What does the tree of life look like as it grows?

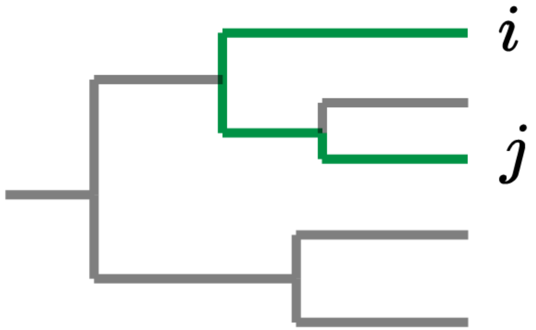
Toward a dynamic theory of phylogenetic space

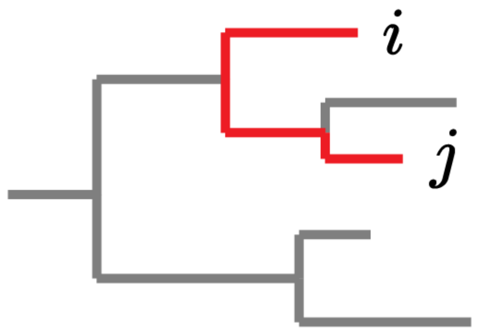
Kevin Hudnall

Biosystems Engineering Graduate Group

Dimensions of phylogenetic space

Branch length ∝ time





Branch length ∝ form

$d : \mathbb{R}^2 \longrightarrow \mathbb{R}$

$d_t(i, j) = \Delta t$ [time]

Chronogram

$d : \mathbb{R}^2 \longrightarrow \mathbb{R}$

$d_\varepsilon(i, j) = \Delta \varepsilon$ [form]

Phylogram

These are **static trees**. We have the quantities Δt and $\Delta\varepsilon$, but not ∂t and $\partial\varepsilon$.

Chronograms and phylograms are **individually inadequate** to model the dynamic case

Requirement of sufficient dimensions:

∂t $\partial \varepsilon$

Chronogram: no $\partial \varepsilon$ Phylogram: no ∂t

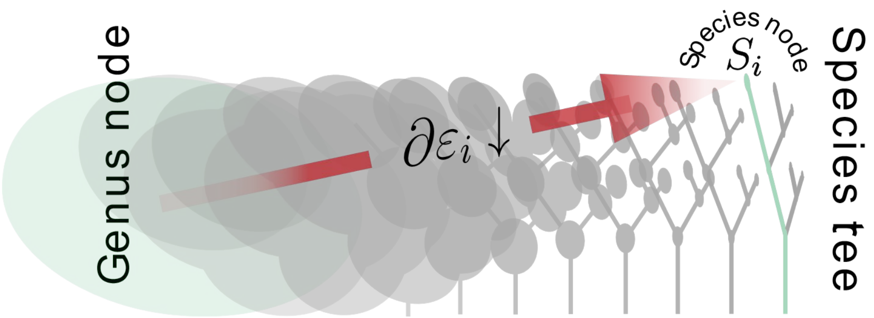
Each has **insufficient dimensions** to represent the dynamic $\frac{\partial \varepsilon}{\partial t}$.

Instead, need **two metrics** $d_t(i,j)$ and $d_\varepsilon(i,j)$ defined on the same space.

That is, to represent the dynamic $\frac{\partial \varepsilon}{\partial t}$, we must **combine chronograms and phylograms**.

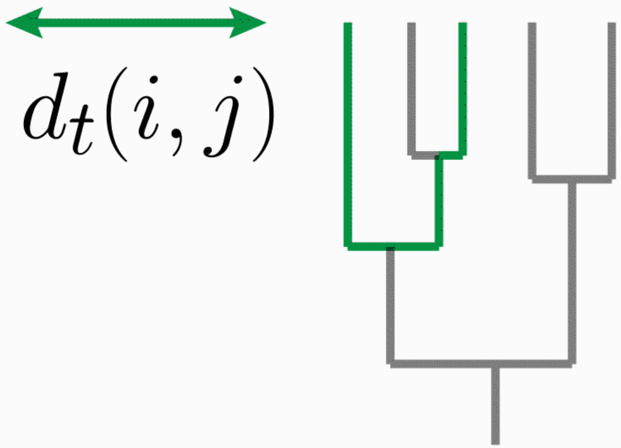
Requirements of a dynamic model

- Sufficient dimensions for the dynamic $\frac{\partial \varepsilon}{\partial t}$
- The random principle
  - ➢ The foundation of modern biology
- The nested principle
  - ➢ The logical structure of descent-with-modification
- The organism-population duality principle
  - ➢ Mantra in biology: “populations, not organisms, evolve”.



A species tree is the **convex hull** of a genus node.

