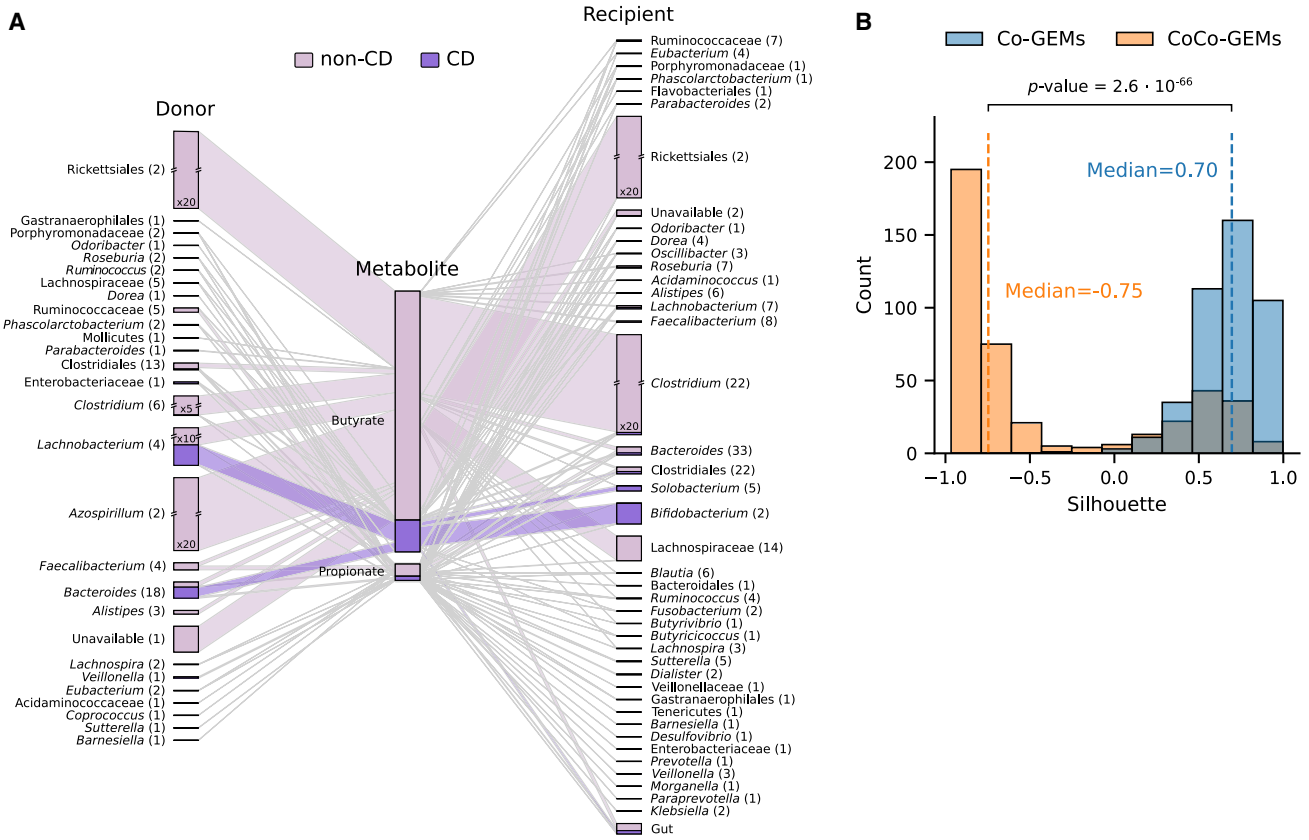


Figure 6. Charting SCFA cross-feeding in the gut microbiota

(A) Cross-species exchange of butyrate and propionate across the considered samples. Line width and box height represent exchange intensity by and to any depicted taxonomic group, while numbers in brackets denote the number of species of each group. Broken-contour boxes were re-scaled by the indicated factors. See also Figure S5.

(B) Silhouette distributions for fluxome clusters defined by the MAG of origin. CoCo-GEMs capture significantly lower silhouette values, reflecting a higher metabolic flexibility in individual MAGs. The p value was obtained by a two-tailed Wilcoxon signed-rank test.

tend to be dominated by a small number of super exchangers that produce and consume most of the SCFA (Figure S5B), while in CoCo-GEMs this pattern is less pronounced (Figure 6A). Moreover, butyrate production was estimated as markedly more intense compared with propionate's, coherently with its higher experimental abundance and with independent estimates in other cohorts.[40]

A large number of species participate in butyrate and propionate cross-feeding across individual communities, as displayed in Figure 6A. Overall, consumers outnumber producers in both subject groups, with a decrease in producers (68 non-CD versus 27 CD) and consumers (145 non-CD versus 80 CD) associated with disease. Some of these organisms belong in taxa known to be involved in butyrate metabolism, such as *Faecalibacterium*, *Roseburia*, *Ruminococcus*, *Bacteroides*, and *Clostridium*.[39,40] Among these, species of the order Rickettsiales and the genera *Clostridium*, *Lachnobacterium*, and *Azospirillum* emerge as the principal butyrate producers, while several *Clostridium* and Rickettsiales species are the main butyrate users. Thus, while microbial functional potential and activity are highly variable across subjects, CoCo-GEMs suggest a reduction of SCFA cross-feeding in CD individuals both in terms of number of participating microbes and overall flux. Finally, we calculated fluxome silhouette coefficients for these communities, finding an even larger separation between Co-GEMs and CoCo-GEMs silhouette distributions than in anaerobic digestion communities (Figure 6B). Such a divide is in line with the highly personal features of the gut microbiota and once again highlights the importance of accounting for functional activity in metabolic models rather than just species co-occurrence and abundance.

DISCUSSION

Computational models of metabolism are invaluable tools for shedding light on complex biological processes occurring on a biochemical reaction scale. Among them, genome-scale models are widely adopted as scaffolds for omics data contextualization and interpretation, both implicitly and explicitly. Yet, upgrading such tools to the metagenome scale remains a largely unexplored area.[7–9] While previous studies have advanced our understanding of microbial interactions,[16,41] such interactions change in response to both dynamic environmental stimuli and biological adaptation.[26,27] In this context, our work takes a step forward

into integrating multi-omics data and knowledge on a metagenome scale in real microbial ecosystems.

Our results provide two alternative use cases where CoCo can be applied: with *de novo* MAG reconstruction or, alternatively, exploiting existing genomes. Both these scenarios have specific advantages and can be evaluated depending on the application. In general, the former strategy is expected to better capture genomic features associated with the presence of specific strains. On the other hand, if an appropriate genome collection is available, the latter solution may be more readily applicable and could be preferred in the cases where a sufficient high-quality MAG representation may not be attained. Clearly, with a decreasing fraction of relevant microbial genomes available, the transcript mapping rate will also diminish. This situation is expected to have a more severe impact where microbiome complexity is high and, crucially, limits the application of GCM and CoCo in extremely complex metagenomes. Nevertheless, large-scale studies are contributing to deciphering such complexity and invest more and more microbial ecosystems, thus supporting the investigation even of complex communities such as host-associated microbiota.[42–44] In this context, it is worth noting that hybrid sequencing can help improve MAG reconstruction,[45] with downstream benefits for GEM building.

