

Metabolic interdependencies in thermophilic communities are revealed using co-occurrence and complementarity networks

Received: 8 March 2024

Accepted: 9 September 2024

Published online: 17 September 2024



Xi Peng^{1,2}, Shang Wang¹, Miaoxiao Wang^{3,4}, Kai Feng¹, Qing He¹, Xingsheng Yang^{1,2}, Weiguo Hou⁵, Fangru Li⁵, Yuxiang Zhao⁶, Baolan Hu^{7,8,9}, Xiao Zou¹⁰ & Ye Deng^{1,2} 

Microbial communities exhibit intricate interactions underpinned by metabolic dependencies. To elucidate these dependencies, we present a workflow utilizing random matrix theory on metagenome-assembled genomes to construct co-occurrence and metabolic complementarity networks. We apply this approach to a temperature gradient hot spring, unraveling the interplay between thermal stress and metabolic cooperation. Our analysis reveals an increase in the frequency of metabolic interactions with rising temperatures. Amino acids, coenzyme A derivatives, and carbohydrates emerge as key exchange metabolites, forming the foundation for syntrophic dependencies, in which commensalistic interactions take a greater proportion than mutualistic ones. These metabolic exchanges are most prevalent between phylogenetically distant species, especially archaea-bacteria collaborations, as a crucial adaptation to harsh environments. Furthermore, we identify a significant positive correlation between basal metabolite exchange and genome size disparity, potentially signifying a means for streamlined genomes to leverage cooperation with metabolically richer partners. This phenomenon is also confirmed by another composting system which has a similar wide range of temperature fluctuations. Our workflow provides a feasible way to decipher the metabolic complementarity mechanisms underlying microbial interactions, and our findings suggested environmental stress regulates the cooperative strategies of thermophiles, while these dependencies have been potentially hardwired into their genomes during co-evolutions.

Microbial ecosystems thrive on intricate webs of interactions, often obscured by their vast diversity and hidden by limitations in conventional culture-based methods^{1,2}. While studying pairwise interactions provides valuable insights^{3,4}, understanding the complete picture within natural complex communities remains challenging⁵. With the

advances in sequencing and computational modeling, microbial co-occurrence networks, constructed from patterns of species co-occurrence across environmental samples, have become a powerful tool to infer interspecies connections^{6–8}. The use of co-occurrence network analysis became popular after several ready-to-use workflows

with user-friendly visualizations were published^{9–14}. However, as several previous studies argue, co-occurrence alone does not provide conclusive evidence of ecological interactions^{15–17}. To address this limitation, we combine co-occurrence data with evidence of metabolic complementarity to infer more robust interactions and explore potential underlying mechanisms.

Metabolic dependencies have been proposed as a major driver of species co-occurrence¹⁸. However, it has also been suggested that species co-occurrence could in turn drive metabolic dependencies². This highlights the need for a nuanced understanding of the dynamic interplay between metabolic interactions and microbial community assembly. Long-read sequencing and improved assembly enable high-quality metagenome-assembled genomes (MAGs), opening doors to map metabolic cooperation and competition in complex communities through nutrient exchange^{19,20}. Although genome-scale metabolic models (GSMMs), along with flux balance analysis (FBA), have shown potential in predicting metabolite exchange and growth under curated laboratory conditions^{21,22}, there remains substantial potential for further exploration in this area. Integrating these models with co-occurrence networks to validate and refine predictions of metabolic dependencies within uncontrolled natural habitats is an untapped frontier.

Geothermal springs, characterized by extreme temperatures, fluctuating ionic strengths, and enriched minerals, harbor a unique community of thermophiles with specialized physiological and metabolic adaptations^{23–25}. These adaptations often involve genome streamlining, leading to reduced genomes and temperature-adapted proteins^{26–31}. Yet, thermophiles exhibit remarkable genomic plasticity and metabolic flexibility³². Metabolic reconstructions of some geothermal phyla suggest dependence on interspecies exchange of amino acids, vitamins, and cofactors, implying close syntrophy^{33,34}. This is further supported by observations of highly cooperative communities formed by small-genome auxotrophs, where cross-feeding promotes mutual benefit^{35–37}. This suggests that harsh geothermal conditions while restricting growth, may simultaneously drive synergistic interactions. Nevertheless, this hypothesis requires a systemic examination. Addressing this question may yield insights into the ongoing debate regarding the prevalence of either antagonistic or synergistic interactions within natural microbial communities, as well as how environmental stresses affect their prevalence^{38–42}.

Here, we leverage the power of random matrix theory (RMT) for robust co-occurrence and metabolic network constructions while tapping into the rich information contained within MAGs^{9,43}. This synergy allows us to identify species co-occurrence patterns and predict potential metabolic interdependencies based on metabolic pathways. Adopting this powerful tool, we show that metabolic complementarity intensifies with rising temperatures, as thermophiles under heat stress increasingly rely on interspecies exchange of essential metabolites for survival. Additionally, we demonstrate that phylogenetic distance dictates cooperative strategies, with closely related thermophiles competing for similar resources while distantly related ones engage in mutually beneficial metabolic exchanges. Furthermore, our findings reveal that species with small genomes, potentially lacking vital metabolic pathways, depend heavily on metabolic partnerships for survival and growth. Our research offers a deeper understanding of microbial synergy in challenging environments, providing valuable insights into the interplay between environmental stress, metabolic dependencies, and the evolution of cooperative strategies within microbial communities.

Results

Network construction and thermal preferences of thermophiles

To comprehensively analyze both co-occurrence and metabolic complementarity networks among thermophiles, we developed a unique

bioinformatic workflow (Fig. 1) and seamlessly integrated it into the iNAP platform (<https://inap.denglab.org.cn>)⁴⁴ (Supplementary Fig. 1), as updated iNAP 2.0⁴⁵. This workflow leverages multiple metabolic complementarity indices, infers the network thresholds of co-occurrence and metabolic complementarity using the random matrix theory (RMT), extracts shared interactions within both network inferences, and identifies potential exchangeable metabolites.

A total of 449 million Illumina short reads (223.35 Gb of raw bases, $N=40$) and 51 million Nanopore long reads (52.01 Gb of raw bases, $N=40$) were obtained from the hot spring sediment samples across a temperature gradient (63.5 °C–85.8 °C), with eight samples each having five replicates. The Nonpareil method confirmed the estimated average coverage of each sample above 80%, indicating sufficient sequencing depth (Supplementary Fig. 2). There is an increasing trend of sequencing diversity (N_d) from the high- to low-temperatures (Fig. 2a), indicating that the environmental conditions of extreme thermal temperatures have reduced the sequence diversity of its microbial communities, and the inferred species diversity might be more specialized. After assembly, binning, and CheckM quality control, 401 medium- and high-quality metagenome-assembled genomes (MAGs) were retained for subsequent network analysis.

The taxonomic assignment revealed that 85.78% of MAGs belonged to 38 bacterial phyla, whereas 57 MAGs were affiliated with seven archaeal phyla (Fig. 2b, Supplementary Data 1). The MAG abundance varied significantly across temperature ranges (PERMANOVA, $p < 0.001$) (Fig. 2c), leading to the classification of sampling sites into three groups: extremely thermal (ET, 78.5–85.8 °C, 2 samples \times 5 replicates), highly thermal (HT, 67.5–73.9 °C, 3 samples \times 5 replicates), and moderately thermal (MT, 63.5–65.8 °C, 3 samples \times 5 replicates). Most MAGs exhibited distinct temperature preferences, with only a few evenly distributed across the gradient. For example, 80.52% of the relative abundance assigned to *Thermoproteota* MAGs originated from the ET group, highlighting this archaea's strong preference for extremely high-temperature habitats. In contrast, *Cyanobacteria* and *Bacteroidota* preferred the cooler group, with 96.28% and 96.98% of relative abundance originating from HT and MT groups (Fig. 2d, Supplementary Data 2).

Archaea also displayed clear temperature preferences despite accounting for only 7.31% of relative abundance across all samples. While *Thermoproteota* favored the ET group, *Micrarchaeota*, the second most abundant, exhibited a unique distribution, with nearly equal contributions from HT and MT groups. These findings underscore the diverse thermal adaptations and preferences of thermophile communities within this hot spring ecosystem.

The co-occurrence patterns of thermophiles within three temperature groups

Three co-occurrence networks were constructed using the relative abundance information, employing the RMT-based method (Fig. 3, Supplementary Data 3). Positive edges constituted the majority in the three networks (ET: 98.82%; HT: 75.54%; MT: 77.99%). Post-extraction of positive edges as subnetworks, the R^2 values of the power-law model for the co-occurrence networks of the HT and MT groups were 0.59 and 0.53, respectively, but the ET group could not fit well to the power-law distribution, losing its scale-free property. We found that the ET group subnetwork exhibited a higher network density, shorter average path distance (GD), lower harmonic geodesic distance (HD), higher average clustering coefficient, and reduced modularity compared to the other two subnetworks (Table 1). These characteristics suggested a tighter interaction structure in the ET group subnetwork than in the other two subnetworks. A harsh environment (high temperature) induces tight and synergistic interactions among thermophiles. Furthermore, the global topological properties of the three networks were significantly different from the respective randomized networks (generated 100 times), demonstrating that all observed interactions

