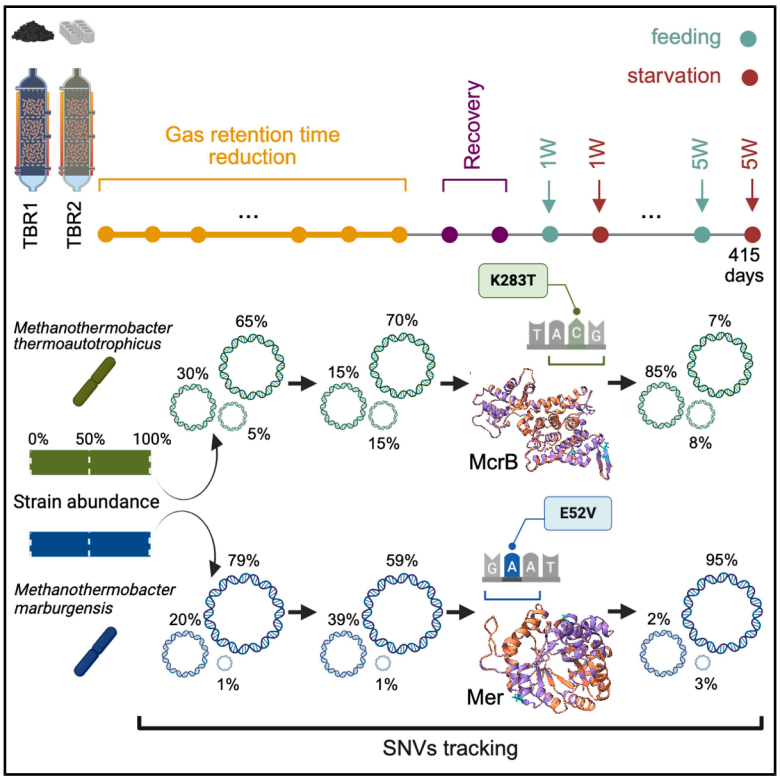
Cell Genomics

Article

# Exploring genetic adaptation and microbial dynamics in engineered anaerobic ecosystems via strain-level metagenomics

## Graphical abstract




## Highlights

- Strain-resolved genomics reveals adaptive population sweeps during $CO_2$ fixation
- SNV trajectories uncover strain replacement triggered by positive selection
- Evolutionary shifts correlate with starvation and increased gas supply turnover
- Amino acid changes in *mer* and *mcrB* genes finely tune methanogenesis efficiency

## Authors

Gabriele Ghiotto, Aikaterini Xirostylidou, Maria Gaspari, Panagiotis G. Kougias, Stefano Campanaro, Laura Treu

## Correspondence

gabriele.ghiotto@phd.unipd.it (G.G.), stefano.campanaro@unipd.it (S.C.)

## In brief

Ghiotto et al. track adaptive evolution over a year in a carbon dioxide-fixing anaerobic microbiota, leveraging strain-resolved metagenomics. They reveal long-term haplotype sweeps in *Methanothermobacter* species and identify *mer* and *mcrB* as potential drivers of strain-level competition, offering insights for engineering microbial communities into methane-based carbon bioconversion systems.

Ghiotto et al., 2025, Cell Genomics *5*, 100949
September 10, 2025 
https://doi.org/10.1016/j.xgen.2025.100949

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Article

# Exploring genetic adaptation and microbial dynamics in engineered anaerobic ecosystems via strain-level metagenomics

Gabriele Ghiotto,[1,3,*] Aikaterini Xirostylidou,[2] Maria Gaspari,[2] Panagiotis G. Kougias,[2] Stefano Campanaro,[1,*] and Laura Treu[1]

[1]Department of Biology, University of Padua, Via U. Bassi 58/b, 35131 Padova, Italy
[2]Soil and Water Resources Institute, Hellenic Agricultural Organisation Dimitra, Thermi, 57001 Thessaloniki, Greece
[3]Lead contact
*Correspondence: gabriele.ghiotto@phd.unipd.it (G.G.), stefano.campanaro@unipd.it (S.C.)
https://doi.org/10.1016/j.xgen.2025.100949

## SUMMARY

Genetic heterogeneity exists within all microbial populations, with sympatric cells of the same species often exhibiting single-nucleotide variations that influence phenotypic traits, including metabolic efficiency. However, the evolutionary dynamics of these strain-level differences in response to environmental stress remain poorly understood. Here, we present a first-of-its-kind study tracking the adaptive evolution of an anaerobic, carbon-fixing microbiota under a controlled engineered ecosystem focused on carbon dioxide bioconversion into methane. Leveraging strain-resolved metagenomics with an ad hoc variant calling and phasing approach, we mapped mutation trajectories and observed that the two dominant *Methanothermobacter* species maintained distinct sweeping haplotypes over time, most likely due to niche-specific metabolic roles. By combining population genetic statistics and peptide reconstruction, *mer* and *mcrB* genes emerged as potential drivers of archaeal strain-level competition. These findings pave the way for targeted engineering of microbial communities to enhance bioconversion efficiency, with significant implications for sustainable energy and carbon management in anaerobic systems.

## INTRODUCTION

Carbon-fixating microbiomes harbor a reservoir of untapped resources that remain hidden within the natural ecosystem.[1] A primary challenge in investigating anaerobic microorganisms that metabolize simple carbon-based compounds arises from their prevalence across diverse biomes, typically within harsh and inaccessible niches (i.e., deep vents, peatlands, etc.).[2–4] Consequently, the need for more controlled systems becomes evident, facilitating the effective modeling of ecological processes and elucidating the underlying microbial dynamics. Synthetic cultivation platforms establish controlled ecosystems where tailored laboratory conditions drive microbial metabolism toward enhanced production of high-energy compounds.[5–8] Leveraging recent advancements in sequencing technologies, novel cultivation-independent molecular methods, such as metagenomics, have emerged over the past two decades.[9] Notably, shotgun sequencing enables direct analysis of genomes and genes sourced from microbial populations, facilitating the extraction of crucial functional insights.[10,11]

Nonetheless, genome-centric metagenomics and the extraction of metagenome-assembled genomes (MAGs) fall short of fully harnessing the intra-population diversity within a microbiome, as its resolving capacity is mainly limited to the species level. Microbial genetic heterogeneity extends beyond this boundary, influenced by *de novo* mutations generating new haplotypes within the populations.[12] Thus, the composition of a microbiome can vary substantially at the strain level, and these shifts are thought to play an important role in the stability and efficiency of microbial metabolism. Understanding how this microbial ecosystem changes from week to week through periods of growth, stress, and treatment is crucial for personalized bioprocess management. Variant frequencies fluctuate over time due to genetic drift or selection processes driven by specific environmental pressures, shaping microbiome evolution. The newly emerged strains, characterized by their unique genetic makeup, may display enhanced metabolic capabilities due to advantageous mutations that rapidly increase in frequency and become predominant in the population through positive selection.[13] At the opposite end of the spectrum is purifying selection, a mechanism that maintains the integrity of functional sequences by eliminating deleterious mutations.[14] Although short-read sequencing does not enable the precise determination of complete haplotypes within a microbial population, novel computational methodologies made possible the reliable identification of genetic variants and their frequencies across constituent species.[15–17] Through the tracking of species-level single-nucleotide variants (SNVs), recent investigations into the gut microbiome have unveiled its propensity for genetic variations over time, even among healthy human hosts.[18,19] These differences

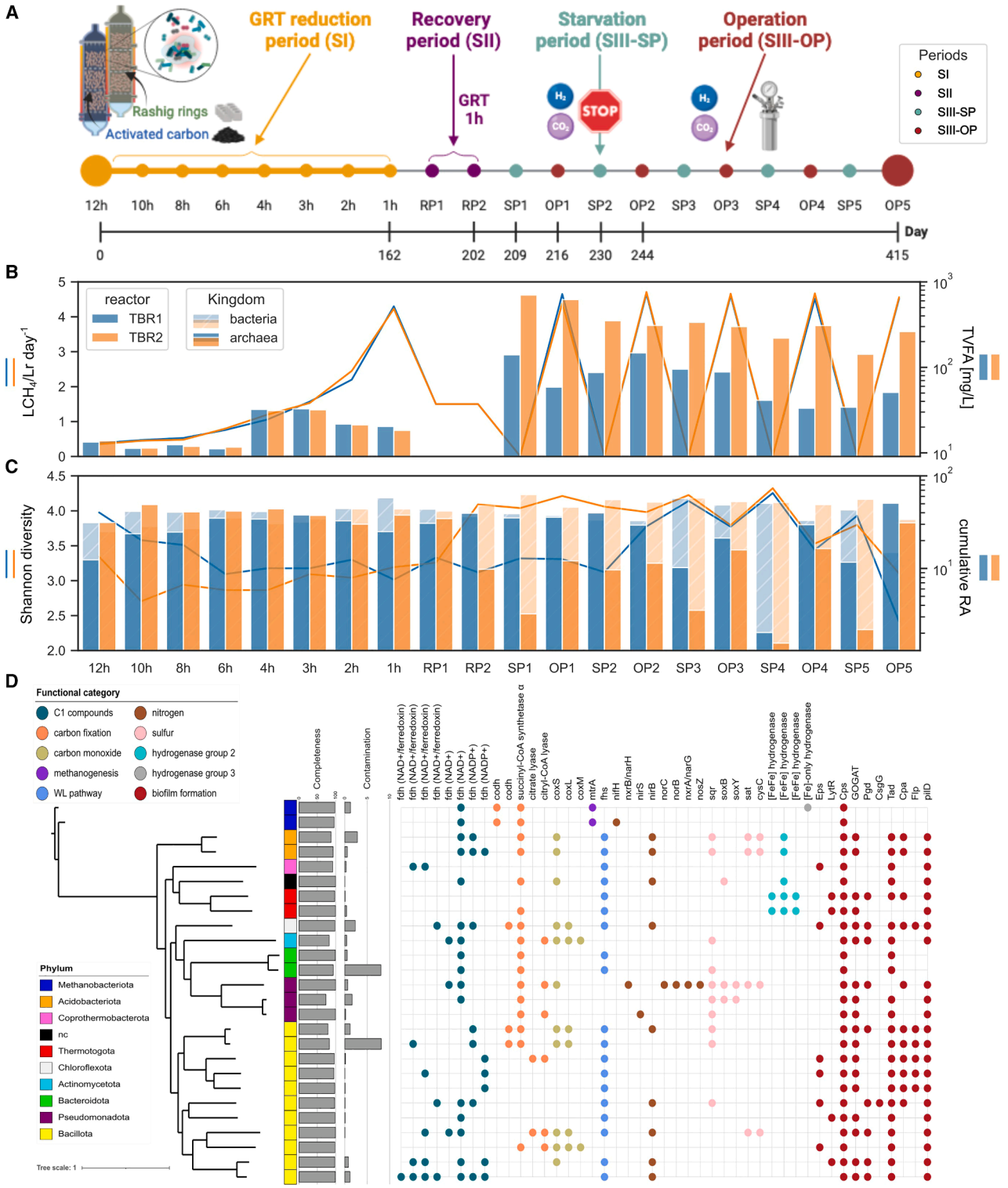



**Figure 1. Phylogenetic composition and functional potential of the microbiome in two vessels packed with carbon particles (TBR1) and a plastic pellet (TBR2)**

(A) Study design. Shotgun metagenomic sequencing was performed on 40 liquid samples collected during a 1-year continuous operation period. The sequential experimental periods are highlighted as follows: SI (stage I; yellow), SII (stage II; purple), SIII-SP (stage III starvation; red), and SIII-OP (stage III operation; green).