Given the momentum in these research areas, new tools could be adopted in any step of the workflow. For instance, CarveMe was selected for its speed, customizability, and ability to generate draft models close to curated ones in terms of reaction and metabolite content.[46] These features allow GEM reconstruction for complex microbial communities while keeping the computation time within reasonable limits, but other tools could alternatively be adopted. As a part of GEM reconstruction, gap-filling is a necessary step that requires some assumptions. In this work, we used general knowledge on archaeal growth capabilities and a minimal amount of expert knowledge. Without such knowledge available, data on the biochemical composition of the environment can be used, or alternatively it is possible to relax nutritional assumptions.[18,47]

Similarly, challenges in modeling RNA abundance in mixed microbial populations add up to recognized uncertainties in defining expression thresholds.[24] Thus, we focused on an approach based on the continuous modulation of expression fold change rather than hard thresholds, using a data-driven approach to fine-tune CoCo-GEMs based on orthogonal observations. Although even this procedure may not be universally optimal, we argue that it can grant flexibility independently from the complexity and the completeness of the metatranscriptome. Ultimately, this allowed us to keep the number of parameters low and simultaneously minimize the computational cost. Moreover, genome-scale modeling of microbial communities is in early development, with new approaches that will likely be developed in the future. One of the main uncertainties is the metabolic objective(s) pursued by a given system, if any. A number of alternatives have been proposed for different scenarios, including community-level and individual-member-level objectives.[7,8] Although here we focused on cooperative trade-off, by acting on the bounds in an objective-agnostic manner, our approach can be used in conjunction with other microbial community modeling approaches. In addition, this strategy does not affect the computational cost, which can be non-negligible with increasing community size.[12] As this field progress and grasps more effective ways to determine biologically meaningful flux states, CoCo can be applied to narrow down the space of feasible solutions and thereby readily support the new tools.

As shown, the proposed approach enables an improved identification of inter-species interactions, and particularly positive ones. Among them, microbial syntrophies are fundamental for the survival of microorganisms in virtually any environment on Earth, including extreme habitats with limited resources or in harsh physical conditions. Such phenomena invest the exchange of nutrients as well the removal of toxic substances.[48] The syntrophy between bacterial fermenters and archaeal methanogens involves the exchange of electron carriers and is one of the most studied paradigmatic examples.[29] In this sense, our first case study functioned as a model for cross-feeding, which is implicated in numerous ecosystems.[16] The recurrent picture of acetoclastic and hydrogenotrophic methanogenesis in a large number of studies on full-scale biogas plants suggests the existence of a delicate balance between the two functional modes whose mechanistic underpinnings are still elusive.[29,49] Here, we investigated the transition from one mode to the other through an approach that provided multi-omic grounds for explaining microbial community adaptation. For instance, in the absence of external hydrogen, *Acetomicrobium* was found to feed *Methanosarcina* with this gas, a phenomenon that was suppressed by its artificial introduction into the system. On the other hand, a consistent flux of alanine was found from all the members of *Acetomicrobium* into *Methanosarcina* in both conditions. These patterns exemplify how the proposed approach can pinpoint both condition-dependent and -independent interactions. Moreover, our results also fall within the records of relevant functional role by rare taxa, which is emerging as a common thread to an increasing number of habitats.[50]

Altogether, this work contributes to the systematic investigation of microbial ecosystems by bridging metagenomics and genome-scale modeling. Given the unculturability of most microorganisms, we believe that this path could be one of the keys to unlocking the principles underlying their organization and evolution.

Limitations of the study

Extending the proposed pipeline to other microbial systems could require some precautions. Primarily, ecosystem complexity determines the number of MAGs that can be reconstructed with a sufficient quality, the fraction of metatranscriptome that maps on them, and in turn the metabolic activity that community models can account for. Final results should be carefully evaluated in light of these factors. Despite that, there are several countermeasures that can be implemented on each individual step, as outlined above. Replication rates estimates, along with other phenotypic measurements such as metabolite accumulation rates, should be used to evaluate the quality of obtained CoCo-GEMs. Ultimately, the proposed methodology is intended to guide the conception of new hypotheses over microbial system functioning and requires support by experimental or independent evidence.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Anaerobic digestion biological system
- METHOD DETAILS
  - DNA extraction and sequencing
  - Human gut microbiota data
  - Genome-centric metagenomics
  - Genome-centric metatranscriptomics
  - Microbial replication rate estimation
  - Genome-scale metabolic model reconstruction
  - Microbial community modeling
  - Genome-centric metatranscriptomic data integration
  - Multivariate flux analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.crmeth.2022.100383.

ACKNOWLEDGMENTS

This work was financially supported by the "Budget Integrato della Ricerca Dipartimentale" (BIRD198423) PRID 2019 of the Department of Biology at the University of Padova, entitled "SyMMoBio: inspection of Syntrophies with Metabolic Modeling to optimize Biogas Production." C.A. acknowledges a Network Development Award from The Alan Turing Institute, grant number TNDC2-100022.

AUTHOR CONTRIBUTIONS

Conceptualization: G.Z., S.C., and L.T.; methodology, G.Z., S.C., and C.A.; software, G.Z.; validation, G.Z.; formal analysis, G.Z.; investigation, G.Z. and S.C.; resources, S.C. and L.T.; data curation, G.Z.; writing – original draft, G.Z.; writing – review & editing, G.Z., S.C., C.A., and L.T.; visualization, G.Z.; supervision, C.A. and L.T.; project administration, L.T.; funding acquisition, L.T.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 28, 2022
Revised: October 7, 2022
Accepted: December 12, 2022
Published: January 6, 2023

REFERENCES

1. Poyet, M., Groussin, M., Gibbons, S.M., Avila-Pacheco, J., Jiang, X., Kearney, S.M., Perrotta, A.R., Berdy, B., Zhao, S., Lieberman, T.D., et al. (2019). A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. Nat. Med. *25*, 1442–1452. https://doi.org/10.1038/s41591-019-0559-3.

2. Pande, S., and Kost, C. (2017). Bacterial unculturability and the formation of intercellular metabolic networks. Trends Microbiol. *25*, 349–361. https://doi.org/10.1016/j.tim.2017.02.015.

3. Almeida, A., Nayfach, S., Boland, M., Strozzi, F., Beracochea, M., Shi, Z.J., Pollard, K.S., Sakharova, E., Parks, D.H., Hugenholtz, P., et al. (2021). A unified catalog of 204, 938 reference genomes from the human gut microbiome. Nat. Biotechnol. *39*, 105–114. https://doi.org/10.1038/s41587-020-0603-3.

4. Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.A., Woodcroft, B.J., Evans, P.N., Hugenholtz, P., and Tyson, G.W. (2017). Recovery of nearly 8, 000 metagenome-assembled genomes substantially expands the tree of life. Nat. Microbiol. *2*, 1533–1542. https://doi.org/10.1038/s41564-017-0012-7.

5. Campanaro, S., Treu, L., Kougias, P.G., De Francisci, D., Valle, G., and Angelidaki, I. (2016). Metagenomic analysis and functional characterization of the biogas microbiome using high throughput shotgun sequencing and a novel binning strategy. Biotechnol. Biofuels *9*, 26. https://doi.org/10.1186/s13068-016-0441-1.

6. Kumar, M., Ji, B., Zengler, K., and Nielsen, J. (2019). Modelling approaches for studying the microbiome. Nat. Microbiol. *4*, 1253–1267. https://doi.org/10.1038/s41564-019-0491-9.