Hint: a tree is not needed if only one metric

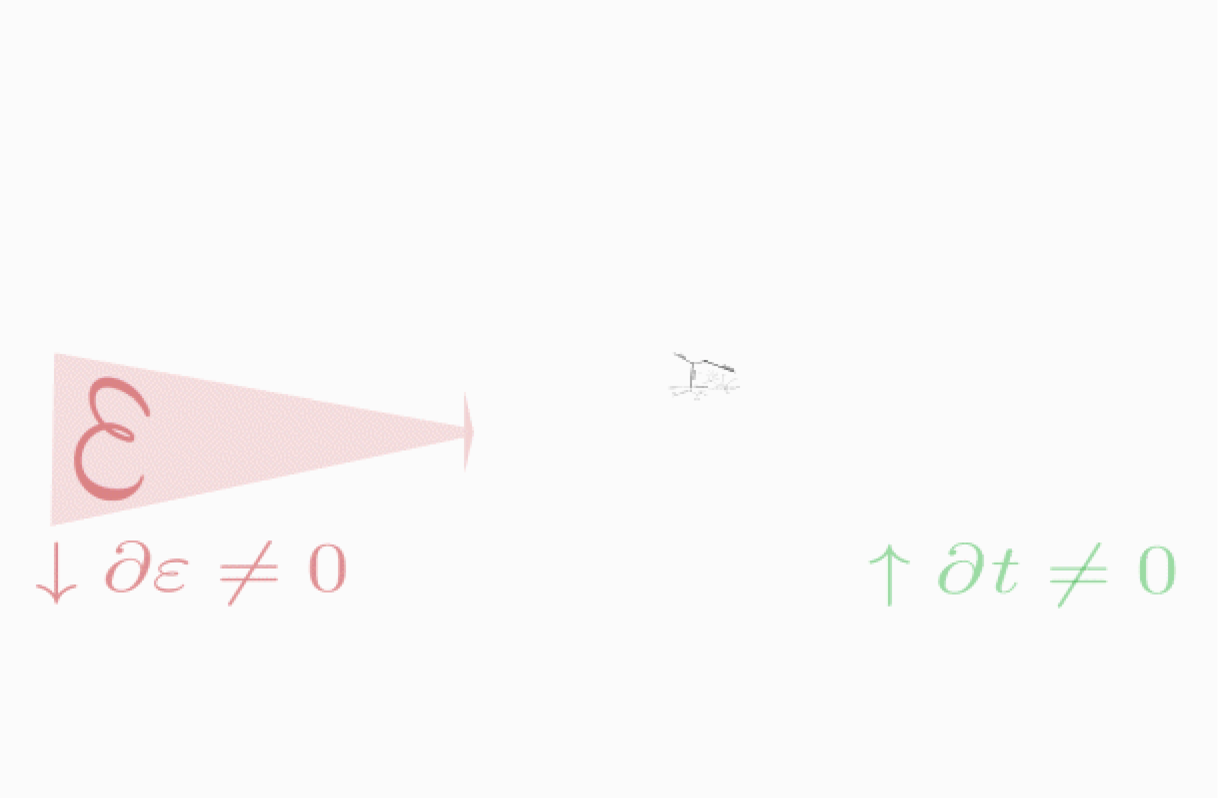


A tree in $\mathbb{R}^2$ with only one metric can be reduced to $\mathbb{R}$ if we utilize **scale**.

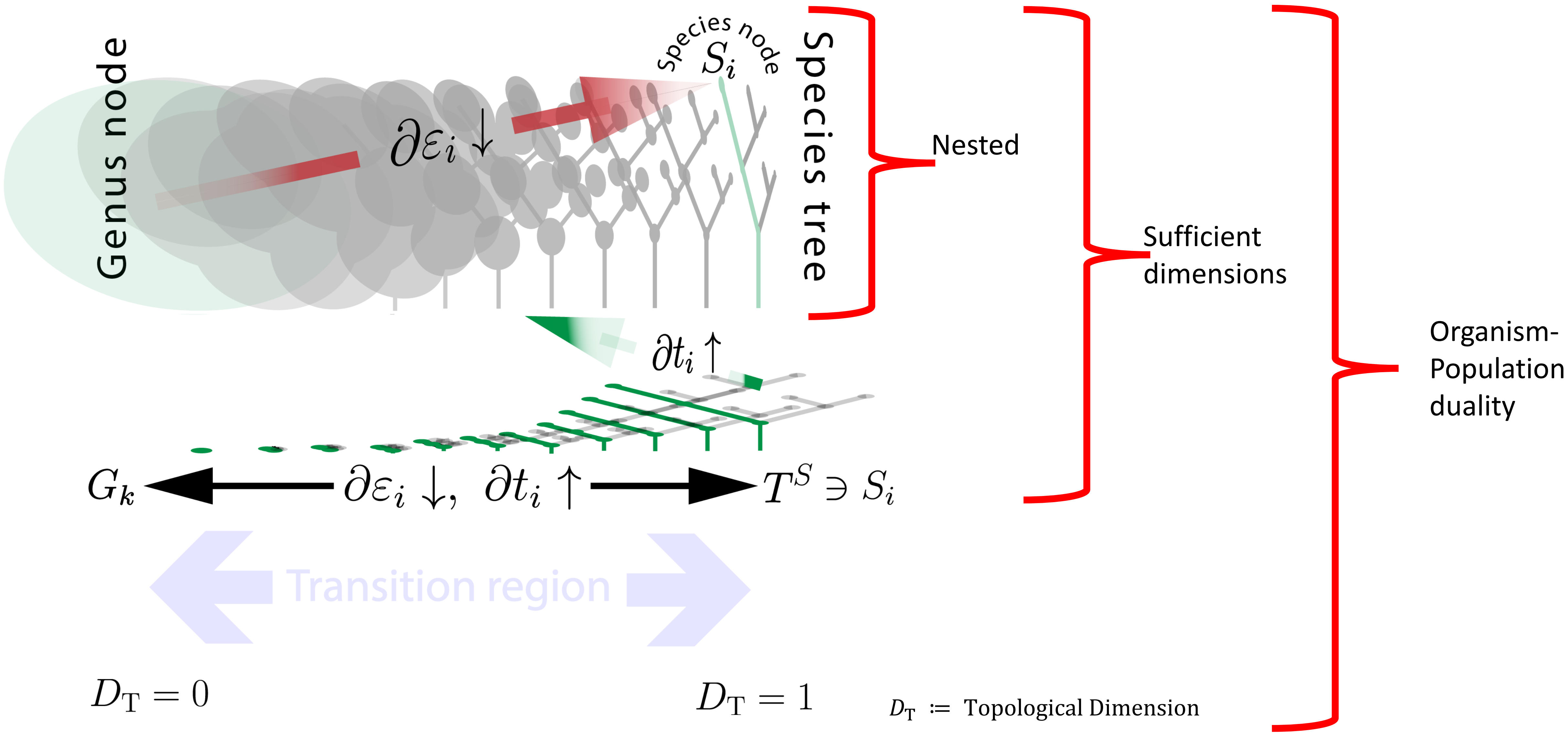
(Think random Cantor dust)