*(legend continued on next page)*

may arise from a combination of external replacement events and the evolutionary dynamics of resident strains.[20] SNVs have been suggested to establish strain-level hierarchies within the same species by altering the 3D structure of key metabolic enzymes,[21,22] thereby influencing the relative fitness of different haplotypes. However, while research on the gut microbiota is driven by the need for personalized responses to drugs or therapies, the population genetic processes underpinning these strain-level dynamics remain poorly characterized in other natural environments. A preliminary study on methanogenic microbiomes revealed strain-level SNV dynamics over a short time period but lacked structural evidence regarding the impact on the proteins encoded by genes under selective pressure.[23] Understanding the mechanisms of mutation, selection, and genetic drift is crucial for predicting ad hoc strategies aimed at optimizing the biotechnological potential inherent in microbiologically relevant processes in nature, such as anaerobic carbon dioxide ($CO_2$) reduction.

To address this gap, this study employed deep shotgun metagenomic sequencing and a tailored SNV tracking approach to monitor the evolution of microbial dynamics within an anaerobic mixed-methanogenic culture. By cultivating the microbiome under thermophilic conditions across two controlled systems, the bioconversion of $CO_2$ into methane ($CH_4$) was explored. Genetic heterogeneity was accessed over a 1-year continuous adaptive laboratory evolution (ALE), examining responses to variations in gaseous substrate flow rate and to carbon and electron source deprivation. The SNV trajectories of key microbial players were reconstructed, enabling the tracing of the evolution of methanogenesis-related genes within the archaeal haplotype population upon imposed stress periods. Increasing the gas supply turnover, as previously hypothesized,[24–26] was shown to be a strong selective pressure, driving the evolution of haplotypes within the existing microbial population. This approach provides a valuable framework for understanding strain-level adaptive genetic mechanisms in anaerobic microbiomes, with implications for optimizing $CH_4$ production and advancing microbial applications in bioenergy and carbon cycling.

## RESULTS

### Community stratification of an anaerobic $CO_2$-fixating microbiome

Genome-resolved metagenomics was employed to study the microbiome in two anaerobic vessels entirely filled with small plastic or carbon support structures. Such vessels are commonly referred to as trickle bed reactors (TBRs) and will hereafter be designated as TBR1 and TBR2. The system was maintained at 55°C under anaerobic conditions for over a 1-year (415-day) period, which encompassed three sequential experimental stages (Figure 1A; see Table S1). First, the community was pushed toward higher bioconversion efficiency through a stepwise increase of the $H_2$ and $CO_2$ supply flow, a process referred to as gas retention time (GRT) reduction (stage I [SI]). Secondly, a 40-day recovery period was employed, where the microbiota were kept at a constant feeding stream of 1-h GRT to favor acclimation and stability after prolonged stress (stage II [SII]). Lastly, stepwise operation (OP) and starvation (SP) periods were tested. Specifically, the response to transient carbon and electron source deprivation events was explored by alternating progressively longer periods of starvation (stage III [SIII-SP]) and feeding (stage III [SIII-OP]), simulating a real-life scenario where $H_2/CO_2$ availability fluctuates. Biomethane daily production ($LCH_4/Lr\ day^{-1}$) and volatile fatty acid (VFA) concentrations were continuously monitored as proxies for microbial activity (Figure 1B).

Microbiome investigation was performed following a co-assembly strategy, and the resulting bins were clustered at 95% average nucleotide identity, allowing us to separate taxa at the species level. Ultimately, a total of 330 medium-to-high-quality MAGs were recovered, which were divided into 17 archaeal and 313 bacterial genomes (see Table S2). The reconstructed MAGs provide a reliable snapshot (see Figure S1A), as profiling of single-copy marker genes identified only 14 species in the unmapped reads (see STAR Methods). The microbiomes underwent substantial perturbation in species-level composition during ALE, especially when comparing the starvation periods with the continuous feeding provision (Figure 1C). The liquid-phase microbial dynamics can be considered representative of the biofilm microbiome, as >70% of reads from triplicate biofilm samples collected at the final time points in both cultivation vessels aligned to the MAGs (see Figure S1B). With the goal of applying population-based analyses to the dominant species-level lineages, only the species with a relative abundance (RA) above 2% in at least one time point of either cultivation vessel were retained for a total of 26 MAGs (Figure 1D). The dominant microbiome fraction was stratified into 26 phyla and governed by two hydrogenotrophic archaea: *Methanothermobacter marburgensis* GRT2.8 and *Methanothermobacter thermoautotrphicus* GRT1.1. The same methanogens were dominant in both liquid and biofilm at the conclusion of the experiment, alongside bacterial taxa such as *Mycobacterium* sp. STV1.135, Bacillota sp. STV1.178, and *Coprothermobacter proteolyticus* STV2.316 (see Figure S1B). While the RA of the two methanogens fluctuated during the experiment, with *M. thermoautotrphicus* GRT1.1 emerging as the winner in the substrate competition, a return to near baseline values by the end of the study was observed (see Table S2). On the other hand, the bacterial fraction of the microbiome was far more plastic. During SI, three bacteria increased progressively in RA, namely Thermotogales sp. GRT1.19, *Candidatus* Saccharicenans sp. STV2.192, and an unclassified species named Bacteria sp. GRT2.72 (see Table S2). These co-occurring species declined over 10-fold during the starvation periods in spite of the rise of new bacterial players. Although these shifts in RA suggest selection, their interpretation requires caution due to the compositional nature of the data, which may not fully reflect changes in total biomass or selection strength.

(B) Line plot of average $CH_4$ production, expressed as $LCH_4/L_r$ produced daily, and bar plot of cumulative VFA (acetate, propionate, and butyrate) in mg/L.
(C) Species-level composition indicated as α-diversity (Shannon index) in the line plot and bar plot of cumulative RA for archaea and bacteria.
(D) Functional profiles of the dominant 26 MAGs with RA > 2% in at least one sample point. Phylogenetic composition and gene-level mining of 45 marker genes were employed to characterize functional guilds based on selected metabolisms of interest.

A gene-level survey expanded the analysis, corroborating insights from phylogenetic classification and identifying specific microbial guilds within the microbiomes. Many taxa showed metabolic deficiencies, which may suggest potential syntrophic partnerships (Figure 1D; see Table S3), as already proposed in previous studies.[27] However, this could also result from incomplete genome reconstruction, despite most MAGs having a completeness of over 90% (Figure 1D). Nonetheless, *Candidatus* Saccharicenans sp. MAGs (phylum Acidobacteria, RA: 5%) were characterized by a wide range of functional potential, including the use of $H_2$ and reduced forms of sulfur compounds as electron donors.[28] Given the nature of the system setup, which relies on the extracellular polymeric substance (EPS) coating, the contribution of the dominant microbiome fraction to biofilm matrix formation was further investigated. The high abundance of *Candidatus* Saccharicenans sp. STV2.192 and Anaerolineae sp. STV2.214 during SI (see Figure S2), along with their genetic background, suggests a pivotal role in EPS synthesis. Although the *eps* operon is absent, both taxa possess the *cps* genes, which collectively facilitate the biosynthesis and export of capsular polysaccharides (Figure 1D). Additionally, the detection of the *tad* operon in Anaerolineae sp. STV2.214, containing genes involved in surface adhesion as well as the *cpa/rcp* pilus assembly genes, further supports the proposed roles of these species in biofilm formation during the initial operational stages.

## Strain-resolved metagenomics reveals shifts in populations genetic composition

The population of 330 reconstructed MAGs was further characterized for variant detection. Detected SNVs, both nonsynonymous and synonymous, were subjected to stringent filtering to enhance confidence in variant calling following a previously established procedure.[29] Considering that MAGs were derived from metagenomic samples of two distinct cultivation systems, each representing a discrete "ecosystem within a vessel," investigations were conducted to assess whether the use of different substrates for supporting the biofilm formation influenced microbial growth and evolution. Observed nucleotide diversity among selected microbiome constituents exhibited notable discrepancies between the two engineered systems (see Figure S2A), displaying values ranging from 0 to 0.007. Notably, *M. marburgensis* sp. GRT2.8 demonstrated the most pronounced significance ($p = 2.2e{-}06$), underscoring the impact of packing materials on the adaptability of methanogens, either enhancing or impairing their fitness (see Figure S2A). Thus, each vessel was analyzed separately to identify relationships between microdiversity and operational parameter alteration, facilitating the tracking of genetic variations within species over time.

Across the sampling time series, cumulative totals of 318,284 and 278,339 high-confidence SNVs from the 26 MAG datasets were identified within the TBR1 and TBR2, respectively. It was evident that the initial levels of genetic diversity exhibited considerable variation across species (Figures 2A–2D). The dominant hydrogenotrophic archaea, *M. marburgensis* sp. GRT2.8 and *M. thermoautotrphicus* sp. GRT1.1, harbored more than ~1,000 and 2,400 nonsynonymous SNVs at intermediate frequencies ($0.1 < f < 0.9$). On the other hand, several bacterial members, including *Fervidobacterium pennivorans* STV1.130 and Tissierellaceae sp. STV2.313, had fewer than ~250 detectable SNVs (or even 0). Of particular interest are SNVs that underwent remarkable changes in frequency between the initial and later time points (e.g., from $0.15 < f$ to $f > 0.75$), henceforth referred to as sweeping SNVs (see STAR Methods). These events indicate a nearly complete sweep within the species of interest where the new allele becomes fixed, as proposed by Rodgar and colleagues.[30] Throughout SI, three species underwent more than 100 frequency shifts, including *M. marburgensis* sp. GRT2.8. Similarly, there was a broad range in the total of sweeping SNVs cataloged during intermittent feeding treatment, ranging from ~400 in certain species (e.g., Bacillota sp. STV1.136) to ~10, or even none, in others (Figure 2B). Finally, among the nonsynonymous SNVs present at the initial time points, a variable fraction is also tracked with a frequency exceeding 0.6 at the end of SI and SIII. It is noteworthy that in this regard, the two archaea exhibit divergent behaviors. While *M. thermautotrophicus* GRT1.1 retains 4% of these SNVs, the percentage retained by *M. marburgensis* sp. GRT8 at the end of SI is around 16% (Figure 2C). Conversely, at the end of the fifth week of starvation, the dynamics are reversed, with nearly identical values of SNVs fixed.

The same fluctuating trends were identified in TBR2, except for *M. marburgensis* sp. GRT8, whose RA sharply declined during the starvation phase. A large spectrum of genetic heterogeneity was evidenced, with *Mycobacterium* sp. STV1.135 and *M. thermautotrophicus* GRT1.1 exhibiting the highest numbers of detected SNVs. Although shorter GRT did not cause an extensive accumulation of sweeping SNVs, several species exceeded the threshold of ~200 during the intermittent feeding (Figure 2E). This includes *M. thermautotrophicus* GRT1.1 and *C. proteolyticus* STV2.316, which exhibit fluctuations in RA corresponding to the feeding regimes (Figure 2E). As expected, the archaeon predominates when $H_2$ and $CO_2$ are provided, whereas the proteolytic species takes over during microbiome starvation. The considerable number of SNVs accumulated by these microbes was also partly retained at a high frequency at later stages. Specifically, 21% and 14% of the SNVs observed during the 1-week starvation for *M. thermautotrophicus* GRT1.1 and *C. proteolyticus* STV2.316, respectively, exceeded 0.6 in the populations (Figure 2F).