7. Frioux, C., Singh, D., Korcsmaros, T., and Hildebrand, F. (2020). From bag-of-genes to bag-of-genomes: metabolic modelling of communities in the era of metagenome-assembled genomes. Comput. Struct. Biotechnol. J. *18*, 1722–1734. https://doi.org/10.1016/j.csbj.2020.06.028.

8. Colarusso, A.V., Goodchild-Michelman, I., Rayle, M., and Zomorrodi, A.R. (2021). Computational modeling of metabolism in microbial communities on a genome-scale. Curr. Opin. Syst. Biol. *26*, 46–57. https://doi.org/10.1016/j.coisb.2021.04.001.

9. García-Jiménez, B., Torres-Bacete, J., and Nogales, J. (2021). Metabolic modelling approaches for describing and engineering microbial communities. Comput. Struct. Biotechnol. J. *19*, 226–246. https://doi.org/10.1016/j.csbj.2020.12.003.

10. Heinken, A., Basile, A., and Thiele, I. (2021). Advances in constraint-based modelling of microbial communities. Curr. Opin. Syst. Biol. *27*, 100346. https://doi.org/10.1016/j.coisb.2021.05.007.

11. Fang, X., Lloyd, C.J., and Palsson, B.O. (2020). Reconstructing organisms in silico: genome-scale models and their emerging applications. Nat. Rev. Microbiol. *18*, 731–743. https://doi.org/10.1038/s41579-020-00440-4.

12. Diener, C., Gibbons, S.M., and Resendis-Antonio, O. (2020). MICOM: metagenome-scale modeling to infer metabolic interactions in the gut microbiota. mSystems *5*, e00606-19. https://doi.org/10.1128/msystems.00606-19.

13. Chan, S.H.J., Simons, M.N., and Maranas, C.D. (2017). Predicting microbial abundances while ensuring community stability. PLoS Comput. Biol. *13*, e1005539. https://doi.org/10.1371/journal.pcbi.1005539.

14. Zomorrodi, A.R., and Maranas, C.D. (2012). A multi-level optimization framework for the metabolic modeling and analysis of microbial communities. PLoS Comput. Biol. *8*, e1002363. https://doi.org/10.1371/journal.pcbi.1002363.

15. Khandelwal, R.A., Olivier, B.G., Röling, W.F.M., Teusink, B., and Bruggeman, F.J. (2013). Community flux balance analysis for microbial consortia at balanced growth. PLoS One *8*, e64567. https://doi.org/10.1371/journal.pone.0064567.

16. Machado, D., Maistrenko, O.M., Andrejev, S., Kim, Y., Bork, P., Patil, K.R., and Patil, K.R. (2021). Polarization of microbial communities between competitive and cooperative metabolism. Nat. Ecol. Evol. *5*, 195–203. https://doi.org/10.1038/s41559-020-01353-4.

17. Basile, A., Campanaro, S., Kovalovszki, A., Zampieri, G., Rossi, A., Angelidaki, I., Valle, G., and Treu, L. (2020). Revealing metabolic mechanisms of interaction in the anaerobic digestion microbiome by flux balance analysis. Metab. Eng. *62*, 138–149. https://doi.org/10.1016/j.ymben.2020.08.013.

18. Zorrilla, F., Buric, F., Patil, K.R., and Zelezniak, A. (2021). Reconstruction of genome scale metabolic models directly from metagenomes. Nucleic Acids Res. *49*, e126. https://doi.org/10.1093/nar/gkab815.

19. Singer, E., Wagner, M., and Woyke, T. (2017). Capturing the genetic makeup of the active microbiome in situ. ISME J. *11*, 1949–1963. https://doi.org/10.1038/ismej.2017.59.

20. Franzosa, E.A., Hsu, T., Sirota-Madi, A., Shafquat, A., Abu-Ali, G., Morgan, X.C., and Huttenhower, C. (2015). Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. Nat. Rev. Microbiol. *13*, 360–372. https://doi.org/10.1038/nrmicro3451.

21. Kakuk, B., Wirth, R., Maróti, G., Szuhaj, M., Rakhely, G., Laczi, K., Kovács, K.L., and Bagi, Z. (2021). Early response of methanogenic archaea to h2 as evaluated by metagenomics and metatranscriptomics. Microb. Cell Fact. *20*, 127. https://doi.org/10.1186/s12934-021-01618-y.

22. Cho, J.S., Gu, C., Han, T.H., Ryu, J.Y., and Lee, S.Y. (2019). Reconstruction of context-specific genome-scale metabolic models using multiomics data to study metabolic rewiring. Curr. Opin. Syst. Biol. *15*, 1–11. https://doi.org/10.1016/j.coisb.2019.02.009.

23. Vijayakumar, S., Conway, M., Lió, P., and Angione, C. (2018). Seeing the wood for the trees: a forest of methods for optimization and omic-network integration in metabolic modelling. Brief. Bioinform. *19*, 1218–1235. https://doi.org/10.1093/bib/bbx053.

24. Richelle, A., Joshi, C., and Lewis, N.E. (2019). Assessing key decisions for transcriptomic data integration in biochemical networks. PLoS Comput. Biol. *15*, e1007185. https://doi.org/10.1371/journal.pcbi.1007185.

25. Embree, M., Liu, J.K., Al-Bassam, M.M., and Zengler, K. (2015). Networks of energetic and metabolic interactions define dynamics in microbial communities. Proc. Natl. Acad. Sci. USA *112*, 15450–15455. https://doi.org/10.1073/pnas.1506034112.

26. Zuñiga, C., Li, C.T., Yu, G., Al-Bassam, M.M., Li, T., Jiang, L., Zaramela, L.S., Guarnieri, M., Betenbaugh, M.J., and Zengler, K. (2019). Environmental stimuli drive a transition from cooperation to competition in synthetic phototrophic communities. Nat. Microbiol. *4*, 2184–2191. https://doi.org/10.1038/s41564-019-0567-6.

27. Blasche, S., Kim, Y., Mars, R.A.T., Machado, D., Maansson, M., Kafkia, E., Milanese, A., Zeller, G., Teusink, B., Nielsen, J., et al. (2021). Metabolic cooperation and spatiotemporal niche partitioning in a kefir microbial community. Nat. Microbiol. *6*, 196–208. https://doi.org/10.1038/s41564-020-00816-5.

28. Zhu, X., Campanaro, S., Treu, L., Seshadri, R., Ivanova, N., Kougias, P.G., Kyrpides, N., and Angelidaki, I. (2020). Metabolic dependencies govern microbial syntrophies during methanogenesis in an anaerobic digestion ecosystem. Microbiome *8*, 22. https://doi.org/10.1186/s40168-019-0780-9.

29. Angelidaki, I., Treu, L., Tsapekos, P., Luo, G., Campanaro, S., Wenzel, H., and Kougias, P.G. (2018). Biogas upgrading and utilization: current status and perspectives. Biotechnol. Adv. *36*, 452–466. https://doi.org/10.1016/j.biotechadv.2018.01.011.

30. Evans, P.N., Boyd, J.A., Leu, A.O., Woodcroft, B.J., Parks, D.H., Hugenholtz, P., and Tyson, G.W. (2019). An evolving view of methane metabolism in the archaea. Nat. Rev. Microbiol. *17*, 219–232. https://doi.org/10.1038/s41579-018-0136-7.

