NSPCS2026

Inflationary metabolic evolution and heterogeneous reaction popularity

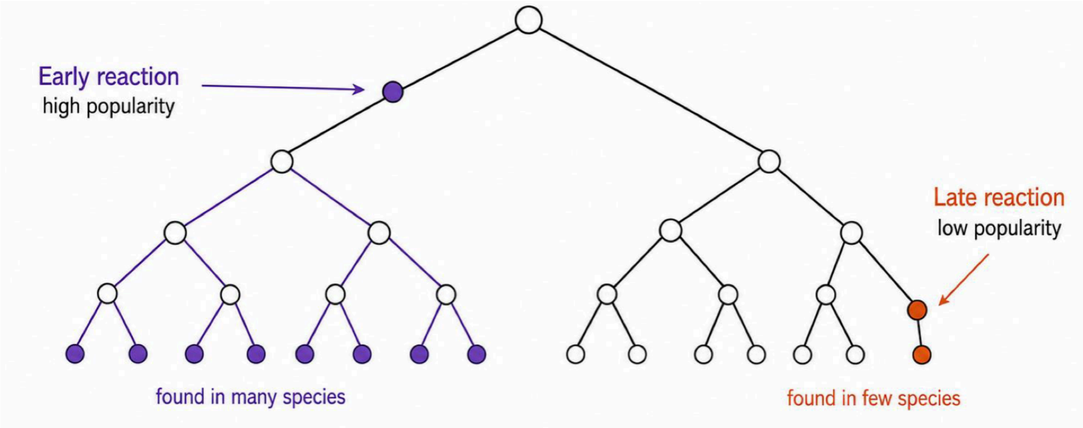
Deok-Sun Lee

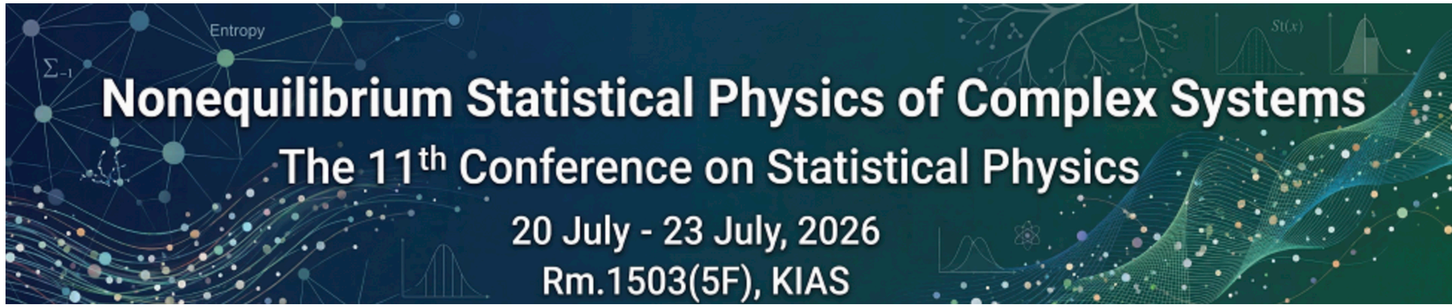
Korea Institute for Advanced Study (KIAS)
Seoul, Korea

in collaboration with

Mi Jin Lee

Pusan National University
Busan, Korea

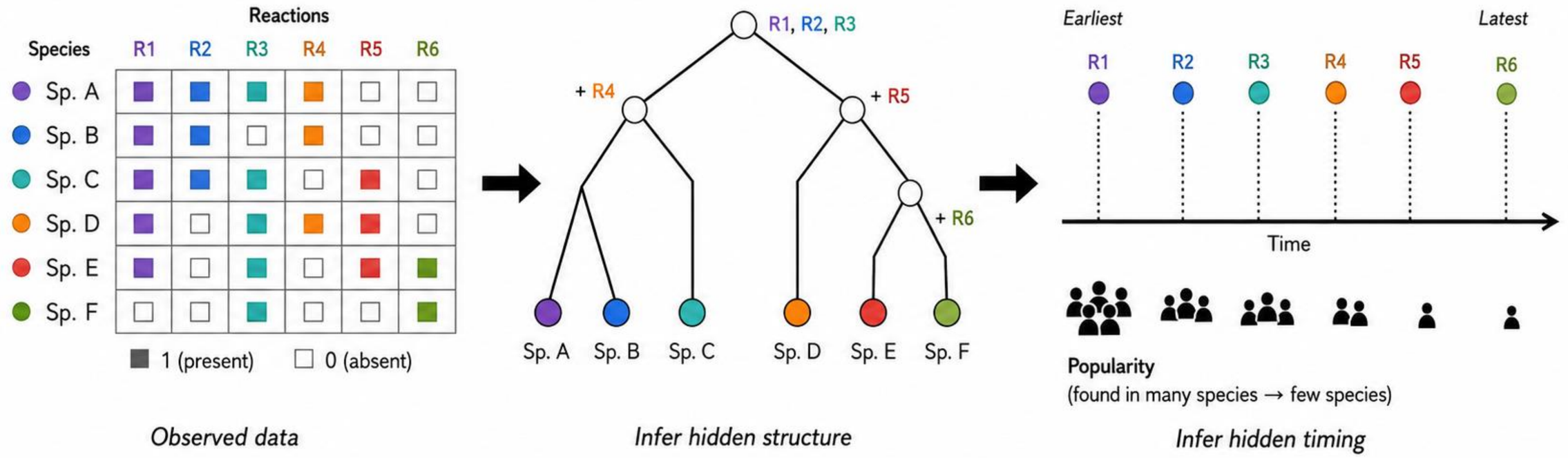




M.J. Lee and DSL, *Heterogeneous popularity of metabolic reactions from evolution*, Phys. Rev. Lett. **132**, 018401 (2024).

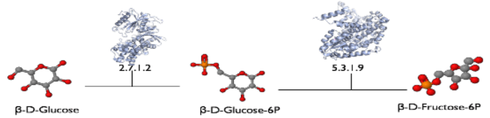
From metabolic repertoires to hidden evolutionary history

Main question of this study

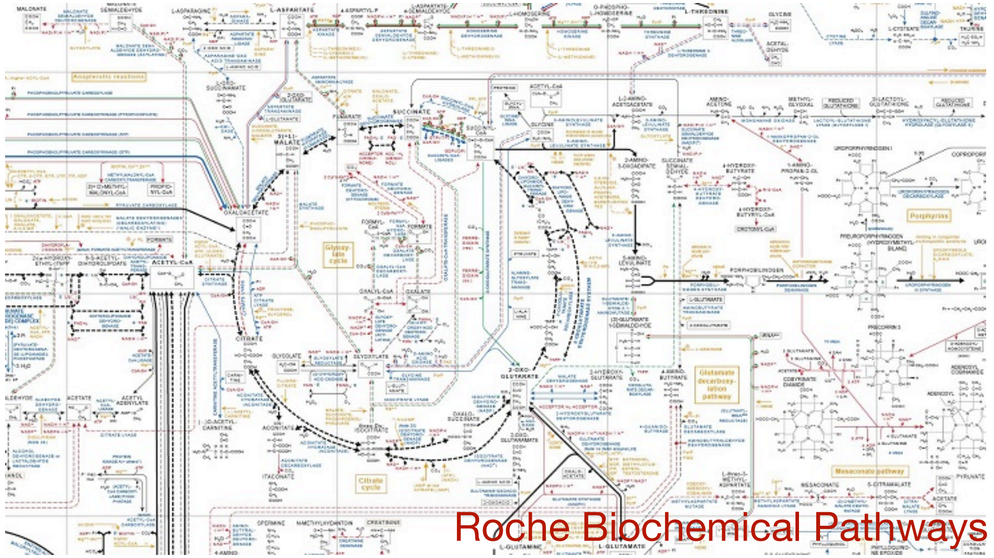


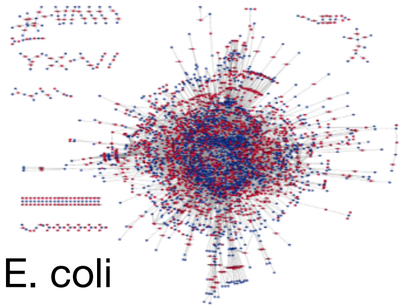
Q: Can present-day reaction sets reveal the evolutionary history of metabolism?