Comparing SNV density across species during experimental stages SI and SIII-SP reveals correlations with environmental changes. Specifically, during the progressive GRT reduction, the two dominant *Methanothermobacter* species, *Candidatus* Saccharicenans sp. GRT2.83 and *Mycobacterium* sp. STV1.135, exhibited SNV densities greater than 1,500 SNVs/Mbp (Figures 2A and 2D). Statistical comparison using the Mann-Whitney U test revealed significant differences in genomic plasticity between these species and the rest of the microbiome across both average ($p_{TBR1} = 0.00052$ and $p_{TBR2} = 0.03110$) and individual time points (see Figure S2B). Under starvation conditions, the two archaea, along with *C. proteolyticus* STV2.316 and Limnochordia sp. STV2.225, displayed the highest SNV densities, which were statistically significant ($p_{TBR1} = 0.00148$ and $p_{TBR2} = 0.00414$) when compared to the rest of the microbiome (see Figure S2B). These findings further support the hypothesis of ecological resilience acquired through ALE, particularly for the archaeal species involved in the bioconversion process.

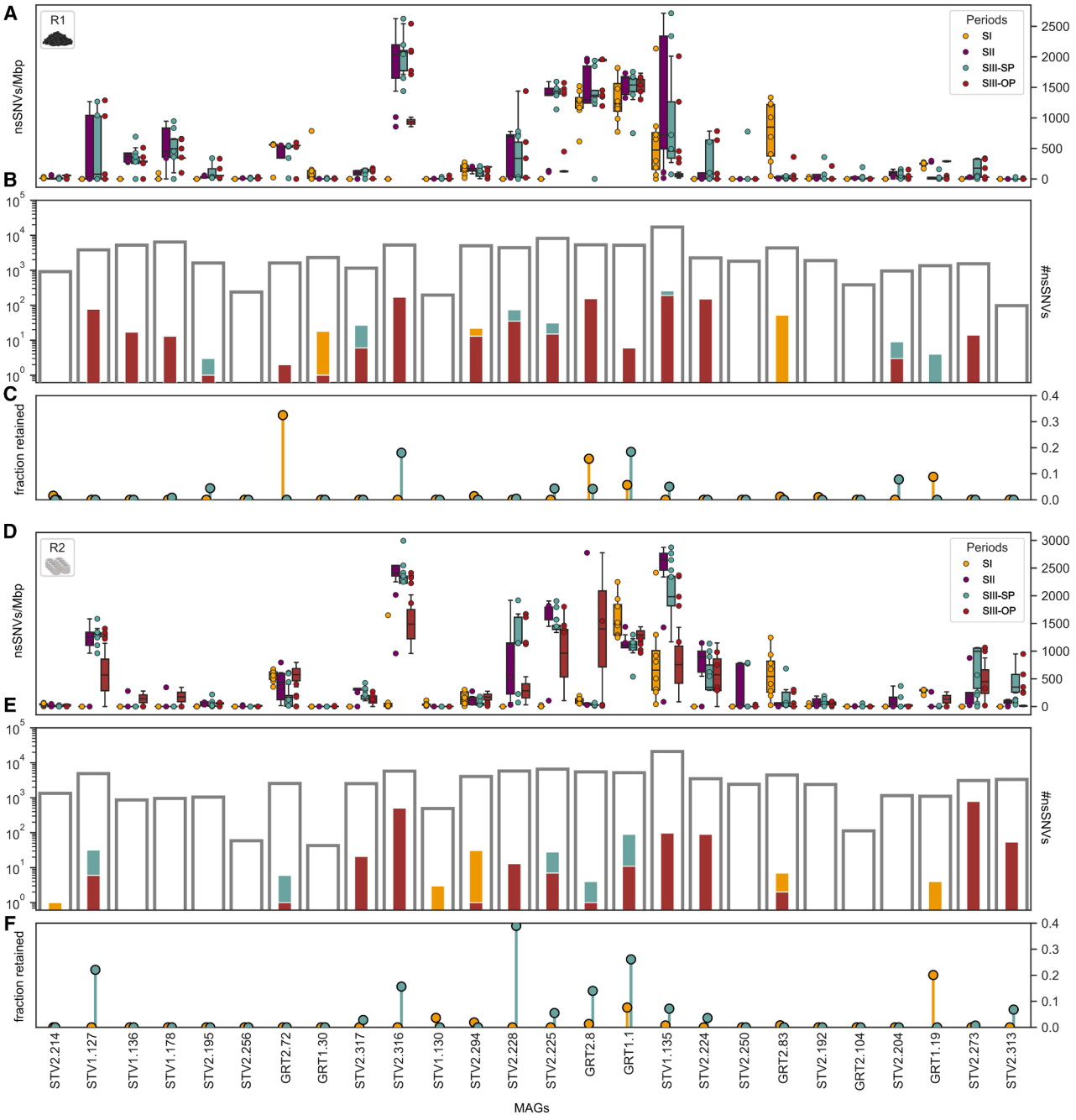



**Figure 2. Varied ecological and genetic responses across 26 abundant species in TBR1 and TBR2**

Varied ecological and genetic responses across 26 abundant species in TBR1 (A–C) and TBR2 (D–F).

(A and D) Within-species nucleotide diversity for each time point, as measured by the fraction of core genome sites with intermediate allele frequencies ($0.1 < f < 0.9$) (see STAR Methods). Each sample is indicated by a point, and the samples from the same epoch are connected by a boxplot to aid in visualization. The sequential experimental epochs are highlighted as follows: SI (stage I; yellow), SII (stage II; purple), SIII-SP (stage III starvation; red), and SIII-OP (stage III operation; green).

(B and E) The total number of sweeping SNVs between a baseline time point and each of the later epochs (see STAR Methods). The height of the white area indicates the total number of polymorphic nonsynonymous SNVs that were tested for temporal variation.

(C and F) Fractions of sweeping SNV that are retained at the final time point ($f > 0.5$). In many species, only a minority of these SNVs gained during SI or SIII are retained.

### Sweeping SNVs reveal strain replacement, evolutionary modification, and selection

The variability in genetic responses cannot be explained by species phylogeny or abundance fluctuations only. No correlation was found between genetic alterations and changes in RA; thus, the source of these sweeps remains unclear. SNV frequencies fluctuated across different time points (Figures 3A–3F). To account for these dynamics, a phasing approach was used to cluster SNVs based on their frequency across the entire experimental period. Nonsynonymous SNVs were selected to separate natural selection from demographic stochasticity, avoiding masking the effect of causative "driver" variations. This technique enabled the simultaneous monitoring of nonsynonymous SNV trajectories within 26 different species, meeting our minimum coverage thresholds (see STAR Methods). The resulting species-level patterns revealed whether the evolutionary dynamics either mirrored or diverged from the trends in their corresponding species abundance. Additionally, the phased haplotypes were confirmed by applying a strain deconvolution algorithm on the MAGs and reconstructing the dominant fraction of the microbiome at the strain level. As an example, Figure 3 illustrates the range of reported behaviors for RA, SNV frequency trajectories, and strain deconvolution for four representative species.

The *M. thermautotrophicus* population exemplifies a strain replacement event, showcasing sweeping haplotypes encompassing over 5,000 nonsynonymous SNVs dispersed throughout the genome and organized into three distinct clusters. Less than 7% of these SNVs persisted from SI to SIII (Figures 2C–2F), supporting the hypothesis of replacement by a genetically distant strain. Specifically, a group of SNVs rapidly became fixed in the populations ($f > 0.75$) and decreased in frequency only later on, during SII. Notably, the SNV trajectories and strain-level abundances depicted in Figures 3A and 3B indicate that the genetically distant strain had already reached substantial frequencies (~10%–25%) well before its substantial fitness advantage became apparent. Furthermore, Figures 3E and 3F illustrate a comparable pattern in the *C. proteolyticus* population, where multiple coexisting strains evolved during the intermittent feeding phase. These findings suggest that the identified replacement events stemmed from a rapid proliferation of pre-existing strains rather than new strains arising from *de novo* SNV accumulation.

On the other side of the spectrum, Haloplasmataceae sp. STV2.228 provides an example of an evolutionary modification event (Figure 3H). In this case, a cluster of SNVs suddenly increased in frequency during 4- and 5-week operational periods, nearly reaching fixation ($f > 0.75$). Unlike the strain replacement mentioned above, this sweep shared nearly 40% of the phased SNVs from the dominant haplotypes at the initial time point (Figure 2F), suggesting that the two strains recently descended from a common ancestor. Beyond these extreme cases, we also observed a third category of events where the minority haplotype was already segregating at substantial frequencies (~5%–20%) before treatment. This pattern suggests that frequency-dependent selection may have initially driven these mutant lineages to intermediate frequencies, keeping them genetically stable until substrate depletion or other environmental shifts triggered their spread throughout the rest of the population. The behavior of *M. marburgensis* in TBR1 mimics this scenario, with two clusters of SNVs gradually increasing in frequency during SI and becoming dominant during SIII (Figure 3C). Additionally, the fraction of synonymous SNVs ($d_N/d_S = 0.46$) falls within the spectrum between conventional purifying selection values ($d_N/d_S = \sim 0.1$) and those indicative of positive adaptive selection ($d_N/d_S \geq 1$). This observation implies a potential prolonged coexistence of lineages within this species population.

### Magnitude of natural selection upon *M. marburgensis* and *M. thermautotrophicus* genomes

The role of natural selection in driving the observed genetic changes within species was investigated, focusing on the adaptation of the two dominant archaeal species that underwent ALE across phases of increased availability of gaseous substrates (SI), recovery (SII), and incremental starvation periods (SIII). MAG contigs were aligned with respect to the reference genomes of both species, and the distribution of SNVs per kilobase (SNVs/kbp) at the genomic level revealed that *M. thermautotrophicus* exhibits greater genomic stability, whereas *M. marburgensis* demonstrates a higher genetic plasticity (Figures 4A and 4B). Although both species exhibit random hotspot regions, *M. marburgensis* contains four scaffolds (GRT2.8_7, GRT2.8_8, GRT2.8_14, and GRT2.8_15) consistently reporting low SNV densities below 5 SNVs/kbp. These regions harbor genes involved in molybdenum cofactor biosynthesis (*moaB*) and transport (*modA/B/C*), as well as genes associated with hydrogenotrophic methanogenesis, including 4Fe-4S ferredoxin (*fwdF*) and two subunits of heterodisulfide reductase (*hdrB2/C2*). Despite nearly half of the predicted genes lacking KEGG orthologs, they remain intriguing and uncharacterized, presenting significant opportunities for further exploration. When comparing methanogenesis gene clusters between the two *Methanothermobacter* species, both exhibit three main groups located on three distinct scaffolds, with similar SNV densities ranging from 15 to 25 SNVs/kbp. However, while the corresponding scaffolds in *M. thermautotrophicus* (GRT1.1_5, GRT1.1_7, and GRT1.1_9) are relatively closely located, one scaffold in *M. marburgensis* is notably more distant from the other two (GRT2.8_1, GRT2.8_3, and GRT2.8_13). This distinct spatial distribution may be shaped by several factors, including operon structure, co-regulation, horizontal gene transfer, minimization of genetic recombination disruption, and/or metabolic efficiency. Even though the two species belong to the same genus and share the same metabolic pathways, further studies are needed to identify the primary drivers behind this spatial organization.

