Evolution of the Earth Microbial Co-occurrence Network

Xiaofei Lv
China Jiliang University
Erinne Stirling
Zhejiang University
Kankan Zhao
Zhejiang University
Yiling Wang
Zhejiang University
Bin Ma (bma@zju.edu.cn)
Zhejiang University https://orcid.org/0000-0003-4807-4992
Jianming Xu
ZJU: Zhejiang University

Short report

Keywords: Earth microbiome, co-occurrence network, topological fitness, evolution, node extinction, node decay

Posted Date: June 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-604225/v1

Title

Evolution of the Earth microbial co-occurrence network

Authors

Xiaofei Lv,[1,2,3] Erinne Stirling,[2,3,4,5] Kankan Zhao,[2,3] Yiling Wang,[2,3] Bin Ma,[2,3,4*] Jianming Xu[2,3]

Affiliations

[1] Department of Environmental Engineering, China Jiliang University, Hangzhou 310018, China

[2] Institute of Soil and Water Resources and Environmental Science, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou, China

[3] Zhejiang Provincial Key Laboratory of Agricultural, Resources and Environment, Zhejiang University, Hangzhou, 310058, China

[4] Hangzhou Global Scientific and Technological Innovation Center, Zhejiang University, Hangzhou, 311200, China

[5] Acid Sulfate Soils Centre, School of Biological Sciences, The University of Adelaide, Adelaide 5005, Australia

[*] Corresponding author's e-mail address: bma@zju.edu.cn

Abstract

Background: Co-occurrence patterns provide vital insights into the complex microbial interactions of microbiomes and although network analysis offers useful tools for describing microbial co-occurrence patterns, the evolution of co-occurrence networks remains largely uncharacterized. Here, we simulated the evolution of the Earth microbial co-occurrence network and estimated topological fitness of its nodes based on the degree growth exponent.

Results: The Earth microbial co-occurrence network evolved following the Bianconi-Barabasi model and reached a stable status at approximately 500 nodes. The degree growth exponent was the major determinant of accumulated degree of taxa and the positive correlation between topological fitness and gene numbers in corresponding genomes suggests an intrinsic feature of topological fitness. The topological fitness gamma distribution suggests the extinction of taxa with low topological fitness even though link acquisition of hub nodes was not affected by node extinction and decay.

Conclusions: This study glimpses into the evolutionary features of the Earth microbial co-occurrence network and provides a framework for predicting potential hubs in future network evolution.

Keywords

Earth microbiome; co-occurrence network; topological fitness; evolution; node extinction; node decay

Background

Microbial interactions are primary factors for microbial community assembly and macroevolution. Microbial interaction networks capture emergent features [1] that help explain the behavior of complex microbial interactions and enable the prediction of perturbation and engineering modification effects on community function [2]. The topological properties of microbial co-occurrence networks provide insights for understanding complex microbial interactions and node topological features, such as degree and betweenness centrality, are known proxies to identify keystone species [3]. Modules in these networks may indicate ecological processes governing community structure, such as niche filtering and habitat preference [4].

Microbial interaction networks evolve simultaneously with the evolution of microbial taxa through time as they are the product of changing taxonomic composition and metabolism [5]. Given that microbial interactions lie at the heart of microbial macroevolution [6], microbial interaction networks might change in predictable ways during macroevolutionary timescales [7]. While we know that microbial interaction networks have particular structural features, the process of network change during evolution is not well known. Even though historical data on microbial interactions in the fossil record are seldom available, phylogenetic trees generated from marker genes might give insight into how microbial interactions change over macroevolutionary timescales [8]. To access how a microbial interaction network evolves using phylogenetic trees, we can track evolution dynamics by integrating microbial interaction network evolution with phylogenetic tree speciation events (**Fig. 1**). Although a short timescale study may not provide strong hypotheses testing to link microbial interactions with macroevolution, the evolution of microbial networks over a large spatial scale that has

experienced long-term evolution provides a novel perspective on the evolution of microbial biodiversity.

Network evolution scenarios such as preferential attachment have been proposed to explain the emergence of hub nodes and a scale-free degree distribution in networks, with the Biaconi-Barabási model suggesting that fitness plays a critical role in network evolution and topology [9][10]. Node topological fitness represents the intrinsic qualities of a node that influence the rate at which it acquires links. In many real-world networks, such as those that describe webpages, social interactions, and academic article citations, it has been shown that a high degree of connectivity is determined by high node topological fitness, even if the node arrives late to the network [11–13]. Nodes with high fitness values have intrinsic properties that cause them to be successful in acquiring links. Of the examples we have used, these properties include a webpage's ability to draw attention, an individual's ability in establishing social links, and an academic article's novelty and importance. In microbial co-occurrence networks, fitness encapsulates the microorganism's capability to form ecological interactions with other microorganisms. It is not clear whether this capability differs across microorganisms and to what degree it influences fitness.

Here, we simulated the evolution of the Earth microbial co-occurrence network that was inferred from the Earth Microbiome Project (EMP) dataset in our previous study [14]. Based on the Bianconi-Barabasi model for evolving networks, we evaluated topological fitness by estimating the node degree growth exponent, β. We further examined the impact of node extinction and decay on the topological fitness estimation. We found that the Earth microbial co-occurrence network topological features are most similar to a scale-free network simulated based on the Bianconi-Barabasi model when compared to the Erdos-Renyi and Barabasi-Albert

models. We observed a positive association between degree number and topological fitness, suggesting that inherent features of microbial taxa mediate potential microbial interaction patterns in microbial communities. In addition, we found a positive association between topological fitness and gene number in corresponding genomes, which suggests that topological fitness is an inherent feature of microbial taxa. Furthermore, the impact of node extinction and decay was limited to taxa with low topological fitness, suggesting that those processes had a relatively narrow impact on hub link acquisition in the Earth microbial co-occurrence network.

Methods

Earth microbial co-occurrence network

The Earth microbial co-occurrence network was inferred from a communal catalog with 23,595 samples and 12,646 exact sequence variants (ESVs) from 14 environments in the Earth Microbiome Project dataset [14]. Briefly, we inferred co-occurrence networks for 14 environments separately and merged them into a single Earth microbial co-occurrence network by overlapping the nodes and links. The Earth microbial co-occurrence network consists of 2,928 ESV nodes and 54,299 links.

Networks based on theorical models

We generated three types of random networks *in silico* according to the Erdos-Renyi, Barabasi-Albert, and Biaconi-Barabási models with *igraph* (version 1.2.6) using the R platform (version 3.6.3). Each model repeated 100 times. The Erdos-Renyi, Barabasi-Albert, and Biaconi-Barabási networks were generated with *erdos.renyi.game*, *sample_pa*, and *sample_pa_age* functions, respectively.

Simulation of the evolution of Earth microbial co-occurrence network

To simulate Earth microbial co-occurrence network evolution, we first built a phylogenetic tree for the ESVs involved in the network. The corresponding 16S rRNA gene segment sequences were retrieved from EMP and aligned using MAFFT [15]. A maximum likelihood (ML) phylogenetic tree was estimated with RAxML [16]. Node ages in the tree were dated using the mean path lengths method function *chronoMPL* of ape package [17]. Network evolution was then simulated by merging nodes according to the divergent sequences of the phylogenetic tree (**Fig. 1**). Parent node links were determined as the intersection of two descendant node links. This process was repeated until the network combined into a single node representing the root ancestor of the phylogenetic tree. After each step, we estimated the degree exponent (λ), average separation (d), and clustering coefficient (C) for the evolved network.