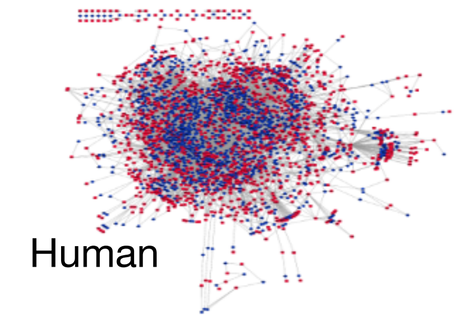
Metabolism across species



[2.7.1.1] ATP + alpha-D-Glucose <=> ADP + alpha-D-Glucose 6-phosphate
[5.3.1.9] alpha-D-Glucose 6-phosphate <=> beta-D-Fructose 6-phosphate
[2.7.1.11] ATP + beta-D-Fructose 6-phosphate <=> ADP + beta-D-Fructose 1,6-bisphosphate
[4.1.2.13] beta-D-Fructose 1,6-bisphosphate <=>
Glycerone phosphate + D-Glyceraldehyde 3-phosphate




E. coli


Human

metabolic network =
bipartite network consisting of reactions and compounds present in a species

BioCyc Database Collection

Species-reaction matrix $A_r^s = 1, 0$ of size
$(S_{\text{emp}} = 5470) \times (R = 11057)$ ver 19.0. & $(S_{\text{emp}} = 19182) \times (R = 12970)$ ver. 29.0

Similar sizes of metabolism

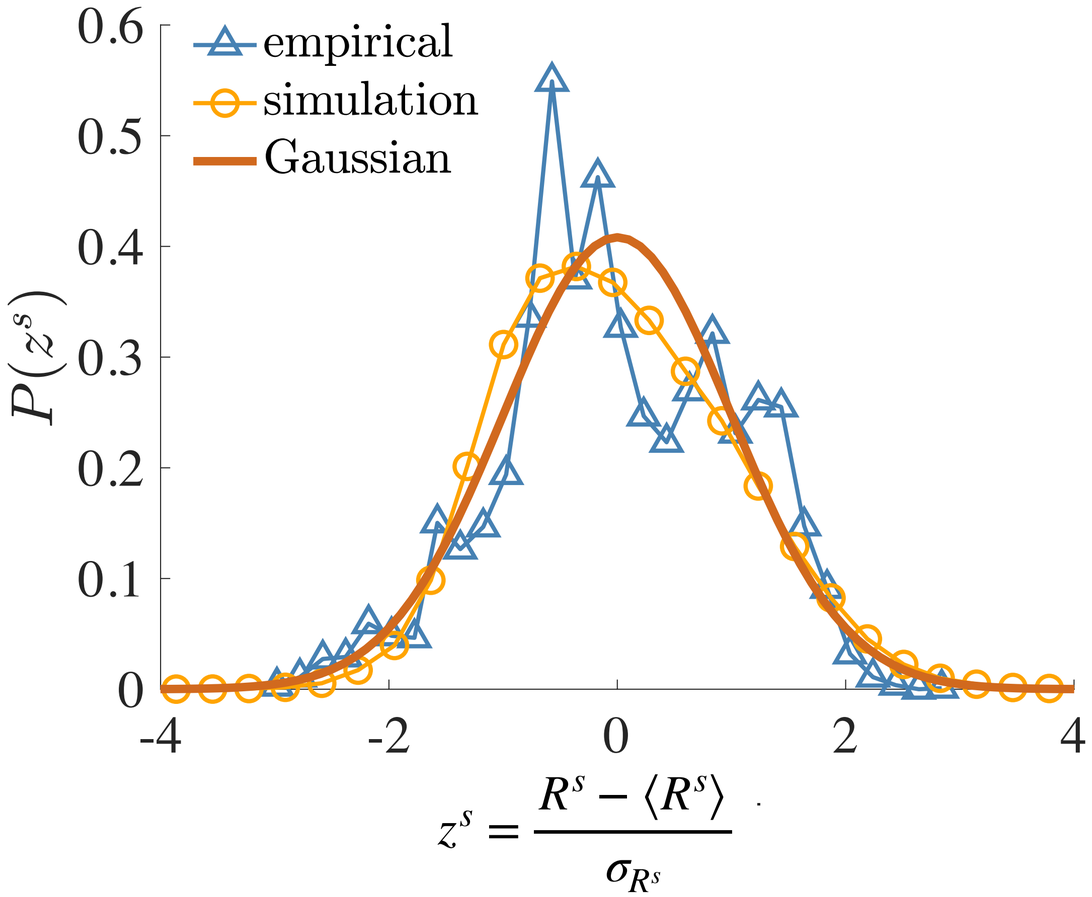
Size of metabolism

Number of reactions in a single species

$$R^s = \sum_{r \in \mathcal{R}} A_r^s \text{ for } s \in \mathcal{S}_{\text{emp}}$$

Mean $\langle R^s \rangle = 1353.5$

Standard deviation $\sigma_{R^s} = 436.2$



Reaction popularity is broadly distributed

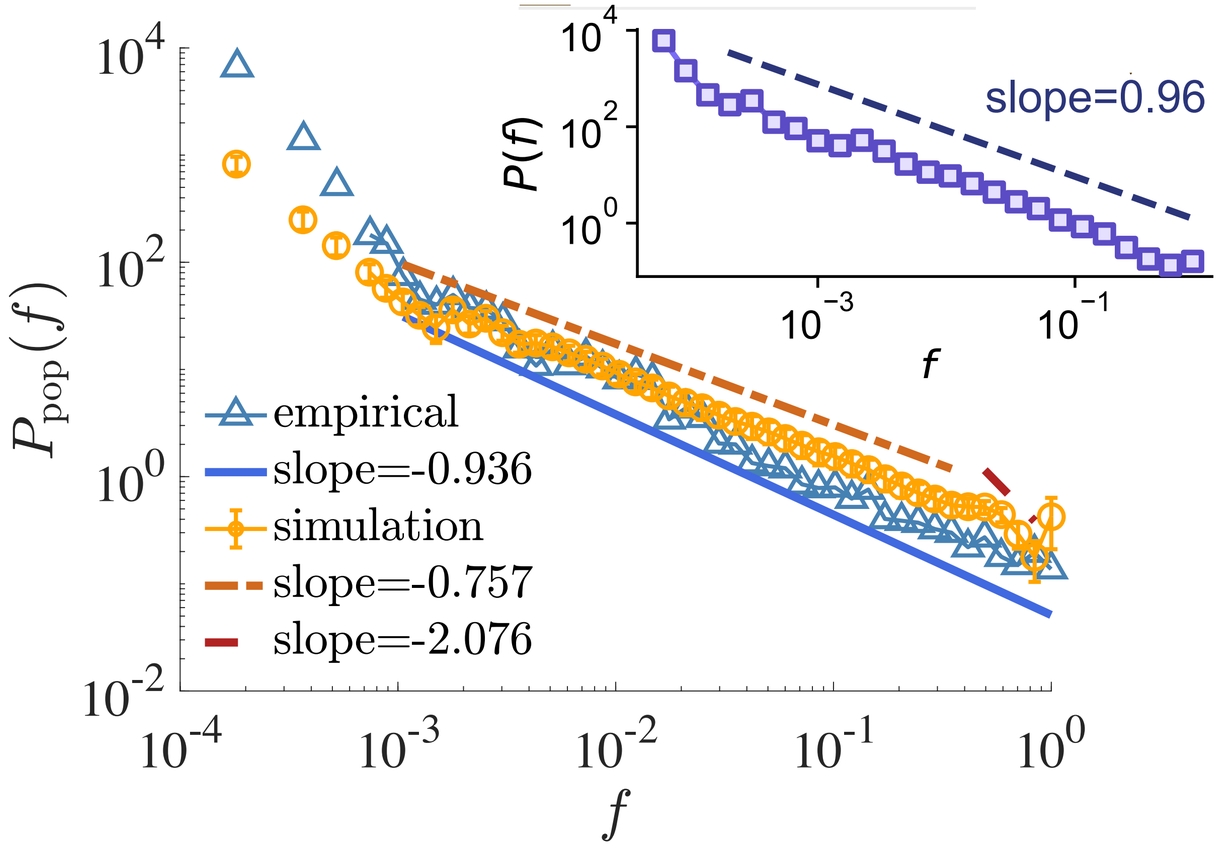
1/f law

Reaction popularity

$$f_r \equiv \frac{\sum_{s \in \mathcal{S}_{\text{emp}}} A_r^s}{S_{\text{emp}}} \text{ for } r \in \mathcal{R}$$

follows a power-law distribution with exponent one

$$P(f) = \frac{1}{R} \sum_r \delta(f_r - f) \sim f^{-1}$$



Bernhardsson, Gerlee, and Lizana, BMC Evol. Biol. 11, 20 (2011)