The effect of the selective pressure imposed by ALE was further assessed by examining the $d_N/d_S$ ratio of genes encoding enzymes crucial to the hydrogenotrophic biomethanation (see Table S4). Eleven of the 50 methanogenesis-related genes annotated on the *M. thermautotrophicus* genome had $d_N/d_S$ ratios over 1. Notably, these include *mtrG*, which encodes a membrane-associated multienzyme complex (Mtr) involved in $CH_4$ biosynthesis along with methyl-coenzyme M reductase (Mcr). Additionally, several subunits of two energy-converting hydrogenases (energy-converting hydrogenase A [EhA] and EhB), which have an anaplerotic role in the methanogenesis pathway,

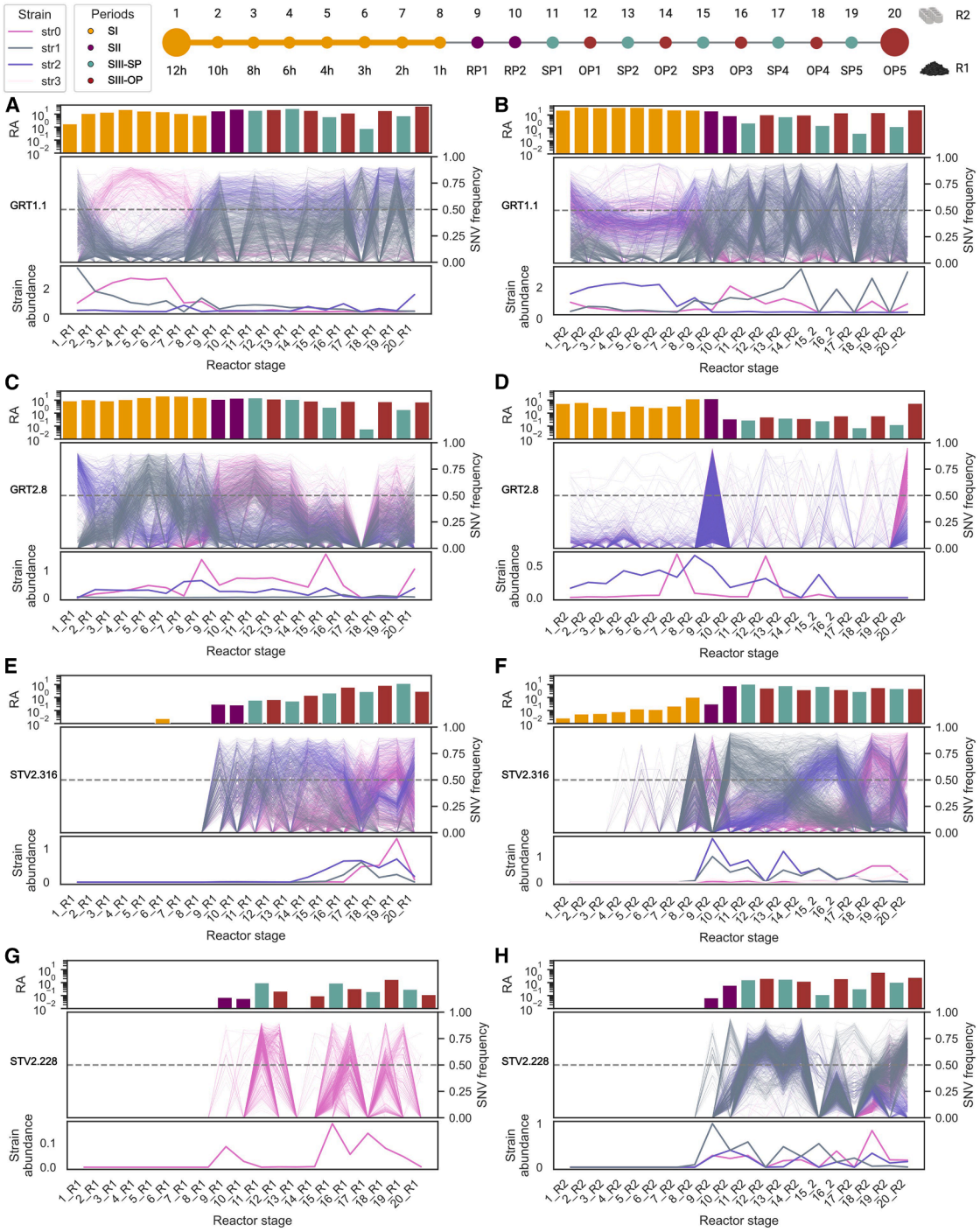



**Figure 3. Ecological and genetic dynamics in four representative species**

A subset of the species reported in Figure 2 was chosen to illustrate a range of strain-level characteristic behaviors in TBR1 and TBR2 vessels: *M. thermautotrophicus* (A and B), *M. marburgensis* (C and D), *C. proteolyticus* (E and F), and Haloplasmataceae sp. STV2.228 (G and H). For each of the four species, the top image illustrates the relative abundance over time, the middle image presents the frequency of nonsynonymous SNVs within the species, and the bottom image depicts the deconvoluted strain composition. The color of the bars is linked to the specific experimental phase. The sequential experimental periods are highlighted as follows: SI (stage I; yellow), SII (stage II; purple), SIII-SP (stage III starvation; red), and SIII-OP (stage III operation; green). Colored lines indicate SNVs clustering together over time in terms of frequency in the population. The colors of temporally varying SNVs were assigned based on hierarchical clustering, and the corresponding strain is colored accordingly (see STAR Methods).

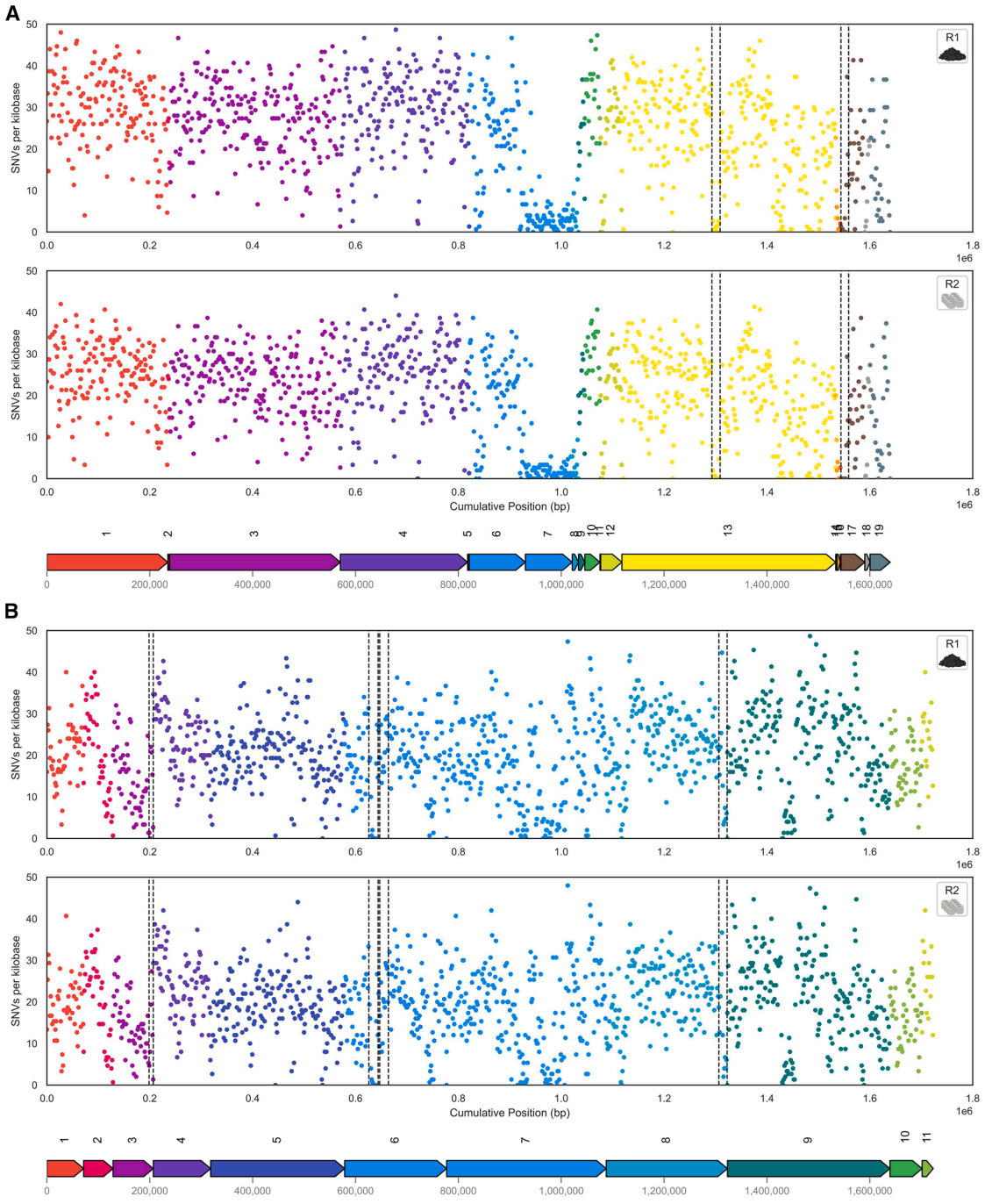



**Figure 4. Frequencies of SNVs/kbp across *M. marburgensis* and *M. thermautotrophicus* genomic scaffolds**

Scaffolds were ordered with respect to the reference genome of the two *Methanothermobacter* species deposited on the NCBI database. Predicted genomic islands are highlighted by vertical lines delimiting the start and the end of the region. SNVs/kbp was calculated at scaffold-level for *M. marburgensis* GRT2.8 (A) and *M. thermautotrophicus* GRT1.1 (B) using a sliding window of 1,500 bp. A different color identifier was assigned to each scaffold.

exhibited $d_N/d_S$ ratios above the threshold. Conversely, the *M. marburgensis* metagenome revealed only two genes meeting the aforementioned threshold, encoding a specific subunit of the formylmethanofuran dehydrogenase (Fwd/Fmd) and EhB, respectively. Furthermore, when considering the role of environmental factors in driving the observed strain-level dynamics, it was found that SNVs also accumulate after the starvation phase in biofilm adhesion genes (i.e., *galE*, *cps*, and *glm*). During this phase, both archaea were required to recolonize the reactor. In *M. thermautotrophicus*, this process occurs through multiple

stages, while in *M. marburgensis*, it is exclusive to the final time point (OP5). This suggests that a newly emerging strain is colonizing the biofilm and outcompeting the previous haplotype (Figure 3D).