Topological fitness estimation

Microbial co-occurrence network evolution was described by the Bianconi-Barabási model [10]; this model assumes that the formation of microbial interactions is driven by the product of a taxon's topological fitness η and its degree k. This fitness is a taxon's capability to form interaction relationships with other taxa. In each time step, a new taxon j with fitness η_j and m interaction links joins the network. The probability that an interaction link of new taxon j directed to taxon i is proportional to the production of taxon i's fitness η_i and its degree k_i:

$$\boldsymbol{\Pi_i} \sim \frac{\boldsymbol{\eta_i k_i}}{\sum_j \boldsymbol{\eta_j k_j}} \quad (1)$$

The dependence of Π_i on k_i and η_i implies that the taxa with more interactions and higher fitness are more likely to build new interactions. Hence, this model assumes that even a new taxon with only a few interactions can build interactions rapidly if it acquires better fitness

during evolution. A taxon's fitness η is proportional to its degree growth exponent $\beta(\eta)$, a parameter indicating the increase rate of link numbers, which is different to biological growth rate of microbial taxon such that,

$$\boldsymbol{\beta(\eta) = \frac{\eta}{C}} \quad (2)$$

where

$$\boldsymbol{C = \int \rho(\eta) \frac{\eta}{1-\beta(\eta)} d\eta} \quad (3)$$

Accordingly, the fitness of a taxon can be reflected by estimating the degree growth exponent $\beta(\eta_i)$ by:

$$\boldsymbol{k(t, t_i, \eta_i) = m \left(\frac{t}{t_i}\right)^{\beta(\eta_i)}} \quad (4)$$

where m is initial degree, t is the age of network, and t_i is the time of node i was added to the network. From Eq. 3 we get the solution of $\beta(\eta_i)$:

$$\boldsymbol{\beta_{(\eta i)} = \frac{ln\,(k) - ln\,(m)}{ln(t) - ln\,(t_i)}} \quad (5)$$

Genome size dataset

We downloaded the genome size data from the NCBI Refseq genome database (https://www.ncbi.nlm.nih.gov/refseq) and calculated the mean genome size for each genus.

Impact of extinction on the fitness estimation

The fitness model was extended to account for node deletion to evaluate the impact of extinction on fitness estimations [11]. At each time step in this model, a randomly selected

node (with probability c) and its links were deleted. The degree growth exponent β is a function of fitness η:

$$\boldsymbol{\beta_{decay}} = \frac{\boldsymbol{\eta}}{\boldsymbol{A}} - \frac{\boldsymbol{c}}{\boldsymbol{1-c}} \quad (6)$$

where A is given by:

$$\boldsymbol{A} = (\boldsymbol{\eta_{max}} + \boldsymbol{\varepsilon})\frac{\boldsymbol{1+c}}{\boldsymbol{1-c}} \quad (7)$$

and η_{max} is the maximum fitness in the system and ε is negligibly small. Thus, the degree growth exponent β_{decay} is given by:

$$\boldsymbol{\beta_{decay}} = \frac{\boldsymbol{\eta(1+c)}}{\boldsymbol{(\eta_{max}+\varepsilon)(1-c)}} - \frac{\boldsymbol{c}}{\boldsymbol{1-c}} \quad (8)$$

The impact of the extinction on the fitness estimation was tested for c between 0.0 and 0.9 using intervals of 0.1.

Impact of fitness aging on the fitness estimation

Microbial co-occurrence network taxa continue to age as their fitness fades with time due to the spread of novel functional genes in the microbial community through horizontal gene transfer. This aging process is captured in the fitness estimation by a tunable parameter v [18], which governs the dependence of attachment probability on the taxon's age:

$$\boldsymbol{\Pi(k, t - t_i) \sim k(t - t_i)^{-v}} \quad (9)$$

From Eq. 4 and Eq. 8 we get the solution of β_{decay}:

$$\boldsymbol{\beta_{decay}} = \frac{\boldsymbol{ln(k) - v \cdot ln(t - t_i) - ln\,(m)}}{\boldsymbol{ln(t) - ln\,(t_i)}} \quad (10)$$

When $v < 0$, a new taxon will prefer to interact with older taxa. In the extreme case of $v \to \infty$, a new taxon will only interact with the oldest taxon. In contrast, when $v > 0$, a new taxon will prefer to interact with younger taxa. When $v > 1$, the aging effect overcomes the role of preferential attachment, leading to the loss of the scale-free property. When $v = 0$, the aging effect is absent. The impact of aging on the fitness estimation was tested by setting v at -1, 0.2, 0.4, 0.6, 0.8 and 1.0.

Results

Earth microbial co-occurrence network evolved following Bianconi-Barabasi model

To reveal the association pattern of the Earth microbial co-occurrence network, we firstly compared its degree distribution with that of a random network generated based on Erdos-Renyi model and with scale-free networks generated based on Barabasi-Albert or Bianconi-Barabasi models. The degree distribution of the Earth microbial co-occurrence network was most similar with the scale-free network simulated with the Bianconi-Barabasi model (**Fig. 2**).

We then simulated the evolution of the Earth microbial co-occurrence network (**Movie S1**) by merging two descendant nodes into a parent node based on the divergence order of the phylogenetic tree. The network was synchronously evolved in different environments (**Fig. S1**) and taxonomic groups (**Fig. S2**). The degree exponent (λ), an indicator for the scale-free property of the network, decreased fast from 6 to 3 and sluggishly converged to a stationary value around 2 during simulated evolution (**Fig. 3**). Average separation (d), denoting the average length of the shortest path over all pairs of nodes, characterizes network interconnectedness. Stationary d values after approximately 500 steps suggest a stationary small-world property of the network (**Fig. 3**). The clustering coefficient (C), a measure of the degree to which the nodes in a graph tend to cluster together, also exhibited a decreasing

tendency and converged to a stationary value around 0.45 after approximately 500 steps (**Fig. 3**).

Degree growth exponent estimation

The Bianconi-Barabasi model assumes that nodes have a different probability of forming interactions, which we refer to as "topological fitness" (η). Given that node topological fitness is related to degree growth exponent (β) by a linear transformation with constant coefficients (see **Eq. 3** in the Methods), we assessed topological fitness by estimating β from **Eq. 4**. The β values of the 50 nodes with the largest degree demonstrate that the degree increased at various rates, supporting that the Earth network evolution follows the Bianconi-Barabasi model. The distribution of β (**Fig. 4a**) shows that it follows a gamma distribution, indicating that topological fitness of microbial taxa varies in a relatively narrow range and that nodes with high fitness and low fitness are rare. The accumulated degree (k_t) increased with β (**Fig. 4b**), suggesting that node topological fitness determines its degree in the microbial co-occurrence network, representing the intrinsic node qualities that influence the rate at which nodes acquire links (for instance, the ease of pathway loss in response to leaky metabolite production). Moreover, β values were not affected by node age (**Fig. 4c**), which also demonstrates that the network does not follow the Barabasi-Alberta model.

To test the dependence of topological fitness due to phylogenetic relationship, we compared the distribution of β among genera (**Fig. 4d**) and calculated phylogenetic signal value of β. The β values varied with microbial genus (ANOVA, DF=120, F=2.38, P<0.001). The genera *Pedobacter*, *Haliangium*, and *Parcubacteria_genera_incertae_sedis* had the highest β values, while *Clostridium_IV, Roseomonas*, and *Mycoplasma* had the lowest β values (**Table S1**). The phylogenetic signal suggests that variation of β among genera are a consequence of their

phylogenetic proximity (Pegel's $\lambda = 0.011$, $P = 0.018$). Furthermore, β was positively correlated with average genomic size (Pearson's $r = 0.11$, $P = 0.05$) and gene number (Pearson's $r = 0.12$, $P = 0.02$) of corresponding genera genomes (**Fig. 4e**).

Impact of node extinction on topological fitness

To address the impacts of extinction of microbial taxa when estimating β, we employed the node deletion model (**Eq. 7**), which considers the node turnover rate (c). It is clear from this model that the impact of extinction was strong on small β values. Taxa became extinct when their topological fitness was lower than a threshold; these threshold values increased with the turnover rate (**Fig. 5a**).