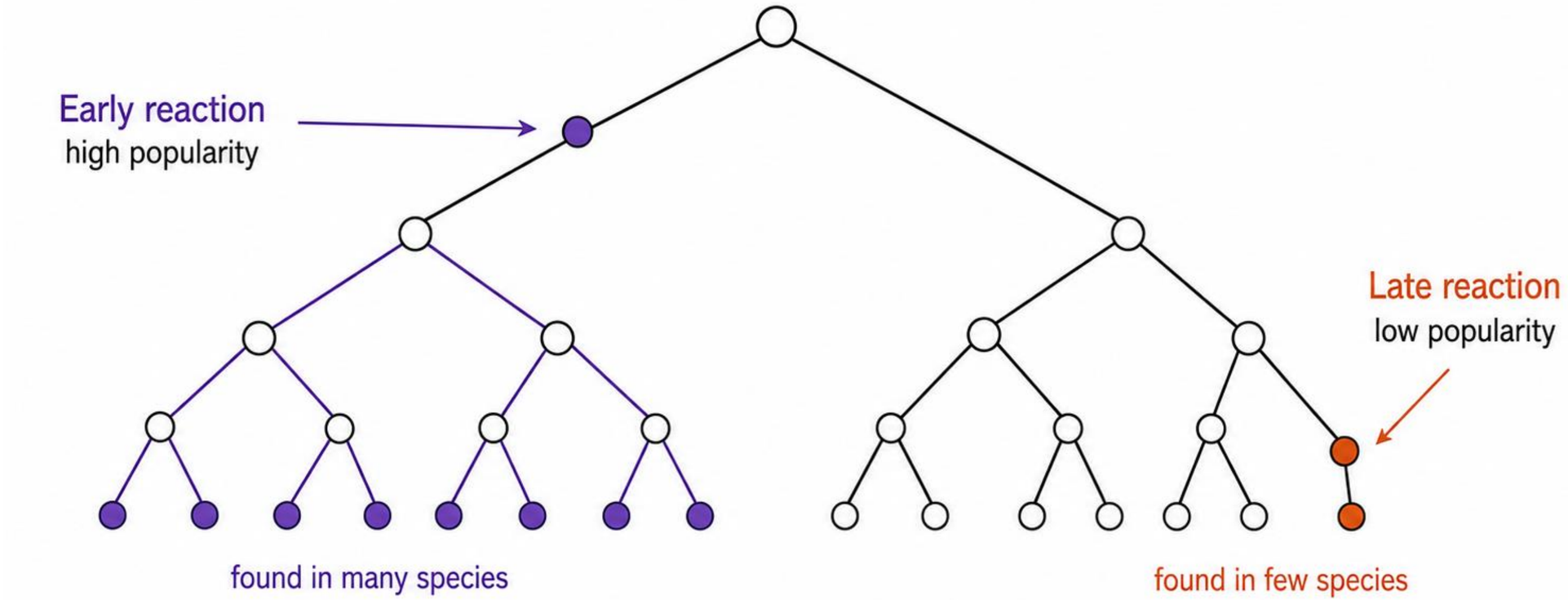
Kim, Lee, Kahng, Sci. Rep (2015); JKPS (2019)

Lee and Lee, PRL (2024)

Why is it such a clear power-law?

- **Question** Why is it such a clear power-law? Any principle there?
- **Answer (1)** f_r reflects the biochemical importance of reaction r for life on earth and the importance is so heterogeneous as to be characterized by the power-law distribution.
- **Answer (2)** f_r is built up over time in the **evolution of metabolism**

Popularity may encode recruitment time

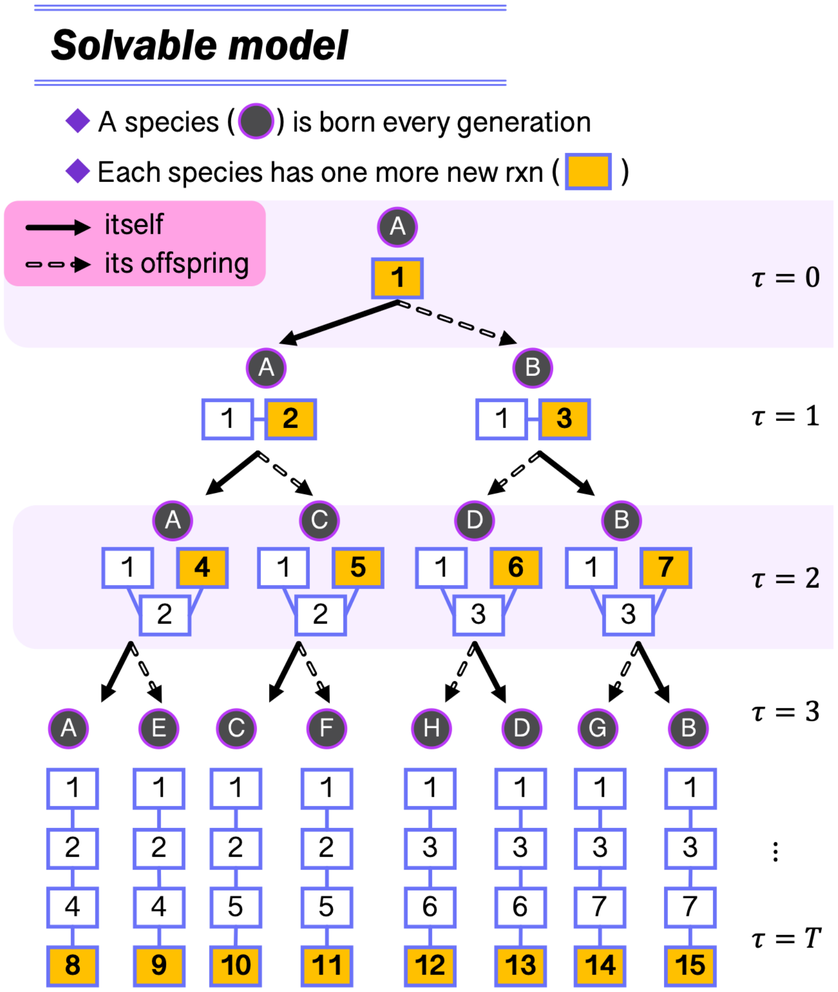


early recruitment → large f_r

late recruitment → small f_r

Let's be more quantitative

How recruitment time determines popularity



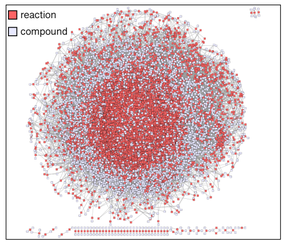
- Number of species $S(\tau) = 2^{\tau}$
- The popularity f_r is determined by recruitment time τ_r

$$f_r = \frac{S_r(\tau)}{S(\tau)} = \frac{2^{\tau-\tau_r}}{2^{\tau}} = 2^{-\tau_r}$$ exponentially decreasing with τ_r

- Popularity distribution $P_{pop}(f) = P_{\text{rec}}(\tau)\left|\frac{d\tau}{df}\right|_{f=2^{-\tau}}$
- First-recruitment time distribution $P_{\text{rec}}(\tau_r) \propto S(\tau_r) = 2^{\tau_r}$
- $P_{pop}(f) = P_{\text{rec}}(\tau)\left|\frac{d\tau}{df}\right|_{f=2^{-\tau}} \sim f^{-1}f^{-1} \sim f^{-2}$

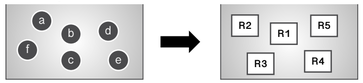
Co-evolution of the phylogenic tree and metabolic networks

