Triads, Dyads, and Gene Functions
When Social Network Analysis meets Phylogenetics

George G Vega Yon, Ph.D.

(with Paul D Thomas Paul Marjoram Huaiyu Mi Duncan Thomas John Morrison)

University of Southern California

Department of Population and Public Health Sciences

Networks 2021

(virtual)

July 10, 2021

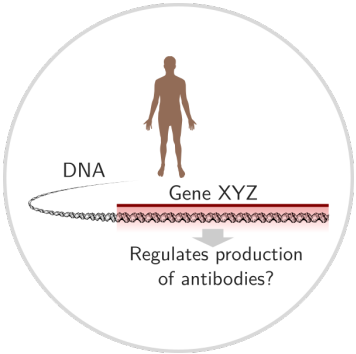
Keck School of Medicine of USC

Contents

You can download the slides from ggv.cl/slides/networks2021

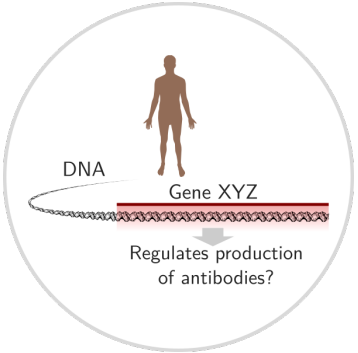
The problem of genes' functions

Is gene *XYZ* involved in process *ABC*?

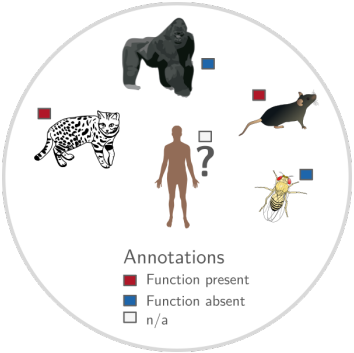


Complex to directly assess

Is gene *XYZ* involved in process *ABC*?

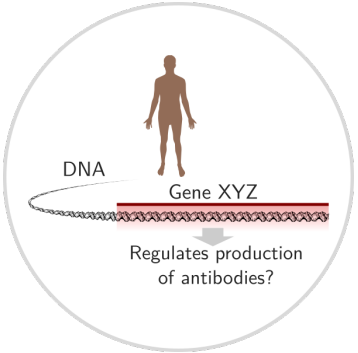


Complex to directly assess

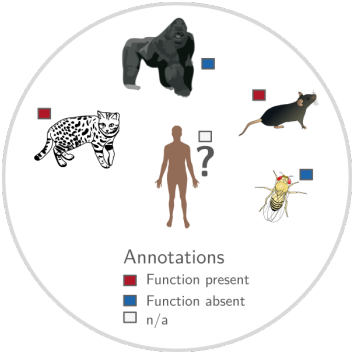


But we may know from other species

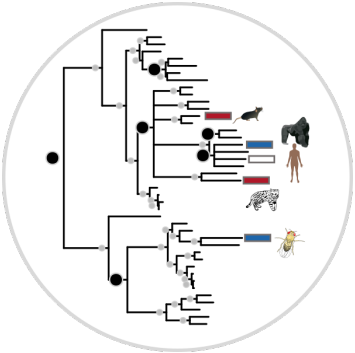
Is gene *XYZ* involved in process *ABC*?



Complex to directly assess

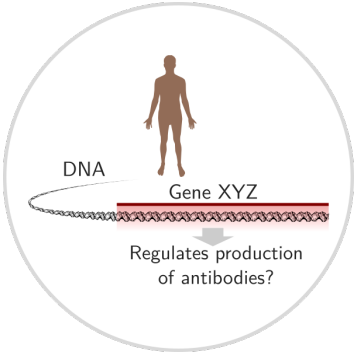


But we may know from other species

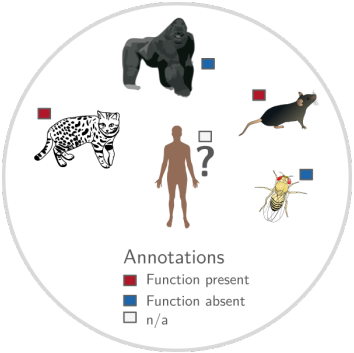


And we further know how these *genetically* connected

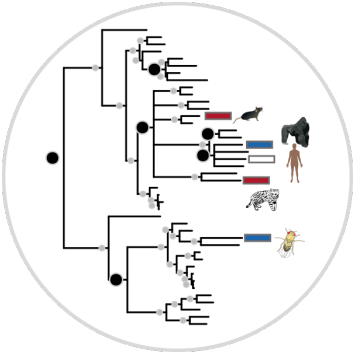
Is gene *XYZ* involved in process *ABC*?