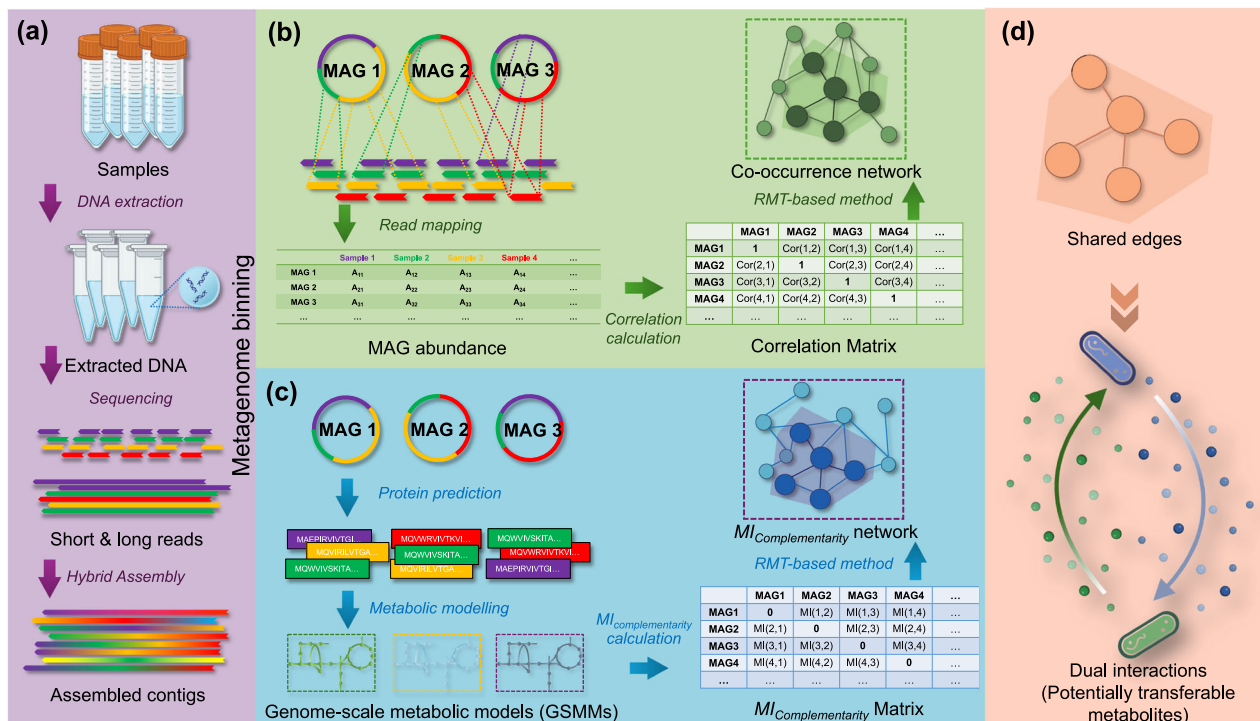


Fig. 1 | A schematic diagram of our study workflow. The entire analytical procedure can be segmented into four distinct parts. **a** Samples were collected, and DNA extraction was performed. Long contigs were obtained by hybrid assembly of short and long reads. **b** MAG abundance was calculated by read-length mapping after genome binning. **c** The genome-scale metabolic model of MAG was constructed by protein prediction after genome binning, and the pairwise metabolic

complementarity index ($MI_{\text{complementarity}}$) of MAGs was calculated. **d** An RMT-based method was used to generate a co-occurrence network and a metabolic complementarity network, and their common edges were analyzed further to provide metabolic clues for potential co-occurrence of species. This figure was created with BioRender.com released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International license.

filtered by RMT cutoff were not randomly connected (Table 1, one sample *t*-test).

The inferred pairwise metabolic complementarity among thermophiles

To assess metabolic dependencies, we reconstructed genome-scale metabolic models for each MAG. We defined the PhyloMint metabolic complementarity index ($MI_{\text{complementarity}}$) to quantitatively assess the degree of metabolic dependencies between each pair of MAGs within three temperature groups (see “Methods” section). Specifically, 29 MAGs were unique to the ET group, while 70 MAGs were detected across all temperatures, likely representing heat-tolerant generalists. Both approaches revealed a surprising rarity of synergistic metabolic interactions (Fig. 4), while the RMT-based threshold yielded an even more stringent classification, with less than 3% of interactions deemed significant (ET: 4.13%; HT: 4.68%; MT: 6.46%) (Fig. 4).

Notably, observed synergistic interactions displayed a marked asymmetry of $MI_{\text{complementarity}}$ values under all three groups. We categorized these into mutualistic (both $MI_{\text{complementarity}}$ values of pairwise MAGs exceeding the threshold) and commensalistic (only one $MI_{\text{complementarity}}$ value of pairwise MAGs exceeding the threshold) metabolic interactions. Regardless of the threshold used, mutualistic interactions were extremely rare, constituting less than 0.3% of all pairs in any group (Fig. 4). This suggests that metabolism-based synergy within the hot spring community is primarily driven by unidirectional feeding, with one thermophile benefitting from the metabolic products of another. In particular, the species pairs involved in commensalistic and mutualistic interactions exhibit dramatically different patterns regarding genome size differences. Across the three temperature groups, the average genome size differences (estimated genome size of the giver minus that of the taker) identified as commensalistic interactions by the RMT threshold were ET: 716 Kbp, HT: 1204 Kbp, and

MT: 879 Kbp. However, the average genome size differences for mutualistic interactions were close to 0 for all three temperature groups. In all commensalistic interactions, the proportion of interactions where the estimated genome size of the taker is less than 2 Mbp and that of the giver is more than 2 Mbp accounted for as high as 31.29% (87 pairs) in the ET group, while in the HT and MT groups, these values were 37.01% (624 pairs) and 29.87% (1057 pairs), respectively. These results indicated that most metabolic dependency between one streamlined genome and another more comprehensive genome is a more prevalent pattern in commensalistic interactions.

Following RMT threshold determination (Fig. 4), the constructed metabolic networks for the ET group posed more robust scale-free properties than that of the occurrence networks (Table 2). In addition, the topological properties of metabolic networks varied across temperatures. The hottest ET network, despite its smaller size, exhibited the highest average clustering coefficient, highest density, and shortest communication paths (Table 2), suggesting a highly interconnected community in this harsh environment. However, the patterns in average degree, average path distance, and harmonic geodesic distance were opposite to those of the co-occurrence networks. Such findings illustrated that metabolic complementarity and co-occurrence networks do not exhibit the same characteristics of microbial communities, which were then reflected in the topology of the network. The hub node identification using $z_i - P_i$ analysis illuminated the key players in the metabolic networks (Supplementary Data 4). The ET network, with 36 hubs, hosted 6 archaeal hubs (3 connector hubs, 1 module hub, and 2 network hubs), including 4 assigned to *Thermoproteota*, 1 to *Asgardarchaeota*, and 1 to *Aenigmataarchaeota*. The larger HT and MT networks revealed diverse hubs (137 and 97), while both contained hubs from *Thermoproteota*, *Micrarchaeota*, and *Methanobacteriota*. These results indicated that despite their low abundance, archaea, particularly *Thermoproteota* and

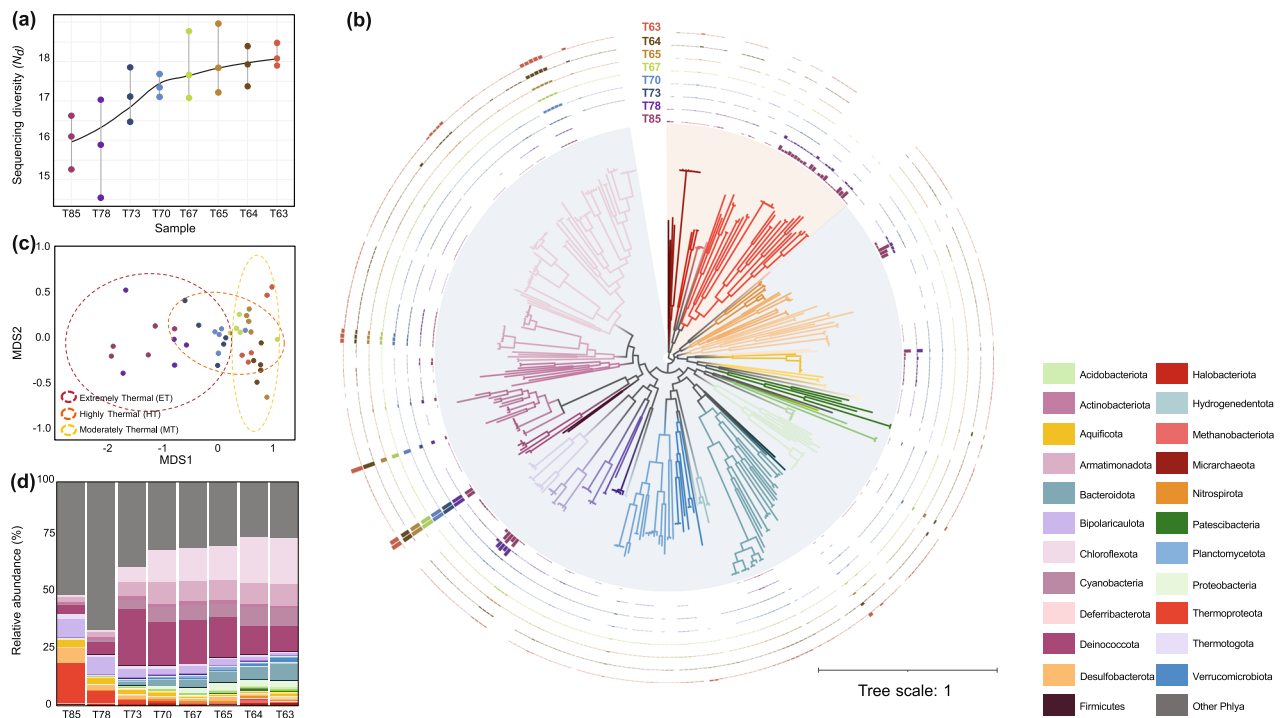


Fig. 2 | Display of sequence diversity and MAG taxonomic distribution across samples. **a** Sequence diversity (N_d) for each temperature group, quantified using Illumina sequencing. Each lollipop plot represents a specific temperature group, with the endpoints indicating the maximum and minimum N_d values observed across five replicates. The central point on each lollipop plot denotes the average N_d value for the respective temperature group, calculated from the same five replicates. The smoothed line was drawn using the loess method to display the

trend. **b** The phylogenetic tree of 401 medium- and high-quality MAGs. The background color of the tree represented the domain of MAGs (light blue for bacteria, light red for archaea). Branches were colored as the representations of the phylum of MAGs. The bar plot in the outer circle represents the relative abundance of MAGs. **c** Non-metric multidimensional scaling (NMDS) analysis was used to show the among-sample dissimilarities, and samples were grouped into three temperature categories. **d** Relative abundance at the phylum level.

Methanobacteriota, emerge as essential hubs across all temperature networks, demonstrating their unique metabolic capabilities and indispensable roles in the thermophile communities.