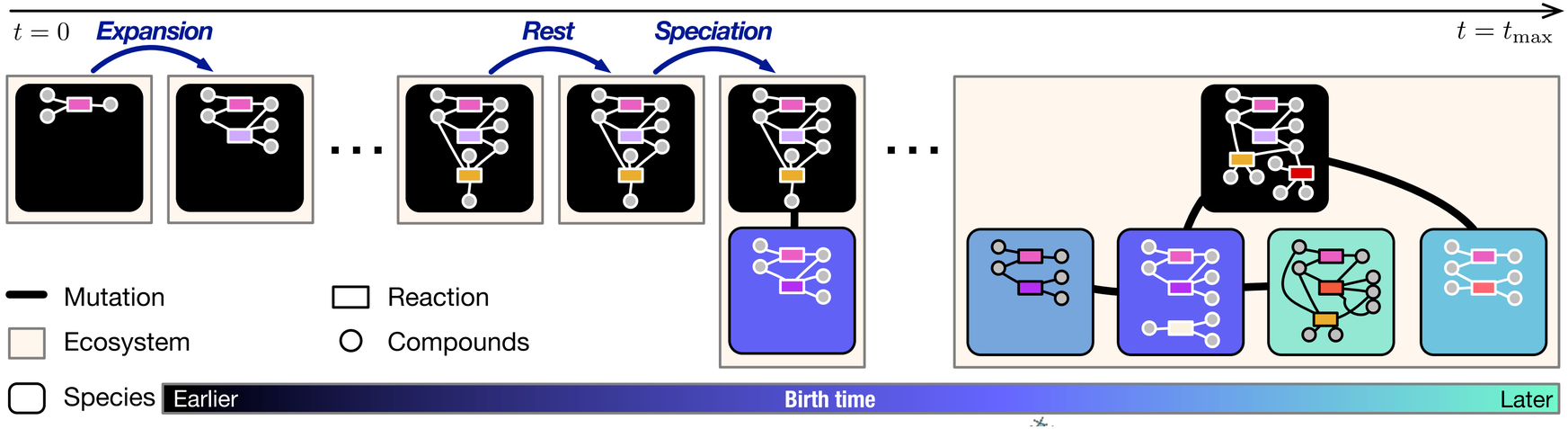
$\tilde{R}$: universal pool of
11057 reactions with 7620 compounds



C^E: 178 externa
available
compounds

$\tilde{R}^{(sa)}$: 440
stand-alone reactions





- Initially ($t = 0$), a single species with a single reaction from $\tilde{R}^{(sa)}$

- From t to $t + 1$, every species s selects a *recruitable* reaction r_{new} from the universal pool $\tilde{R}$, and
  - If there is no *similar* reaction to r_{new} in the current metabolic network of s, then r_{new} is added to the network of s **(expansion)**
  - if a similar reaction r_{sim} is found in s, then
    a species s' is born with probability μ from s with r_{new} in place of r_{sim} **(speciation)**
    Nothing happens otherwise **(rest)**

- The above is repeated until the total number of species reaches the empirical value $S_{\text{emp}} = 5470$.

Simulation

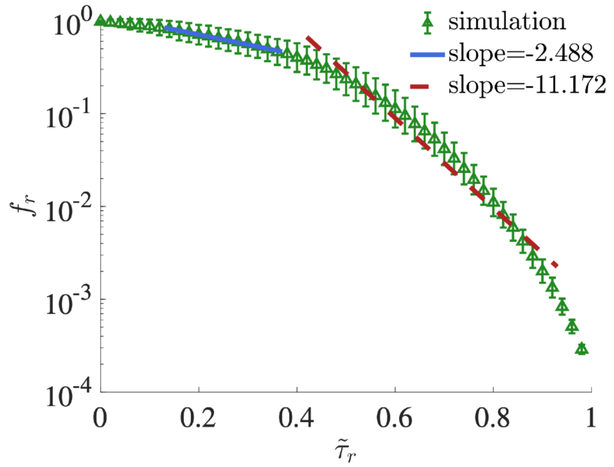
Network expansion + speciation

Our model

Finite pool of reactions

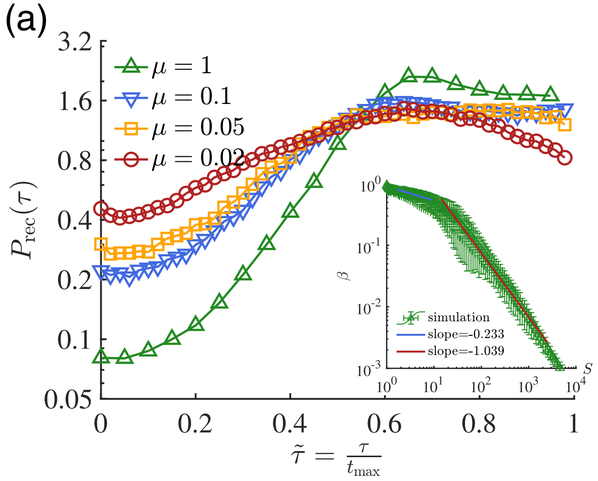
$$P_{pop}(f) = P_{\mathrm{rec}}(\tau)\left|\frac{d\tau}{df}\right|_{f=2^{-\tau}}$$

Popularity vs. First-recruitment time



$$f_r = \frac{S_r}{S} \sim \frac{e^{(1/2)\mu\alpha(\tau-\tau_r)}}{e^{(1/2)\mu\alpha\tau}} \sim e^{-(1/2)\mu\alpha\tau_r}$$

First-recruitment time distribution

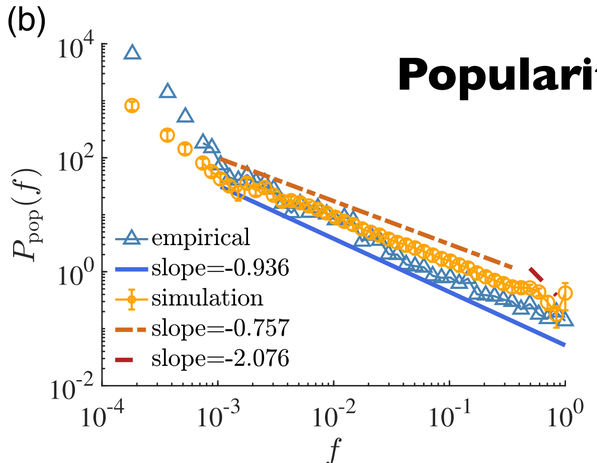


$$P_{\mathrm{rec}}(\tau_r) \sim \beta(\tau_r)S(\tau_r)$$

β : prob. a recruited reaction is brand new

$$\beta(\tau) \sim S(\tau)^{-\theta} \text{ with } \theta \simeq \begin{cases} -0.2 & \text{for } \tau \lesssim \tau_* \\ -1.04 & \text{for } \tau \gtrsim \tau_* \end{cases}$$

$$P_{\mathrm{rec}}(\tau_r) \sim \begin{cases} e^{(const)\tau_r} & \text{for } \tau_r \lesssim \tau_* \\ \text{const} & \text{for } \tau_r \gtrsim \tau_* \end{cases}$$



Popularity distribution

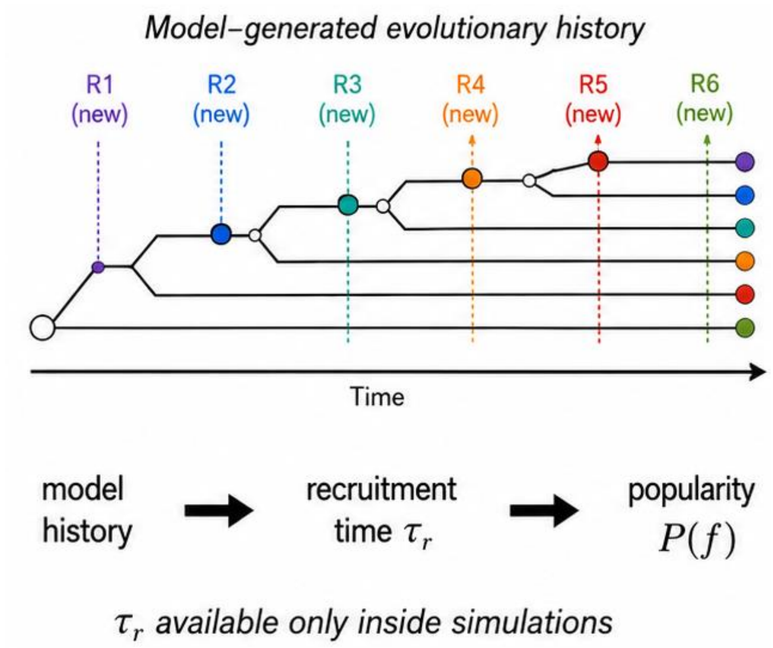
$$P_{\mathrm{pop}}(f) = P_{\mathrm{rec}}(\tau)\left|\frac{d\tau}{df}\right|_{f=2^{-\tau}} \sim \begin{cases} f^{-1} & \text{for } f \lesssim f_* \\ f^{-2} & \text{for } f \gtrsim f_* \end{cases}$$