Impact of node decay on topological fitness

To illustrate the impact of the aging effect on node topological fitness, we introduced a decay rate (v) into the model (**Eq. 9**). The fitness decreased with time when $v > 0$, and *vice versa*. Given that there is a scale-free feature and decaying fitness in this network, we can predict that $0 < v < 1$. The v values were decreasingly skewed when $0 < v < 1$, especially for smaller β values. The skewness level was increased by tuning the v value from 0.2 to 1 (**Fig. 5b**).

Discussion

Our study indicates that the evolution of the Earth microbial co-occurrence network followed the Bianconi-Barabasi model, suggesting that inherent features of microbial taxa, such as genome gene numbers, mediate the potential microbial interaction patterns in microbial communities. The framework employed here assumes that taxa have different propensities for engaging in ecological interactions, that these different propensities affect their topological fitness and that ecological interactions evolve on the same time scale as the taxa themselves.

253 The second assumption deserves further exploration, while the last assumption is likely
254 unrealistic given the fast emergence of ecological interactions in response to altered conditions
255 [19]. In addition, the phylogenetic tree and microbial co-occurrence network are likely of low
256 accuracy as they are both inferred. Regardless, the simulation is not meant to recapitulate the
257 evolution of the Earth microbial co-occurrence network, but rather to investigate how evolved
258 network properties relate to different assumptions about network evolution. The comparison of
259 different evolutionary scenarios helps to understand how these network properties arise. In the
260 future, more realistic evolutionary scenarios will be tested and the resulting network properties
261 compared to observed properties; the simple model applied here is a first step in this direction.
262
263 Although the Earth microbial co-occurrence network evolved into a complex system consisting
264 of 2,928 nodes in our simulations, its topological features were stationary after evolving into a
265 network consisting of approximately 500 nodes. The decreasing λ values during the initial
266 stages, from 6 to 2, suggests that the metacommunity scale-free network evolved from a
267 random network without network hubs [20]. We speculate that early microbial communities
268 did not yet have many interactions or that they were unable to form hubs, but given the
269 methodological limitations, we cannot investigate this question further. If extinct branches
270 were able to be included, we expect that λ would stabilize later and decrease more gradually.
271 The sluggishly decreasing tendency of λ values after $\lambda < 3$ indicates that the network hubs
272 attracted more links during simulated network evolution as $k_{max} \sim N_1/\lambda\text{-}1$ in a scale-free network
273 [18]. Since λ converged to 2, the hubs tend to associate with a large proportion of nodes and
274 the network tends to reduce the distance between any two nodes. This relationship also explains
275 why the average distance was maintained at a low level when λ reached a stationary stage. Note
276 though that λ values should not be lower than 2 as self-loops and multiple links were not
277 counted [18].

Average degree increased with β in the Earth microbial co-occurrence network. However, there was a large degree of variation for nodes with the same β which was caused by different node emergence times. Small differences can escalate over time and a distinction in degree for nodes with slight β differentiations can became quite significant into the future. The gamma distribution of β suggests that both high and low topological fitness nodes were rare and that the β values for most nodes were ~1. Rare nodes with high topological fitness were expected due to the scale free feature of the Earth microbial co-occurrence network. Simultaneously, the decreasing number of nodes with fitness when $\beta < 1$ indicates that there are few nodes with small topological fitness values. One explanation for the deficiency in nodes with small fitness values is that lower fitness leads to node deletion from the network as microbial species cannot compensate for a loss of associations by gaining new associations. We examined the impact of extinction on fitness estimation and proved that the impact of extinction indeed increased as fitness decreased. Another possible explanation is that rare biospheres were not included in the present Earth microbial co-occurrence network; however, although rare biospheres might play a critical role in microbial communities [21], the topological fitness of rare biospheres is expected to be small due to their limited chances of interacting with other taxa.

Our model does not provide evidence that topological fitness has genetic roots or that it is linked to fitness in the evolutionary sense [7]. However, if such a link exists, natural selection would enhance topological fitness through mutations, genetic recombination, gene loss and horizontal gene transfer events [5]. Accordingly, the gene profile of a microbial genome might describe the essential aspects of topological fitness in an evolving microbial ecological network. However, we assume that gene loss is a rare interaction mechanism, and that topological fitness is reduced over time by vertical transmission of interaction mechanisms to

descendants, leading to a “dilution” of topological fitness [22]. Topological fitness decay with node age shows an aging effect ($v > 0$) of nodes in the Earth microbial co-occurrence network; gradual aging homogenizes the network by impeding the older hubs, which are less likely to form new links. However, given the merging strategy used, an aging effect is the only possible outcome when $v > 0$ as older nodes can only have higher fitness than younger nodes when $v < 0$. Similar to the impact of extinction, the impact of aging also increased as topological fitness decreased. In this case, nodes with small topological fitness values failed to persist in the network when the topological fitness was smaller than a threshold value. This phenomenon is also supported by the distribution of β as many nodes are absent when $\beta < 1$. Note though that the definition and measurement of topological fitness in a microbial ecological network is not straightforward; many interaction mechanisms have not yet been deciphered and genomic information for most microbial species is unavailable due to a failure to culture the dominant microbial species in many environments [23]. There remains further knowledge gaps in the association between functional gene profiles and these networks.

We have not inferred the timing of divergence events because bacteria and archaea lack sufficient evidence (such as a comprehensive fossil record) to calibrate the phylogenetic tree. Whole genome sequences provide a more reliable resource to dating than the molecular clock, but a greater effort is needed to sequence the genomes in uncultured microbial “dark matter”. Other life on Earth (for example viruses, plants, animals, and fungi) is also essential to the microbial interactome due to their influences on microbial environments, and we anticipate that novel studies within the EMP framework will fill the gaps for microbial eukaryotes and viruses. Given the increasing recognition of the value of communal microbial biodiversity monitoring and the current global advance in sequencing efforts, we can infer in the future an

ever more comprehensive global microbial co-occurrence network from the rapid accumulation of microbial sequence data.

Conclusions

The extent of our understanding of microbial communities is reflected in our ability to predict future trends. Given the vital role of microbial interaction in community assembly, predicting potential interaction features and tendency is important for microbial community controllability. In this study, we simulated the evolution of the Earth microbial co-occurrence network following divergent sequences in the phylogenetic tree generated from EMP datasets, and assessed topological fitness of nodes in the network by estimating the degree growth exponent. Our study provides a new framework for predicting potential tendency of evolving microbial ecology networks. It remains to be seen what the underlying mechanisms are that form topological fitness, and how inherent features can be manipulated to control microbial community function.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

All data generated or analysed during this study are included in this published article and its supplementary information files. All R code is available at github.com/EMPs.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China [42090060, 41991334], Zhejiang Natural Science Foundation [LD19D060001, LQ20C030006], grants from CAS Key Laboratory of Soil Environment and Pollution Remediation, Institute of Soil Science, Chinese Academy of Sciences, and the Key Laboratory of Coastal Environmental Processes and Ecological Remediation, YICCAS [2020KFJJ08].

Authors' contributions

BM and JX designed this study. XL, KZ, and YW analyzed the network evolution. XL, ES, and BM were a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Figures

Figure 1. Simulated evolution of microbial co-occurrence network based on divergence order of phylogenetic tree.

Figure 2. Comparison of degree distrubution of the Earth microbial co-occurrence network and simulated network based on (a) Erdos-Renyi model, (b) Barabasi-Albert model, and (c) Bianconi-Barabasi model.

Figure 3. Dynamic of degree exponent (λ), average separation (d), and clustering coefficient (C) of evolving Earth microbial co-occurrence network.

Figure 4. Degree growth exponont (β) of the Earth microbial co-occurrence network. (a) distribution of β values. (b) The relationship between accumulated degree and β. (c) The relationship between node age and β. (d) The boxplot of β values for various genera wherein

red dots show mean β values of corresponding genera. (e) The relationship between mean genome size and mean β values of various genera.

Figure 5. The impact of node turnover and node decay on β value estimation. (a) The impact of node turnover on β value estimation when turnover rate (c) changed from 0.0 to 0.9. (b) The impact of node decay on β value estimation when decay exponent (v) changed from -1 to 1.