31. Angione, C., and Lió, P. (2015). Predictive analytics of environmental adaptability in multi-omic network models. Sci. Rep. *5*, 15147. https://doi.org/10.1038/srep15147.

32. Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., Schulz, F., Jarett, J., Rivers, A.R., Eloe-Fadrosh, E.A., et al. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat. Biotechnol. *35*, 725–731. https://doi.org/10.1038/nbt.3893.

33. Lieven, C., Beber, M.E., Olivier, B.G., Bergmann, F.T., Ataman, M., Babaei, P., Bartell, J.A., Blank, L.M., Chauhan, S., Correia, K., et al. (2020). MEMOTE for standardized genome-scale metabolic model testing. Nat. Biotechnol. *38*, 272–276. https://doi.org/10.1038/s41587-020-0446-y.

34. Joseph, T.A., Chlenski, P., Litman, A., Korem, T., and Pe'er, I. (2022). Accurate and robust inference of microbial growth dynamics from metagenomic sequencing reveals personalized growth rates. Genome Res. *32*, 558–568. https://doi.org/10.1101/gr.275533.121.

35. Fontana, A., Kougias, P.G., Treu, L., Kovalovszki, A., Valle, G., Cappa, F., Morelli, L., Angelidaki, I., and Campanaro, S. (2018). Microbial activity response to hydrogen injection in thermophilic anaerobic digesters revealed by genome-centric metatranscriptomics. Microbiome *6*, 194. https://doi.org/10.1186/s40168-018-0583-4.

36. Walker, C.B., Redding-Johanson, A.M., Baidoo, E.E., Rajeev, L., He, Z., Hendrickson, E.L., Joachimiak, M.P., Stolyar, S., Arkin, A.P., Leigh, J.A., et al. (2012). Functional responses of methanogenic archaea to syntrophic growth. ISME J. *6*, 2045–2055. https://doi.org/10.1038/ismej.2012.60.

37. Lloyd-Price, J., Arze, C., Ananthakrishnan, A.N., Schirmer, M., Avila-Pacheco, J., Poon, T.W., Andrews, E., Ajami, N.J., Bonham, K.S., Brislawn, C.J., et al. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature *569*, 655–662. https://doi.org/10.1038/s41586-019-1237-9.

38. Almeida, A., Mitchell, A.L., Boland, M., Forster, S.C., Gloor, G.B., Tarkowska, A., Lawley, T.D., and Finn, R.D. (2019). A new genomic blueprint of the human gut microbiota. Nature *568*, 499–504. https://doi.org/10.1038/s41586-019-0965-1.

39. Oliphant, K., and Allen-Vercoe, E. (2019). Macronutrient metabolism by the human gut microbiome: major fermentation by-products and their impact on host health. Microbiome *7*, 91. https://doi.org/10.1186/s40168-019-0704-8.

40. Tanca, A., Abbondio, M., Palomba, A., Fraumene, C., Manghina, V., Cucca, F., Fiorillo, E., and Uzzau, S. (2017). Potential and active functions in the gut microbiota of a healthy human cohort. Microbiome *5*, 79. https://doi.org/10.1186/s40168-017-0293-3.

41. Pacheco, A.R., Moel, M., and Segrè, D. (2019). Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems. Nat. Commun. *10*, 103. https://doi.org/10.1038/s41467-018-07946-9.

42. Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., et al. (2019). Extensive unexplored human microbiome diversity revealed by over 150, 000 genomes from metagenomes spanning age, geography, and lifestyle. Cell *176*, 649–662.e20. https://doi.org/10.1016/j.cell.2019.01.001.

43. Campanaro, S., Treu, L., Rodriguez-R, L.M., Kovalovszki, A., Ziels, R.M., Maus, I., Zhu, X., Kougias, P.G., Basile, A., Luo, G., et al. (2020). New insights from the biogas microbiome by comprehensive genome-resolved metagenomics of nearly 1600 species originating from multiple anaerobic digesters. Biotechnol. Biofuels *13*, 25. https://doi.org/10.1186/s13068-020-01679-y.

44. Stewart, R.D., Auffret, M.D., Warr, A., Walker, A.W., Roehe, R., and Watson, M. (2019). Compendium of 4, 941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. Nat. Biotechnol. *37*, 953–961. https://doi.org/10.1038/s41587-019-0202-3.

45. Ciuffreda, L., Rodríguez-Pérez, H., and Flores, C. (2021). Nanopore sequencing and its application to the study of microbial communities. Comput. Struct. Biotechnol. J. *19*, 1497–1511. https://doi.org/10.1016/j.csbj.2021.02.020.

46. Mendoza, S.N., Olivier, B.G., Molenaar, D., and Teusink, B. (2019). A systematic assessment of current genome-scale metabolic reconstruction tools. Genome Biol. *20*, 1158. https://doi.org/10.1186/s13059-019-1769-1.

47. Zimmermann, J., Kaleta, C., and Waschina, S. (2021). gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. Genome Biol. *22*, 81. https://doi.org/10.1186/s13059-021-02295-1.

48. Morris, B.E.L., Henneberger, R., Huber, H., and Moissl-Eichinger, C. (2013). Microbial syntrophy: interaction for the common good. FEMS Microbiol. Rev. *37*, 384–406. https://doi.org/10.1111/1574-6976.12019.

49. Westerholm, M., Moestedt, J., and Schnürer, A. (2016). Biogas production through syntrophic acetate oxidation and deliberate operating strategies for improved digester performance. Appl. Energy *179*, 124–135. https://doi.org/10.1016/j.apenergy.2016.06.061.

50. Jousset, A., Bienhold, C., Chatzinotas, A., Gallien, L., Gobet, A., Kurm, V., Küsel, K., Rillig, M.C., Rivett, D.W., Salles, J.F., et al. (2017). Where less may be more: how the rare biosphere pulls ecosystems strings. ISME J. *11*, 853–862. https://doi.org/10.1038/ismej.2016.174.

51. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

52. Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017). A new versatile metagenomic assembler. Genome Res. *27*, 824–834. https://doi.org/10.1101/gr.213959.116.

53. Bertrand, D., Shaw, J., Kalathiyappan, M., Ng, A.H.Q., Kumar, M.S., Li, C., Dvornicic, M., Soldo, J.P., Koh, J.Y., Tong, C., et al. (2019). Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. Nat. Biotechnol. *37*, 937–944. https://doi.org/10.1038/s41587-019-0191-2.

54. Wick, R.R., Judd, L.M., Gorrie, C.L., and Holt, K.E. (2017). Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput. Biol. *13*, e1005595. https://doi.org/10.1371/journal.pcbi.1005595.

55. Olm, M.R., Brown, C.T., Brooks, B., and Banfield, J.F. (2017). A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. *11*, 2864–2868. https://doi.org/10.1038/ismej.2017.126.

56. Uritskiy, G.V., DiRuggiero, J., and Taylor, J. (2018). MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome *6*, 158. https://doi.org/10.1186/s40168-018-0541-1.

57. Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ *7*, e7359. https://doi.org/10.7717/peerj.7359.

58. Wu, Y.-W., Tang, Y.-H., Tringe, S.G., Simmons, B.A., and Singer, S.W. (2014). MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome *2*, 26. https://doi.org/10.1186/2049-2618-2-26.

59. Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015). Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. *25*, 1043–1055. https://doi.org/10.1101/gr.186072.114.

60. Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2019). {GTDB}-tk: a toolkit to classify genomes with the genome taxonomy database. Bioinformatics *36*, 1925–1927. https://doi.org/10.1093/bioinformatics/btz848.

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

62. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for {RNA}-seq data with DESeq2. Genome Biol. *15*, 1550. https://doi.org/10.1186/s13059-014-0550-8.

63. Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinf. *11*, 119. https://doi.org/10.1186/1471-2105-11-119.

64. Machado, D., Andrejev, S., Tramontano, M., and Patil, K.R. (2018). Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Res. *46*, 7542–7553. https://doi.org/10.1093/nar/gky537.

65. Ebrahim, A., Lerman, J.A., Palsson, B.O., and Hyduke, D.R. (2013). COBRApy: constraints-based reconstruction and analysis for python. BMC Syst. Biol. *7*, 74. https://doi.org/10.1186/1752-0509-7-74.

66. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). Author correction: scipy 1.0: fundamental algorithms for scientific computing in python. Nat. Methods *17*, 352. https://doi.org/10.1038/s41592-020-0772-5.

67. Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., and Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. Front. Neuroinform. *8*, 14. https://doi.org/10.3389/fninf.2014.00014.

68. Wick, R.R., Judd, L.M., and Holt, K.E. (2019). Performance of neural network basecalling tools for oxford nanopore sequencing. Genome Biol. *20*, 1129. https://doi.org/10.1186/s13059-019-1727-y.

69. Sisk-Hackworth, L., and Kelley, S.T. (2020). An application of compositional data analysis to multiomic time-series data. NAR Genom. Bioinform. *2*, lqaa079. https://doi.org/10.1093/nargab/lqaa079.

70. Human Microbiome Jumpstart Reference Strains Consortium; Nelson, K.E., Weinstock, G.M., Highlander, S.K., Worley, K.C., Creasy, H.H., Wortman, J.R., Rusch, D.B., Mitreva, M., Sodergren, E., et al. (2010). A catalog of reference genomes from the human microbiome. Science *328*, 994–999. https://doi.org/10.1126/science.1183605.

71. Forster, S.C., Kumar, N., Anonye, B.O., Almeida, A., Viciani, E., Stares, M.D., Dunn, M., Mkandawire, T.T., Zhu, A., Shao, Y., et al. (2019). A human gut bacterial genome and culture collection for improved metagenomic analyses. Nat. Biotechnol. *37*, 186–192. https://doi.org/10.1038/s41587-018-0009-7.

72. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.

73. Klingenberg, H., and Meinicke, P. (2017). How to normalize metatranscriptomic count data for differential expression analysis. PeerJ *5*, e3859. https://doi.org/10.7717/peerj.3859.

74. Noronha, A., Modamio, J., Jarosz, Y., Guerard, E., Sompairac, N., Preciat, G., Daníelsdóttir, A.D., Krecke, M., Merten, D., Haraldsdóttir, H.S., et al. (2019). The virtual metabolic human database: integrating human and gut microbiome metabolism with nutrition and disease. Nucleic Acids Res. *47*, D614–D624. https://doi.org/10.1093/nar/gky992.

75. Goldansaz, S.A., Guo, A.C., Sajed, T., Steele, M.A., Plastow, G.S., and Wishart, D.S. (2017). Livestock metabolomics and the livestock metabolome: a systematic review. PLoS One *12*, e0177675. https://doi.org/10.1371/journal.pone.0177675.

76. Campanaro, S., Treu, L., Kougias, P.G., Luo, G., and Angelidaki, I. (2018). Metagenomic binning reveals the functional roles of core abundant microorganisms in twelve full-scale biogas plants. Water Res. *140*, 123–134. https://doi.org/10.1016/j.watres.2018.04.043.

77. Munro, L.J., and Kell, D.B. (2021). Intelligent host engineering for metabolic flux optimisation in biotechnology. Biochem. J. *478*, 3685–3721. https://doi.org/10.1042/bcj20210535.

78. Bar-Joseph, Z., Gifford, D.K., and Jaakkola, T.S. (2001). Fast optimal leaf ordering for hierarchical clustering. Bioinformatics *17*, S22–S29. https://doi.org/10.1093/bioinformatics/17.suppl_1.s22.

79. Kim, P.-J., Lee, D.Y., Kim, T.Y., Lee, K.H., Jeong, H., Lee, S.Y., and Park, S. (2007). Metabolite essentiality elucidates robustness of *escherichia coli* metabolism. Proc. Natl. Acad. Sci. USA *104*, 13638–13642. https://doi.org/10.1073/pnas.0703262104.

80. Rousseeuw, P.J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

81. Steiger, J.H. (1980). Tests for comparing elements of a correlation matrix. Psychol. Bull. *87*, 245–251. https://doi.org/10.1037/0033-2909.87.2.245.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Liquid samples from anaerobic digestion reactors	This paper	N/A
Critical commercial assays		
PowerSoil© DNA Isolation Kit	Mo Bio Laboratories	12888-50; 12888-100
Rapid sequencing kit	Oxford Nanopore Technologies	SQK-RAD004
Flow cell	Oxford Nanopore Technologies	FLO-MIN106D (R9)
Deposited data		
Long raw DNA sequencing reads	This paper	SRA: PRJNA814623
Short raw DNA and RNA sequencing reads	Zhu et al.[28]	SRA: PRJNA525781
Raw DNA and RNA sequencing reads and metabolite abundances for the human gut	Lloyd-Price et al.[37]	https://ibdmdb.org
Metagenome-assembled genomes and reference genomes for the human gut	Almeida et al.[38]	https://ftp.ebi.ac.uk/pub/databases/metagenomics/umgs_analyses/
Metagenome-assembled genomes, genome-centric metatranscriptomes, and other processed data and associated code	This paper	Zenodo: https://doi.org/10.5281/zenodo.7404716
Software and algorithms		
Guppy (v2.3.7 + e041753)	Oxford Nanopore	https://community.nanoporetech.com
Trimmomatic (v0.39)	Bolger et al.[51]	http://www.usadellab.org/cms/index.php?page=trimmomatic
metaSPAdes (v3.13.0)	Nurk et al.[52]	https://github.com/ablab/spades
OPERA-MS	Bertrand et al.[53]	https://github.com/CSB5/OPERA-MS
Unicycler (v0.4.8)	Wick et al.[54]	https://github.com/rrwick/Unicycler
MEGAHIT (v1.2.9)	Li et al., 2016	https://github.com/voutcn/megahit
dRep (v2.3.2)	Olm et al.[55]	https://github.com/MrOlm/drep
MetaWRAP (v0.8/1.3)	Uritskiy et al.[56]	https://github.com/bxlab/metaWRAP
MetaBAT (v2.12.1)	Kang et al.[57]	https://bitbucket.org/berkeleylab/metabat
Maxbin (v2.2.4/2.2.6)	Wu et al.[58]	https://sourceforge.net/projects/maxbin/
CheckM (v1.1.2)	Parks et al.[59]	https://github.com/Ecogenomics/CheckM
GTDB-Tk (v0.1.3)	Chaumeil et al.[60]	https://github.com/Ecogenomics/GtdbTk
Bowtie2 (v2.2.6)	Langmead and Salzberg,[61]	https://github.com/BenLangmead/bowtie2
DESeq2	Love et al.[62]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
CoPTR (v1.0.0)	Joseph et al.[34]	https://github.com/tyjo/coptr
ProDiGal (v2.6.3)	Hyatt et al.[63]	https://github.com/hyattpd/Prodigal
CarveMe (v1.4.1)	Machado et al.[64]	https://github.com/cdanielmachado/carveme
Memote (v0.13.0)	Lieven et al.[33]	https://github.com/opencobra/memote
Cobrapy (v0.20.0)	Ebrahim et al.[65]	https://opencobra.github.io/cobrapy/
Micom (v0.21.3)	Diener et al.[12]	https://github.com/micom-dev/micom
CoCo	This paper	Zenodo: https://doi.org/10.5281/zenodo.7404716
SciPy (v1.5.2)	Virtanen et al.[66]	https://github.com/scipy/scipy
Scikit-learn (v0.24.1)	Pedregosa et al.[67]	https://github.com/scikit-learn/scikit-learn
Scikit-bio (v0.5.6)	The scikit-bio development team	https://github.com/biocore/scikit-bio