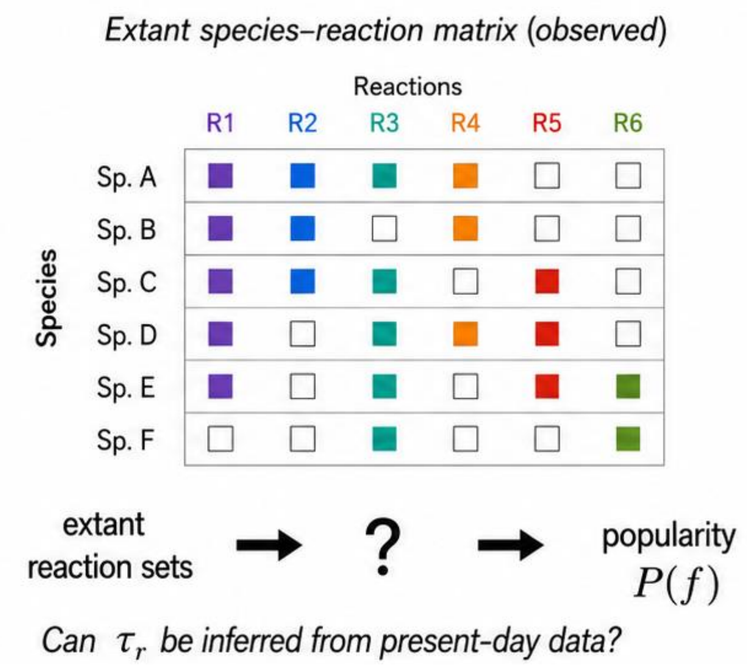
Lee and Lee, PRL (2024)

Timing of reaction recruitment?

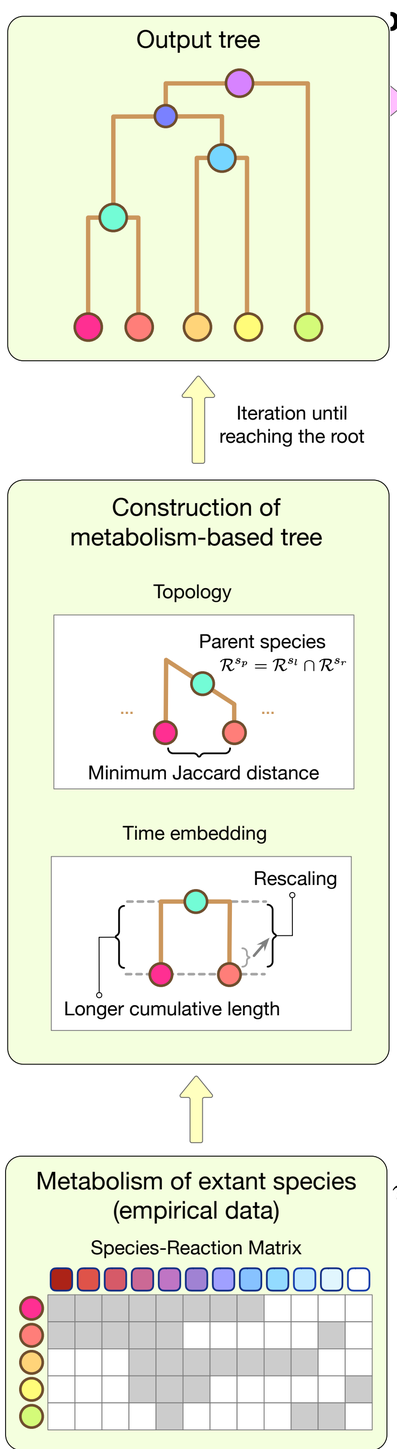
No such crossover seen in the empirical popularity distribution.

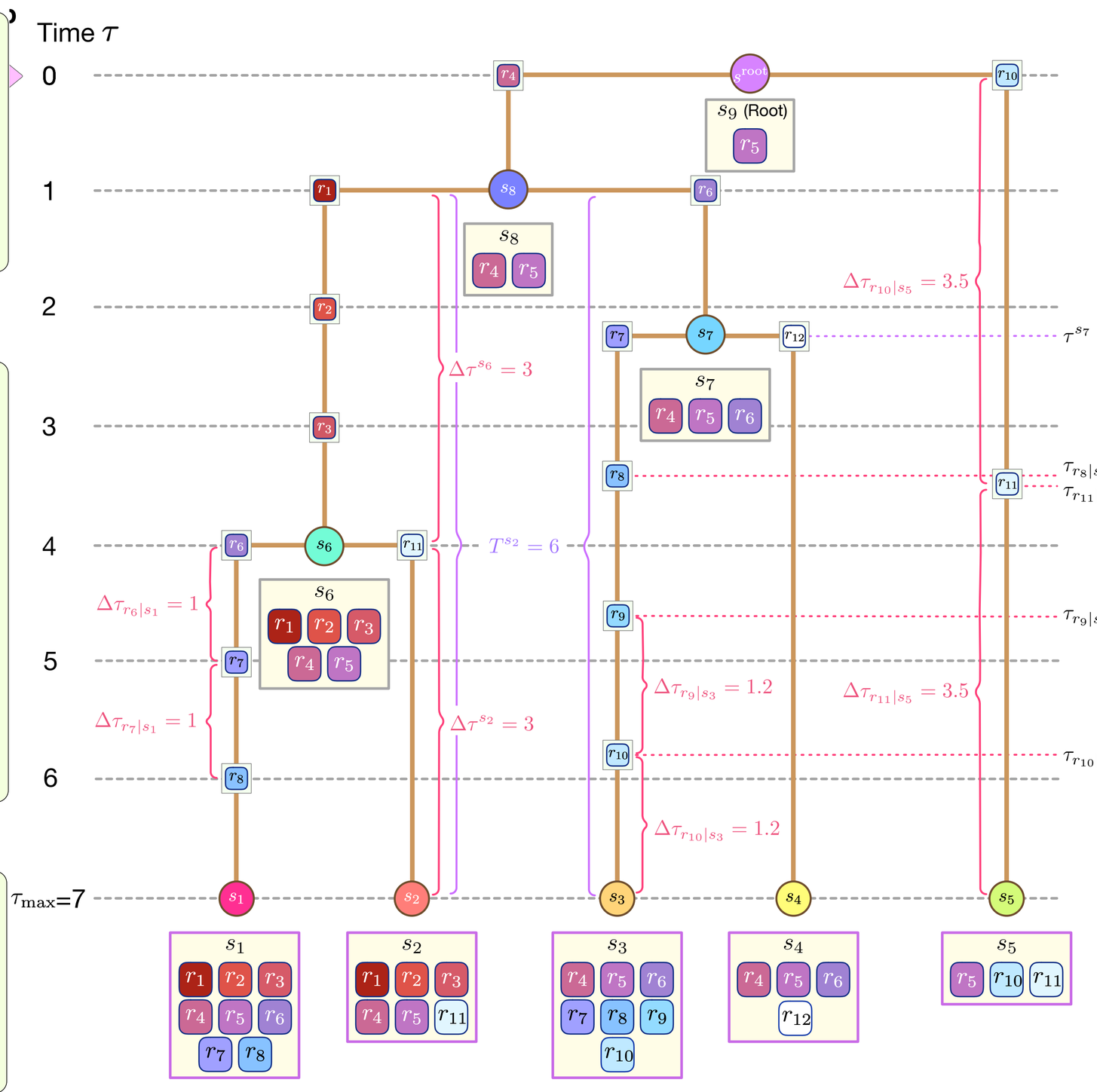
The previous model assumes that every species recruits reactions regularly.





Can we infer hidden recruitment times from present-day metabolic repertoires?