To understand how microbial interactions arise, we investigated the relationships between phylogenetic distance (Faith's PD) and the co-occurrence patterns/ metabolic complementarity. The co-occurrence strength, measured by Spearman's correlation, negatively correlated to PD in all temperature groups (Fig. 5a–c). However, a fascinating pattern emerged after RMT filtering (cutoff = 0.830). The density distribution of PD exhibited two distinct peaks: one near 0 and another near 2. This suggested two potential drivers of co-occurrence: the close association of similar species with shared resource requirements (PD at peak 0) and the co-occurrence of phylogenetically distant species with complementary metabolic needs (PD at peak 2). Besides, significant positive correlations between $MI_{complementarity}$ and PD were observed in all temperature groups (Fig. 5d–f), indicating that distantly related MAGs were more likely to exhibit strong metabolic dependencies. This trend was particularly pronounced in the hottest temperature. Notably, after RMT filtering, those MAG pairs of PD lower than 1 became rare, with most values clustering around 2. It suggested that the analysis of metabolic complementarity omitted co-occurrences based on phylogenetic proximity, highlighting the importance of metabolic complementarity between distantly related species.

Genomic clues to species metabolic synergies

Genome sizes and contents were shown to be crucial connecting links for metabolic interactions within the hot spring community, which also were intricately linked to nutrient requirements. The estimated genome sizes of MAGs distributed at different temperatures

underwent a transition from negative to positive correlation with temperature from the ET group (T85 and T78) to lower temperature (Fig. 6a). In harsh environments with extremely high temperatures, small-genome species tend to occupy higher abundances. However, when the environmental temperature drops to a range where most species can adapt, the small-genome species no longer hold this abundance advantage and tend to occupy only a small abundance in the environment. Genome size was also confirmed to correlate with the proportion of genes in the genome that perform different functions (Fig. 6b). Estimated genome size showed a significant positive correlation with the proportion of genes involved in *Secondary metabolites biosynthesis, transport and catabolism* (COG-Q, Spearman's $Rho = 0.670$, $p < 0.001$), *Carbohydrate transport and metabolism* (COG-G, Spearman's $Rho = 0.473$, $p < 0.001$), *Lipid transport and metabolism* (COG-I, Spearman's $Rho = 0.234$, $p < 0.001$) and *Inorganic ion transport and metabolism* (COG-P, Spearman's $Rho = 0.125$, $p < 0.05$). Conversely, smaller genomes prioritized housekeeping functions like *Translation, ribosomal structures and biogenesis* (COG-J, Spearman's $Rho = -0.932$, $p < 0.001$), *Nucleotide transport and metabolism* (COG-F, Spearman's $Rho = -0.634$, $p < 0.001$) and *Replication, recombination and repair* (COG-L, Spearman's $Rho = -0.462$, $p < 0.001$). Furthermore, a linear mixed-effect model was constructed to reveal that when genome size was set to be random effect, differences in biological functions such as energy production and conversion, metabolism of nucleotide, amino acid and lipid had a strong positive effect on $MI_{complementarity}$ (Fig. 6c). The $MI_{complementarity}$ of pairwise MAGs significantly increased with the difference in genome sizes (Fig. 6d–f). This correlation was strongest in the ET group, suggesting that genome size plays more crucial role in shaping metabolic partnerships in harsh environments. These results reflected a trade-off of smaller size

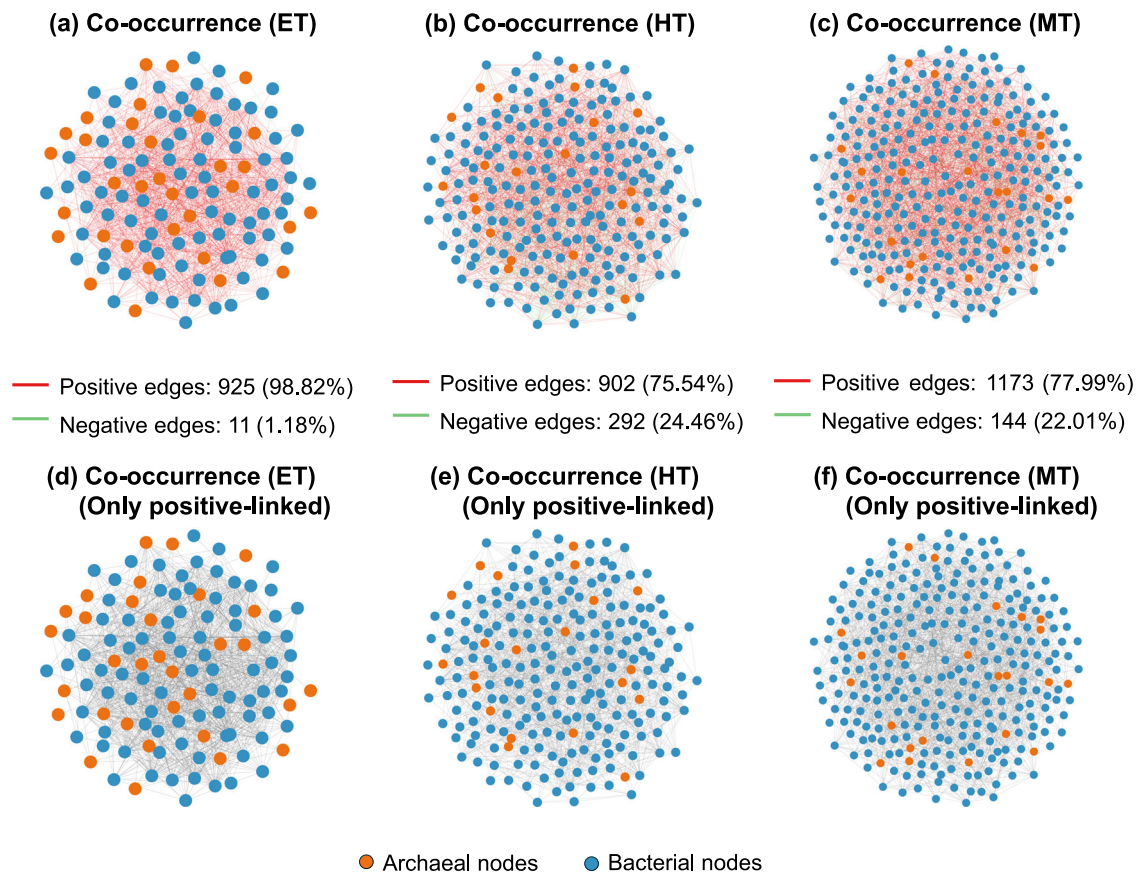


Fig. 3 | The co-occurrence networks of three temperature groups and their corresponding positively-linked subnetworks. a–c Network visualizations of co-occurrence networks of three temperature groups. **d–f** Corresponding positively-linked subnetworks of three temperature groups.

Table 1 | The statistical properties of the co-occurrence networks (positive links) of MAGs within three temperature groups

Global Properties	ET	HT	MT
RMThreshold	0.830		
Total nodes	119	230	294
Total links	925	902	1173
R square of power-law	0.100	0.587	0.529
Average degree (avgK)	15.546	7.843	7.980
Average clustering coefficient (avgCC)	0.745	0.740	0.710
Average path distance (GD)	2.280	2.459	4.476
Harmonic geodesic distance (HD)	1.738	1.818	3.050
Density	0.132	0.034	0.027
Modularity (No. of modules)	0.387 (9)	0.752 (30)	0.757 (33)
Randomized Network Indexes (100 iterations, data are presented as mean values ± standard deviation)			
Average clustering coefficient (avgCC)	0.331 ± 0.016***	0.07 ± 0.008***	0.042 ± 0.006***
Average path distance (GD)	2.206 ± 0.023***	2.904 ± 0.034***	3.003 ± 0.021***
Harmonic geodesic distance (HD)	0.512 ± 0.004	0.382 ± 0.003	0.366 ± 0.002
Modularity	1.953 ± 0.014***	2.62 ± 0.02***	2.729 ± 0.014***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ ($N = 100$, One sample t-test, two-sided).

genomes, where the genes associated with genetic information storage and processing were retained to a greater degree while genes related to various metabolisms underwent loss, relying on synergistic interactions to compensate for their limited metabolic repertoire. However, it is essential to consider that the absence of specific genes might also be due to incomplete genome reconstruction, given the accepted cutoff for MAG completeness at 50%.

More genomic clues of species synergies were found in the overlapped partnerships between co-occurrence and metabolic complementarity networks. Only a handful partnerships (7, 49, and 58 in the ET, HT, and MT groups, respectively) exhibited strong metabolic complementarity alongside persistent co-occurrence (Supplementary Data 5). Among these partnerships, 274 metabolites were detected as potentially transferable, while 58 were