Supporting information

Table S1. Degree growth exponent (β) of dominant genera.

Figure S1. The divergence order of nodes from different environments.

Figure S2. The divergence order of nodes from different taxonomic groups.

Movie S1. Evolving Earth co-occurrence network. The green nodes and links highlight the latest 20 nodes and associated links.

Reference

1. Ma B, Wang H, Dsouza M, Lou J, He Y, Dai Z, et al. Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. ISME J. 2016;10:1891–901.

2. Röttjers L, Faust K. From hairballs to hypotheses-biological insights from microbial networks. FEMS Microbiol Rev. 2018;42:761–80.

3. Berry D, Widder S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Front Microbiol [Internet]. 2014 [cited 2019 Jul 10];5. Available from: http://journal.frontiersin.org/article/10.3389/fmicb.2014.00219/abstract

4. Lima-Mendez G, Faust K, Henry N, Decelle J, Colin S, Carcillo F, et al. Determinants of community structure in the global plankton interactome. Science. 2015;348:1262073.

400 5. Cordero OX, Polz MF. Explaining microbial genomic diversity in light of evolutionary
401 ecology. Nature Reviews Microbiology. 2014;12:263–73.
402 6. Weber MG, Wagner CE, Best RJ, Harmon LJ, Matthews B. Evolution in a Community
403 Context: On Integrating Ecological Interactions and Macroevolution. Trends in Ecology &
404 Evolution. 2017;32:291–304.
405 7. Harmon LJ, Andreazzi CS, Débarre F, Drury J, Goldberg EE, Martins AB, et al. Detecting
406 the macroevolutionary signal of species interactions. Journal of Evolutionary Biology
407 [Internet]. [cited 2019 Jul 5]; Available from:
408 https://onlinelibrary.wiley.com/doi/abs/10.1111/jeb.13477
409 8. Evlampiev K, Isambert H. Conservation and topology of protein interaction networks
410 under duplication-divergence evolution. Proceedings of the National Academy of Sciences.
411 2008;105:9863–8.
412 9. Barabasi A. Emergence of Scaling in Random Networks. Science. 1999;286:509–12.
413 10. Bianconi G, Barabási A-L. Competition and multiscaling in evolving networks. EPL.
414 2001;54:436.
415 11. Kong JS, Sarshar N, Roychowdhury VP. Experience versus talent shapes the structure of
416 the Web. Proceedings of the National Academy of Sciences. 2008;105:13724–9.
417 12. Barabási A-L, Song C, Wang D. Publishing: Handful of papers dominates citation.
418 Nature. 2012;491:40.
419 13. Fowler JH, Dawes CT, Christakis NA. Model of genetic variation in human social
420 networks. Proc Natl Acad Sci USA. 2009;106:1720–1724.
421 14. Ma B, Wang Y, Ye S, Liu S, Stirling E, Gilbert JA, et al. Earth microbial co-occurrence
422 network reveals interconnection pattern across microbiomes. Microbiome. 2020;8:82.

15. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution. 2013;30:772–780.

16. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–1313.

17. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 2004;20:289–290.

18. Barabási A-L. Network science. Cambridge, UK: Cambridge university press; 2016.

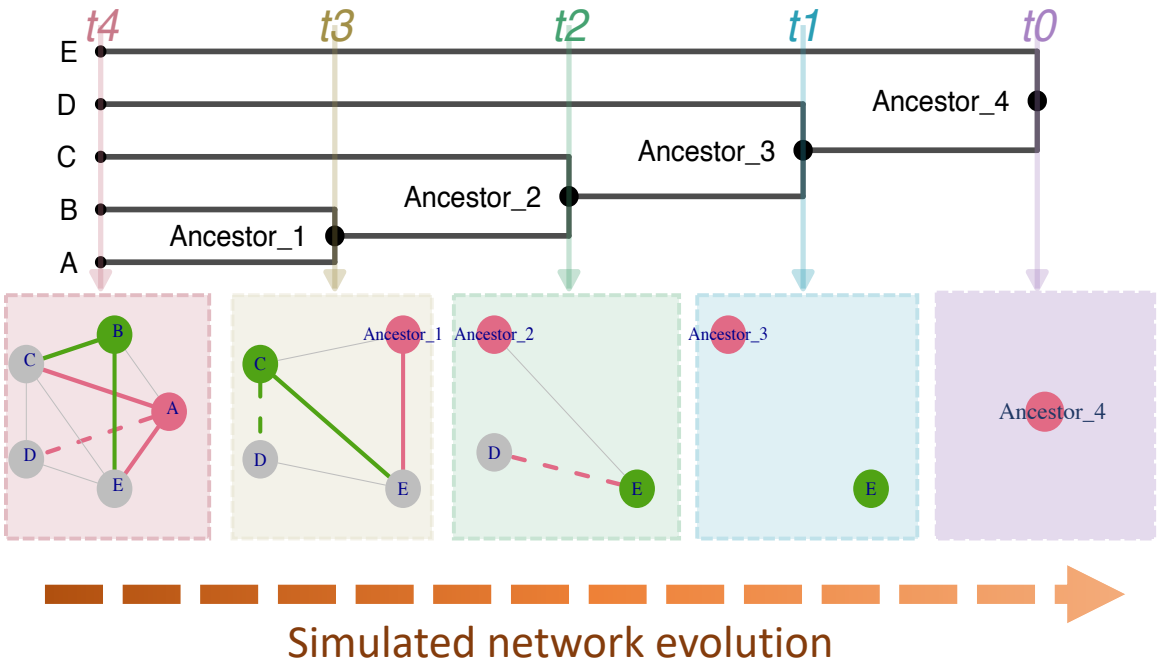
19. Lawrence D, Fiegna F, Behrends V, Bundy JG, Phillimore AB, Bell T, et al. Species Interactions Alter Evolutionary Responses to a Novel Environment. Ellner SP, editor. PLoS Biol. 2012;10:e1001330.

20. Liu Y-Y, Slotine J-J, Barabási A-L. Controllability of complex networks. Nature. 2011;473:167–73.

21. Carlos P-A. Dipping into the rare biosphere. Science. 2007;315:192–193.

22. Gogarten JP, Townsend JP. Horizontal gene transfer, genome innovation and evolution. Nature Reviews Microbiology. 2005;3:679–87.

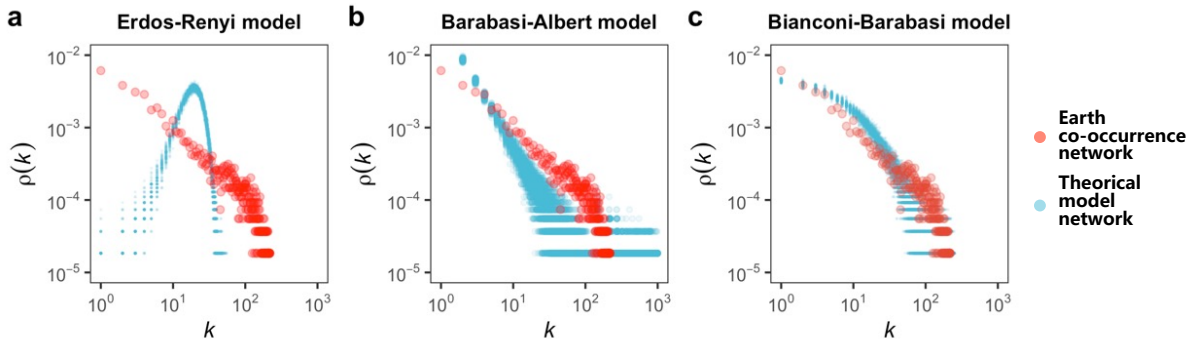
23. Lok C. Mining the microbial dark matter. Nature News. 2015;522:270.



441

442 **Figure 1**. Simulated evolution of microbial co-occurrence network based on divergence order

443 of phylogenetic tree.