### 3D reconstruction of mutated enzymes undergoing positive selective pressure in *Methanothermobacter* species

After establishing that genetic adaptation likely drove the increased $d_N/d_S$ ratios in the two archaea, the putative effect of fixed SNVs on the function of relevant proteins was investigated. The Grantham distance (Gd) was used as a proxy to assess the impact of nonsynonymous SNVs on the protein function (see Tables S5 and S6), measuring the physical and chemical differences between the reference and mutated amino acids (aa).[31] The Gd values range from 5 (L > I) to 215 (C > W), indicating varying degrees of functional impact. In *M. marburgensis*, the dominant haplotype during SI harbored three variants with medium to high Gd. Two SNVs were localized on *mer*, encoding the 5,10-methylenetetrahydromethanopterin reductase (Mer), and correspond to aa substitutions with Gd values of 99 (A192S) and 121 (E52V), respectively. Both switches entail a change of the aa polarity. Another variant was identified within *fwdE* (G58E), which encodes subunit E of Fwd, the enzyme catalyzing the initial step of methanogenesis from $CO_2$. The substitution of glycine, a nonpolar aa, with glutamate, which has a negatively charged side chain, resembles changes observed in Mer. Putative high-impact SNVs were less common within the dominant haplotype of *M. thermautotrophicus* during SI. Among the noteworthy variants were changes on the *mcrB* gene, encoding a subunit integral to Mcr, which resulted in aa substitutions K283T, K285N, and S346G. Conversely, the sweeping haplotypes in SIII encompass multiple SNVs with a frequency near fixation (f > 0.75) and moderate Gd. The tetrahydromethanopterin S-methyltransferase (*mtrB*) and Fwd (*fwdC*) genes were mutated in three and two positions, respectively, all with Gd values above 89. Similarly to the other archaeon, all the aa switches determine a polarity change, thus possibly determining structural rearrangements for Mtr, Mcr, and Fwd complexes. The presence of pathway-specific fixed SNVs further supports the previously proposed hypothesis of selection. However, structural evidence demonstrating the impact of these changes on peptide conformation and function is still required.

Leveraging recent advancements in protein structure prediction,[32] which facilitate rapid and accurate structural characterization of protein sequences from metagenomic data, we conducted a 3D reconstruction of ten peptide structures, both wild type and mutated, for five different proteins (Mer, McrB, MtrB, FwdC, and FwdE). Major structural impacts resulting from aa substitutions were predominantly observed in Mer (Figures 5A–5C) and McrB (Figures 5D–5F). The analysis of phi/psi angles via Ramachandran plot revealed the clustering of main-chain torsion angles within the favored α- and β-regions, aligning with expectations for a properly folded protein (Figures 5A and 5D). In *M. marburgensis*, the mutated Mer exhibited changes at residue positions 52 and 192, which are located in the α2 and α6 helices, respectively (Figures 5B and 5C). Sequence alignments revealed that the terminal portion of the α2 helix is conserved within the *Methanothermobacter* genera, whereas in other species, such as *Methanopyrus kandleri*, a higher presence of hydrophobic aa is observed (see Figure S3). The 3D reconstruction of MtrB and McrB in *M. thermautotrophicus* reveals changes that may have affected their binding site functionality. The cluster of three high-impact aa changes was detected at positions 63, 65, and 66 of the MtrB sequence, falling inside one of the transmembrane helices of the Mtr complex.[33] Additionally, the three aa switches on McrB were located on the C-terminal domain of the subunit β (Figures 5E and 5F; see Figure S4), where the binding site for coenzymes F430, B, and M are located.[34] Finally, the structural analysis of FwdC and FwdE in both archaea did not elucidate the critical positions of the aa changes encoded within their genes.

## DISCUSSION

The emergence of metagenomics has increased our understanding of the structure and functions of microbial communities, offering new insights into their evolutionary and functional diversity.[35] Notably, extensive *de novo* assembly of reads into metagenomes can provide invaluable knowledge into the existence of previously uncharacterized organisms and their genetic potential.[1,10,11] The response of microbial communities to environmental selective pressure is known to be associated with their bioconversion efficiency,[36] the spread of stress-resistance mechanisms,[37] and their long-term stability.[38] While numerous studies have described the taxonomical or functional resilience of these communities, strain-level fluctuations contributing to ecological features are still poorly characterized. The knowledge gap is even more pronounced when considering microbiomes other than those colonizing the human gut,[39] as investigations on ALE and the patterns of natural genetic variation through selection are lacking.

In this pioneering study, we provide the first in-depth characterization of within-species dynamics by monitoring the microbiome of two anaerobic cultivation vessels. Gene presence screening enabled the identification of key species, including hydrogenotrophic archaea responsible for biomethanation, as well as bacterial species that secrete EPS, thereby contributing to biofilm formation. When comparing biofilm and liquid samples, the dominant planktonic bacterial species (i.e., *Mycobacterium* sp. STV1.135 and *C. proteolyticus* STV2.316) are also the most abundant within the EPS matrix, and the same holds for the archaeal population. This further reinforces the relevance of our downstream results, which are based on the microbiome analysis of the liquid phase, given the vessel-related limitations. Utilizing deep shotgun metagenomic sequencing alongside an ad hoc variant calling and phasing methodology, we tracked the trajectories of SNVs within 26 distinct species and compared these trends with broader ecological shifts at the species level. Consistent with our expectations, we demonstrated that ALE can lead to widespread shifts in the genetic composition of individual $CO_2$-fixating species, as reported in the antibiotically treated gut,[12,40] upon heavy metal perturbation in the soil[41] or increased ammonia load[29] in biotechnological systems. However, these within-species variations rarely supported the hypothesis of extinction followed by recolonization under more

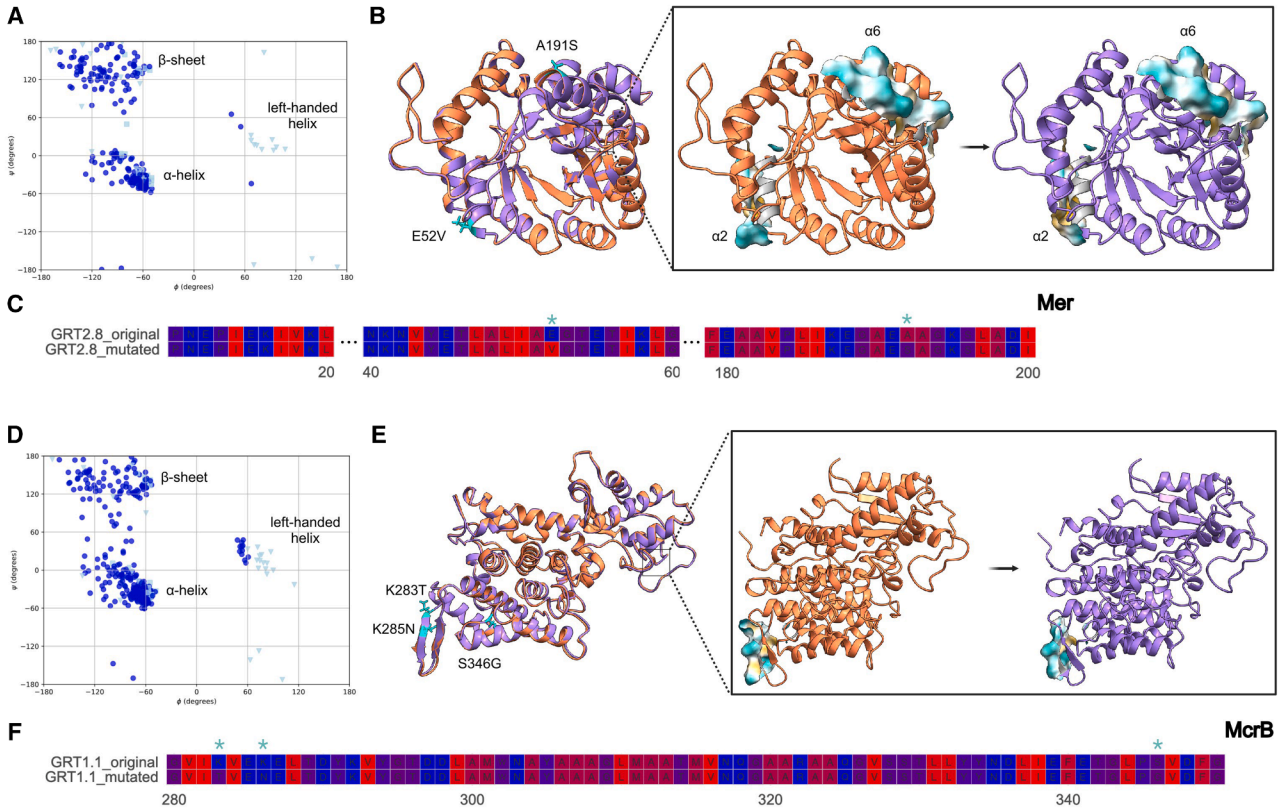



**Figure 5. Structural characterization of enzymes involved in methanogenesis under selective pressure in *M. marburgensis* and *M. thermautotrophicus***

For each selected protein, a Ramachandran plot and the superposition of the predicted AlphaFold original (coral) and mutated structure (purple) were generated. In the Ramachandran plot, the main-chain torsion angles (phi and psi) are depicted as light blue triangles for glycine, light blue squares for proline, and blue circles for all other residues. The aa changes are colored in turquoise. Specifically, (A) and (D) represent the Ramachandran plot, while (B) and (E) depict the Mer and McrB structure prediction in *M. marburgensis* and *M. thermautotrophicus*, respectively. In the zoomed regions of (B) and (E), protein segments impacted by aa changes in Mer and McrB are highlighted in gray, with their hydrophobic surfaces superimposed. (C) and (F) present a section of the alignment of peptide sequences, with mapped aa shifts indicated by an asterisk (*). aa are colored according to their hydrophobicity, from red (hydrophobic) to blue (hydrophilic).

selective conditions. Instead, the genetic responses varied widely across species, with some resident populations showing a large number of variants while others exhibited only a handful. Observed SNVs may represent either *de novo* mutations or the emergence of pre-existing strains that diverged over generations before eventually becoming dominant, resulting in strain replacement. These genetic changes were frequently witnessed in taxa without substantial changes in abundance. Conversely, large abundance fluctuations were not always accompanied by widespread genetic alterations. Furthermore, we found that some of the utmost fluctuations in RA and the appearance of SNVs occurred at lower GRTs and during the starvation weeks, especially in *Methanothermobacter* species. Together, these findings suggest that the response to more selective conditions was not driven by discrete recolonization events but rather by the subtler processes of strain-level competition and evolution, as previously documented in antibiotic-treated gut microbiomes.[30]

At this population's genetic level, strain dynamics were considerably more intricate than those described by the neutral theory of molecular evolution.[42] This theory posits that most molecular-level variations have negligible effects on fitness, thus suggesting that stochastic processes best explain the evolutionary fate of genetic variation.[42,43] Conversely, a Tajima's D value above +5 for *M. marburgensis* and *M. thermautotrophicus* indicates an excess of rare genetic variants relative to neutrality expectations,[44] possibly due to factors like population expansion, positive selection, or purifying selection against deleterious mutation. Our results reveal comparable behaviors in mixtures of distantly related strains, as well as in haplotypes that presumably evolved *de novo*. While similar instances of strain replacement have been previously reported,[23,29] the current comprehensive time-series analysis conducted over a 1-year period uncovers unprecedented, condition-driven evolutionary events in microbial communities. The observed partial genome-wide sweeps contained multiple linked nonsynonymous SNVs, clustering together according to their frequency. Within these clusters, SNVs had previously segregated at substantial frequencies ($f > 5\%$) before transitioning to either a lower GRT or entering a starvation phase. In the activated carbon-packed vessel, a haplotype of *M. thermautotrophicus* experienced a rapid increase in frequency over weekly timescales but later on reverted the sweep. These observations align with temporally fluctuating selection

pressures, wherein reversions do not culminate in extinction but rather stabilize close to their initial frequency. A similar sweep was also evidenced for *M. marburgensis*; however, the haplotype became fixed in the population during SI and was maintained throughout SIII. The same fluctuations were also identified within the other vessel despite the fact that the two systems are independent and exhibit structural differences, with the exception of *M. marburgensis*. Altogether, these evolutionary events imply that the sweeping haplotypes may have been consistently maintained within their respective populations over time, possibly due to the presence of stable metabolic or spatial niches. The findings emphasize, for the first time, broader microbial dynamics that extend beyond specific material influences, underscoring the utility of this novel approach for understanding microbial structure and function in engineered ecosystems.