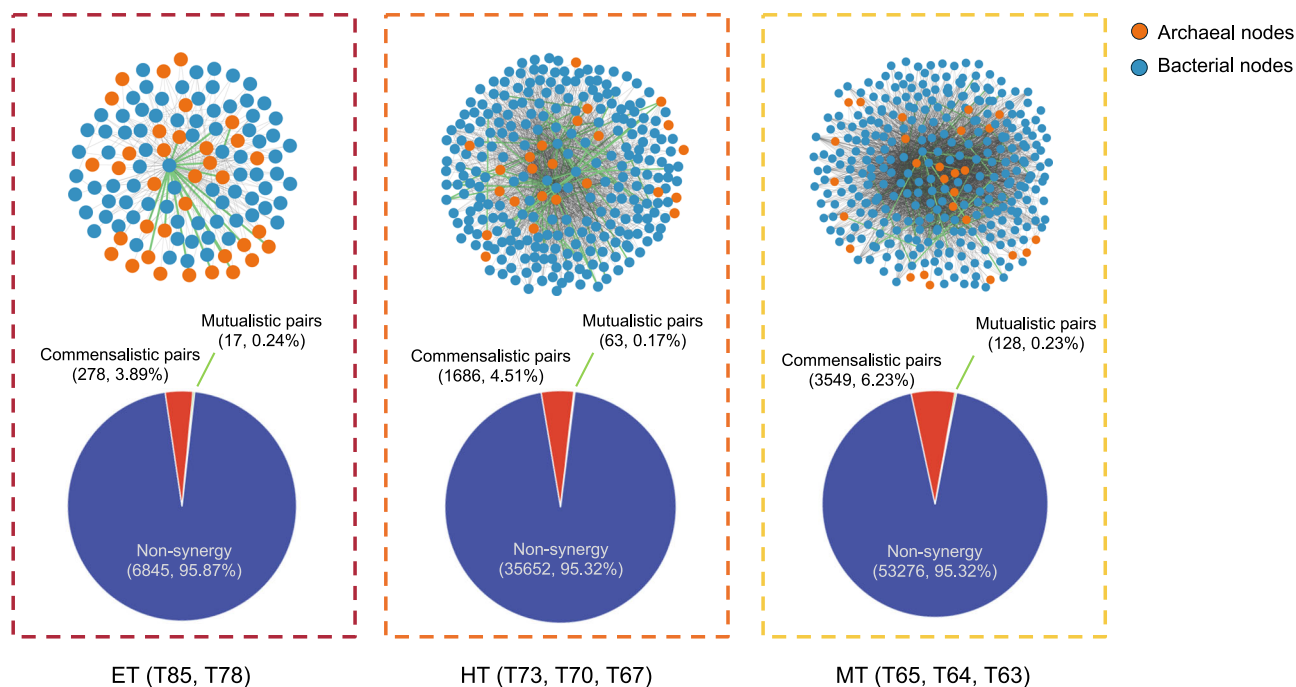


Fig. 4 | The distributions of three categories of interactions based on the pairwise $MI_{complementarity}$ values. In network graphs, gray links represent asymmetrical synergy (commensalism), while green links represent symmetric synergy (mutualism).

coenzyme A derivatives. Amino acids (with peptides and analogs) and carbohydrates (and carbohydrate conjugates) followed, with 43 and 36, respectively. There were also 30 molecules classified as nucleosides, nucleotides, and analogs deemed to be potentially transferrable. Notably, between those paired species, the types of metabolites potentially transferred from one to the other were quite different (Supplementary Data 5).

In exploring the intricacies of species metabolic synergies, our focus was drawn to 17 archaea-bacteria interactions, which were supported by both metabolic networks and co-occurrence networks, indicating that inter-domain synergies were prevalent in harsh environments such as hot springs. In these interactions, there were some MAGs that exhibited the properties of co-occurrence and metabolic complementarity with multiple other species. For instance, an archaeal MAG (T64.bin.92, *Micrarchaeota* phylum) had seven distinct bacterial partners (Fig. 7). This centrally positioned MAG, characterized by its smaller estimated genome size of 0.857 Mbp, predominantly assumed the role of a taker in these interactions. This was evidenced by its substantial reliance on obtaining several types of metabolites, such as some coenzyme A derivatives and other crucial metabolic need that it cannot fulfill endogenously, from other species. Despite its taker centric position, T64.bin.92 contributed to these commensalistic relationships by providing potential surplus materials like carbohydrates and nucleoside/nucleotide-related substances, embodying a reciprocal dynamic in these interactions. This necessity for external coenzyme A derivatives, vital for biosynthesis reactions involving acyl transfers, underlines the significant dependence of T64.bin.92 and its partners. Furthermore, its requirement for certain inorganic substances, including various metal ion and non-metal compounds from its metabolic counterparts, further illustrated the complexity and significance of its role as a taker. However, it is important to note that these conclusions are based on genetic potential inferred from genomic data. To confirm the actual metabolic exchanges and interactions, additional validation with functional omics and metabolomics data is necessary.

The validation of our study on composting system experiment

In hot spring habitats, we noticed that symmetric interactions (mutualism) were rare. Besides, in commensalistic and mutualistic interactions, the genome size differences between species pair varied. In a similar vein, our findings from a composting system where temperature is a key stress factor mirror those from hot springs. A low occurrence of mutualistic interactions were observed across different sampling time (temperature) groups: D00 ($T_{avg} = 26.02^\circ\text{C}$), D05 ($T_{avg} = 64.29^\circ\text{C}$), and D25 ($T_{avg} = 41.35^\circ\text{C}$), with mutualism rates at 0.23%, 0.087%, and 0.23%, respectively. Asymmetrical interactions, suggesting commensalism, were slightly more common at 4.23%, 2.73%, and 4.18%. The genome size differences in commensalistic interactions at D05 were around 1.36 Mbp, similar to findings in hot springs, whereas mutualistic interactions across all groups showed negligible genome size differences (Supplementary Fig. 3). The correlation between phylogenetic distance and pairwise $MI_{complementarity}$ was also observed to be significant in composting system. Specifically, D05, as the highest temperature sample in the experiment, showed the strongest positive correlation. Furthermore, the pattern that the greater the genome size difference between two species corresponds to higher their complementarity potential was confirmed by testing the significance of the correlation between $MI_{complementarity}$ and genome size difference (D00: Spearman's $Rho = 0.122$, $p < 0.001$; D05: Spearman's $Rho = 0.213$, $p < 0.001$; D25: Spearman's $Rho = 0.090$, $p < 0.001$).

The sequencing data from the composting experiment recovered few overly streamlined genomes, suggesting that such genomes might arise from long-term selection rather than short-term temperature changes. Nevertheless, we observed asymmetrical synergistic interactions, with a significant proportion being commensalistic (Supplementary Fig. 4). These interactions showed a considerable difference in genome size, particularly in the highest temperature group (genome size difference: D00, 815 Kbp; D05, 1368 Kbp; D25, 525 Kbp), similar to the hot spring habitats. This implies that high temperatures are a crucial factor driving synergy between genomes of vastly different sizes.

Table 2 | The statistical properties of the metabolic complementarity networks of MAGs within three temperature groups

Global Properties	ET	HT	MT
RMThreshold	0.338		
Total nodes	116	273	338
Total links	295	1749	3677
R square of power-law	0.395	0.495	0.418
Average degree (avgK)	5.086	12.813	21.757
Average clustering coefficient (avgCC)	0.652	0.415	0.378
Average path distance (GD)	2.003	1.990	1.969
Harmonic geodesic distance (HD)	1.944	1.933	1.899
Density	0.044	0.047	0.065
Modularity (No. of modules)	0.237 (5)	0.190 (4)	0.151 (3)
Randomized Network Indexes (100 iterations, data are presented as mean values ± standard deviation)			
Average clustering coefficient (avgCC)	0.003 ± 0.004***	0.004 ± 0.005***	0.014 ± 0.006***
Average path distance (GD)	1.995 ± 0.006***	1.983 ± 0.006***	1.96 ± 0.004***
Harmonic geodesic distance (HD)	0.516 ± 0.001	0.519 ± 0.001	0.528 ± 0.001
Modularity	1.939 ± 0.004***	1.929 ± 0.004***	1.893 ± 0.002***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ ($N = 100$, One sample t-test, two-sided).

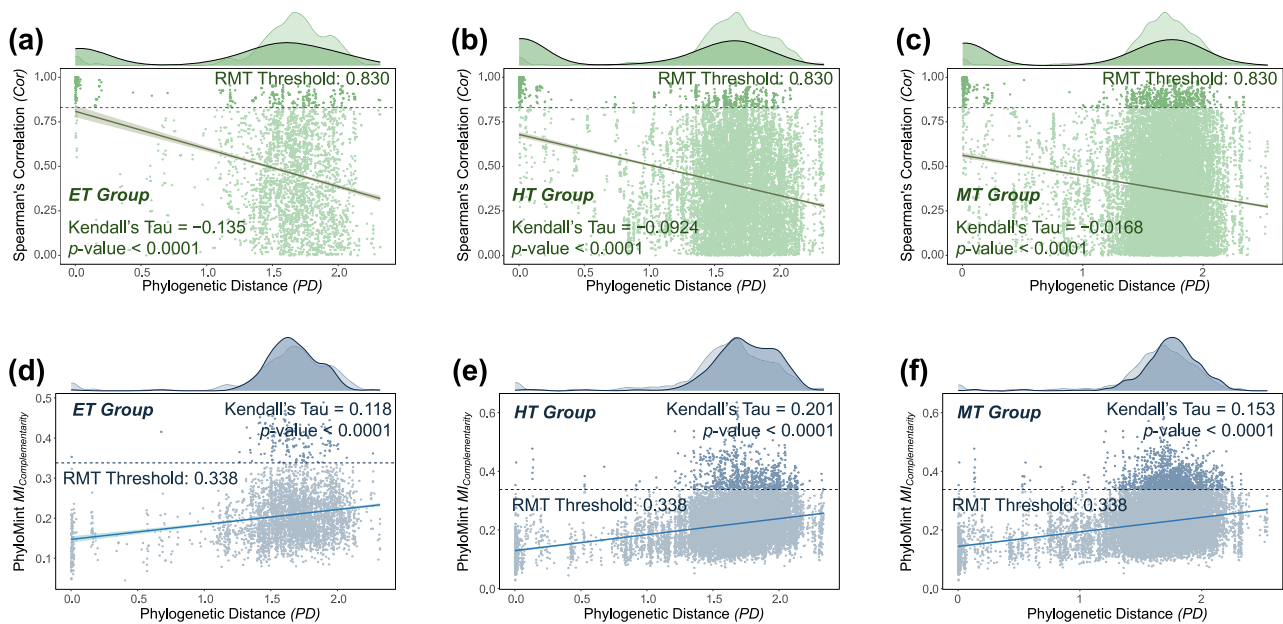


Fig. 5 | Scatter Plot of relationships between phylogenetic distance (PD) and Spearman's correlation (Cor)/PhyloMint Complementarity Index ($MI_{complementarity}$). **a–c** The correlations between Spearman's correlation (Cor) and phylogenetic distance (PD) of three temperature groups. **d–f** The correlations between PhyloMint complementarity index ($MI_{complementarity}$) and phylogenetic distance (PD) of three temperature groups. **a–f** The dashed lines indicate the threshold of RMT. Dark points represent the interactions (edges) selected into

networks, and light points are not. The fitted curve was modeled using the generalized linear model and the shaded area indicates the 95% confidence interval. The correlation was statistically tested using Kendall's rank correlation test (two-sided). The density charts describe the distribution of interactions. The density curves above each plot indicate the distribution of interactions along PD. The inset graphs present the 2-dimensional density distribution of nodes.