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CPLEX Optimization Studio (v12.8.0)	IBM	https://www.ibm.com/products/ilog-cplex-optimization-studio
Other		
Nanodrop 2000	ThermoFisher Scientific	RRID: SCR_018042
MinION	Oxford Nanopore Technologies	RRID: SCR_017985

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Stefano Campanaro (stefano.campanaro@unipd.it).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Raw long-read DNA sequences generated in this study have been deposited at Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra/) under accession code SRA: PRJNA814623 and are publicly available as of the date of publication.
- All original code and the main associated processed data files have been deposited on GitHub (https://github.com/gzampieri/coco_paper) and are publicly available as of the date of publication. The DOI can be found in the key resources table.
- Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Anaerobic digestion biological system

The system was composed of three 1.8 L continuous stirred-tank reactors inoculated with digestate from biogas plants fed with 70-90% animal manure and 10–30% food industrial organic waste.[28] Reactors were fed with basal anaerobic medium (BA), with acetic acid as the only carbon source.

The process was divided into two stages. First, BA was the only feed until biogas production reached a steady state. At that point, hydrogen injection was activated, triggering a biological response that eventually adjusted on a new steady state. Liquid samples were taken during the two steady states and shortly after the hydrogen shock. Details over sample collection and sequencing were described by Zhu et al.[28]

METHOD DETAILS

DNA extraction and sequencing

Liquid samples were acquired from the triplicate reactors before, 18 h after, and 36 days after initiating hydrogen addition. For all the samples, the genomic DNA was extracted with PowerSoil © DNA Isolation Kit (Mo Bio Laboratories, Inc., Carlsbad, USA) with additional phenol cleaning steps to improve the extraction quality. Nanodrop 2000 (ThermoFisher Scientific, Waltham, MA) was used to evaluate the quality of the extracted DNA. Library preparation was performed using the SQK rapid sequencing kit (Oxford Nanopore Technologies, Oxford, UK) and the libraries were sequenced with a paired-end FLO-MIN106D R9 flow cell on a MinION device (Oxford Nanopore Technologies, Oxford, UK) at the CRIBI Biotechnology Center sequencing facility (University of Padova, Italy). Oxford Nanopore Technologies base-calling for translating raw electrical signals to nucleotide sequences was performed using Guppy (v2.3.7 + e041753).[68] Generated DNA sequences amount to 3′908′766′452 total bases and were deposited at Sequence Read Archive under accession code SRA: PRJNA814623.

Human gut microbiota data

Quality- and host-filtered DNA and RNA reads were downloaded from the Inflammatory Bowel Disease Multi'omics Database website (https://ibdmdb.org) along with untargeted metabolomics data. A total of five subjects diagnosed with Crohn's disease (CD) and five control subjects were selected based on their dysbiosis profile, so that CD patients presented a dysbiotic gut microbiome. The definition of dysbiosis is based on the similarity with a core set of microbiomes in terms of species abundance profile and is described in detail in the original publication.[37] Samples were taken from those having both metagenomics, metatranscriptomics, and metabolomics data available and such that they were a representative subset in terms of SCFA distributions. Metabolomic abundances

were centred-log-ratio transformed with scikit-bio v0.5.6 (https://github.com/biocore/scikit-bio) to eliminate compositionality effects.[69]

Genome-centric metagenomics

Long DNA reads obtained from the anaerobic reactors were co-assembled together with short reads previously generated with an Illumina platform as described before,[28] with the aim of increasing metagenome-assembled genome (MAG) quality. Low-quality short DNA reads were filtered out by Trimmomatic v0.39[51] and those passing the filter were assembled with multiple tools to further support MAG reconstruction. Metagenome assembly was thus performed in parallel with metaSPAdes v3.13.0,[52] OPERA-MS,[53] and Unicycler v0.4.8.[54] MetaSPAdes was executed specifying `-k 21,29,39,49,59,69,79,89,99,109,119,127`, while OPERA-MS and Unicycler (in bold mode) were run with a contig threshold length of 1 kb. All other arguments were left as default.

Starting from each of the three assemblies, DNA contig binning was executed through MetaWRAP,[56] including MetaBAT v2.12.1[57] and Maxbin v2.2.6[58] and discarding the contigs below 1 kb in length. Next, bin refinement was performed with parameters `-c 50 -x 10` so as to get a consolidated bin set for each assembly. Final MAG were obtained by dereplicating together the three MAG groups so as to remove redundant genomes while retaining those with the highest quality. In this process, the most complete and uncontaminated MAGs were selected based on a 95% average nucleotide identity threshold as estimated by dRep v2.3.2 using the following parameters: `-comp 50 -con 25 -sa 0.95 -nc 0.5`.[55] As a result, the final MAG set comprised 32 MAGs generated by metaSPAdes, 29 by OPERA-MS, and 8 by Unicycler. Their taxonomical classification was performed via GTDB-Tk v0.1.3,[60] while quality and completeness were estimated through CheckM v1.1.2 with default parameters.[59] To determine MAG coverage, short reads were first aligned on the concatenated genome sequences with Bowtie2 v2.2.6,[61] and the alignment results were processed by CheckM.

As regards the human gut microbiota, a collection of 1,952 high-quality MAGs and 553 reference genomes was collected from a previous study[38] where genomes from several studies were unified and dereplicated into representative species.[70,71] The presence and coverage of these 2,505 genomes in the considered samples was estimated as described above.

SAMtools v1.9/10 was used for all sam and bam file processing, including the procedures reported below.[72] DNA mapping rates for all the samples are provided in Table S1.

Genome-centric metatranscriptomics

RNA sequencing of anaerobic reactor samples was performed and described previously,[28] while human gut microbiota metatranscriptomes were retrieved along with the associated omic data. In both datasets, reads were aligned against concatenated gene sequences for all the genomes with Bowtie2 v2.2.6,[61] which were identified using Prodigal v2.6.3.[63] Alignment counts were summarised by SAMtools' `idxstats` command and later normalised with respect to library size with the `poscounts` estimator in DESeq2.[62] In this process, genes having total counts over all the samples equal to or less than 5 were excluded and normalised counts were transformed into a $\log_2$ scale. Moreover, count normalisation was independently performed for each taxon, as appropriate for metatranscriptomic data.[28,73] In contrast to whole-metatranscriptome normalisation, such an approach eliminates the variation associated with changing taxonomic abundances and better retains gene expression differences reflecting functional variation in individual taxa.