Strain-level investigations documented in the literature have already demonstrated that newly emerging strains arise and attain dominance due to their capacity to withstand inhibitory conditions, including antibiotic treatment[30] or environmental stressors.[29] However, an unresolved question remains as to whether the variations accumulated and maintained by these fluctuating haplotypes serve as the primary determinant of their superior fitness. To address this gap, the current study aims to uniquely complement the variant phasing approach adopted with gene-level statistics and a 3D reconstruction of the peptide sequences targeted by SNV-induced aa alterations. Mapping the SNVs onto the genes and calculating the $d_N/d_S$ unveiled selective pressure on genes associated with the hydrogenotrophic methanogenesis metabolic pathway. Notably, within both *M. marburgensis* and *M. thermautotrophicus*, multiple subunits of the Mtr and Mcr complexes exhibited a ratio exceeding 1. These findings corroborate how environmental alterations in bioconversion systems can induce selection, resulting in the fixation of specific haplotypes distinguished by their unique genetic makeup compared to other strains within the populations.[29] Furthermore, the potential of these nonsynonymous SNVs to positively influence the activity of the enzymes was revealed by the superposition of the predicted structures. The positioning of aa changes within the conserved binding sites clefts of McrB in *M. thermautotrophicus* indicated reduced steric hindrance, potentially enhancing substrate docking.[45] Conversely, the shift from a hydrophilic to hydrophobic aa in *M. marburgensis* has the potential to alter the tetrameric state of Mer. Previous studies have shown that the interface region involved in tetramer formation in *Methanothermobacter* species exhibits lower hydrophobicity compared to *M. kandleri*,[46] disturbing the protein complex assembly. Therefore, the observed switch from Glu52 to Val52, corresponding to Val54 in *M. kandleri*,[46] enhances the region's hydrophobic character, probably promoting the tetrameric state of Mer. These alterations, bolstered by moderate Gd metrics, may have provided a phenotypic advantage to the strains harboring them, determining a shift in dominance among archaeal population haplotypes during SI and SIII.

### Limitations of the study

There are several limitations inherent in the metagenomic data and methodology used in this study. A key limiting factor is the current inability to directly link the identified SNVs to the strain of origin. Emerging approaches leveraging single-cell technologies are presently under development based on the concept that single-amplified genomes with closely related SNVs can be regarded as belonging to the same strain.[47] Although our ad hoc SNV phasing approach represents a first in allowing the observation of distinct behavioral responses across individual species within the same community, even at the strain level, further research is needed to assess the prevalence of these patterns across haplotypes and upon different stressors. Nonetheless, our high-resolution time series provides a valuable framework that can guide future classification efforts in larger but lower-resolution studies. Unlike conventional metagenomic analyses, which primarily focus on identifying key microbial taxa at the taxonomic level and inferring their functional roles, we developed a first-of-its-kind population genetic workflow to quantitatively track fine-scale genetic changes within microbial populations over time. By monitoring the microbiome through periods of increased $H_2$ and $CO_2$ flow rate, recovery, and intermittent feeding provision, we uncovered new evidence suggesting that the ecological resilience in microbial communities extends down to the strain level. Understanding how this resilience arises from the intricate interplay between haplotype populations and fluctuations in operational parameters, as well as its broader implications for microbiome evolution and engineering, remains an exciting avenue for future research.

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Gabriele Ghiotto (gabriele.ghiotto@phd.unipd.it).

### Materials availability

This study did not generate new unique reagents.

### Data and code availability

- All original code has been deposited at GitHub (https://github.com/gabrieleghiotto/Microbial-SNVs-analysis) and is publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## ACKNOWLEDGMENTS

G.G. acknowledges the project LIFE20 CCM/GR/001642 – LIFE CO2toCH4 of the European Union LIFE+ program for financially supporting his salary.

## AUTHOR CONTRIBUTIONS

Conceptualization, G.G., S.C., and L.T.; software, G.G.; formal analysis, G.G.; data curation, G.G., A.X., and M.G.; writing – original draft, G.G.; visualization, G.G.; investigation, A.X. and M.G.; resources, P.G.K., S.C., and L.T.; supervision, P.G.K., S.C., and L.T.; funding acquisition, P.G.K., S.C., and L.T.; project administration, L.T.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## STAR★METHODS

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

- KEY RESOURCES TABLE
- METHOD DETAILS
  - Experimental setup
  - Analytical monitoring of biochemical parameters
  - DNA extraction and sequencing
  - Metagenomics reconstruction of the microbiome
  - SNVs profiling over time within species
  - SNVs filtering, SNVs sweeps identification and $d_N/d_S$ ratio
  - 3D reconstruction of mutated protein

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2025.100949.

Received: November 27, 2024
Revised: March 28, 2025
Accepted: June 16, 2025
Published: July 15, 2025

## REFERENCES

1. Pavlopoulos, G.A., Baltoumas, F.A., Liu, S., Selvitopi, O., Camargo, A.P., Nayfach, S., Azad, A., Roux, S., Call, L., Ivanova, N.N., et al. (2023). Unraveling the functional dark matter through global metagenomics. Nature *622*, 594–602.
2. Ver Eecke, H.C., Butterfield, D.A., Huber, J.A., Lilley, M.D., Olson, E.J., Roe, K.K., Evans, L.J., Merkel, A.Y., Cantin, H.V., and Holden, J.F. (2012). Hydrogen-limited growth of hyperthermophilic methanogens at deep-sea hydrothermal vents. Proc. Natl. Acad. Sci. *109*, 13674–13679.
3. Chen, X., Xue, D., Wang, Y., Qiu, Q., Wu, L., Wang, M., Liu, J., and Chen, H. (2023). Variations in the archaeal community and associated methanogenesis in peat profiles of three typical peatland types in China. Environ. Microbiome *18*, 48.
4. Tyne, R.L., Barry, P.H., Lawson, M., Byrne, D.J., Warr, O., Xie, H., Hillegonds, D.J., Formolo, M., Summers, Z.M., Skinner, B., et al. (2021). Rapid microbial methanogenesis during CO2 storage in hydrocarbon reservoirs. Nature *600*, 670–674.
5. Bozell, J.J., and Petersen, G.R. (2010). Technology development for the production of biobased products from biorefinery carbohydrates—the US Department of Energy's "Top 10" revisited. Green Chem. *12*, 539.
6. Angenent, L.T., Richter, H., Buckel, W., Spirito, C.M., Steinbusch, K.J.J., Plugge, C.M., Strik, D.P.B.T.B., Grootscholten, T.I.M., Buisman, C.J.N., and Hamelers, H.V.M. (2016). Chain Elongation with Reactor Microbiomes: Open-Culture Biotechnology To Produce Biochemicals. Environ. Sci. Technol. *50*, 2796–2810.
7. Cantera, S., Tamarit, D., Strong, P.J., Sánchez-Andrea, I., Ettema, T.J.G., and Sousa, D.Z. (2022). Prospective CO2 and CO bioconversion into ectoines using novel microbial platforms. Rev. Environ. Sci. Biotechnol. *21*, 571–581.
8. 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.
9. Riesenfeld, C.S., Schloss, P.D., and Handelsman, J. (2004). Metagenomics: Genomic Analysis of Microbial Communities. Annu. Rev. Genet. *38*, 525–552.
10. 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.
11. Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., Wu, D., Paez-Espino, D., Chen, I.-M., Huntemann, M., et al. (2021). A genomic catalog of Earth's microbiomes. Nat. Biotechnol. *39*, 499–509.
12. Zhao, S., Lieberman, T.D., Poyet, M., Kauffman, K.M., Gibbons, S.M., Groussin, M., Xavier, R.J., and Alm, E.J. (2019). Adaptive Evolution within Gut Microbiomes of Healthy People. Cell Host Microbe *25*, 656–667.e8.
13. Fay, J.C., and Wu, C.-I. (2000). Hitchhiking Under Positive Darwinian Selection. Genetics *155*, 1405–1413.
14. Charlesworth, B., Morgan, M.T., and Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. Genetics *134*, 1289–1303.
15. Olm, M.R., Crits-Christoph, A., Bouma-Gregson, K., Firek, B.A., Morowitz, M.J., and Banfield, J.F. (2021). inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. Nat. Biotechnol. *39*, 727–736.
16. Zhao, C., Dimitrov, B., Goldman, M., Nayfach, S., and Pollard, K.S. (2023). MIDAS2: Metagenomic Intra-species Diversity Analysis System. Bioinformatics *39*, btac713.
17. Van Rossum, T., Costea, P.I., Paoli, L., Alves, R., Thielemann, R., Sunagawa, S., and Bork, P. (2022). metaSNV v2: detection of SNVs and subspecies in prokaryotic metagenomes. Bioinformatics *38*, 1162–1164.
18. Dethlefsen, L., and Relman, D.A. (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc. Natl. Acad. Sci. USA *108*, 4554–4561.
19. Ng, K.M., Aranda-Díaz, A., Tropini, C., Frankel, M.R., Van Treuren, W., O'Loughlin, C.T., Merrill, B.D., Yu, F.B., Pruss, K.M., Oliveira, R.A., et al. (2019). Recovery of the Gut Microbiota after Antibiotics Depends on Host Diet, Community Context, and Environmental Reservoirs. Cell Host Microbe *26*, 650–665.e4.
20. Garud, N.R., Good, B.H., Hallatschek, O., and Pollard, K.S. (2019). Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. PLoS Biol. *17*, e3000102.
21. Ghiotto, G., De Bernardini, N., Giangeri, G., Tsapekos, P., Gaspari, M., Kougias, P.G., Campanaro, S., Angelidaki, I., and Treu, L. (2024). From microbial heterogeneity to evolutionary insights: A strain-resolved metagenomic study of H2S-induced changes in anaerobic biofilms. Chem. Eng. J. *485*, 149824.
22. Ma, C., Zhang, Y., Jiang, S., Teng, F., Huang, S., and Zhang, J. (2023). Cross-cohort single-nucleotide-variant profiling of gut microbiota suggests a novel gut-health assessment approach. mSystems *8*, e00828-23.
23. Ghiotto, G., Zampieri, G., Campanaro, S., and Treu, L. (2024). Strain-resolved metagenomics approaches applied to biogas upgrading. Environ. Res. *240*, 117414.
24. Ebrahimian, F., De Bernardini, N., Tsapekos, P., Treu, L., Zhu, X., Campanaro, S., Karimi, K., and Angelidaki, I. (2022). Effect of pressure on biomethanation process and spatial stratification of microbial communities in trickle bed reactors under decreasing gas retention time. Bioresour. Technol. *361*, 127701.
25. Tsapekos, P., Treu, L., Campanaro, S., Centurion, V.B., Zhu, X., Peprah, M., Zhang, Z., Kougias, P.G., and Angelidaki, I. (2021). Pilot-scale biomethanation in a trickle bed reactor: Process performance and microbiome functional reconstruction. Energy Convers. Manag. *244*, 114491.
26. Chatzis, A., Orellana, E., Gaspari, M., Kontogiannopoulos, K., Treu, L., Zouboulis, A., and Kougias, P.G. (2023). Comparative study on packing materials for improved biological methanation in trickle Bed reactors. Bioresour. Technol. *385*, 129456.
27. Zampieri, G., Santinello, D., Palù, M., Orellana, E., Costantini, P., Favaro, L., Campanaro, S., and Treu, L. (2025). Core cooperative metabolism in low-complexity CO2-fixing anaerobic microbiota. ISME J. *19*, wraf017.
28. Kadnikov, V.V., Mardanov, A.V., Beletsky, A.V., Karnachuk, O.V., and Ravin, N.V. (2019). Genome of the candidate phylum Aminicenantes bacterium from a deep subsurface thermal aquifer revealed its fermentative saccharolytic lifestyle. Extrem. Extremophiles *23*, 189–200.