Discussion

Although some evidence has indicated that metabolic exchanges are ubiquitous, detecting those synergistic metabolic interactions in natural communities is still highly challenging². Here, we built a bioinformatic workflow integrating co-occurrence with metabolic network approaches using metagenomics sequencing datasets. There are two primary considerations for this workflow construction. First, either co-occurrence or metabolic network approaches have their advantages and defects. The current metabolic network approaches mainly infer synergistic interaction by measuring the metabolic complementarity

dependency or metabolite exchange potential between any two microbial genomes^{18,46,47}. However, those two species may be out of synergy in time and space due to lack of contact. Conversely, co-occurrence network approaches infer species relationships in a series of real natural communities, but they only reflect superficial co-occurrence patterns that might not indicate true ecological interactions^{16,17,48}. Therefore, we aim to combine the strengths of both approaches to address their limitations (Fig. 1). Second, we applied the random matrix theory (RMT) to construct the metabolic network models, alleviating unreliable synergistic interactions. The RMT

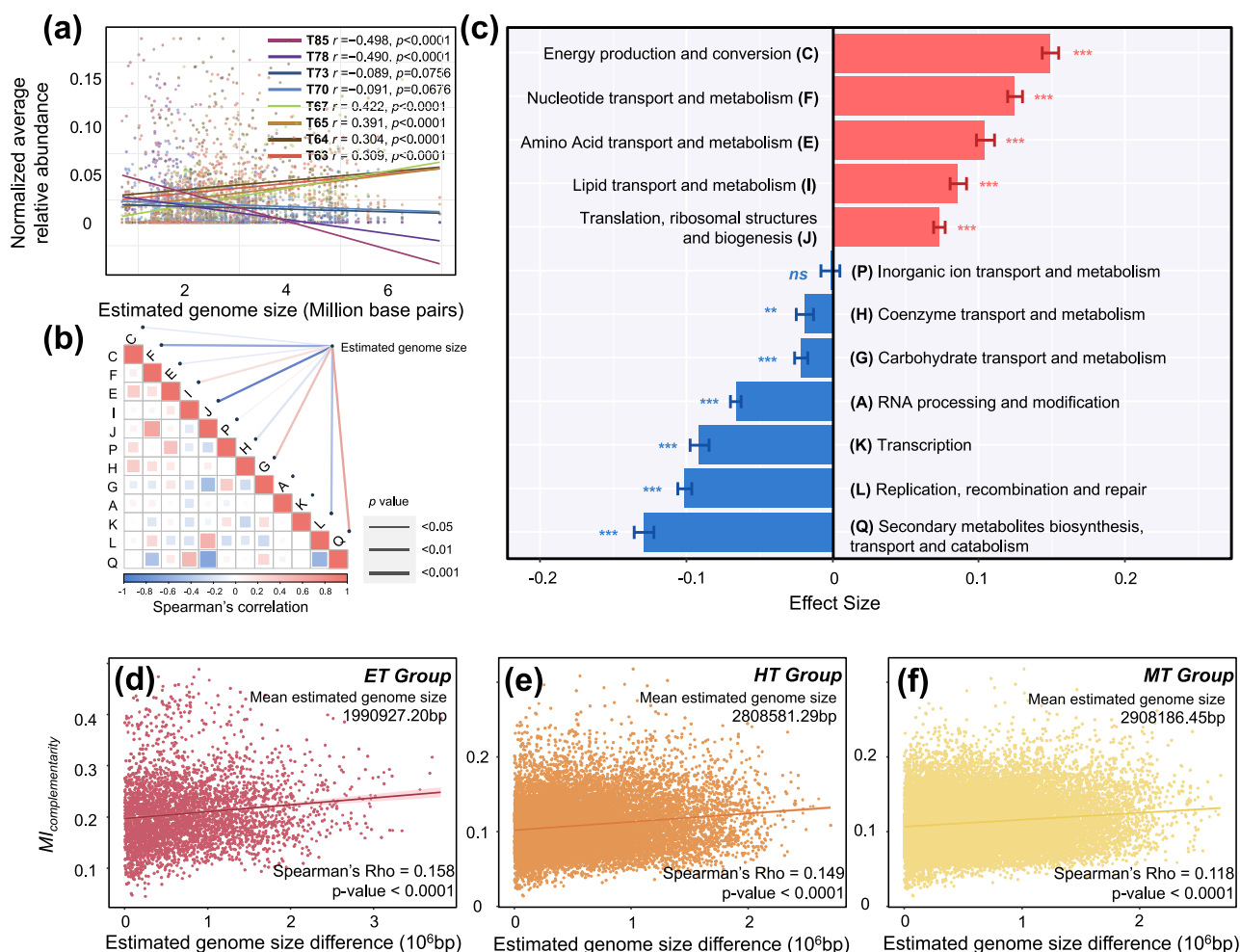


Fig. 6 | Relationships between $MI_{\text{complementarity}}$ and estimated genome size and biological function proportion. **a** Correlation (Spearman's rank correlation test, two-sided) between normalized average abundance of each temperature group and the estimated genome size. **b** Correlation matrix on proportions of twelve COG biological functions in MAGs. The lines between estimated genome size and each COG biological function indicated Spearman's correlation (two-sided). The color shade of the lines represented the strength of the correlation, and the thickness represented the confidence level. **c** Effect of COG biological function proportion on $MI_{\text{complementarity}}$ by linear mixed-effects models (LMMs) fit by restricted maximum

likelihood (REML) estimation. Data are presented as mean values \pm standard errors of the estimated effect sizes. Bar length and error bar indicated the mean values and standard errors of the estimated effect sizes. Significance test was conducted using a t -test of Satterthwaite's methods (number of observations $n = 80,200$ for non-self-loop pairwise interactions between 401 MAGs). Significant effects are represented by asterisks: *** $p < 0.001$, ** $p < 0.01$, 'ns' stands for not significant. **d–f** Correlation between $MI_{\text{complementarity}}$ and genome size difference. The fitted curve was modeled using the generalized linear model. The correlation was tested using Spearman's rank correlation test (two-sided).

method could obtain the lowest false positives in correlation-based networks^{49,50}. In a synergistic metabolic network, there is a similar requirement for threshold determination on metabolic complementarity dependencies. Therefore, we established a publicly available workflow integrated into our iNAP pipeline to facilitate analyses based on both co-occurrence and metabolic networks.

Using this workflow, we yield some insights into how thermophiles within geothermal ecosystems adapt to the scorching heat through their synergies. Both co-occurrence and metabolic complementarity networks revealed that the network density increased significantly with the rising temperatures (Figs. 3 and 4). It suggested that thermophiles may form tighter connections under extreme heat and engage in more frequent material exchanges. These collaborations could serve as a crucial adaptation strategy, fostering community survival and enhancing resistance stability^{51,52}. Some previous studies have demonstrated that thermophiles may prioritize amino acid assimilation to cope with nitrogen limitation⁵³, while some may even adapt by exchanging or acquiring DNA from the environment⁵⁴. Our metabolic networks provided additional evidence that the species

achieved high community function stability at high temperatures by enhancing the efficiency of resource and information transfer (Fig. 6c, Supplementary Data 5).

Our analysis also revealed some genomic clues for metabolic complementarity among thermophiles. As species diverged further on the evolutionary tree, their metabolic synergy intensified, particularly at extreme temperatures (Fig. 4). This aligns with previous findings that species with less overlap in their metabolic abilities find greater benefit in partnering with distant relatives⁴². This observation likely stems from underlying genomic differences. Distant phylogenetic relationships often translate to greater genomic disparities, leading to variations in essential compounds needed for survival^{35,55,56}. Furthermore, we also found that thermophiles inhabiting extremely heat niches harbored the smallest average genomes (Fig. 6c) that streamlined to minimize substrate and energy requirements (Supplementary Data 5), consistent with the Black Queen Hypothesis^{26,57}. Consequently, to maintain their vital metabolic activities, species with reduced genome sizes were compelled to engage in more frequent metabolic interactions with other species. Finally, differences in genomic

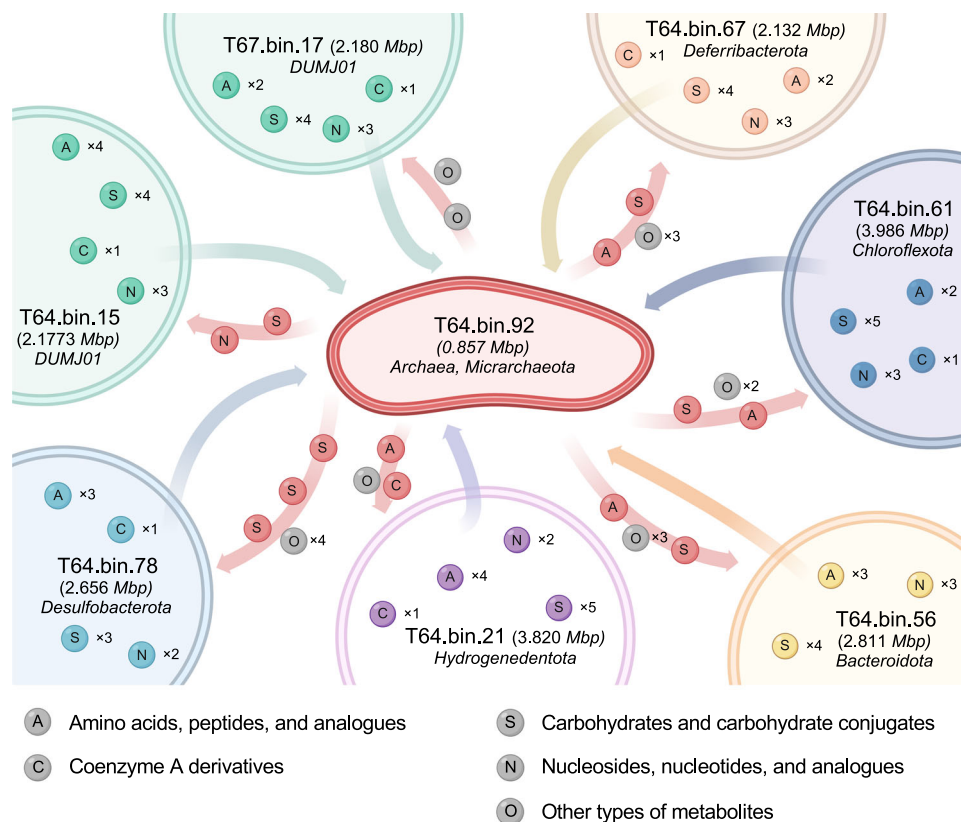


Fig. 7 | The potentially transferable metabolite map of the ET group. The arrows illustrate the direction of metabolite transfer, and each sphere represents a class of metabolites. Here, T64.bin.92 serves as a metabolic taker while the other four serve as metabolic givers Sankey diagram displaying the categorical correspondence of

metabolites in the HMDB database and in this study. This figure was created with BioRender.com released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International license.

features, particularly in gene function distribution, are more likely to occur between distantly related species with complementarity metabolisms (Fig. 6c). We observed that species with smaller genomes prioritize genes crucial for genetic information storage and processing, potentially reflecting a strategy to conserve essential functions while relying on external sources for metabolic needs^{33,34,58}. This suggested a symbiotic approach to survival under harsh stress, where core genetic elements are preserved while costly metabolic tasks are outsourced to their synergistic partners.