Complex to directly assess



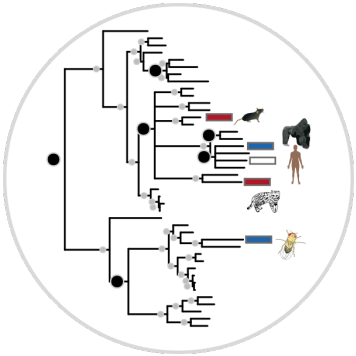
But we may know from other species



And we further know how these *genetically* connected

... let's rephrase the question.

Is the human gene **XYZ** involved in process **ABC**, given what we know about that for other *related* species?

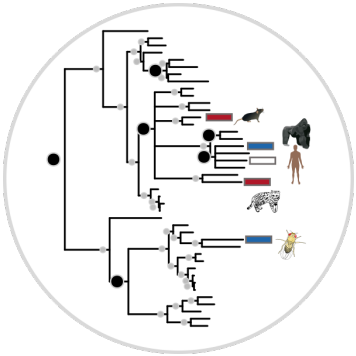


GENEONTOLOGY
Unifying Biology

- $\sim$ 15,000 phylogenetic trees
- $\sim$ 8 million annotations
- $\sim$ 600 thousand on human genes
- $\sim <$ 10% are based on experimental evidence...

Only on 2020, 2,000+ COVID-19 papers using the GO (Google Scholar)

more

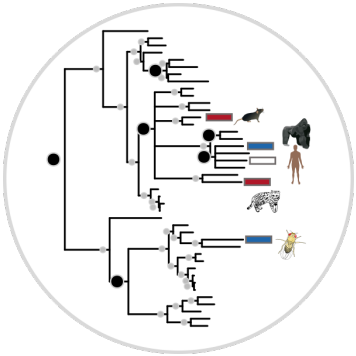


GENEONTOLOGY
Unifying Biology

- $\sim$ 15,000 phylogenetic trees
- $\sim$ 8 million annotations
- $\sim$ 600 thousand on human genes
- $\sim <$ 10% are based on experimental evidence...

Only on 2020, 2,000+ COVID-19 papers using the GO (Google Scholar)

more

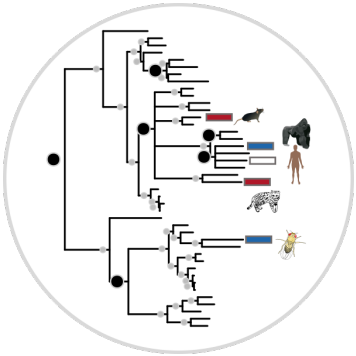


GENEONTOLOGY
Unifying Biology

- $\sim$ 15,000 phylogenetic trees
- $\sim$ 8 million annotations
- $\sim$ 600 thousand on human genes
- $\sim <$ 10% are based on experimental evidence...

Only on 2020, 2,000+ COVID-19 papers using the GO (Google Scholar)

more

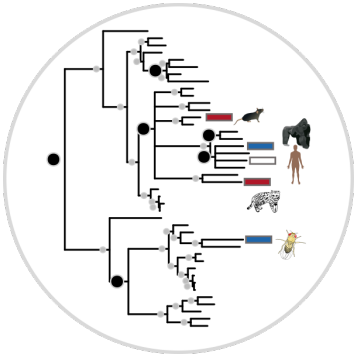


GENEONTOLOGY
Unifying Biology

- $\sim$ 15,000 phylogenetic trees
- $\sim$ 8 million annotations
- $\sim$ 600 thousand on human genes
- $\sim <$ 10% are based on experimental evidence...

Only on 2020, 2,000+ COVID-19 papers using the GO (Google Scholar)

more



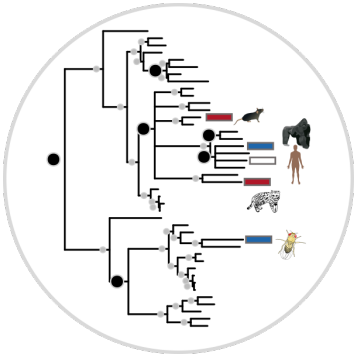
GENEONTOLOGY
Unifying Biology

- $\sim$ 15,000 phylogenetic trees
- $\sim$ 8 million annotations
- $\sim$ 600 thousand on human genes
- $\sim <$ 10% are based on experimental evidence...