29. Bucci, L., Ghiotto, G., Zampieri, G., Raga, R., Favaro, L., Treu, L., and Campanaro, S. (2023). Adaptation of Anaerobic Digestion Microbial Communities to High Ammonium Levels: Insights from Strain-Resolved Metagenomics. Environ. Sci. Technol. *58*, 580–590.

30. Roodgar, M., Good, B.H., Garud, N.R., Martis, S., Avula, M., Zhou, W., Lancaster, S.M., Lee, H., Babveyh, A., Nesamoney, S., et al. (2021). Longitudinal linked-read sequencing reveals ecological and evolutionary responses of a human gut microbiome during antibiotic treatment. Genome Res. *31*, 1433–1446.

31. Grantham, R. (1974). Amino Acid Difference Formula to Help Explain Protein Evolution. Science *185*, 862–864.

32. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589.

33. Gottschalk, G., and Thauer, R.K. (2001). The Na+-translocating methyltransferase complex from methanogenic archaea. Biochim. Biophys. Acta *1505*, 28–36.

34. Ermler, U., Grabarse, W., Shima, S., Goubeaud, M., and Thauer, R.K. (1997). Crystal Structure of Methyl-Coenzyme M Reductase: The Key Enzyme of Biological Methane Formation. Science *278*, 1457–1462.

35. Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.-F., Darling, A., Malfatti, S., Swan, B.K., Gies, E.A., et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. Nature *499*, 431–437.

36. Konstantinidis, D., Pereira, F., Geissen, E.M., Grkovska, K., Kafkia, E., Jouhten, P., Kim, Y., Devendran, S., Zimmermann, M., and Patil, K.R. (2021). Adaptive laboratory evolution of microbial co-cultures for improved metabolite secretion. Mol. Syst. Biol. *17*, e10189.

37. Soucy, S.M., Huang, J., and Gogarten, J.P. (2015). Horizontal gene transfer: building the web of life. Nat. Rev. Genet. *16*, 472–482.

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

39. Wolff, R., Shoemaker, W., and Garud, N. (2023). Ecological Stability Emerges at the Level of Strains in the Human Gut Microbiome. mBio *14*, e02502-22.

40. Sharon, I., Morowitz, M.J., Thomas, B.C., Costello, E.K., Relman, D.A., and Banfield, J.F. (2013). Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. Genome Res. *23*, 111–120.

41. Coyte, K.Z., Stevenson, C., Knight, C.G., Harrison, E., Hall, J.P.J., and Brockhurst, M.A. (2022). Horizontal gene transfer and ecological interactions jointly control microbiome stability. PLoS Biol. *20*, e3001847.

42. Kimura, M. (1968). Evolutionary Rate at the Molecular Level. Nature *217*, 624–626.

43. Kimura, M. (1991). The neutral theory of molecular evolution: A review of recent evidence. Jpn. J. Genet. *66*, 367–386.

44. Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics *123*, 585–595.

45. Ermler, U. (2005). On the mechanism of methyl-coenzyme M reductase. Dalton Trans., 3451–3458.

46. Shima, S., Warkentin, E., Grabarse, W., Sordel, M., Wicke, M., Thauer, R. K., and Ermler, U. (2000). Structure of Coenzyme F420 Dependent Methylenetetrahydromethanopterin Reductase from Two Methanogenic Archaea. J. Mol. Biol. *300*, 935–950.

47. Zheng, W., Zhao, S., Yin, Y., Zhang, H., Needham, D.M., Evans, E.D., Dai, C.L., Lu, P.J., Alm, E.J., and Weitz, D.A. (2022). High-throughput, single-microbe genomics with strain resolution, applied to a human gut microbiome. Science *376*, eabm1483.

48. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

49. Bushnell, B., Rood, J., and Singer, E. (2017). BBMerge – Accurate paired shotgun read merging via overlap. PLoS One *12*, e0185056.

50. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., and Korobeynikov, A. (2020). Using SPAdes De Novo Assembler. Curr. Protoc. Bioinforma *70*, e102.

51. Wu, Y.-W., Simmons, B.A., and Singer, S.W. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics *32*, 605–607.

52. Kang, D.D., Froula, J., Egan, R., and Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ *3*, e1165.

53. 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.

54. Nissen, J.N., Johansen, J., Allesøe, R.L., Sønderby, C.K., Armenteros, J.J. A., Grønbech, C.H., Jensen, L.J., Nielsen, H.B., Petersen, T.N., Winther, O., and Rasmussen, S. (2021). Improved metagenome binning and assembly using deep variational autoencoders. Nat. Biotechnol. *39*, 555–560.

55. Sieber, C.M.K., Probst, A.J., Sharrar, A., Thomas, B.C., Hess, M., Tringe, S. G., and Banfield, J.F. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat. Microbiol. *3*, 836–843.

56. Olm, M.R., Brown, C.T., Brooks, B., and Banfield, J.F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. *11*, 2864–2868.

57. Chklovski, A., Parks, D.H., Woodcroft, B.J., and Tyson, G.W. (2023). CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Nat. Methods *20*, 1203–1212.

58. Aroney, S.T.N., Newell, R.J.P., Nissen, J.N., Camargo, A.P., Tyson, G.W., and Woodcroft, B.J. (2025). CoverM: read alignment statistics for metagenomics. Bioinformatics *41*, btaf147.

59. Woodcroft, B.J., Aroney, S.T.N., Zhao, R., Cunningham, M., Mitchell, J.A. M., Blackall, L., and Tyson, G.W. (2024). SingleM and Sandpiper: Robust microbial taxonomic profiles from metagenomic data. Preprint at bioRxiv. https://doi.org/10.1101/2024.01.30.578060.

60. Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2022). GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics *38*, 5315–5316.

61. Asnicar, F., Thomas, A.M., Beghini, F., Mengoni, C., Manara, S., Manghi, P., Zhu, Q., Bolzan, M., Cumbo, F., May, U., et al. (2020). Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nat. Commun. *11*, 2500.

62. 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.

63. Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol. Biol. Evol. *38*, 5825–5829.

64. Quince, C., Nurk, S., Raguideau, S., James, R., Soyer, O.S., Summers, J.K., Limasset, A., Eren, A.M., Chikhi, R., and Darling, A.E. (2021). STRONG: metagenomics strain resolution on assembly graphs. Genome Biol. *22*, 214.

65. Camargo, A.P., Roux, S., Schulz, F., Babinski, M., Xu, Y., Hu, B., Chain, P. S.G., Nayfach, S., and Kyrpides, N.C. (2023). Identification of mobile genetic elements with geNomad. Nat. Biotechnol. *42*, 1303–1312.

66. Banerjee, P., Eulenstein, O., and Friedberg, I. (2024). Discovering genomic islands in unannotated bacterial genomes using sequence embedding. Bioinform. Adv. *4*, vbae089.

67. Sievers, F., and Higgins, D.G. (2018). Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. *27*, 135–145.

68. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.

69. Meng, E.C., Goddard, T.D., Pettersen, E.F., Couch, G.S., Pearson, Z.J., Morris, J.H., and Ferrin, T.E. (2023). UCSF ChimeraX: Tools for structure building and analysis. Protein Sci. *32*, e4792.

70. Xirostylidou, A., Gaspari, M., Kontogiannopoulos, K.N., Ghiotto, G., Treu, L., Campanaro, S., Zouboulis, A.I., and Kougias, P.G. (2024). Biomethanation on demand: Continuous and intermittent hydrogen supply on biological CO2 methanation. Chem. Eng. J. *495*, 153677.

71. 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.

72. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., and Ishiguro-Watanabe, M. (2023). KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. *51*, D587–D592.

73. Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bansal, P., Bridge, A.J., Poux, S., Bougueleret, L., and Xenarios, I. (2016). UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View. Methods Mol. Biol. *1374*, 23–54.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Biological samples | | |
| Liquid samples from biomethanation reactors | This paper | N/A |
| Biofilm samples from biomethanation reactors | This paper | N/A |
| Critical commercial assays | | |
| PowerSoil© DNA Isolation Kit | Mo Bio Laboratories | 12888-50; 12888-100 |
| Deposited data | | |
| Short raw DNA and RNA sequencing reads | This paper | SRA: PRJNA1179988 |
| Software and algorithms | | |
| Trimmomatic (v0.39) | Bolger et al.[48] | https://github.com/timflutre/trimmomatic |
| BBMerge (v38.53) | Bushnell et al.[49] | https://sourceforge.net/projects/bbmap/ |
| metaSPAdes (v3.13.0) | Prjibelski wt al.[50] | https://github.com/ablab/spades |
| Maxbin2 (v2.2.7) | Wu et al.[51] | https://sourceforge.net/projects/maxbin/ |
| Metabat1 | Kang et al., 2015[52] | https://bitbucket.org/berkeleylab/metabat/src/master/ |
| Metabat2 (v2.15) | Kang et al., 2019[53] | https://bitbucket.org/berkeleylab/metabat/src/master/ |
| Vamb (v4.13) | Nissen et al.[54] | https://github.com/RasmussenLab/vamb |
| Das_Tool (v1.1.6) | Sieber et al.[55] | https://github.com/cmks/DAS_Tool |
| dRep (v3.4.0) | Olm et al.[56] | https://github.com/MrOlm/drep |
| checkM2 (v1.0.2) | Chklovski et al.[57] | https://github.com/chklovski/CheckM2 |
| coverM (v0.6.1) | Aroney et al.[58] | https://github.com/wwood/CoverM |
| SingleM (v0.18.3) | Woodcroft, B. J. et al.[59] | https://github.com/wwood/singlem |
| GTDB-Tk (v2.1.1) | Chaumeil et al.[60] | https://github.com/Ecogenomics/GTDBTk |
| PhyloPhlAn (v3.0) | Asnicar et al.[61] | https://github.com/biobakery/biobakery/wiki/PhyloPhlAn3 |
| Prodigal (v2.6.3) | Hyatt et al.[62] | https://github.com/hyattpd/Prodigal |
| eggNOG-mapper (v2.1.10) | Cantalapiedra et al.[63] | https://github.com/eggnogdb/eggnog-mapper |
| InStrain (v1.8.0) | Olm et al.[15] | https://github.com/MrOlm/inStrain |
| STRONG | Quince et al.[64] | https://github.com/chrisquince/STRONG |
| geNomad (v1.8.0) | Camargo et al.[65] | https://github.com/apcamargo/genomad |
| TreasureIsland | Banerjee et al.[66] | https://github.com/FriedbergLab/GenomicIslandPrediction |
| ClustalOmega (v1.2.2) | Sievers et al.[67] | http://www.clustal.org/omega/ |
| Jailview (v2.11.3.0) | Waterhouse et al.[68] | https://www.jalview.org/ |
| ChimeraX (v1.9) | Meng et al.[69] | https://www.cgl.ucsf.edu/chimerax/ |
| Scripts used for analysis in this study | This study | https://github.com/gabrieleghiotto/Microbial-SNVs-analysis; Zenodo: https://doi.org/10.5281/zenodo.15672092 |