Microbial replication rate estimation

Microbial replication rate in each condition was obtained through CoPTR v1.0.0 utilising short reads.[34] Such tool allows an accurate peak-to-trough (PTR) estimation even with very low coverage. Compared to previous methods, CoPTR achieves a better accuracy and has more robust theoretical bases. CoPTR was run with default parameters in the case of anaerobic digestion genomes and by lowering the minimum number of samples required for each genome to 3 in the case of human gut genomes, in order to compensate for the higher compositional heterogeneity and estimate more values per sample.

Genome-scale metabolic model reconstruction

Starting from MAGs, draft genome-scale metabolic models (GEMs) were built with CarveMe v0.14.1.[64] We filtered poorly represented MAGs based both on their DNA and RNA abundance so as to select those community members that sensibly contribute to the overall activity. Specifically, we defined a multi-omic abundance as $a_m \cdot c_m$, where a_m is the relative abundance of MAG m and c_m is the fraction of its gene expression counts, log-scaled. A multi-omic abundance threshold was applied to select community members in each condition, while less abundant MAGs were excluded. Threshold values of 1 and 0.1% yielded a total of 28 and 289 unique taxa in anaerobic digestion and human gut communities, respectively, for which GEMs were built.

To obtain biologically meaningful GEMs, draft models were gap-filled so as to allow growth on relevant media based on MAG taxonomy. Following previous observations, archaeal MAGs were imposed growth on BA via hydrogenotrophic methanogenesis.[28] For *Methanosarcina*, we also required it to grow on BA with acetate and BA with methanol as carbon sources, in order to reflect its generalist metabolism.[28] Bacterial GEMs were instead gap-filled to prevent oxygen dependency under less strict assumptions. In the case of anaerobic digestion taxa, as bacterial MAGs comprised a wide range of metabolic capabilities, they were required to grow on a rich pool of metabolites representing a mix of animal manure and food waste, which correspond to the feed if the source biogas plant of the inoculum. Food waste was modeled by using the metabolites listed on the Virtual Metabolic Human (VMH) portal in the Nutrition

section.[74] Animal manure was assembled by taking all the metabolites provided by the Livestock Metabolome Database that were either identified or quantified in feces and urine.[75] The union of these compounds was used to represent such biogas plant feedstock for the gap-filling. For human gut taxa, an average western diet composition was assumed as compiled previously,[12,74] based on broad-level food intake questionnaire information collected for the considered subjects.[37] The composition of all growth media utilised is provided along with the code.

Further, we integrated the model of the *Limnochordia* species with transport and exchange reactions for acetate along with glycine reductase and serine dehydratase using COBRApy v0.20.0.[65] These integrations are supported by previous evidence of this taxon's acetate utilisation.[76]

All GEMs were benchmarked with MEMOTE v0.13.0, a test suite to identify gaps and inconsistencies in annotations and metabolic networks,[33] as shown in Figures S1 and S5.

Microbial community modeling

Microbial community metabolic modeling was based on MICOM v0.21.3, a recent approach that explicitly accounts for a varying degree of microbial cooperation.[12] Traditional flux balance analysis applied to multi-species models can be biased toward a few fast-growing members of a community, and it can be difficult to balance *a priori* individual growth rates. MICOM works around this issue by explicitly parametrising the level of cooperation through a user-defined parameter. The approach entails solving two sequential optimisation problems that jointly define the cooperative trade-off algorithm. First, the community growth rate is maximised, thus obtaining $\mu_c^{\max}$. Second, the following quadratic minimisation problem is solved:

$$\begin{aligned} &\text{minimise} \sum_i \mu_i^2 \\ &\text{such that } \mu_c \geq \alpha \mu_c^{\max} \\ &\text{and community constraints,} \end{aligned} \qquad \text{(Equation 1)}$$

where α is a parameter controlling the degree of cooperation among species. Community constraints include the following:

$$\begin{aligned} \mathbf{S}\quad \mathbf{v} &= 0 \\ \mu_i &\geq \mu_i^{\min} \\ \mathbf{v}_i^{lb} &\leq \mathbf{v}_i \leq \mathbf{v}_i^{ub} \\ \mathbf{v}_i^{lb,tr} &\leq \mathbf{v}_i^{tr} \leq \mathbf{v}_i^{ub,tr} \\ \mathbf{v}_i^{lb,ex} &\leq a_i \mathbf{v}_i^{ex} \leq \mathbf{v}_i^{ub,ex}. \end{aligned} \qquad \text{(Equation 2)}$$

In these equations, a_i represents the relative abundance of the community member *i*, $\mu_i^{\min}$ is a user-specified lower bound on μ_i, while $\mathbf{v}_i$, $\mathbf{v}_i^{lb}$, and $\mathbf{v}_i^{ub}$ are the fluxes and the respective lower and upper bounds. The superscripts *m* and *ex* flag exchange fluxes of community member *i* with the reactor and the external environment, respectively. Thus, relative abundance, which is commonly estimated in metagenomics analysis, can be directly used to rescale member-specific fluxes within the model, which were here calculated by setting `pFBA=True`.

On top of the stoichiometric constraints defined above, reactor working parameters and biochemical measurements were used to set community-level bounds for key compounds in anaerobic digestion models, thus well defining the boundary conditions for microbial community metabolism. For the two main steady states, acetate and hydrogen uptake rates were obtained by taking the set feeding rates and subtracting undigested amounts that were measured inside the bioreactors, whose concentration was verified to be constant over time. Similarly, biogas measurements were used to set methane and carbon dioxide production rates. The obtained rates were converted in $mmol/g_{DW}/h$ units by dividing by the total amount of volatile soluble solids in each process phase and their SD was used to define the exchange lower and upper bounds around mean values. In the case of SAH samples, biochemical and gas measurements immediately after the start of hydrogen injection were used. Additionally, a set of other metabolites were monitored over the process, including volatile fatty acids and ethanol, with no detected accumulation in the medium. We thus blocked the overall production of these compounds by setting the corresponding community-level exchange upper bound to 0.

Secondly, we set all the remaining uptake bounds so as to represent the nutritional conditions given by the BA composition in an anaerobic environment. Assuming that hydrogen was the compound with the most rapid consumption, we used its minimum uptake rate as a lower bound for all the BA components. Given that the bioreactors were fed also with small amounts of yeast extract and that cell death makes available a range of molecules, we allowed a minimal uptake rate for non-BA compounds with a lower bound of 1/100 the value for BA components. Oxygen, carbon dioxide and methane were excluded from this evaluation, and their net consumption was blocked.

As regards gut community models, only average western diet composition was used to impose community-level nutritional constraints. Associated bounds were taken from a previous study on gut metagenome modeling upon conversion and homogenisation of compound identifiers.[12] In the homogenisation process, we replaced unmappable compounds with equal molar amounts of the corresponding monomers where possible by using VMH metabolite definitions.[74]

All the simulations were performed in Python 3.6 with the CPLEX 12.8 solver (https://cplex.org). Obtained flux values lower than the numerical tolerance of the solver (10^{-6}) were set to zero.