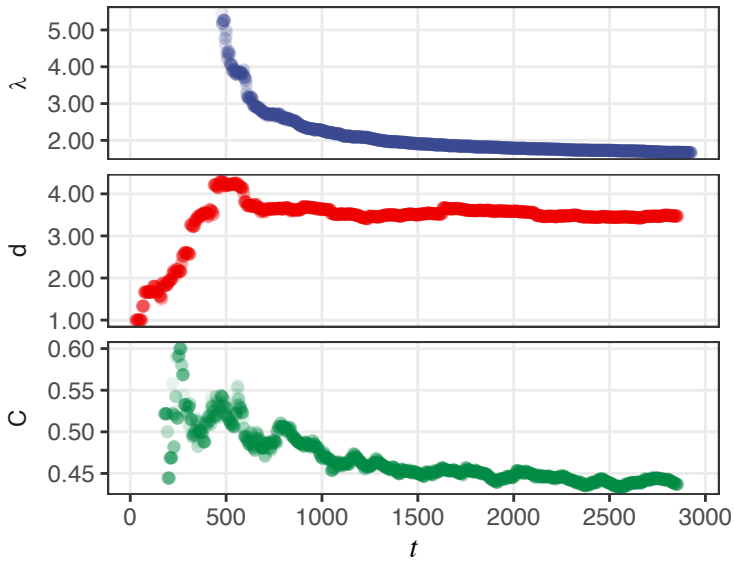
444

445

446 **Figure 2**. Comparison of degree distrubution of the Earth microbial co-occurrence network

447 and simulated network based on (a) Erdos-Renyi model, (b) Barabasi-Albert model, and (c)

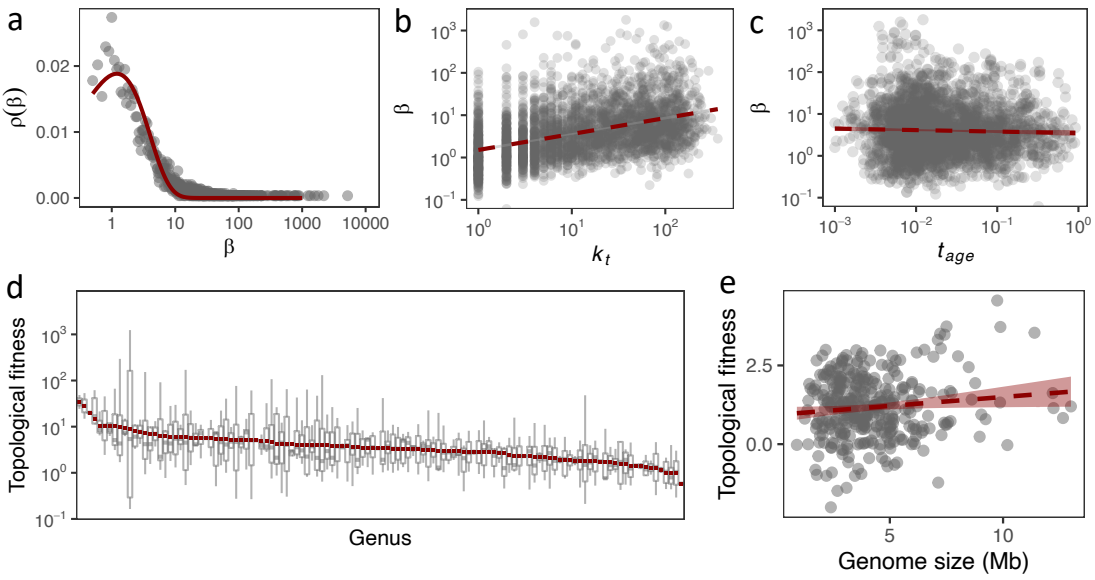
448 Bianconi-Barabasi model.



449

450 **Figure 3**. Dynamic of degree exponent (λ), average separation (d), and clustering coefficient

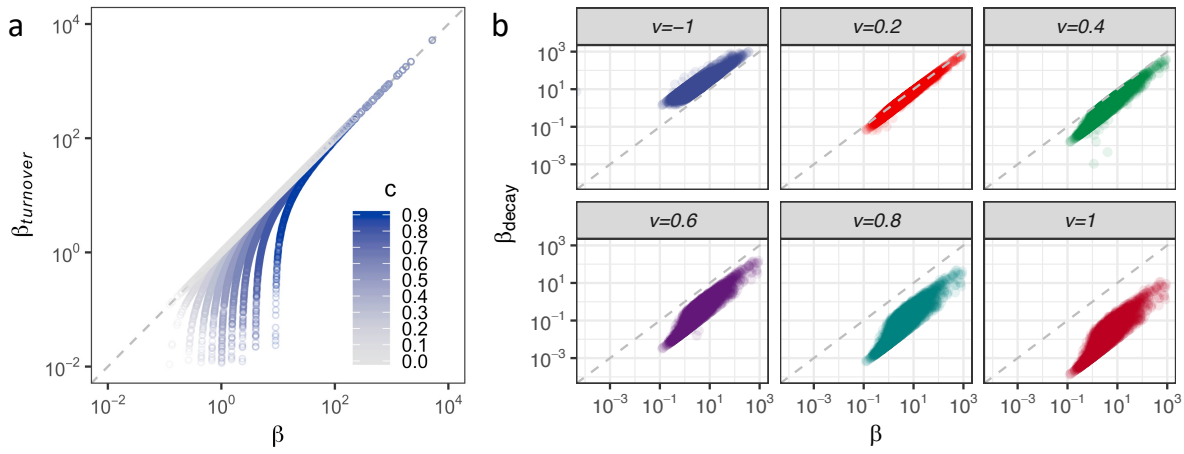
451 (C) of evolving Earth microbial co-occurrence network.



452

453

454 **Figure 4**. Degree growth exponont (β) of the Earth microbial co-occurrence network. (a)
455 distribution of β values. (b) The relationship between accumulated degree and β. (c) The
456 relationship between node age and β. (d) The boxplot of β values for various genera. The red
457 dot showed mean β values of corresponding genera. (e) The relationship between mean genome
458 size and mean β values of various genera.



459

460 **Figure 5**. The impact of node turnover and node decay on β value estimation. (a) The impact

461 of node turnover on β value estimation when turnover rate (c) changed from 0.0 to 0.9. (b) The

462 impact of node decay on β value estimation when decay exponent (v) changed from -1 to 1.

463 **Table S1**. Degree grwoth exponent values of various genera.