Approach: assign one variable to scale, the other to branch length

Form ε ⟶ scale
Time t ⟶ branch length
Evolution $\partial\varepsilon$ ⟶ scale contraction
Time expansion ∂t ⟶ branch elongation



The dynamic: $\partial\, \varepsilon_i / \partial\, t_i$



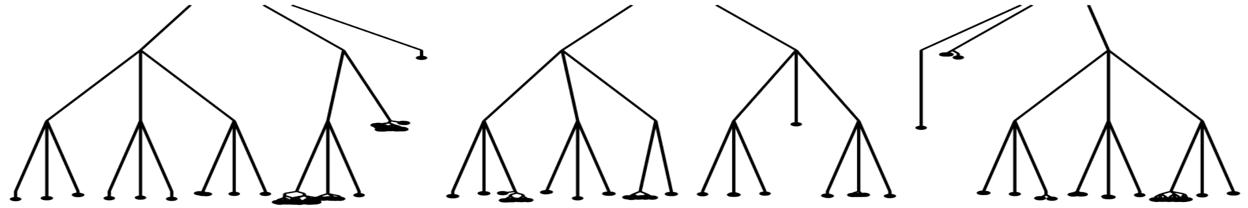
The model: trees within trees within trees…

$\partial \varepsilon_{\ddots i}$: path-specific contraction on the unit square to extant/living leaf

$\partial t_{\ddots i}$: path-specific branch elongation

$$\frac{\partial \varepsilon_{\ddots i}}{\partial t_{\ddots i}}$$

Stochastic Iterated Function System



Auxiliary functions

Let $R\colon [0,1] \times [0,1] \longrightarrow [0,1] \times [0,1]$ such that

$$R(\mathcal{T}_{0_{\cdots i}}) = (a, b), \text{ the coordinates of the root node of tree } \mathcal{T}_{0_{\cdots i}}.$$

Let $L\colon [0,1] \longrightarrow \mathbb{Z}$ such that

$$L(\mathcal{T}_{0_{\cdots i}}) = \text{the number of leaves (terminal nodes) of tree } \mathcal{T}_{0_{\cdots i}}, \text{ and}$$

Let l_{i_j} denote the jth leaf of tree $\mathcal{T}_{0_{\cdots i}}$.

Ratio list:

For tree $\mathcal{T}_{0_{\cdots i}}$ in the $K-1^{\text{th}}$ system iterate:

$$\left(\varepsilon_{0_{\cdots i_1}}, \ \ldots, \varepsilon_{0_{\cdots i_{L\left(\mathcal{T}_{0_{\cdots i}}^{K-1}\right)}}}\right)$$

Assigned to biological form

Where each $\varepsilon_{0_{\cdots i_j}}$ is a random variable with $\varepsilon_{0_{\cdots i_j}} \sim U\left(0, \varepsilon_{0_{\cdots i}}\right)$, and $\varepsilon_1 \sim U(0,1)$.

Function list:

The leaves l_{i_j} of tree $\mathcal{T}_{0_{\cdots i}}$ in the K^{th} system iterate E_K, are comprised of $L\left(\mathcal{T}_{0_{\cdots i}}^{K-1}\right)$ executions of function $T_{0_{\cdots i_j}}^{K}$:

$$\left(T_{0_{\cdots i_1}}^{K}, \ldots, T_{0_{\cdots i_{L\left(\mathcal{T}_{0_{\cdots i}}^{K-1}\right)}}}^{K}\right)$$

Where $i \in \mathbb{Z}$ with range $1 \leq i \leq L(\mathcal{T}^{K-1})$

Here, $T_{0_{\cdots i_j}}^{K}\colon [0,1] \times [0,1] \longrightarrow [0,1] \times [0,1]$ is given by:

$$T_{0_{\cdots i_j}}^{K} = \boldsymbol{Z} * \varepsilon_{0_{\cdots i_j}} + \left\| R\left(T_{0_{\cdots i_j}}^{K}\right) - l_{i_j} \right\|_2, \ l_{i_j} \in \mathcal{T}_{0_{\cdots i}}^{K-1}$$

GW Scale Shift

Where $\boldsymbol{Z}$ is a sequence generated by the Galton-Watson branching process:

$$Z_{p+1} = \begin{cases} \xi_1^{p+1} + \cdots + \xi_{Z_p}^{p+1}, & Z_p > 0 \\ 0, & Z_p = 0 \end{cases}$$

Static approximation tree

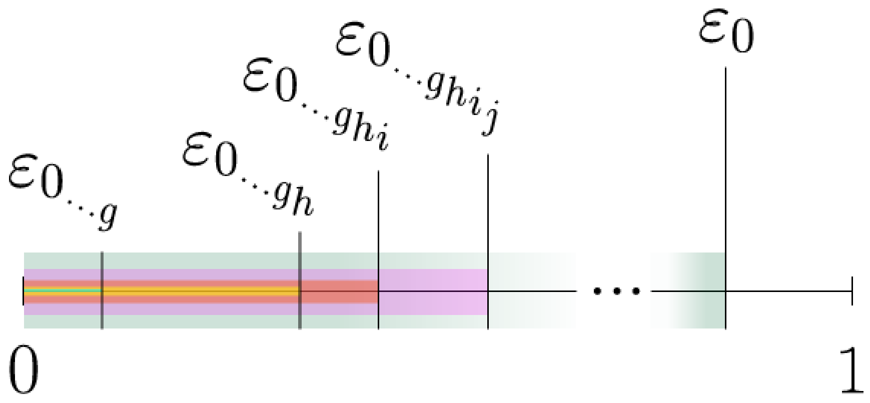
With $\xi_r^p \in \mathbb{Z}$ $i.i.d.$ Nonnegative random variables, and $p, r \geq 0$.