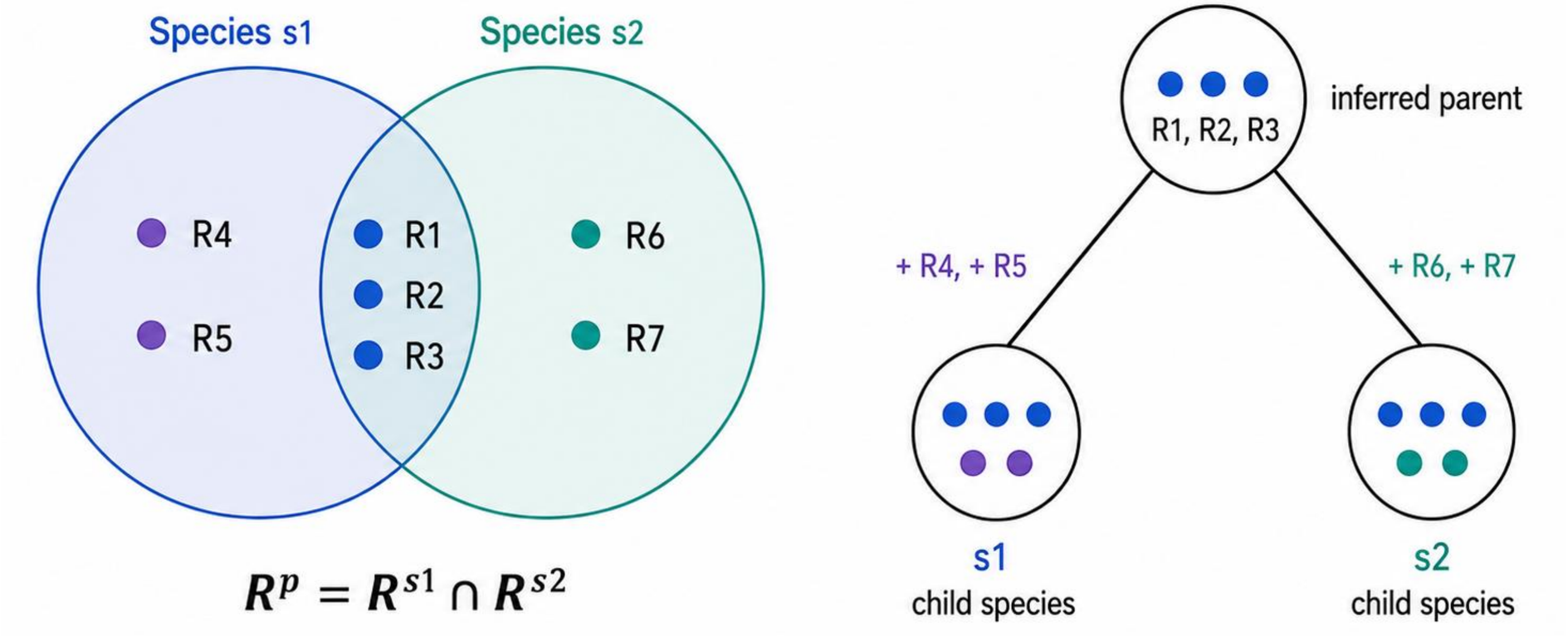




New idea to build a phylogeny and embed it in time

- Merge the closest reaction sets and assign the intersection to each internal node
- Read branch-specific reaction gains as recruitment events
- Place all extant species at equal time and equalize branch durations.

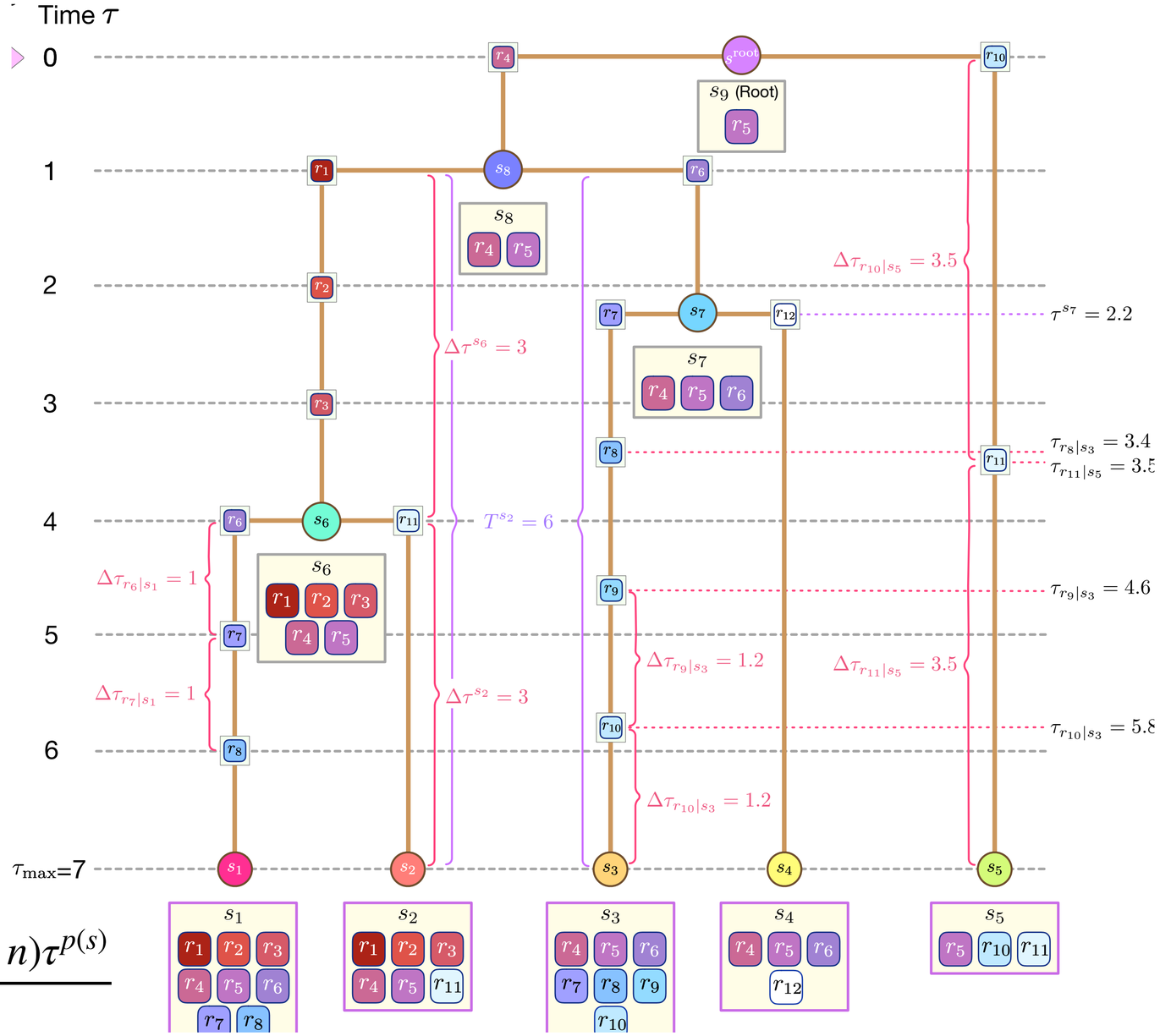
The parent keeps the shared metabolism



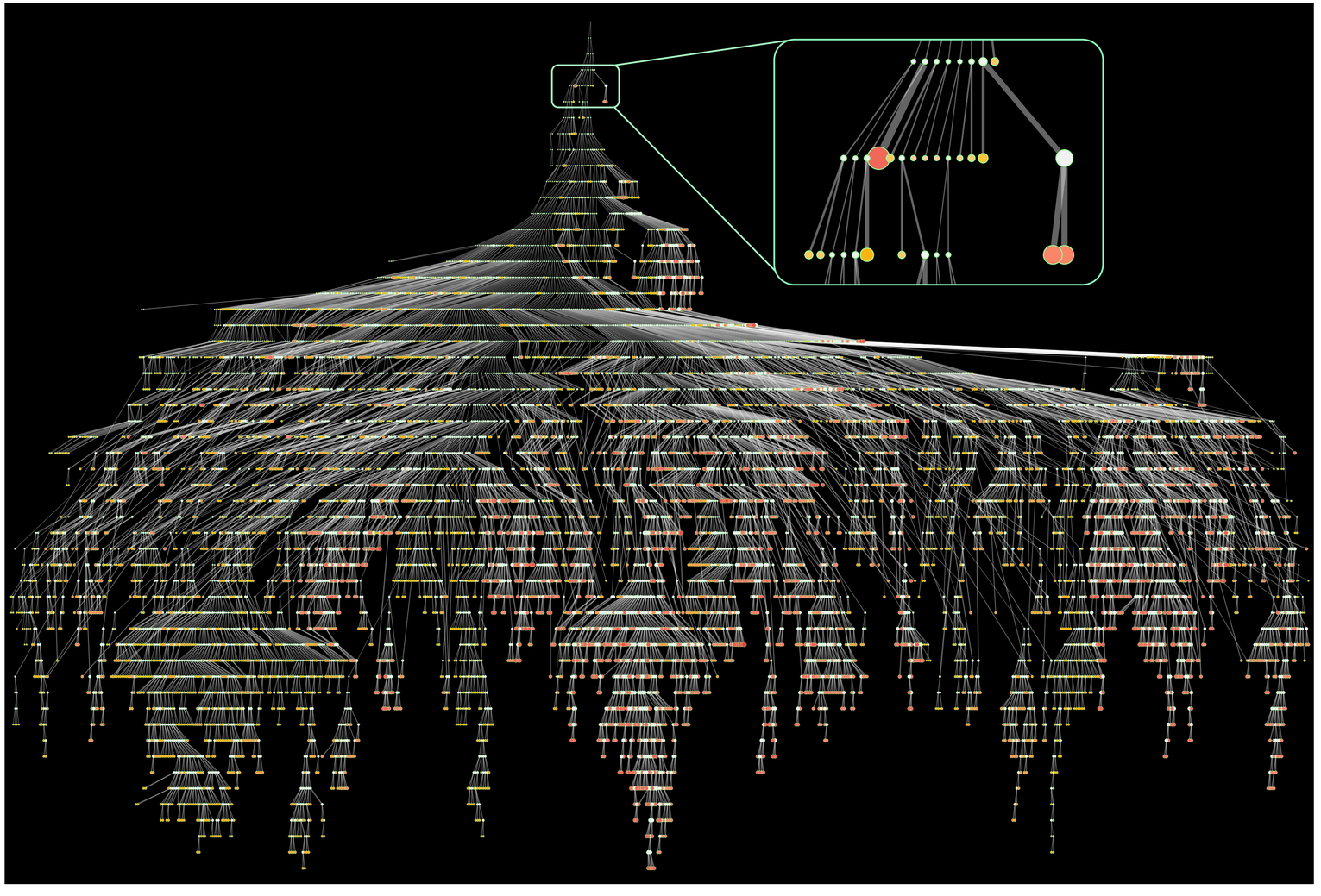
Shared reactions define the ancestral core; differences define branch-level recruitment events.