Genus	Mean	SD	*N*	Genus	Mean	SD	*N*
Pedobacter	330.5	607.2	4	*Lactobacillus*	5.5	4.7	8
Haliangium	199.9	595.0	9	*Clostridium_sensu_stricto*	5.5	7.1	8
Parcubacteria_genera_incertae_sedis	194.6	848.8	41	*Chryseobacterium*	5.4	7.7	5
Geopsychrobacter	76.5	145.4	4	*Pirellula*	4.9	4.5	4
Woesearchaeota_Incertae_Sedis_AR16	63.4	184.6	9	*Anaerolinea*	4.9	4.9	2
Desulfovibrio	54.2	135.2	11	*Pelobacter*	4.8	3.5	5
WPS-1_genera_incertae_sedis	36.9	80.0	6	*Saccharibacteria_genera_incertae_sedis*	4.7	5.3	2
Chondromyces	36.5	17.2	2	*Opitutus*	4.6	4.0	6
Ruminococcus	34.7	67.3	6	*Geobacter*	4.4	3.3	3
Terrimicrobium	34.1	27.2	2	*Armatimonadetes_gp2*	4.3	4.3	4
Spartobacteria_genera_incertae_sedis	32.9	58.9	10	*Kofleria*	4.2	3.0	4
Planctopirus	26.6	24.6	5	*Gp17*	4.2	5.8	6
Rubrobacter	25.6	48.2	5	*Ornatilinea*	4.2	2.5	8
Sphaerochaeta	25.4	44.3	4	*GpIIa*	4.1	2.4	3
Clostridium_XlVa	25.0	52.1	6	*Desulfobulbus*	4.0	2.5	5
BRC1_genera_incertae_sedis	24.9	36.1	3	*Fibrobacter*	4.0	3.7	3
Syntrophorhabdus	23.9	30.3	2	*Gp6*	3.9	3.5	20
Microgenomates_genera_incertae_sedis	18.9	38.8	7	*Parachlamydia*	3.8	3.2	2
Sphingobacterium	17.7	30.8	4	*Methanobacterium*	3.8	0.4	3
Desulfuromonas	16.2	25.3	6	*Halobacteriovorax*	3.8	3.3	5
Gimesia	16.0	22.3	8	*Latescibacteria_genera_incertae_sedis*	3.7	2.6	5
Chlorophyta	14.7	21.0	6	*Intestinimonas*	3.7	1.8	7
Zavarzinella	14.3	27.7	16	*Anaeromyxobacter*	3.5	4.8	3
Syntrophus	13.4	18.2	7	*Cystobacter*	3.5	1.7	3
Gemmata	13.1	24.0	5	*Sediminibacterium*	3.5	3.6	2
Acholeplasma	12.6	18.8	6	*Limisphaera*	3.4	3.4	2
Subdivision3_genera_incertae_sedis	12.5	20.2	12	*Bdellovibrio*	3.3	0.7	4
Gaiella	12.3	9.1	4	*Spiroplasma*	3.2	1.8	3
Porphyromonas	12.1	8.1	3	*Roseibacillus*	3.1	2.1	2
Desulfohalobium	11.9	14.8	2	*Clostridium_III*	3.1	3.2	9
Neochlamydia	11.4	14.5	3	*Blastopirellula*	2.9	2.2	4
Sphingomonas	10.9	13.0	4	*Legionella*	2.8	1.7	10
Subdivision5_genera_incertae_sedis	10.1	21.1	13	*Pararhodospirillum*	2.8	0.5	2
Desulfonatronum	9.9	11.9	5	*Gp16*	2.6	2.1	5
Algoriphagus	9.9	14.4	3	*Gp5*	2.6	2.4	4
Polycladomyces	9.8	4.1	5	*Gp25*	2.6	2.3	3
Byssovorax	9.5	11.2	3	*Armatimonadetes_gp4*	2.5	2.2	5
Gp10	9.4	13.9	3	*Spirochaeta*	2.3	0.0	2
Jahnella	9.3	11.4	7	*Pseudoflavonifractor*	2.2	0.5	4
Bacillariophyta	9.2	7.1	2	*Gemmatimonas*	2.2	1.3	11
Phaselicystis	8.4	6.7	2	*Thermomarinilinea*	2.2	1.6	3
Gp13	8.1	8.3	2	*Telmatocola*	2.1	1.4	2
Gp3	7.5	12.1	11	*Eubacterium*	2.1	2.1	3
Bacteroides	7.4	10.8	3	*SR1_genera_incertae_sedis*	2.0	1.4	3
Treponema	7.2	10.3	9	*Gp2*	2.0	1.9	4
Gp4	6.8	14.1	11	*Poribacteria_genera_incertae_sedis*	1.9	0.4	3
Aquicella	6.7	4.6	7	*Paenibacillus*	1.9	1.1	5
Azospirillum	6.6	4.6	3	*Taibaiella*	1.9	0.6	2
Hymenobacter	6.6	3.5	5	*Thermodesulfovibrio*	1.8	1.0	4
Gp1	6.6	5.7	6	*Microvirga*	1.8	0.2	2
Elusimicrobium	6.5	3.3	5	*Pseudenhygromyxa*	1.5	1.4	3
Rubinisphaera	6.4	7.1	5	*Armatimonas/Armatimonadetes_gp1*	1.4	0.7	4
Zymobacter	6.3	6.2	6	*Prevotella*	1.4	0.3	2
Saccharofermentans	6.1	6.6	3	*Pseudomonas*	1.4	0.7	2
Solirubrobacter	5.8	3.0	4	*Mycoplasma*	1.3	1.3	6
Omnitrophica_genera_incertae_sedis	5.5	10.0	11	*Clostridium_IV*	1.2	0.7	5
Dysgonomonas	5.5	6.2	2	*Roseomonas*	1.2	0.3	2

464

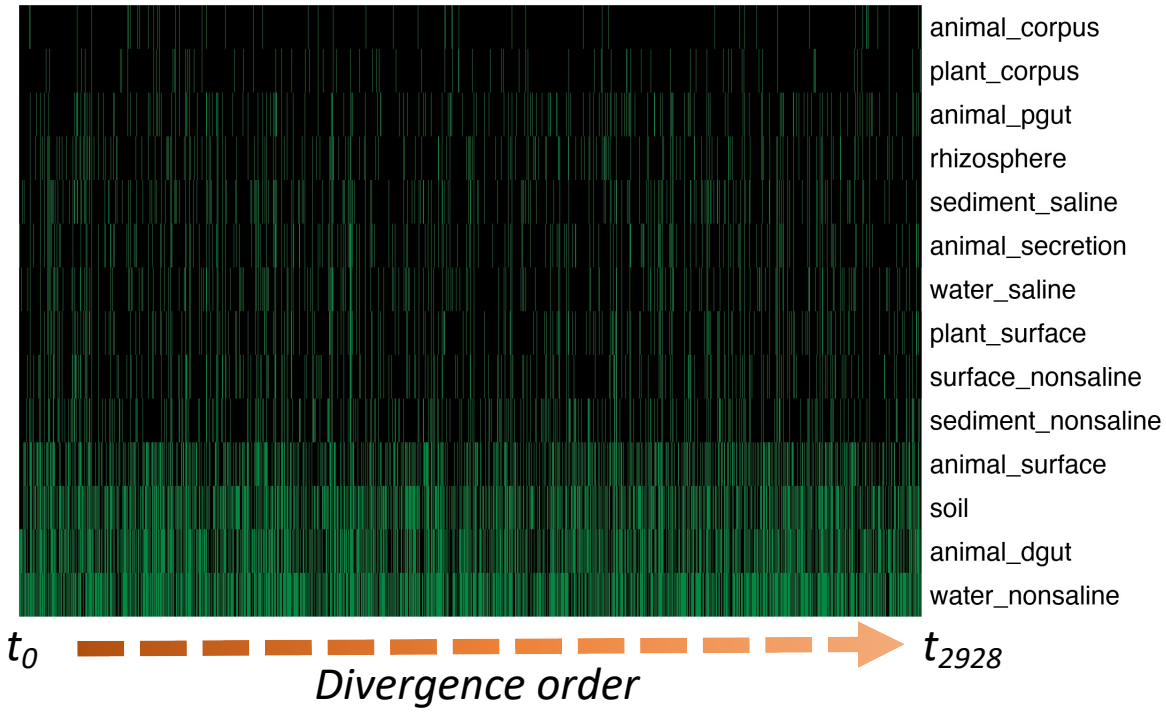


Figure S1. The divergence order of nodes from different environments.