Every scale path $\varepsilon_{\ddots i_j}$ is a Markov chain

Each $\varepsilon_{\ddots i_j}$ is a random variable with $\varepsilon_{\ddots i_j} \sim U\left(0, \varepsilon_{\ddots i}\right)$, and $\varepsilon_1 \sim U(0,1)$. So $\varepsilon_{\ddots i_j} < \varepsilon_{\ddots i}$ for all $\varepsilon_{\#}$

$$\mathrm{PR}\left(\mathcal{E}_{0\ddots h_{i_j}} = \varepsilon_{0\ddots h_{i_j}} \middle| \mathcal{E}_{0\ddots h_i} = \varepsilon_{0\ddots h_i}, \mathcal{E}_{0\ddots h} = \varepsilon_{0\ddots h}, \dots, \mathcal{E}_0 = \varepsilon_0\right) = PR\left(\mathcal{E}_{0\ddots h_{i_j}} = \varepsilon_{0\ddots h_{i_j}} \middle| \mathcal{E}_{0\ddots h_i} = \varepsilon_{0\ddots h_i}\right)$$

Non-stationary.
But very well-behaved.



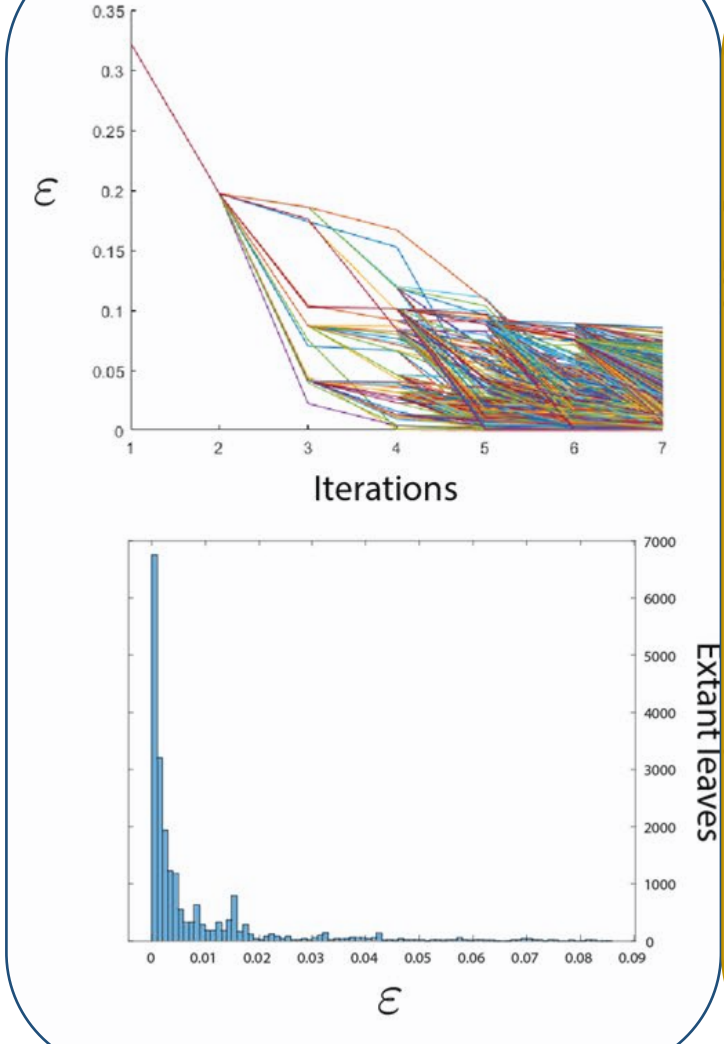
Therefore, the entropy is:

$$\mathrm{H}\left[\varepsilon_{\ddots i_j}\right] = log\left(\varepsilon_{\ddots i}\right) \leq 0$$