Evolutionary time

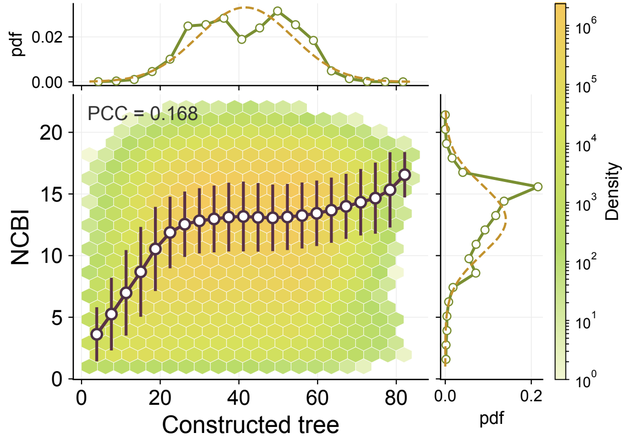
- **Species age** $T^s \equiv \tau_{\max} - \tau^s$
- Reactions recruited along a branch $\mathscr{M}^s = \mathscr{R}^s \backslash \mathscr{R}^{p(s)}$
- 0) $T^s = \begin{cases} 0 & (s \in \mathscr{S}^{\mathrm{emp}}) \\ \to \infty & (s \notin \mathscr{S}^{\mathrm{emp}}) \end{cases}$ initially
- i) Parent's age $T^s = \max(T^{s_1} + m^{s_1}, T^{s_2} + m^{s_2})$
- ii) Rescaling time for all descendants s' and s''
  $T^{s'} \to T^{s'} \dfrac{T^s}{T^{s_1} + m^{s_1}}$ and $T^{s''} \to T^{s''} \dfrac{T^s}{T^{s_2} + m^{s_2}}$
- i) and ii) are repeated until all ages become finite
- **Reaction recruitment time** $\tau_{r|s} = \dfrac{n\tau^s + (m^s - n)\tau^{p(s)}}{m^s}$
- Inter-recruitment time $\Delta\tau_{r|s} = \dfrac{\Delta\tau^s}{m^s}$



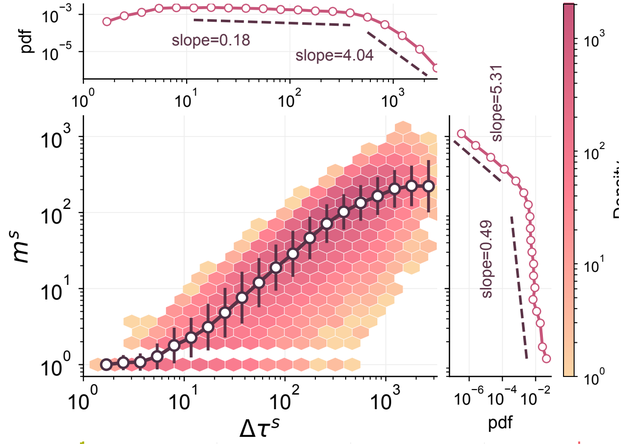
Metabolism-based phylogeny



Shortest path length between species



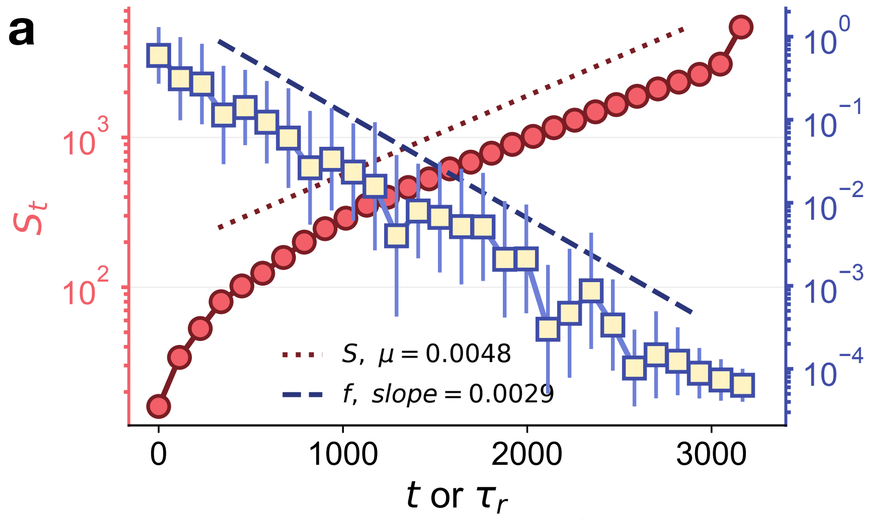
Branch lengths (no. rxns and time durations)



First-recruitment time distribution is flat

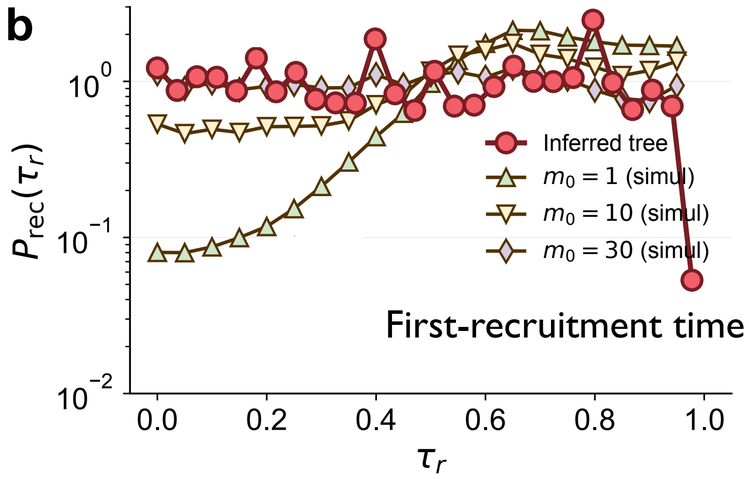
$\tau_r \equiv \min_s \tau_{r|s}$ the time when r is recruited for the first time in the reconstructed tree

Popularity vs. First-recruitment time



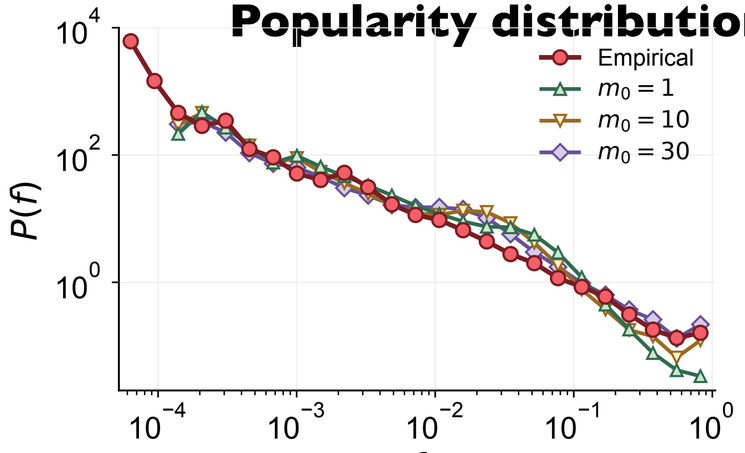
$$\ln f_r \sim -\tau_r \Rightarrow \left|\frac{d\tau_r}{df_r}\right| \sim f^{-1}$$

