Heterogeneous Popularity of Metabolic Reactions from Evolution

Mi Jin Lee[1] and Deok-Sun LEE[2]
[1]Department of Physics, Pusan National University, Busan, Republic of Korea
[2]School of Computational Sciences and Center for AI and Natural Sciences, Korea Institute for Advanced Study, Seoul, Korea

Metbolic Network, Evolution Model, Co-evolution dynamics, Power-law Distributions, Reaction Recruitment

Extended Abstract

Motivation. Metabolic networks inferred from genomic data are now available for thousands of species, enabling large-scale comparative analyses of cellular metabolism. For example, the BioCyc database contains metabolic information for 5470 bacterial species comprising more than 11000 reactions. Empirical analyses reveal a striking contrast: while different species contain similar numbers of metabolic reactions, the reactions themselves are unevenly distributed across species (Fig. 1). In particular, the number f of species containing a given reaction follows a heavy-tailed distribution close to $P(f) \sim f^{-1}$. Explaining the evolutionary origin of this distribution is important for understanding how metabolic diversity emerges across species.

Approach and Methodology. To investigate the evolutionary origin of the heterogeneous popularity of metabolic reactions, we propose an evolutionary model describing the coevolution of species and their metabolic networks [1], illustrated in Fig. 2. Species expand their metabolism by recruiting new reactions and diversify through speciation events, while descendant species inherit reactions from their ancestors. Our key hypothesis is that the popularity of a reaction is determined by the time τ at which it is first recruited during evolution. If reactions recruited at time τ spread to descendant species so that $f_\tau \sim e^{-c\tau}$, the popularity distribution follows from $P(\tau)\,d\tau = P(f)\,df$, which yields $P(f) \sim f^{-1}$ when $P(\tau)$ varies weakly.

Results. The model reproduces the heavy-tailed distribution of reaction popularity with an exponent close to 1 [Fig. 1(b)], consistent with empirical observations. Reactions recruited earlier during evolution are inherited by many descendant species through successive speciation events, whereas later reactions remain restricted to fewer lineages. As a result, the model generates the heterogeneous occurrence patterns of metabolic reactions observed in large metabolic databases.

Conclusions and Outlook. These results show that the heterogeneous popularity of metabolic reactions can arise from the coevolution of species diversification and reaction recruitment during metabolic evolution. The proposed framework provides a simple evolutionary mechanism underlying the statistical organization of metabolic reactions across species.

References

[1] Mi Jin Lee and Deok-Sun Lee. Heterogeneous popularity of metabolic reactions from evolution. *Phys. Rev. Lett.*, 132:018401, Jan 2024.

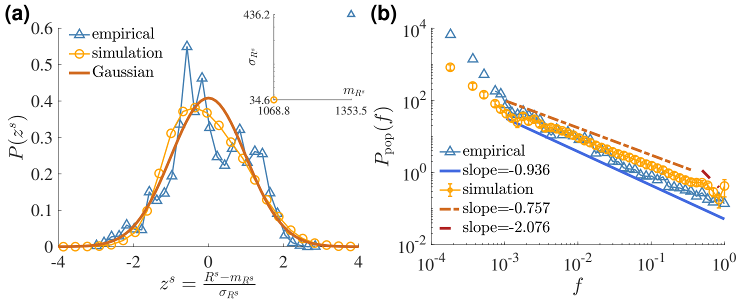

Figure 1: **Statistics of the species-reaction association in empricial data and the network evolution model.** (a) Standardized distributions of the number of reactions R^s in a species s. The standardized variable $z^s = \frac{R^s - m_{R^s}}{\sigma_{R^s}}$ is used with the mean m_{R^s} and standard deviation σ_{R^s} from empirical data (triangle) and from the model (circle). The solid line shows the Gaussian distribution $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$. (b) Distributions of the reaction popularity.

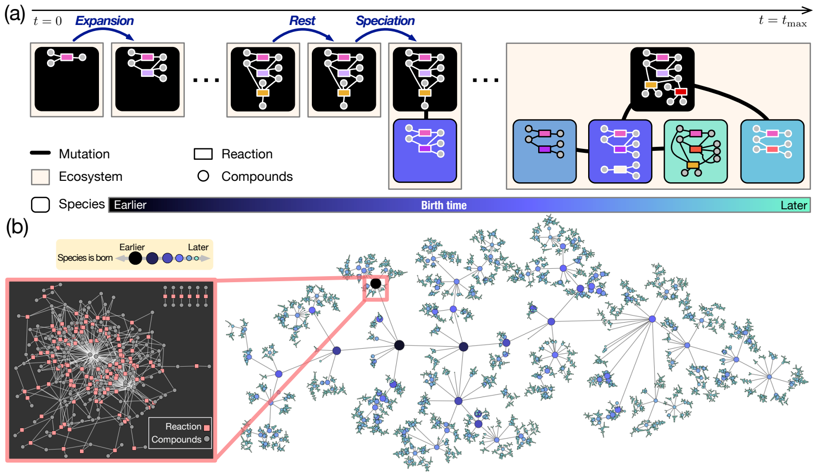

Figure 2: **Network evolution model.** (a) A bipartite network of reactions (rectangles) and compounds (circles) represents each species. At every time step, each species may do nothing (rest) or evolve by either gaining a new reaction (expansion) or giving birth to a daughter species inheriting active components formed by a new reaction replacing an old one (speciation). (b) The species tree from a simulation is shown, where nodes represent 5660 species and links represent the parent-daughter relationship. Node size and color vary with the birth time of the corresponding species. The metabolic network of the oldest species is shown in the left box.