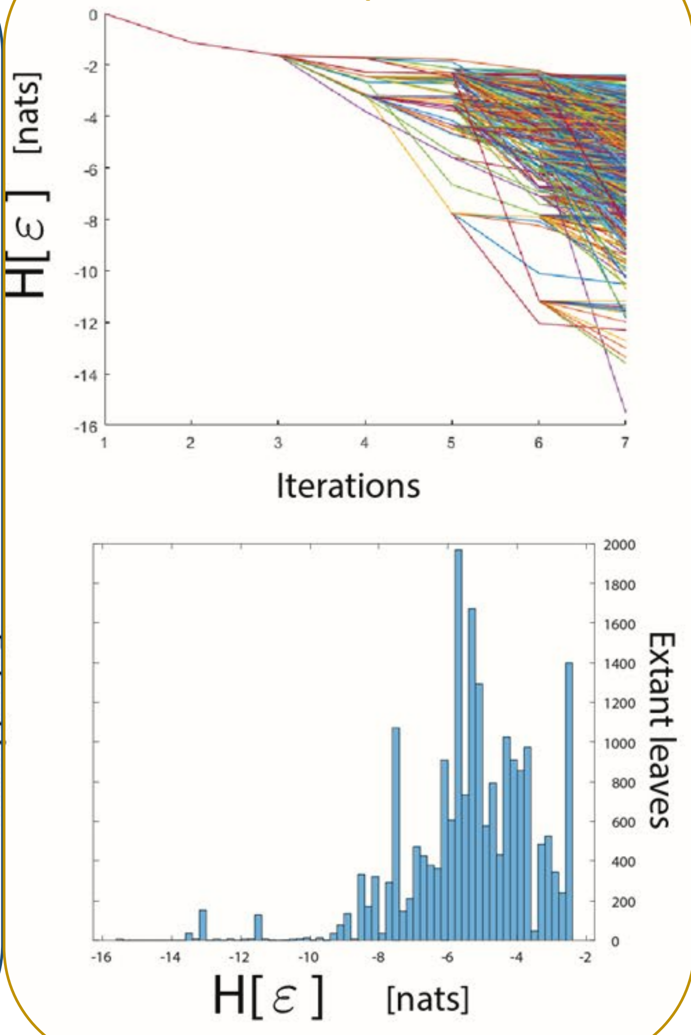
Comparing information measures

Example results: 7 iterations, 20643 extant leaves

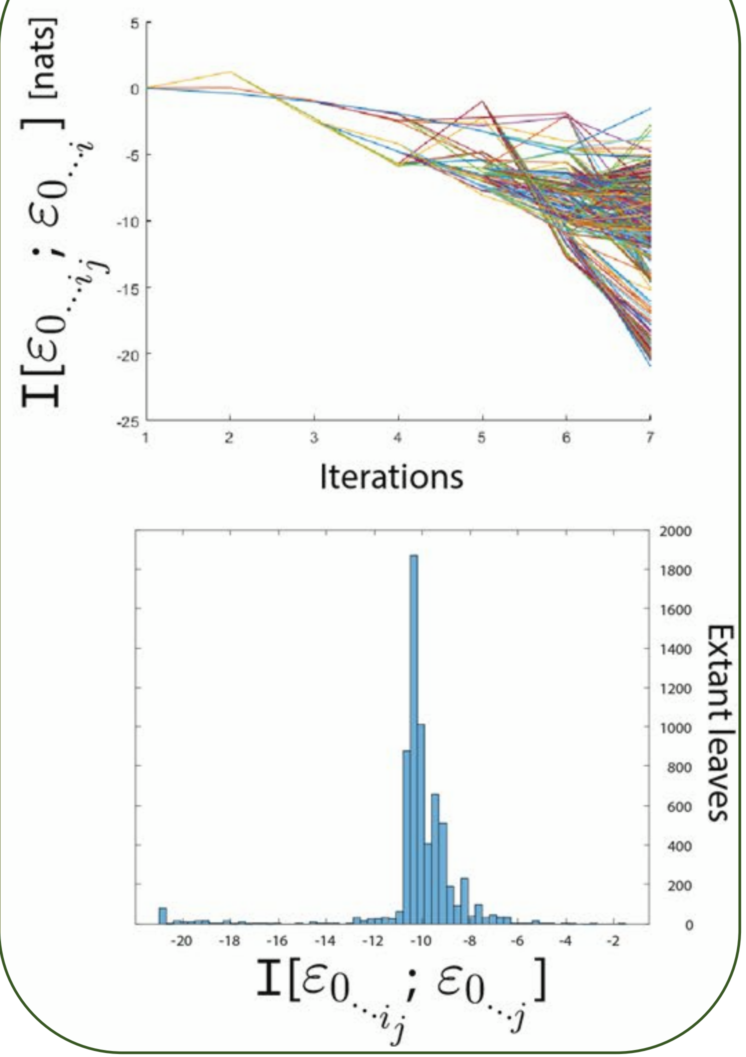
Scale RV



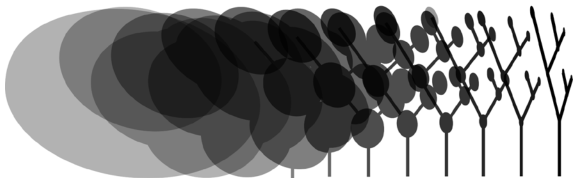
Entropies



Pathwise mutual information



Example results: Local channel capacity densities

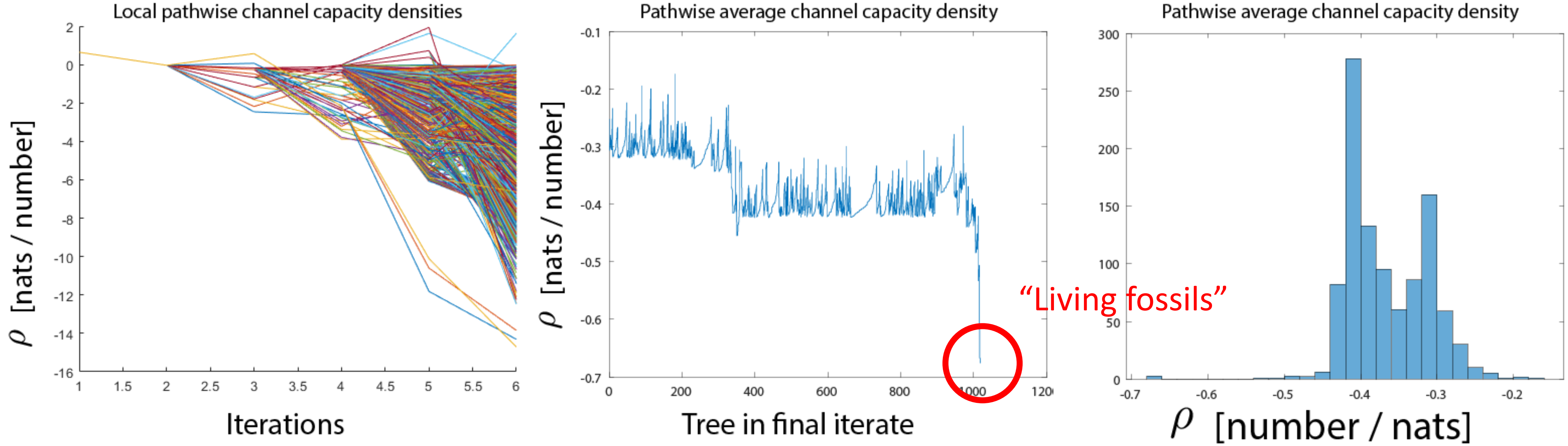


i —— Local channel ——→ j

Channel capacity: $$\mathcal{C}_{ij} = max_{p\left(\varepsilon_{0\cdots i_j}\right)} I\left[\mathcal{E}_{0\cdots i_j};\, \mathcal{E}_{0\cdots i}\right] = H\left[\mathcal{E}_{0\cdots