First-recruitment time distribution



First-recruitment time distribution $P_{\mathrm{rec}}(\tau_r) \sim \mathrm{const}.$

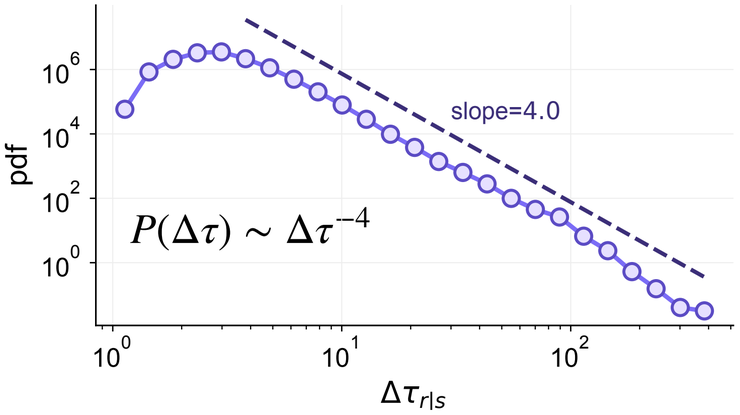
Popularity distribution



$$P_{pop}(f) = P_{\mathrm{rec}}(\tau)\left|\frac{d\tau}{df}\right|_{f=2^{-\tau}} \sim \mathrm{const}.f^{-1} \sim f^{-1}$$

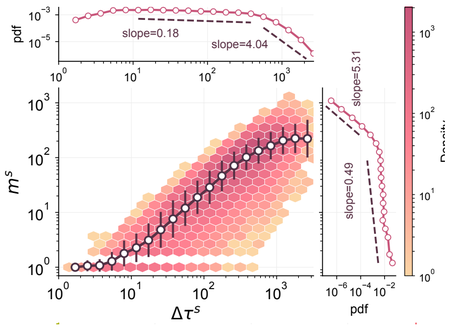
Early inflationary recruitment: data and model

Inter-recruitment time distribution

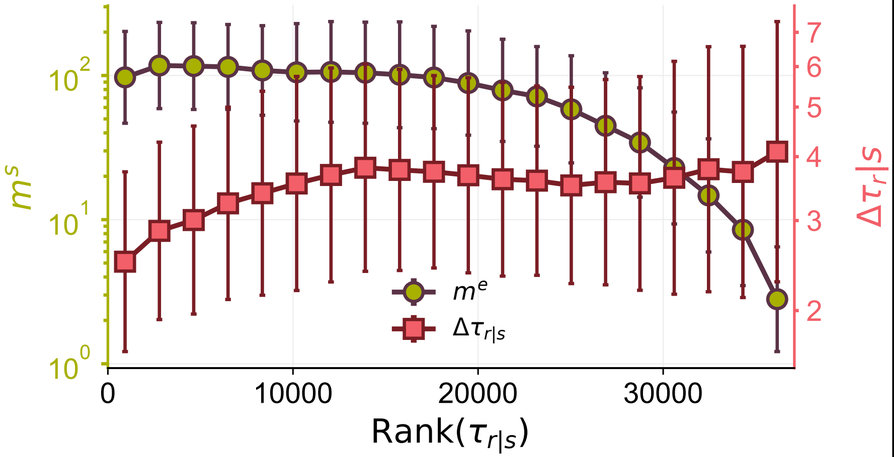


Bursty behavior of reaction recruiting

$$\Delta\tau_{r|s} = \frac{\Delta\tau^s}{m^s}$$



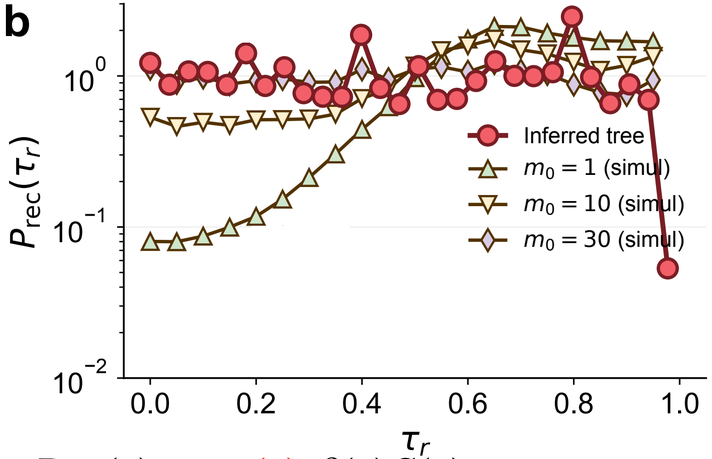
Early fast recruitment



Number of rxns recruited along a branch m^s decreases with time

Inter-recruitment time $\Delta\tau_{r|s}$ increases with time

Model with multiple recruitment



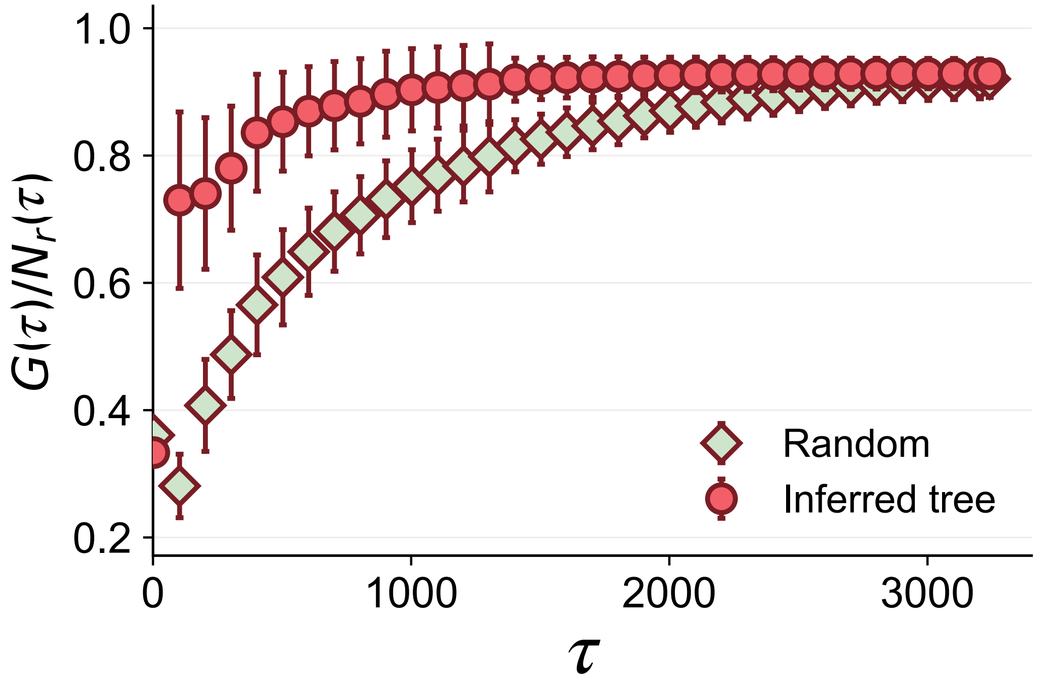
$$P_{\mathrm{rec}}(\tau) \sim m(\tau)\,\beta(\tau)S(\tau)$$

$$m(\tau) \sim \lfloor \frac{m_0}{S(\tau)} + 1 \rfloor$$

$$P_{\mathrm{rec}}(\tau) \sim \mathrm{const}\,.$$

(Early: $m \sim S^{-1}$ late: $\beta \sim S^{-1}$)

A connected metabolic backbone



Inferred ordering
rapidly reaches a giant connected component

Randomized ordering
connects only gradually

Summary

- $P_{\mathrm{pop}}(f) \sim f^{-2}$ due to **inheritance** : A reaction found in many species may be popular not only because it is biochemically important, but also because it was recruited early and inherited by many descendants.
- Crossover from f^{-1} to f^{-2} due to **finiteness** of reaction pool : Recruited reactions are brand new in the early period but not so in the late phase.
- We proposed the method of inferring ancestral reaction sets and reaction recruitment times to reconstruct the metabolism-based phylogenetic tree.
- f^{-1} over the entire range due to **early inflationary expansion**: The rate of reaction recruitment was higher in the early phase than in the late one, resulting in the flat distribution of the first-recruitment time.
- Present-day bacterial metabolic repertoires retain signatures of the tempo and organization of metabolic evolution.