Only on 2020, 2,000+ COVID-19 papers using the GO (Google Scholar)

more

Keck School of Medicine of USC



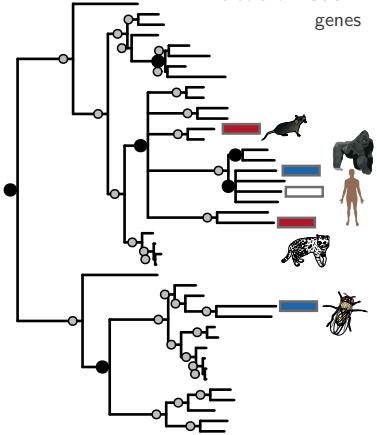
GENEONTOLOGY
Unifying Biology

- $\sim$ 15,000 phylogenetic trees
- $\sim$ 8 million annotations
- $\sim$ 600 thousand on human genes
- $\sim <$ 10% are based on experimental evidence... Improving our knowledge on genetics is fundamental for advancing Biomedical Research

Only on 2020, 2,000+ COVID-19 papers using the GO (Google Scholar)

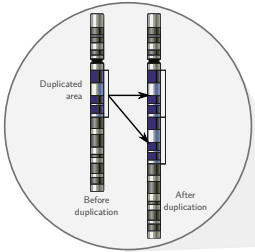
more

The tips of the tree are **modern** genes

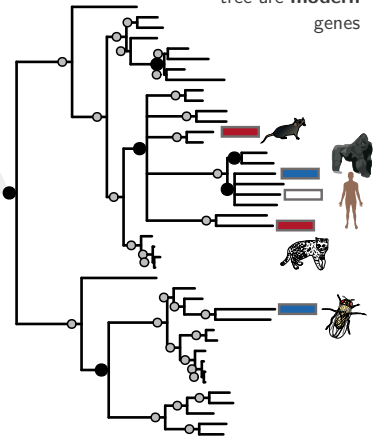


example

● nodes are **Duplication** Events

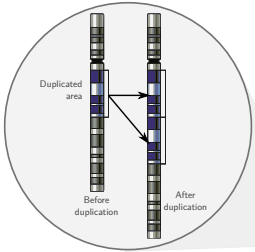


The tips of the tree are **modern** genes

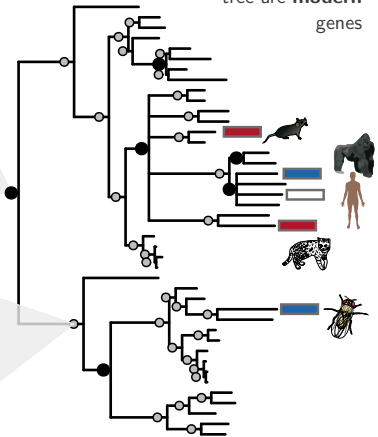


example

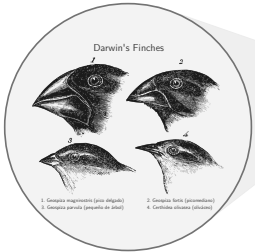
● nodes are **Duplication** Events



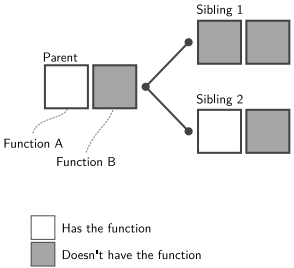
The tips of the tree are **modern** genes

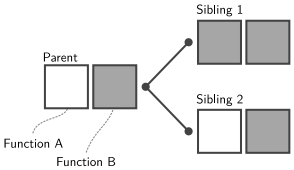


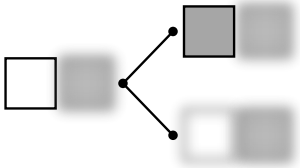
○ nodes are **Speciation** Events



▸ example

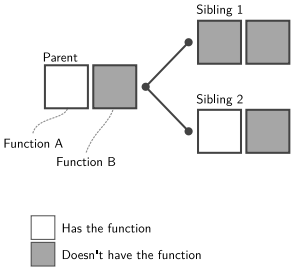

Sibling 1
Parent
Sibling 2
Function A
Function B
Has the function
Doesn't have the function