i_j}\right] + H\left[\mathcal{E}_{0\cdots i}\right] - H\left[\mathcal{E}_{0\cdots i_j}, \mathcal{E}_{0\cdots i}\right] = log\left(\mathcal{E}_{0\cdots i_j}\right) + log\left(\mathcal{E}_{0\cdots i}\right) - log\left(\frac{\varepsilon_{\cdots i_j}}{\varepsilon_{\cdots i}}\right)$$

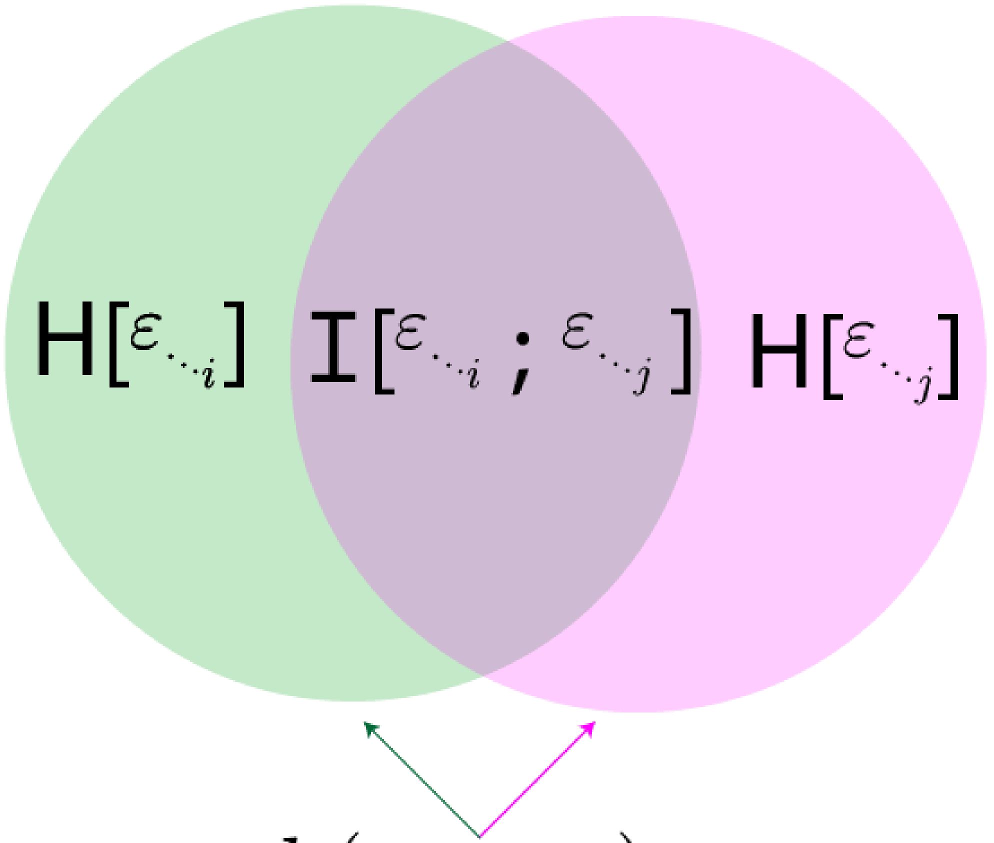
Channel capacity density: $$\rho_{ij} = \frac{\mathcal{C}_{ij}}{L\left(\mathcal{T}_{i_j}\right)}$$

7 iterations, 4,430 extant leaves, 1,018 trees in 7th iterate

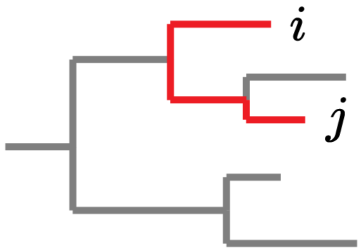


Mutual information and information distance:

How do extant leaves relate?