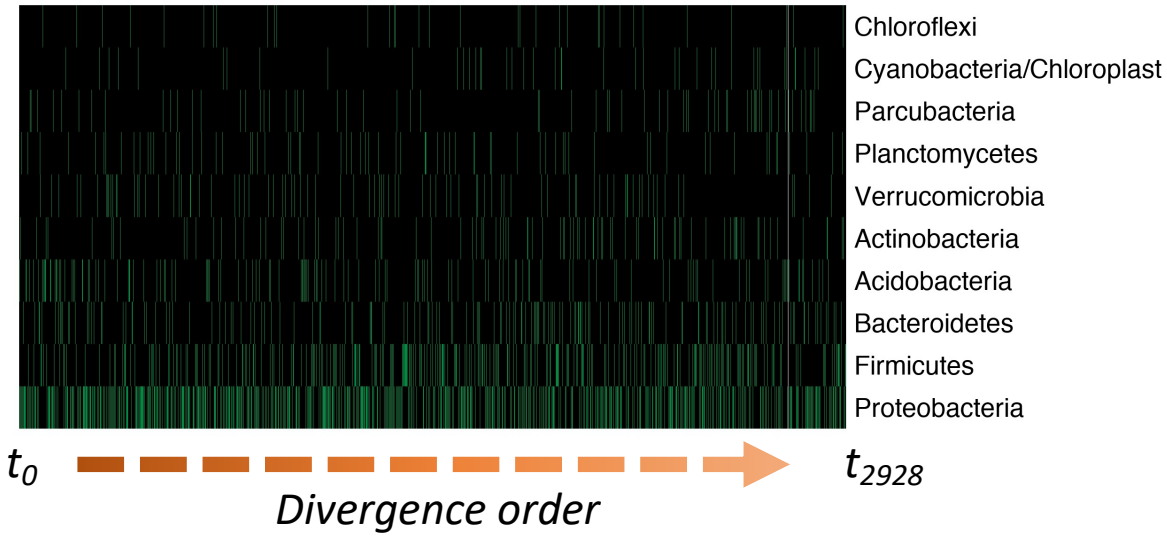


Figure S2. The divergence order of nodes from different taxonomic groups.