Genome-centric metatranscriptomic data integration

Transcriptomics data are commonly used for creating condition-specific models (also referred to as context-specific models), where the condition may be defined by varying external stresses, genetic engineering modifications or given cell types.[23,24] In GEMs, gene-protein-reaction (GPR) rules define the relationship between metabolic reactions and their enzymes' activity, which in turn depends on gene expression. Here, GCM counts were used to quantify gene expression levels and, taking the mean expression over all the samples as a reference, to estimate the corresponding expression fold changes. Following previous work,[31] GPR expressions were evaluated as minimum/maximum operations as follows:

$$\begin{aligned}\Theta(g_{i,1}\wedge g_{i,2}) &= \min\{\theta(g_{i,1}),\theta(g_{i,2})\}\\ \Theta(g_{i,1}\vee g_{i,2}) &= \max\{\theta(g_{i,1}),\theta(g_{i,2})\},\end{aligned} \quad \text{(Equation 3)}$$

where $\theta(g_{i,1})$ represents the expression fold change of gene $g_{i,1}$ in community member i and $\Theta(g_{i,1}\wedge g_{i,2})$, $\Theta(g_{i,1}\vee g_{i,2})$ identify the effective expression fold change of gene set $\{g_{i,1},g_{i,2}\}$. These expressions were recursively evaluated until a single value $\Theta_{i,r}$ was obtained, representing an effective gene-set expression fold change for reaction r in community member i, which was used to alter the respective flux bounds. For any genes undetected by GCM, a fold change equal to 1 was employed so as to avoid regulatory assumptions altogether.

To map gene-set expression changes onto flux bounds, we generalised a previous method called METRADE (MEtabolic and TRanscriptomics ADaptation Estimator).[31] METRADE considers expression fold-changes to adjust flux bounds according to the following non-linear map:

$$h(\Theta_{i,r}) = (1+\gamma\,|\log(\Theta_{i,r})|)^{\mathrm{sign}(\Theta_{i,r}-1)}, \quad \text{(Equation 4)}$$

where γ is a user-specified parameter that controls the impact of gene-set expression changes on the flux bounds. This map accounts for saturation phenomena at high levels of expression, dampening the effect of elevate up-regulation.[77] However, metatranscriptomics profiles generally have highly variable count distributions, which only partially reflect species abundance and in general impairs the reliability of such profiles.

Starting from Equation 4, we thus introduced a modified map that takes into account coverage heterogeneity across different GCM profiles. Specifically, we assumed that the more uniform expression coverage is across MAGs, and the more gene expression changes should have even an impact across community members. On the contrary, the less uniform the coverage, the less should low-coverage members be affected by gene expression changes, because of the uncertainty bound to estimating such changes at low coverage. To model this effect, we calculated the ratio between each member's transcript count sum c_i and the maximum transcript count sum over all community members $c_{\max}$, hereby denoted as $\nu = c_i/c_{\max}$. Such a ratio was used to weigh the impact of gene-set expression changes across members as by the following relation:

$$h(\Theta_{i,r}) = \nu_i(1+\gamma\,\nu_i\,|\log(\Theta_{i,r})|)^{\mathrm{sign}(\Theta_{i,r}-1)} + (1-\nu_i). \quad \text{(Equation 5)}$$

In this way, we limited the flux bound interval reachable by reactions in low-coverage members while also tuning the effect of γ in a MAG-specific manner. Additionally, given that the magnitude of transcriptional activity can vary even between equally-abundant community members, in principle more active species should be able to achieve larger fluxes and vice versa. Therefore, we introduced MAG-specific reference bounds that are modulated by a scaling factor δ based on the transcript count sums c_i, as follows:

$$\begin{aligned}d^{lb}_{i,r} &= \begin{cases}-\delta\log(c_i+1) & \text{if } r \text{ has GPR},\\ -1000 & \text{otherwise}.\end{cases}\\ d^{ub}_{i,r} &= \begin{cases}\delta\log(c_i+1) & \text{if } r \text{ has GPR},\\ 1000 & \text{otherwise}.\end{cases}\end{aligned} \quad \text{(Equation 6)}$$

By considering the combined effect of the gene-level map $h(\Theta_{i,r})$ (Equation 5) and of the MAG-level coefficients $d^{lb}_{i,r}, d^{ub}_{i,r}$ (Equation 6), the resulting flux bounds for any community member i were thus obtained as follows:

$$\begin{aligned}\mathbf{v}^{lb}_i &= \mathbf{d}^{lb}_i \odot \mathbf{h}_i\\ \mathbf{v}^{ub}_i &= \mathbf{d}^{ub}_i \odot \mathbf{h}_i,\end{aligned} \quad \text{(Equation 7)}$$

where $\odot$ identifies the Hadamart product between any pair of arrays $\mathbf{d}^{lb}_i$, $\mathbf{d}^{ub}_i$, and $\mathbf{h}_i$ including the coefficients for all the member's reactions.

In general, parameters γ and δ can be varied and optionally fine-tuned with the use of independent measurements, such as replication rates for individual microbial species.[34] Here, we explored ranges of values by grid search, studying the Pearson's correlation between predicted growth rates and reference replication rates, obtained as described above. In anaerobic digestion communities, for δ we explored the values $\{1,2,\ldots,10\}$, while for γ we considered the values $\{1,2,\ldots,7\}$. In the human gut communities, we instead

tested the range $[0.001, 1]$ and the values $\{1, 1.5, 2\}$, respectively. In practice, we found that $\gamma = 1$ tends to generally work well, whereas δ has a more severe impact on the results. Such a tuning strategy also has the advantage of allowing a better harmonisation with absolute flux measurements, like community-level metabolite exchange rates, if available.

Multivariate flux analysis

PCA was used to visualise the temporal trajectory of the microbial communities in terms of their net metabolite consumption and production rates, while t-SNE was utilised to represent microbial metabolic activity across conditions. These methods were applied via scikit-learn v0.24.1 (https://github.com/scikit-learn/scikit-learn).[67] Amino acid exchanges were clustered by using hierarchical clustering with the complete method and Euclidean distance as implemented in SciPy v1.5.2 (https://github.com/scipy/scipy).[66,78]

QUANTIFICATION AND STATISTICAL ANALYSIS

The concordance between microbial proliferation rates and predicted growth rates was evaluated by Pearson's correlation, as well as the agreement between microbial export rates and metabolomic abundances in the human gut. In the latter case, the underlying assumption was that increasing export rates across a community would increase the likelihood of detecting an extracellular metabolite in larger abundance. This assumption is essentially analogous to the flux sum approach.[79] The separation between fluxomic profiles belonging to individual species were quantified in terms of the silhouette coefficient, which considers the mean intra-cluster distance and the mean nearest-cluster distance for any given sample.[80] The differences between silhouette coefficient distributions were evaluated by two-tailed Wilcoxon signed-rank tests, while the differences between gas production rates were assessed by two-tailed t-tests for dependent samples. All these metrics and tests were applied as implemented in SciPy v1.5.2 (https://github.com/scipy/scipy)[66] and scikit-learn v0.24.1 (https://github.com/scikit-learn/scikit-learn).[67] We assessed the significance of the difference between Pearson's correlations by using a custom Python implementation of the Steiger method for dependent groups,[81] available along with the rest of the code.