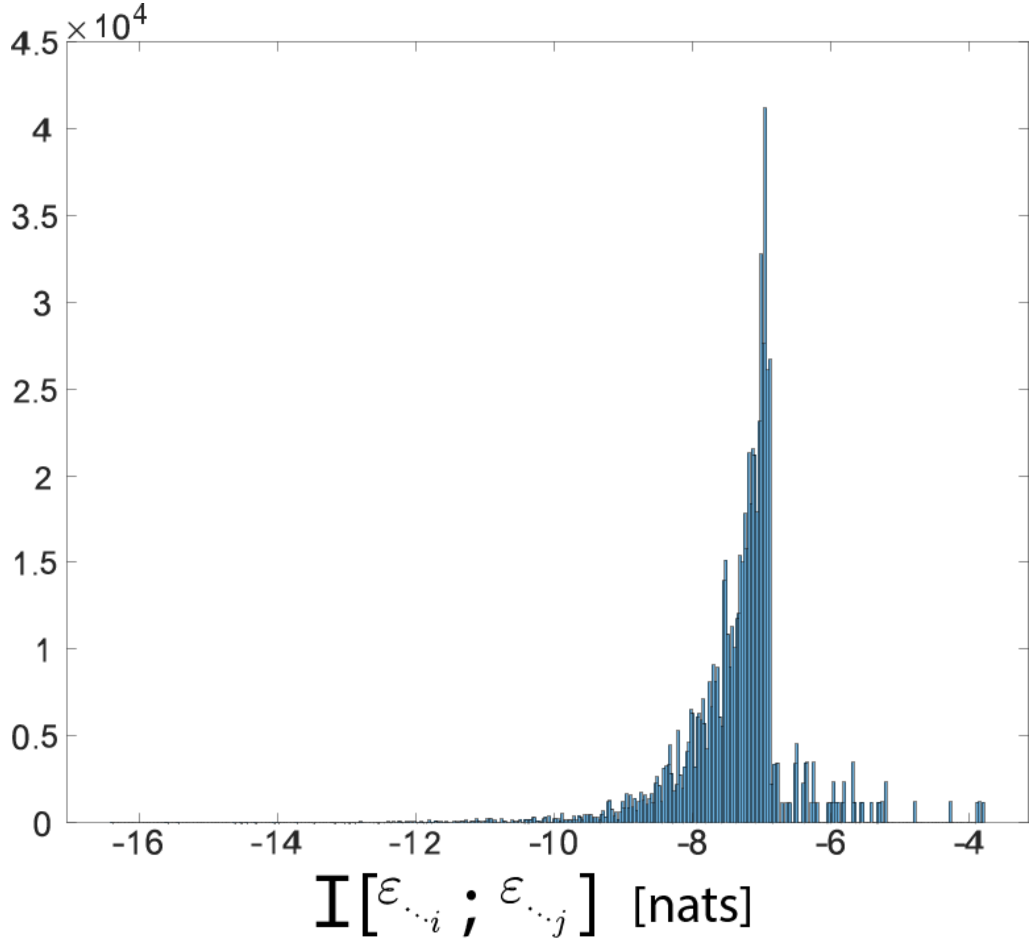
$$d_\varepsilon(\varepsilon_{\cdots i}\ ;\varepsilon_{\cdots j}) \coloneqq \Delta\varepsilon_{ij} \text{ [form]}$$



Assign **information distance** d_ε to be **the metric of biological form**.

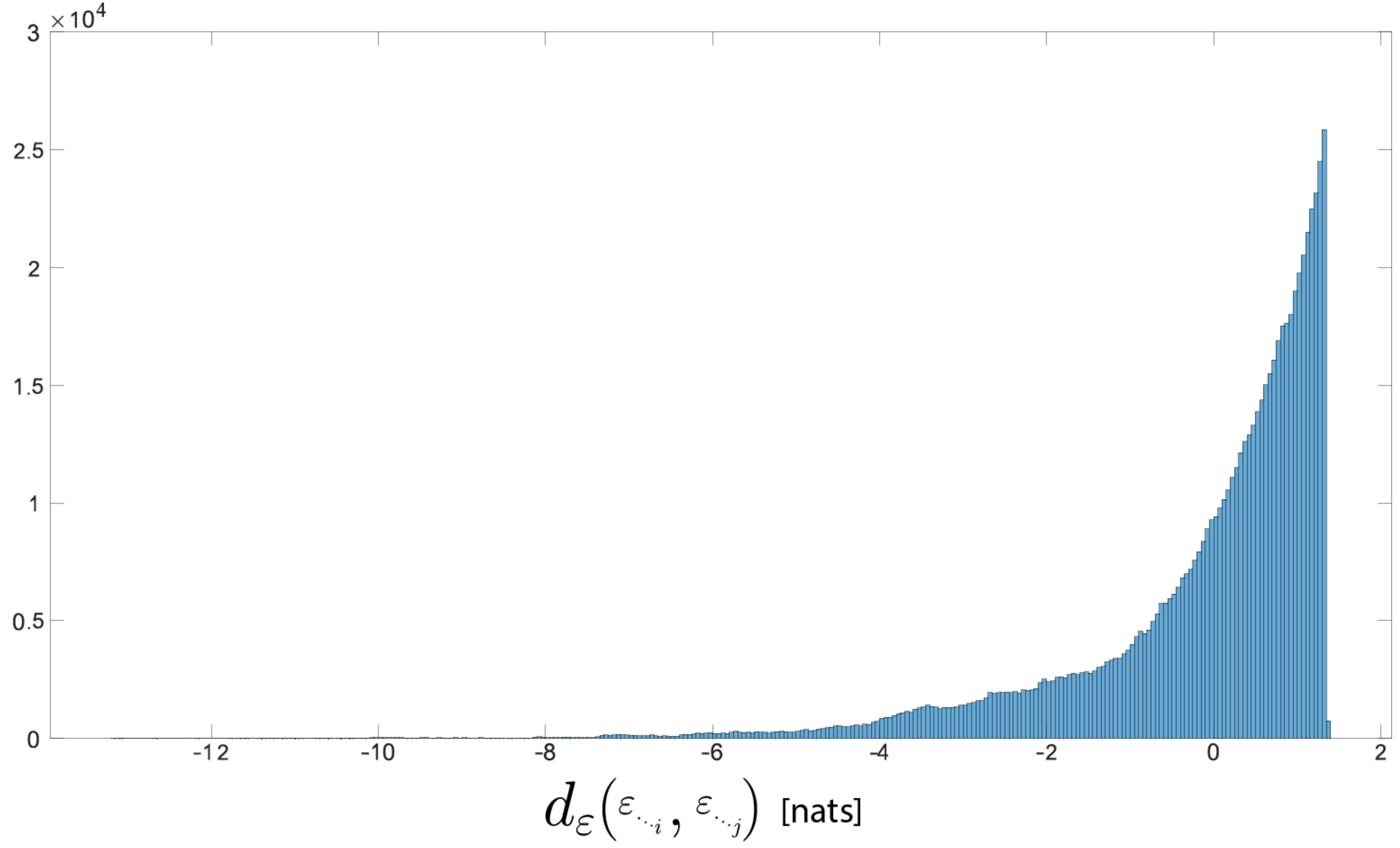
Example results: cross-path mutual information