### METHOD DETAILS

#### Experimental setup

The study employed TBRs consisting of 1L packed-bed biochambers, gas and liquid delivery systems, and a gas-metering system. Briefly, TBRs provide a large surface area for the development of microbial biofilms, a configuration that facilitates the absorption of $CO_2$ from the gaseous phase and its subsequent biological conversion into $CH_4$. A detailed schematic of the experimental setup is

provided in the associated publication.[70] Specifically, TBR-1 contained activated carbon pellets (dimensions: 20 × 4 mm, surface area 20 $m^2$/g, total pore volume 6 x 10-2 $cm^3$/g, micropore volume 1 x 10-3 $cm^3$/g), while TBR-2 was filled with high-density polyethylene Raschig rings (dimensions: 10 × 10 mm, surface area 3.3 × 10-3 $m^2$/g). Both systems operated at a stable thermophilic temperature (55 ± 1°C) using thermal jackets. Each cultivation vessel was connected to a dedicated liquid reservoir maintained in a temperature-stabilized water bath. Nutrient medium delivery was achieved using a peristaltic pump (Sci-Q 300, Watson Marlow, United Kingdom) at a flow rate of 95 mL/min. The growth substrate providing vitamins and cofactors was composed by exhaust biomass (see Table S1). The biochambers were fed an 80% $H_2$ and 20% $CO_2$ synthetic gas mixture through a dual-head peristaltic pump (Ismatec Reglo, Masterflex, United States), injected concurrently with the liquid flow. The 1:4 of $CO_2$ to $H_2$ represents a stoichiometry that facilitates methanogenesis. Gas production was monitored daily using automated water-displacement gas meters with a 100 mL reversible cycle. Additionally, throughout the entire experiment, metabolites and environmental conditions, including pH, VFA concentrations, and atmospheric composition, to assess microbiome's stability (Figure 1B). The experimental process consisted of three stages: Stage I involved continuous operation with progressively reduced GRT over 162 days, Stage II maintained a stable 1-h GRT to allow recovery lasting 40 days, and Stage III introduced intermittent starvation conditions (213 days). During Stage I the GRT was gradually reduced by increasing gas supply until reaching stable $CH_4$ output (>95%). The two TBRs were tested at eight GRTs, from 12 to 1 h (data regarding the 45 min GRT time point were excluded from the current analysis but are reported in the associated publication).[70] Stage III involved five starvation phases, lasting from 1 to 5 weeks (SP1-SP5), alternated with equal-duration normal operation periods (OP1-OP5). During starvation (SIII-SP), feeding gas was withheld, while temperature conditions were maintained. Conversely, during the regular operational periods (SIII-OP), the bioreactors were kept at 1-h GRT. The impact of starvation on recovery time during subsequent normal operation was evaluated. Figure 1A outlines the experimental design, incorporating operational and starvation phases.

### Analytical monitoring of biochemical parameters

Gas phase composition was analyzed using a gas chromatograph (GC-2014187 Pro, Shimadzu, Japan) equipped with a thermal conductivity detector and Porapak Q Column (1.8 m length, 1/8-inch inner diameter, film thickness 2 mm), followed by an Agilent J&W CP-Molsieve 5 Å column (1.8 m length, 1/8-inch inner diameter, film thickness 2 mm) with helium as the carrier gas at a flow rate of 20 mL/min. Both the column and detector were set at 100°C. Liquid samples were collected twice weekly for pH and VFA analysis. pH was measured with an HI2020 EDGE pH meter (HANNA Instruments, USA). VFAs concentration was determined using gas chromatography (GC-2010 Pro, Shimadzu, Japan) with a flame ionization detector and Agilent J&W Capillary Column (30 m length, 0.5 mm inner diameter, film thickness 1.020 μm). Helium, hydrogen, and synthetic air were used as carrier gasses.

### DNA extraction and sequencing

Genomic samples were collected from the liquid phase of the cultivation vessels on the last day of each operation change, which was either the day prior to the GRT reductions (Stage I) or the day before the starts/stops of $H_2$/$CO_2$ supply (Stage II). DNA extraction was carried out employing the DNeasy PowerSoil Kit (QIAGEN, Hilden, Germany), following the manufacturer's recommended procedure. The library was prepared using the Nextera DNA Flex Library Prep Kit (Illumina Inc. in San Diego, CA.) and sequencing was conducted on an Illumina NextSeq 500 platform (Illumina Inc. in San Diego, CA.). The sequencing process was performed at the CRIBI Biotechnology Center sequencing facility (University of Padova, Italy). Sequencing depth was ranging between 9 and 28 millions of reads (average 14 million), depending on the sample considered. Raw reads were deposited to the SRA portal under the BioProject identifier PRJNA1179988.

### Metagenomics reconstruction of the microbiome

The sequencing reads (150 bp paired) underwent preprocessing using Trimmomatic v0.39.[48] To check for contamination, BBDuck v38.93[49] was utilized. Paired reads were merged with BBMerge v38.93,[49] and short-read co-assembly was conducted using Spades v3.15.5,[50] resulting in an assembly size of 1.39 Gb. The metagenome's percentage of alignment on the reads varies across samples, ranging from 60 to 90% (see Figure S1A). The MAGs acquisition involved multiple binning software: MaxBin v2.2.7,[51] MetaBAT1,[52] MetaBAT2 v2.15,[53] and VAMB v4.13.[54] Subsequently, all recovered bins were refined and dereplicated utilizing DAS_Tool v1.1.6[55] followed by dRep v3.4.0[56] (parameters: -pa 0.9 -sa 0.95), ensuring the recovery of reliable species-level MAGs according to average nucleotide identity. Only medium to high-quality MAGs were retained for subsequent analyses, according to MIMAG standards.[71] Quality and completeness of MAGs were evaluated using checkM2 v1.0.2,[57] and their RA was estimated through coverM v0.6.1,[58] using a range from 0 to 100%. SingleM v0.18.3[59] was used to evaluate the representation of the microbial community by our set of 330 MAGs, considering all marker genes available in the software. Species were classified as not reconstructed only if they had >100 unassembled reads mapping to a marker gene but were absent from the MAGs. Taxonomic assignment was carried out using GTDB-TK v2.1.1[60] and database version R214. Phylogenetic tree was reconstructed using PhyloPhlAn v3.0.[61] Gene prediction was executed using Prodigal v2.6.3[62] and functional annotation with eggNOG-mapper v2.1.10.[63] The KEGG database[72] was used for manual metabolic reconstruction of pathways of interest, including microbial carbon metabolism, hydrogenotrophic methanogenesis and biofilm formation. Plasmids from assembled archaeal sequences were identified using geNomad v1.8.0,[65] while genomic islands were detected using TreasureIsland.[66]

### SNVs profiling over time within species

SNVs analysis was performed using the software InStrain v1.8.0[15] (parameters: –min_read_ani 0.98, –min_mapq 2, –min_cov 5, –min_freq 0.05, –fdr 1e−06, –min_genome_coverage 1, –skip_plot_generation, –skip_mm_profiling) on the high-quality MAGs recovered. Species identified in samples with less than 50% breadth of coverage were excluded from subsequent downstream analysis. Then, MAGs with RA > 2% were selected based on variant metrics, and the STRONG[64] strain deconvolution pipeline was applied. The abundances of the deconvoluted strains were defined taking into account the RA of the corresponding MAG. A variant phasing approach was performed by clustering the SNVs frequency and then comparing this information with the abundance of predicted strains, as previously described.[23] The approach was based on a hierarchical clustering algorithm, using the Ward distance as a metric for calculating variant similarity. This procedure allowed to group together SNVs most probably belonging to the same strain, enabling to track the transmission of genetic modifications over time.

### SNVs filtering, SNVs sweeps identification and $d_N/d_S$ ratio

The results obtained from the variant calling using InStrain were processed in order to remove low confidence results. A total of 366,298 and 327,035 SNVs were originally detected across TBR1 and TBR2 reactors, respectively. First, all the SNVs located within 150 bp from the 3′ and 5′-end of each scaffold were removed, since in those genomic regions the coverage tends to drop along with the results reliability. Then, all the variants where the difference between the coverage of the SNV and the average coverage of the scaffold exceeded the interval [−100; +100] were discarded. Lastly, the ratio between the number of reads supporting the variant allele and the one supporting the reference allele was computed and all the SNVs where this ratio was lower than 0.1 in all time points were discarded. This process resulted in the removal of an average of 15.37% of the initial SNV total: 58,014 and 48,696 from TBR1 and TBR2 reactors, respectively. The detected SNVs were used to estimate the overall levels of genetic diversity in Figure 2, defined as the fraction of sites in the ‘‘core genome’’ in which the alternate allele was present at an intermediate frequency ($0.1 < f < 0.9$). Additionally, allele frequency trajectories were used to identify subsets of SNVs that experienced large changes in frequency between pairs of sequenced time points. We refer to these as ‘‘sweeping SNVs,’’ because they indicate a full or partial ‘‘sweep’’ of the allele through the resident population of interest. For each ordered pair of time points, we identified SNVs with minor allele frequencies that increased from $f < 0.3$ at the initial time point to $f > 0.60$ at the subsequent time point. Finally, the $d_N/d_S$ ratio was calculated at the pathway and gene level, including in the calculation all the synonymous and nonsynonymous SNV that passed the filtering step. Ratios above +1 were considered as indicators of selective pressure.

### 3D reconstruction of mutated protein

Genes harboring high Gd distance (>70) and $d_N/d_S$ ratios (>1) in the two dominant archaeal species were selected for 3D reconstruction through machine learning approaches. The predicted sequence of the proteins[62] were modeled with AlphaFold[32] to obtain the 3D structure of the mutated polypeptides and compare them with the original ones. Moreover, the sequence of the proteins was aligned using the multiple sequence alignment tool Clustal Omega,[67] based on the ClustalW algorithm, against reference crystallographic structure deposited in the Swiss-Prot database.[73] This comparison allowed us to evaluate the potential impact of the SNVs on the investigated protein, localizing the SNV-encoded AA changes along the 3D model. Alignments were visualized using the Jalview v2.11.3.0 software.[68] ChimeraX v1.9 was employed for PDB visualization.[69]