Figures

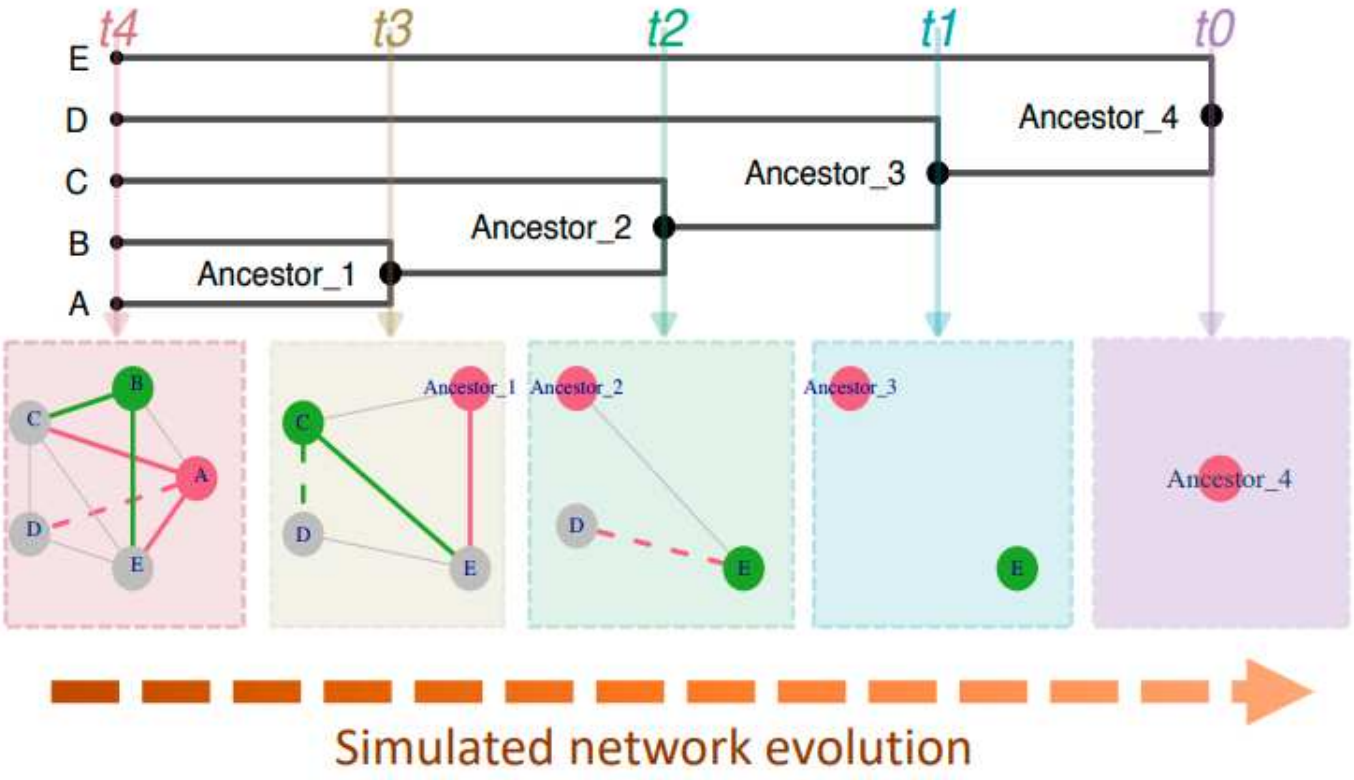


Figure 1

Please see the Manuscript PDF file for the complete figure caption

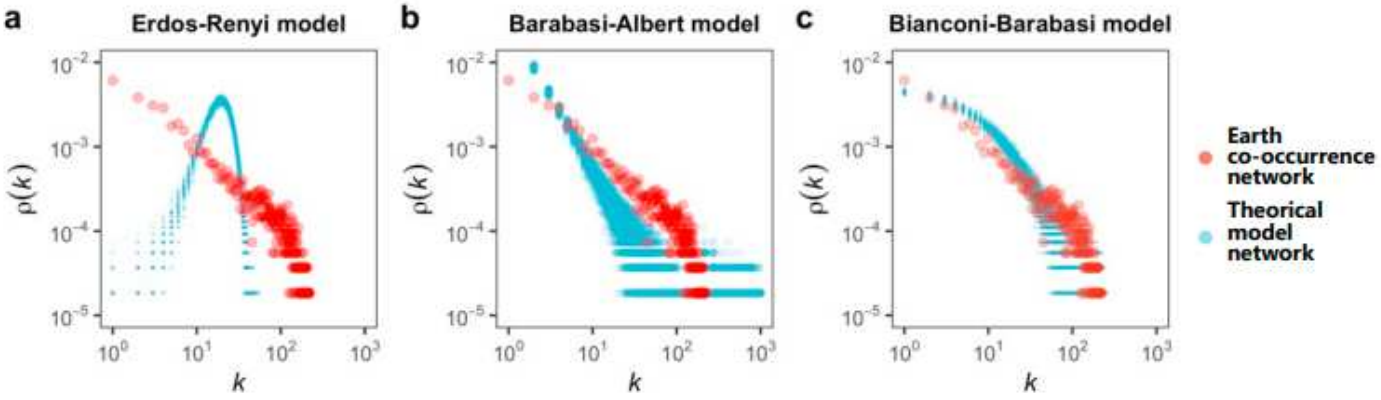


Figure 2

Please see the Manuscript PDF file for the complete figure caption

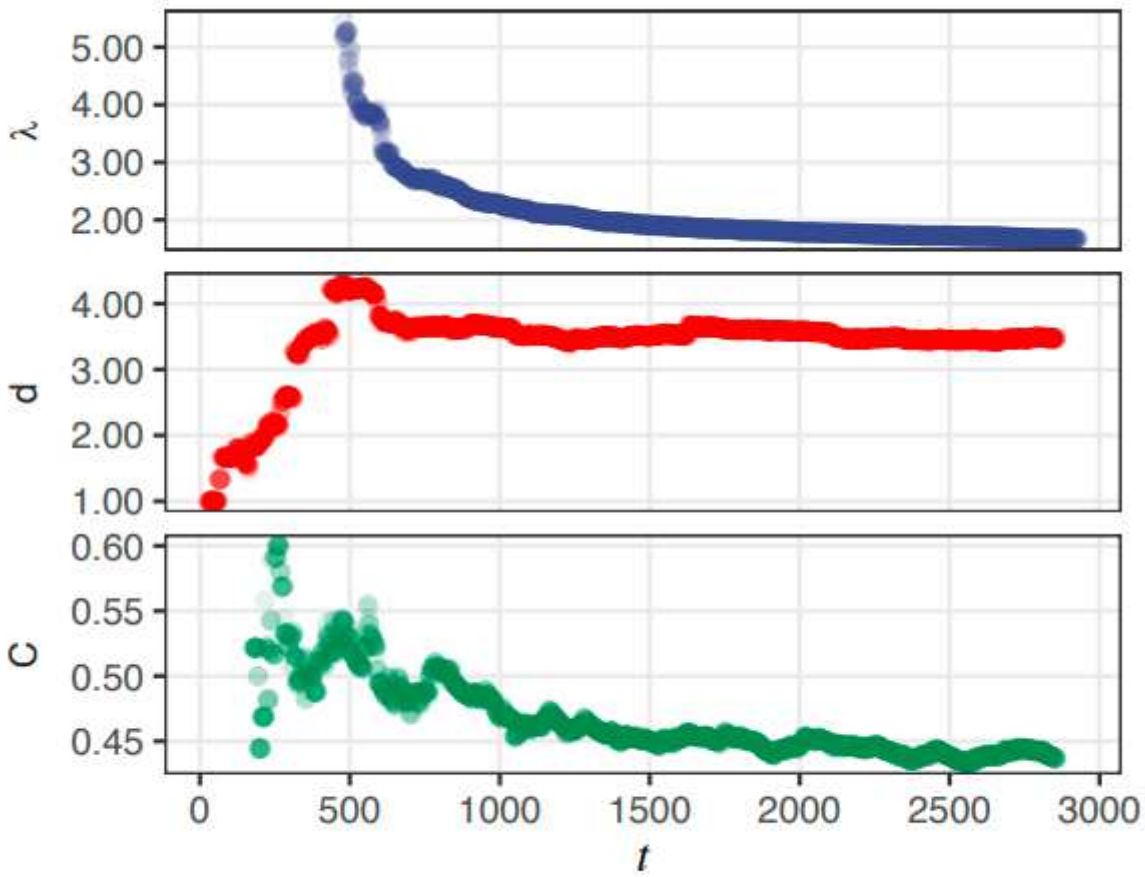


Figure 3

Please see the Manuscript PDF file for the complete figure caption

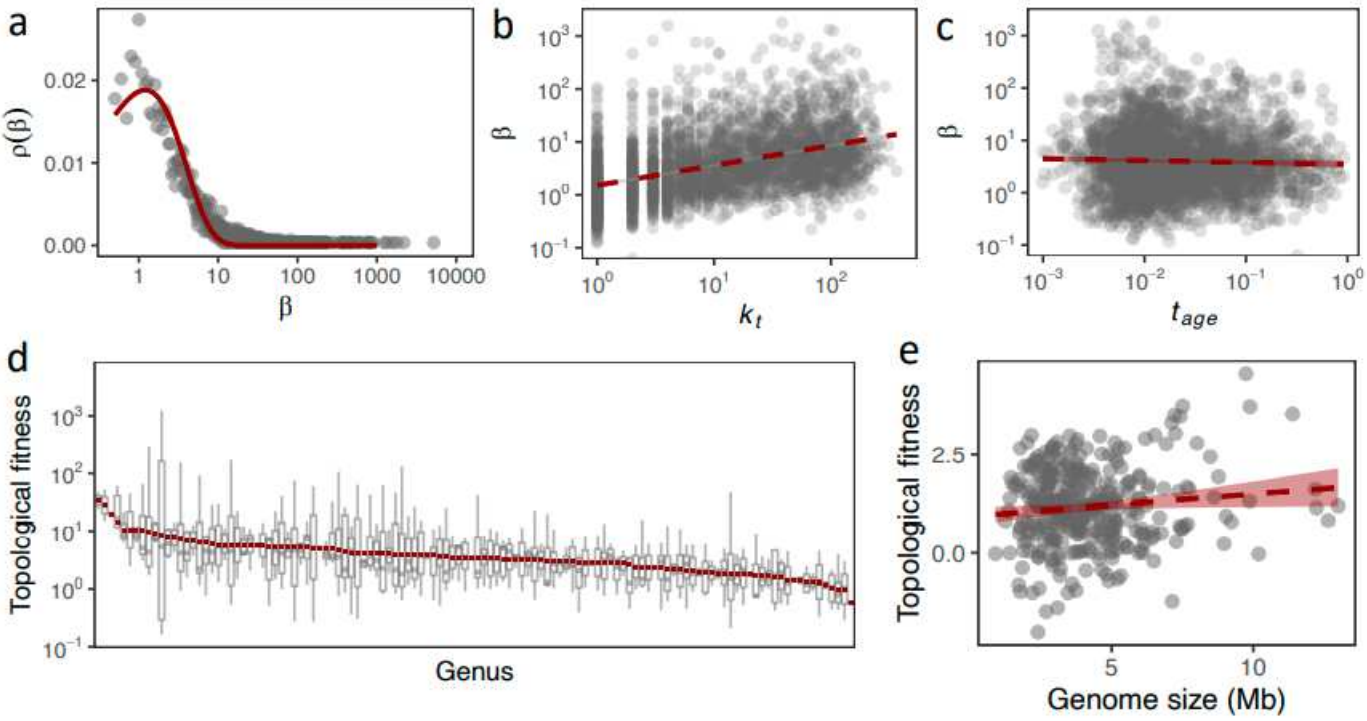


Figure 4

Please see the Manuscript PDF file for the complete figure caption

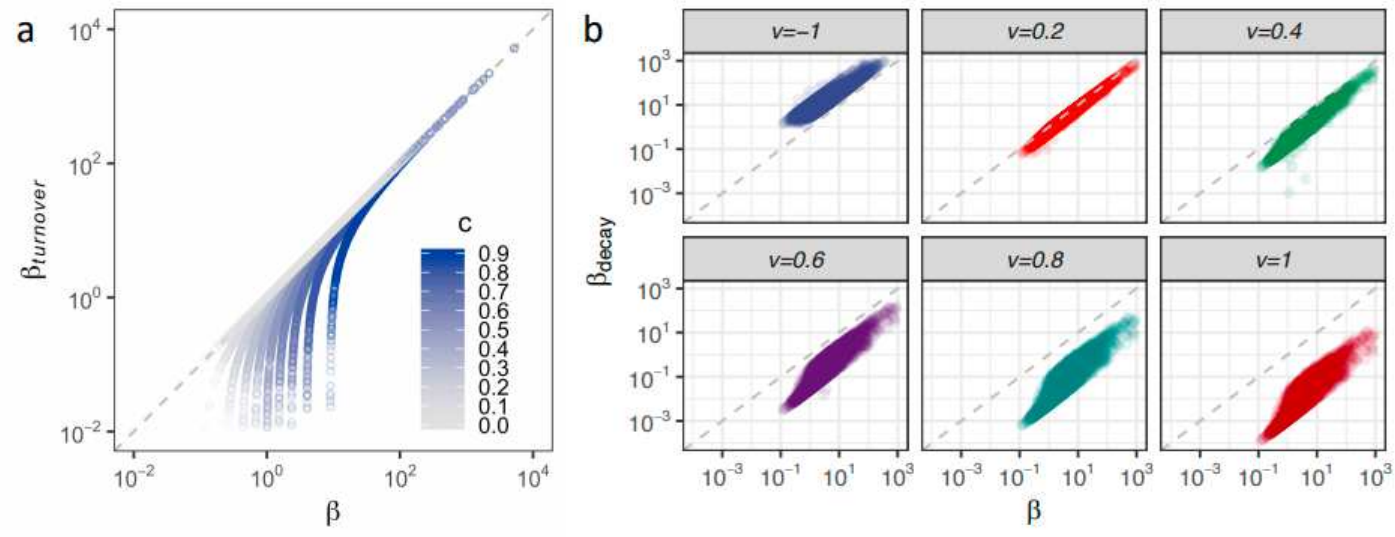


Figure 5

Please see the Manuscript PDF file for the complete figure caption

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- s2.png
- s1.png
- movie.mp4