All unique pairwise combinations of mutual information for 6,616 extant leaves (7 iterations)



Example results: cross-path information distance

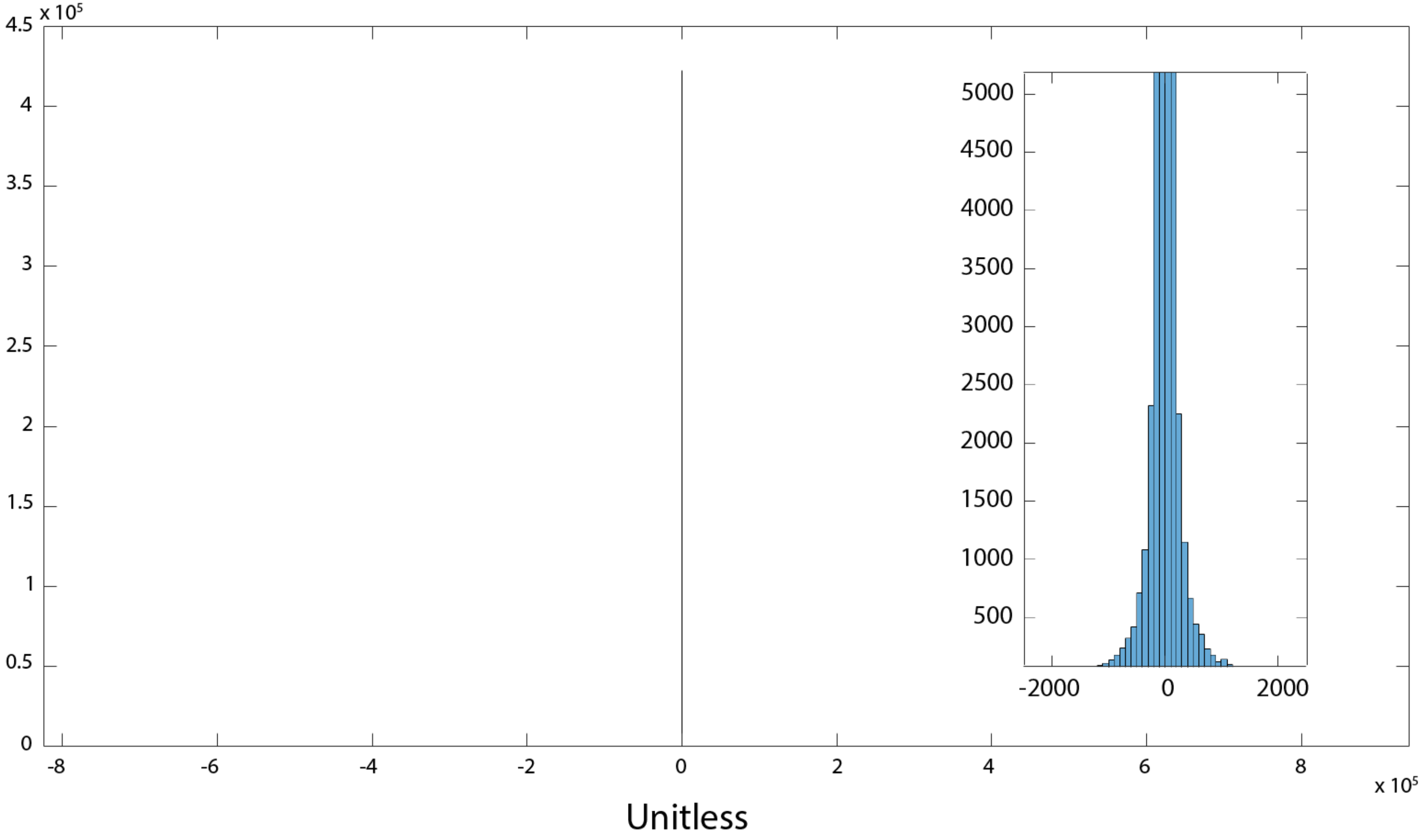
All unique pairwise combinations of information distance for 6,616 extant leaves (7 iterations)



Interested in:

$$\frac{I\left[\varepsilon_{\cdot\cdot\cdot i}, \varepsilon_{\cdot\cdot\cdot j}\right]}{d_\varepsilon\left(\varepsilon_{\cdot\cdot\cdot i}, \varepsilon_{\cdot\cdot\cdot j}\right)}$$

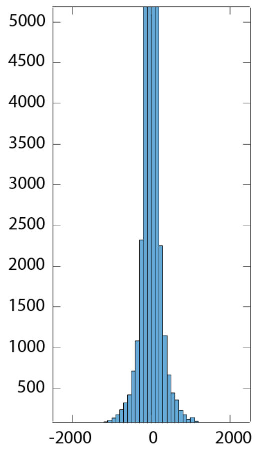
All unique pairwise combinations of $\frac{I}{d_\varepsilon}$ for 6,616 extant leaves (7 iterations)



The metric $d_t(\varepsilon_{\cdot\cdot\cdot i}, \varepsilon_{\cdot\cdot\cdot j})$? For another day

Multifractal structure $\Longrightarrow$ not a simple task

Need $\dfrac{I[\varepsilon_{\cdot\cdot\cdot i}, \varepsilon_{\cdot\cdot\cdot j}]}{d_\varepsilon(\varepsilon_{\cdot\cdot\cdot i}, \varepsilon_{\cdot\cdot\cdot j})}$ and the multifractal spectrum



Thanks


Genus A.b.i
Genus A.a.ii
Genus A.a.i
Family A.a
Family A.b
Species A.b.ii.j