Notably, these synergistic relationships can be asymmetrical⁵⁹. The distinct patterns observed in genome size differences between commensalistic and mutualistic interactions underscore this imbalance. Our results illuminated that a considerable proportion of synergistic relationships are commensalistic (Fig. 4). In these commensalistic interactions, there is a significant tendency for one participant (the taker) to have a streamlined genome, while the other (the giver) possesses a comparatively larger genome. This is particularly evident in the extremely thermal group, where over half of the commensalistic interactions involve a taker with a genome size streamlined to lower than 2 Mbp (Fig. 7). Similar trends, though less pronounced, are observed in the HT and MT groups. These findings suggest an ecological strategy among geothermal microbial communities. The commensalistic relationships often formed between microorganisms where one has undergone genome streamlining, forming loss of function (LOF), suggesting a specialized, especially energy-efficient or metabolism-conserving role, while their partners maintain a larger, potentially more versatile genomic repertoire. This dynamic could reflect an evolutionary optimization, where streamlined genomes reduce metabolic redundancy, relying instead on the metabolic

versatility of their partners with more complete genomes and various capabilities. In contrast, mutualistic interactions, where both participants benefit, showed different genome patterns. These mutualistic relationships often involve species with more balanced genomic capabilities, suggesting that both partners contribute to and benefit from their combined metabolic activities. By accurately distinguishing between commensalism and mutualism, our findings provide a clearer understanding of the ecological and evolutionary dynamics within geothermal microbial communities.

We applied a novel approach to confirm that metabolic complementarity is significantly associated with high temperatures in hot spring habitats. To extend the applicability of our findings, we also examined a composting system, another environment where high temperatures are a primary stressor. Here, similar to hot springs, mutualistic interactions, which involve symmetric synergies, were notably infrequent (Supplementary Fig. 4). Among asymmetrical synergistic interactions, i.e., commensalistic ones, the group subjected to the highest temperature (D05) displayed the most pronounced genome size differences between interacting parties (Supplementary Fig. 4). However, compared to hot springs, fewer extremely streamlined genomes were identified in composting samples. Furthermore, we observed a positive correlation between the complementarity index ($M_{\text{complementarity}}$) and phylogenetic distance, particularly strong in the highest temperature group (Supplementary Fig. 3). This suggests that temperature significantly influences the establishment of metabolic complementarity between distantly related species in the composting system as well. Although the retrieved MAGs from the composting system did not exhibit as significant a range in genome sizes as those from hot springs, a clear correlation between metabolic

complementarities, as determined by the RMT approach, and genome size differences was evident, especially at higher temperatures. These findings underscore the critical role of temperature in shaping metabolic complementarity across diverse environments.

While the developed metabolic pipeline marks progress in microbial network analysis, it also highlights areas that require further refinement for practical application in microbiome studies. As observed in this study, there were only a few shared connections between co-occurrence networks and metabolic networks (Supplementary Data 5), and there are several possible reasons for this. First, the basis behind the co-occurrence networks is the measured abundance of microbial species obtained by metagenome sequencing, which means that the statistical correlation between pairwise species is itself biased, or even erroneous. In complex microbial communities, the higher-order interaction (HOI) might be a non-negligible factor that cannot be achieved in pairwise interaction models. The modeling calculations and mechanistic studies associated with HOI must consider much more than substance exchange alone. One prime example is that the production of a substance by one species may require the coexistence of two or more other species^{60,61}. Second, the metabolic approach only demonstrated specific exchanged metabolites between species, while fundamental species interactions are much more varied than just the exchange of metabolites. For instance, the exchanges of information systems or signal molecules, such as quorum sensing, may mediate species interactions^{62,63}. Since microbial interactions in natural habitats are difficult to reproduce under laboratory conditions, our approach provides insights into observing co-occurrence and interpreting metabolic interactions.

Methods

Sample collection, DNA extraction, and sequencing

Sediment samples for our study were collected from a hot spring in Tengchong, Yunnan Province, China (N24°56′ - 25°27′, E98°26′ - 98°27′) in June 2020. The temperature range measured at this hot spring was between 63.5 °C and 85.8 °C, and the water was slightly alkaline (pH range = 8.36–8.70) (Supplementary Fig. 5a). Eight sampling sites displaying a gradual temperature decrease were selected along the spring flow, with five replicate samples collected at each site. According to the measured temperature of each sample, eight temperature sites were labeled as T85, T78, T73, T70, T67, T65, T64, and T63 (Supplementary Fig. 5b). After sample collection, they were promptly placed in a liquid nitrogen container and transported to laboratory within 2 days. Subsequently, they were stored at -80 °C. Before storage, sediment samples for DNA extraction were pre-divided to prevent DNA damage due to repeated freeze-thaw cycles in later experiments. After 48 h of lyophilization, total DNA within the microbial community was extracted from 1.5 g of freeze-dried sediment using the grind plus kit method as previously described⁶⁴.

The acquired DNA was used for Illumina NovaSeq6000 PE250 metagenomic sequencing (DNA library insertion size: 450 bp) and Oxford Nanopore sequencing (the PromethION R9.4 flow cells FLO-PR0002, Oxford Nanopore PromethION sequencer). The Illumina sequencing was conducted on all 40 samples by Magigene Biotechnology Co., Ltd. (Guangzhou, China). Replicate samples of each site were mixed and sequenced using Nanopore sequencing by Bena-gen Technology Co., Ltd (Wuhan, China).

Metagenome assembly, genome binning, and MAG classification

Raw Illumina metagenomic reads were quality trimmed using Trimmomatic (v0.39, *LEADING:3 TRAILING:3 MINLEN:50 SLIDINGWINDOW:4:20*)⁶⁵. The remaining reads of each sample were de novo assembled using IDBA-UD (v1.1.3) with default parameters⁶⁶. Here, the reads of 5 replicates from one sampling site were co-assembled to improve the robustness of assembly and the possibility of obtaining a higher diversity of genomes within the homogeneous environment. The

sequencing coverage and read diversity were estimated using the Nonpareil method^{67,68}. Then, contigs of samples from the same sampling site were pooled together and used for co-assembly with Nanopore long reads using OPERA-MS (v0.9.0) with default parameters⁶⁹. Longer contigs generated by this hybrid assembly, with a minimum length of 1000 bp (1500 bp for metaBAT2), were used for genome binning using metaWRAP (v1.3.2)⁷⁰, with metaBAT2 (v2.12.1)⁷¹, MaxBin2 (v2.2.6)⁷², and CONCOCT (v1.0.0)⁷³ as the core binning tools. Draft bins were quality controlled using metaWRAP bin_refinement module with parameters *-c 50 -x 10*, indicating only MAGs with completeness higher than 50% and contamination lower than 10% were retained for the following analysis, declared to be medium- and high-quality MAGs⁷⁴. The refined bins were then replicated using dRep (v3.5.0)⁷⁵ with parameters *-pa 0.9 -sa 0.99*. The estimated genome size was calculated by dividing the genome size by the sum of completeness and contamination⁷⁶. The taxonomic classification of the selected MAGs was conducted with GTDB-Tk classify_wf workflow against the GTDB genome database (Release 202)⁷⁷. The phylogenetic trees for bacterial and archaeal MAGs were constructed using multiple sequence alignment results generated by GTDB-Tk workflow. The unrooted bacterial and archaeal phylogenetic trees were rooted using midpoint rooting method, performed by midpoint.root function in *phytools* R package⁷⁸, and were then visualized using the online tree display tool iTOL (v6.6)⁷⁹. The phylogenetic distance was then calculated using the *cophenetic.phylo* function in R.

Quantification of MAGs and co-occurrence network construction

The refined bins were quantified by using CoverM (v0.7.0, <https://github.com/wwood/CoverM>, genome mode and coupled reads as input) with parameter: *-method relative_abundance*. In the output relative abundance table (referred as MAG table), every column representing a sample contains unmapped reads percentage, making it reasonable to compare relative abundance across samples. A non-metric multidimensional scaling analysis (NMDS) was conducted with an encapsulated function metaMDS in the vegan package (R) to investigate the community-level difference among sampling sites. A dissimilarity test based on the Bray–Curtis distance was then applied to divide samples into three groups. MAGs were assigned to three groups (ET, HT, MT) based on their relative abundance in samples. Each group comprised multiple samples with five replicates. A MAG was considered present in a group if it had non-zero abundance in more than half of the samples (≥ 6 for ET; ≥ 8 for HT and MT). If a MAG did not meet this criterion in any group, it was assigned to the group with the highest total relative abundance. Additionally, MAGs were ensured to be in the group corresponding to their sample prefix (e.g., T85 and T78 in ET). With the abundance table, the pairwise Spearman's rank correlation of each two MAGs was calculated. Since the relative abundance table originated from the distribution of hypothetical species in a natural environment, it can be regarded as the matrix that the random matrix theory (RMT)-based approach requires. Before using RMT cutoff tool on iNAP website⁴⁴, majority_selection tool was used to filter MAGs that had zero abundance in less than half of samples. This step is to exclude biases from correlation calculations due to the effect of too many zeros. The separated noise formed a network representing MAG pairs with a high abundance correlation, inferring their co-occurrence. The network's general properties were calculated using the igraph R package and visualized using Cytoscape (v3.10.0)⁸⁰.

Genome-scale metabolic models and interaction network construction

Genome-scale metabolic models (GSMMs) for all MAGs were constructed using CarveMe (v1.4.1) with default parameters⁸¹. The input for CarveMe was coding sequences of each MAG, which were predicted using Prokka⁸². The interactions between pairwise GSMMs were predicted using PhyloMint⁴⁷. All modeling software mentioned above

required the installation of an optimization solver, which for our study was IBM CPLEX Optimizer (ILOG COS 20.10 Linux x86-64 version). PhyloMint metabolic complementarity index ($MI_{\text{Complementarity}}$) was used to represent the cooperation potential of pairwise GSMMs, or MAGs. For a given MAG pair, two $MI_{\text{Complementarity}}$ values, ranging from 0 to 1, were calculated due to the asymmetry of the index that a MAG could be both the metabolic giver and taker. Therefore, a higher $MI_{\text{Complementarity}}$ value was regarded as the maximum metabolic cooperation ability between pairwise MAGs and was used for the following analysis. $MI_{\text{Complementarity}}$ values were determined using the RMT cutoff tool on iNAP website, and the corresponding pairwise interactions were selected to form the PhyloMint complementarity network. In the network, synergistic interactions are determined based on whether the original PhyloMint values for all species pairs represented by the edges exceed the RMT screening threshold. If only the unidirectional PhyloMint value exceeds the RMT threshold, the edge is defined as representing commensalistic interactions; if the bidirectional PhyloMint values both exceed the RMT threshold, the edge is defined as representing mutualistic interactions. Similarly, the network attributes were also computed. The predicted coding sequences of each MAG were searched against the database of Clusters of Orthologous Genes (COGs, <https://www.ncbi.nlm.nih.gov/research/cog/>)⁸³ and were categorized into various biological functions. The effects of biological functions on the complementarity potential between pairwise MAGs were estimated using linear mixed-effects models (LMMs) to eliminate the impact of genome size on functions. This step was performed using the R package lme4. Pairwise $MI_{\text{Complementarity}}$ was regarded as the response variable, and gene proportion differences of each biological function were regarded as fixed effects. The genome size difference was termed a random intercept effect.

Extraction of shared interactions and determination of potential exchangeable metabolites

If a specific pairwise MAG interaction possessed both co-occurring properties, as detected by the co-occurrence network and the metabolic cooperation potential detected by the PhyloMint $MI_{\text{Complementarity}}$ network, they were defined as a “dual interaction”. A dual interaction indicated that two MAGs shared a strong correlation between environmental co-occurrence patterns and metabolite profiling complementarity. Using the PhyloMint algorithm, seed metabolites were defined by the strongly connected component and represented substrates acquired exogenously. When considering two MAGs, A and B, a metabolite found in A's seed set but not in B's seed set indicated a potential transferability from B to A. This suggests that B could synthesize the metabolite while A could utilize it. This definition aligns with the computation method of the PhyloMint complementarity index. In line with the previously described definitions, the potential of one MAG to utilize the metabolites produced by another was determined by the overlap between one MAG's seed metabolite set and the other MAG's non-seed metabolite set. The metabolite profiles were taken directly from the corresponding genome-scale metabolic models generated by CarveMe with the BiGG Models database (Version 1.6)⁸⁴. The metabolites were categorized using the Human Metabolome Database (HMDB Version 5.0)⁸⁵. Due to the miscellaneous nomenclature of diverse metabolites, we manually browsed all related metabolites by their name in the BiGG Models database to recheck the classification in HMDB and made some changes to the metabolite classification rules of HMDB as follows (at “super class” level): (i) fatty acyl CoAs were separated from lipids and lipid-like molecules, and were then grouped with other CoA derivatives to form the class “Coenzyme A derivatives”; (ii) compounds identified as amino acids, peptides, and analogs were separated from organic acids and derivatives into a separate class; (iii) Carbohydrates and carbohydrate conjugates were separated from organic oxygen compounds into a separate class; (iv) all

inorganic compounds were combined into one class, “Inorganics”. This class included homogeneous metal and non-metal compounds (inorganic compounds that contain solely metal or non-metal elements, respectively); (v) Organic nitrogen compounds, organic oxygen compounds, organoheterocyclic compounds, benzenoids, and organosulfur compounds were combined into one class “Organic compounds with specific atoms or structures”; (vi) Oxidized ferredoxin, reduced ferredoxin, protoheme, staphyloferrin B, and pyoverdine *P. putida* specific were categorized into “Other metabolites”. With these modified rules, all metabolites involved in our study were classified into nine categories (Supplementary Fig. 6, Supplementary Data 5).

Analysis of the composting system

Zhao et al. conducted a 30-day composting experiment in 10 composting piles at a food waste composting facility. As the temperature changed during the experiment, they selected six sampling time points, with three replicates from each composting pile for each sampling, obtaining a total of 180 samples for metagenomic sequencing. After co-assembly and binning, they obtained 159 high-quality MAGs (completeness >90%, contamination <5%) for further study. Three sampling time points were selected, Day 0, 5, and 25 (tagged as D00, D05, and D25), with average temperatures of 26.02 °C, 64.29 °C, and 41.35 °C. These three time points were regarded as three states of temperature change in the composting experiment (Day 0: the beginning of the experiment, the coolest temperature; Day 5: the highest temperature; Day 25: the temperature dropped in the late stage of the experiment). The relative abundance table of MAGs at these three sampling time points was summarized as a proxy for the state of these MAGs at that time point. The metabolic models and metabolic complementarity indices of MAGs were constructed and calculated using the methods described above. In the three groups, MAGs with relative abundance greater than 0 were assigned to the corresponding group for the construction of metabolic complementarity networks. The threshold for identifying metabolic complementarity within networks was established at an RMT cutoff of 0.280.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The metagenomic sequence data for this study have been deposited in the Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (<https://ngdc.cnbc.ac.cn/gsa>). The Illumina sequencing data are archived under the GSA accession [CRA013105](#). The Nanopore sequencing data are archived under the GSA accession [CRA013130](#). All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Information. Source Data are provided as a source data file. Source data are provided with this paper.

References

1. Faust, K. & Raes, J. Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* **10**, 538–550 (2012).
2. Kost, C., Patil, K. R., Friedman, J., Garcia, S. L. & Ralser, M. Metabolic exchanges are ubiquitous in natural microbial communities. *Nat. Microbiol.* **8**, 2244–2252 (2023).
3. Mee, M. T., Collins, J. J., Church, G. M. & Wang, H. H. Syntrophic exchange in synthetic microbial communities. *Proc. Natl Acad. Sci. USA* **111**, E2149–E2156 (2014).
4. Shou, W., Ram, S. & Vilar, J. M. Synthetic cooperation in engineered yeast populations. *Proc. Natl Acad. Sci. USA* **104**, 1877–1882 (2007).

5. O'Toole, G. A. We have a community problem. *J. Bacteriol.* **206**, e00073–00024 (2024).
6. Barberan, A., Bates, S. T., Casamayor, E. O. & Fierer, N. Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* **6**, 343–351 (2012).
7. Riera, J. L. & Baldo, L. Microbial co-occurrence networks of gut microbiota reveal community conservation and diet-associated shifts in cichlid fishes. *Anim. Microb.* **2**, 36 (2020).
8. Ishimoto, C. K. et al. Microbial co-occurrence network and its key microorganisms in soil with permanent application of composted tannery sludge. *Sci. Total Environ.* **789**, 147945 (2021).
9. Deng, Y. et al. Molecular ecological network analyses. *BMC Bioinform.* **13**, 113 (2012).
10. Ruan, Q. et al. Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics* **22**, 2532–2538 (2006).
11. Xia, L. C. et al. Extended local similarity analysis (eLSA) of microbial community and other time series data with replicates. *BMC Syst. Biol.* **5**, S15 (2011).
12. Faust, K. et al. Microbial co-occurrence relationships in the human microbiome. *PLoS Comput. Biol.* **8**, e1002606 (2012).
13. Friedman, J. & Alm, E. J. Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* **8**, e1002687 (2012).
14. Kurtz, Z. D. et al. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput. Biol.* **11**, e1004226 (2015).
15. Faust, K. Open challenges for microbial network construction and analysis. *ISME J.* **15**, 3111–3118 (2021).
16. Goberna, M. & Verdu, M. Cautionary notes on the use of co-occurrence networks in soil ecology. *Soil Biol. Biochem.* **166**, 108534 (2022).
17. Blanchet, F. G., Cazelles, K. & Gravel, D. Co-occurrence is not evidence of ecological interactions. *Ecol. Lett.* **23**, 1050–1063 (2020).
18. Zelezniak, A. et al. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc. Natl Acad. Sci. USA* **112**, 6449–6454 (2015).
19. Almeida, O. G. G. & De Martinis, E. C. P. Metagenome-assembled genomes contribute to unraveling of the microbiome of cocoa fermentation. *Appl. Environ. Microbiol.* **87**, e0058421 (2021).
20. Garza, D. R. & Dutilh, B. E. From cultured to uncultured genome sequences: metagenomics and modeling microbial ecosystems. *Cell. Mol. Life Sci.* **72**, 4287–4308 (2015).
21. Levy, R. & Borenstein, E. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proc. Natl Acad. Sci. USA* **110**, 12804–12809 (2013).
22. Pinto, F., Medina, D. A., Perez-Correa, J. R. & Garrido, D. Modeling metabolic interactions in a consortium of the infant gut microbiome. *Front. Microbiol.* **8**, 2507 (2017).
23. He, Q. et al. Temperature and microbial interactions drive the deterministic assembly processes in sediments of hot springs. *Sci. Total Environ.* **772**, 145465 (2021).
24. Hou, W. et al. A comprehensive census of microbial diversity in hot springs of Tengchong, Yunnan Province China using 16S rRNA gene pyrosequencing. *PLoS ONE* **8**, e53350 (2013).
25. He, Q. et al. High speciation rate of niche specialists in hot springs. *ISME J.* **17**, 1303–1314 (2023).
26. Sabath, N., Ferrada, E., Barve, A. & Wagner, A. Growth temperature and genome size in bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation. *Genome Biol. Evol.* **5**, 966–977 (2013).
27. Alcorta, J., Alarcon-Schumacher, T., Salgado, O. & Diez, B. Taxonomic novelty and distinctive genomic features of hot spring cyanobacteria. *Front. Genet.* **11**, 568223 (2020).
28. Venton, D. Highlight: bacteria shed DNA as they adapt to hot temperatures. *Genome Biol. Evol.* **5**, 987–988 (2013).
29. Sriaporn, C., Campbell, K. A., Van Kranendonk, M. J. & Handley, K. M. Genomic adaptations enabling *Acidithiobacillus* distribution across wide-ranging hot spring temperatures and pHs. *Microbiome* **9**, 135 (2021).
30. Sorensen, J. W., Dunivin, T. K., Tobin, T. C. & Shade, A. Ecological selection for small microbial genomes along a temperate-to-thermal soil gradient. *Nat. Microbiol.* **4**, 55–61 (2019).
31. Zhang, Y. et al. Hot spring distribution and survival mechanisms of thermophilic comammox *Nitrospira*. *ISME J.* **17**, 993–1003 (2023).
32. Kostesic, E. et al. Microbial diversity and activity of biofilms from geothermal springs in Croatia. *Microb. Ecol.* **86**, 2305–2319 (2023).
33. Elae-Fadrosch, E. A. et al. Global metagenomic survey reveals a new bacterial candidate phylum in geothermal springs. *Nat. Commun.* **7**, 10476 (2016).
34. Beam, J. P. et al. Ecophysiology of an uncultivated lineage of Aigarchaeota from an oxic, hot spring filamentous 'streamer' community. *ISME J.* **10**, 210–224 (2016).
35. Machado, D. et al. Polarization of microbial communities between competitive and cooperative metabolism. *Nat. Ecol. Evol.* **5**, 195–203 (2021).
36. Zhao, Y. et al. Inter-bacterial mutualism promoted by public goods in a system characterized by deterministic temperature variation. *Nat. Commun.* **14**, 5394 (2023).
37. Ona, L. et al. Obligate cross-feeding expands the metabolic niche of bacteria. *Nat. Ecol. Evol.* **5**, 1224–1232 (2021).
38. Palmer, J. D. & Foster, K. R. Bacterial species rarely work together. *Science* **376**, 581–582 (2022).
39. Kehe, J. et al. Positive interactions are common among culturable bacteria. *Sci. Adv.* **7**, eabi7159 (2021).
40. Wang, M. et al. Even allocation of benefits stabilizes microbial community engaged in metabolic division of labor. *Cell Rep.* **40**, 111410 (2022).
41. D'Souza, G. et al. Ecology and evolution of metabolic cross-feeding interactions in bacteria. *Nat. Prod. Rep.* **35**, 455–488 (2018).
42. Giri, S. et al. Metabolic dissimilarity determines the establishment of cross-feeding interactions in bacteria. *Curr. Biol.* **31**, 5547–5557 (2021).
43. Deng, Y. et al. Network succession reveals the importance of competition in response to emulsified vegetable oil amendment for uranium bioremediation. *Environ. Microbiol.* **18**, 205–218 (2016).
44. Feng, K. et al. iNAP: an integrated network analysis pipeline for microbiome studies. *iMeta* **1**, e13 (2022).
45. Peng, X. et al. iNAP 2.0: Harnessing metabolic complementarity in microbial network analysis. *iMeta*. e235 (2024).
46. Magnusdottir, S. et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* **35**, 81–89 (2017).
47. Lam, T. J., Stambouliau, M., Han, W. T. & Ye, Y. Z. Model-based and phylogenetically adjusted quantification of metabolic interaction between microbial species. *PLoS Comput. Biol.* **16**, e1007951 (2020).
48. Hirano, H. & Takemoto, K. Difficulty in inferring microbial community structure based on co-occurrence network approaches. *BMC Bioinform.* **20**, 329 (2019).
49. Weiss, S. et al. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME J.* **10**, 1669–1681 (2016).
50. Guseva, K. et al. From diversity to complexity: microbial networks in soils. *Soil Biol. Biochem.* **169**, 108604 (2022).
51. Mougi, A. & Kondoh, M. Diversity of interaction types and ecological community stability. *Science* **337**, 349–351 (2012).
52. Morriën, E. et al. Soil networks become more connected and take up more carbon as nature restoration progresses. *Nat. Commun.* **8**, 14349 (2017).

53. Lai, D. et al. Resource partitioning and amino acid assimilation in a terrestrial geothermal spring. *ISME J.* **17**, 2112–2122 (2023).
54. van Wolferen, M., Ajon, M., Driessen, A. J. M. & Albers, S.-V. How hyperthermophiles adapt to change their lives: DNA exchange in extreme conditions. *Extremophiles* **17**, 545–563 (2013).
55. Maigne, V., Vannier, N., Vandenkoornhuyse, P. & Hacquard, S. Multi-genome metabolic modeling predicts functional inter-dependencies in the Arabidopsis root microbiome. *Microbiome* **10**, 217 (2022).
56. Pan, X. et al. Larger phylogenetic distances in litter mixtures: lower microbial biomass and higher C/N ratios but equal mass loss. *Proc. R. Soc. B* **282**, 20150103 (2015).
57. Giovannoni, S. J., Cameron Thrash, J. & Temperton, B. Implications of streamlining theory for microbial ecology. *ISME J.* **8**, 1553–1565 (2014).
58. Cobo-Simon, M. & Tamames, J. Relating genomic characteristics to environmental preferences and ubiquity in different microbial taxa. *BMC Genom.* **18**, 499 (2017).
59. Kern, L., Abdeen, S. K., Kolodziejczyk, A. A. & Elinav, E. Commensal inter-bacterial interactions shaping the microbiota. *Curr. Opin. Microbiol.* **63**, 158–171 (2021).
60. Ludington, W. B. Higher-order microbiome interactions and how to find them. *Trends Microbiol.* **30**, 618–621 (2022).
61. Morin, M. A., Morrison, A. J., Harms, M. J. & Dutton, R. J. Higher-order interactions shape microbial interactions as microbial community complexity increases. *Sci. Rep.* **12**, 22640 (2022).
62. Lowery, C. A., Dickerson, T. J. & Janda, K. D. Interspecies and interkingdom communication mediated by bacterial quorum sensing. *Chem. Soc. Rev.* **37**, 1337–1346 (2008).
63. Zeng, X. et al. Quorum sensing-mediated microbial interactions: mechanisms, applications, challenges and perspectives. *Microbiol. Res.* **273**, 127414 (2023).
64. Zhou, J. Z., Bruns, M. A. & Tiedje, J. M. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62**, 316–322 (1996).
65. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
66. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).
67. Rodriguez, R. L. & Konstantinidis, K. T. Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. *Bioinformatics* **30**, 629–635 (2014).
68. Rodriguez, R. L., Gunturu, S., Tiedje, J. M., Cole, J. R. & Konstantinidis, K. T. Nonpareil 3: fast estimation of metagenomic coverage and sequence diversity. *mSystems* **3**, e00039–18 (2018).
69. Bertrand, D. et al. Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. *Nat. Biotechnol.* **37**, 937–944 (2019).
70. Uritskiy, G. V., DiRuggiero, J. & Taylor, J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 158 (2018).
71. Kang, D. W. D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
72. Wu, Y. W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2016).
73. Alneberg, J. et al. Binning metagenomic contigs by coverage and composition. *Nat. Methods* **11**, 1144–1146 (2014).
74. Bowers, R. M. et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).
75. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).
76. Parks, D. H. et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* **2**, 1533–1542 (2017).
77. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2020).
78. Revell, L. J. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ* **12**, e16505 (2024).
79. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
80. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
81. Machado, D., Andrejev, S., Tramontano, M. & Patil, K. R. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.* **46**, 7542–7553 (2018).
82. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
83. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33–36 (2000).
84. King, Z. A. et al. BiGG models: a platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* **44**, D515–D522 (2016).
85. Wishart, D. S. et al. HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic Acids Res.* **50**, D622–D631 (2022).

Acknowledgements

This project was supported by the National Nature Science Foundation of China (U23A2043 to S.W. and K.F., 92351303 to W.H.), and the National Key Research and Development Program of China (grant No. 2022YFE0114000 to Y.D.).

Author contributions

Conceptualization: Y.D., X.P., S.W., K.F., and Q.H. Methodology: X.P., Y.D., K.F., X.S.Y., Y.X.Z., and B.L.H. Investigation: S.W., Q.H., W.G.H., F.R.L., and X.Z. Visualization: X.P. and X.S.Y. Supervision: Y.D. and S.W. Writing—original draft: X.P. and Y.D. Writing—review & editing: X.P., Y.D., S.W., M.X.W., K.F., and Q.H.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-52532-x>.

Correspondence and requests for materials should be addressed to Ye Deng.

Peer review information *Nature Communications* thanks Sarahi Garcia, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024

¹CAS Key Laboratory for Environmental Biotechnology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences (CAS), Beijing, China.

²College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China. ³Department of Environmental Systems Science, ETH Zürich, Zürich, Switzerland. ⁴Department of Environmental Microbiology, Eawag, Dübendorf, Switzerland. ⁵State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Beijing, China. ⁶Department of Environmental Engineering, Zhejiang University, Hangzhou, China.

⁷College of Environmental and Resource Sciences, Zhejiang University, Hangzhou, China. ⁸Zhejiang Province Key Laboratory for Water Pollution Control and Environmental Safety, Hangzhou, China. ⁹Key Laboratory of Environment Remediation and Ecological Health, Ministry of Education, College of Environmental Resource Sciences, Zhejiang University, Hangzhou, China. ¹⁰Department of Ecology/Key Laboratory of Plant Resource Conservation and Germplasm Innovation in Mountainous Region (Ministry of Education), College of Life Sciences, Guizhou University, Guiyang, China. ✉e-mail: yedeng@rcees.ac.cn