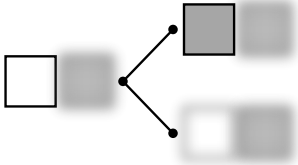




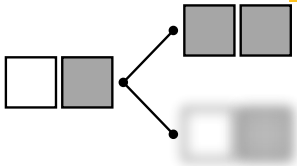
(a) Sibling and Function Conditional Independence

Keck School of Medicine of USC

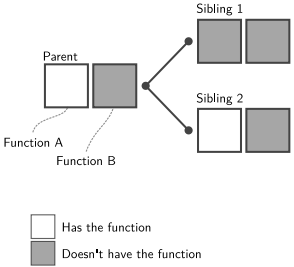


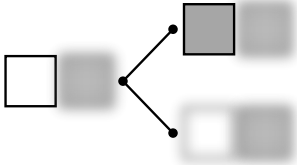


(a) Sibling and Function Conditional Independence

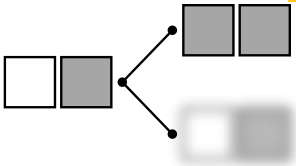


(b) Sibling Conditional Independence

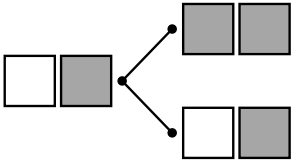




(a) Sibling and Function Conditional Independence



(b) Sibling Conditional Independence

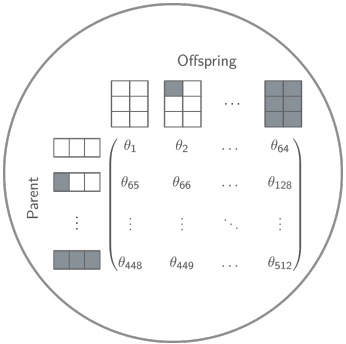


(c) No conditional independence

Avoiding the Curse of dimensionality

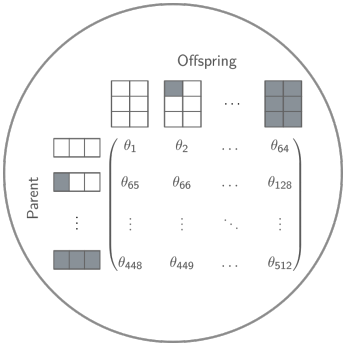
If we wanted to build a model with 3 functions, we would need to estimate...

Full Markov Transition Matrix



If we wanted to build a model with 3 functions, we would need to estimate...

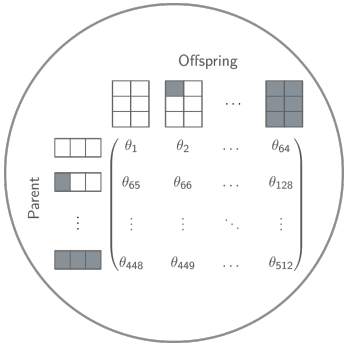
Full Markov Transition Matrix



- 512 parameters

If we wanted to build a model with 3 functions, we would need to estimate...

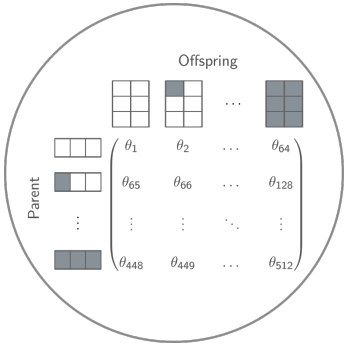
Full Markov Transition Matrix



- 512 parameters
- Finding this many parameters not easy.

If we wanted to build a model with 3 functions, we would need to estimate...

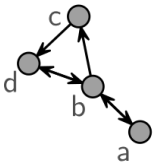
Full Markov Transition Matrix



- 512 parameters
- Finding this many parameters not easy.
- Even if you can, interpretation is awkward.

Keck School of Medicine of USC

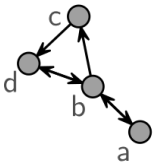
Social Network



	a	b	c	d
a		■		
b	■		■	■
c				■
d		■		

Keck School of Medicine of USC

Social Network

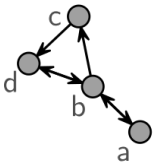


	a	b	c	d
a		■		
b	■		■	■
c				■
d		■		

- Not about individual ties.

Keck School of Medicine of USC

Social Network

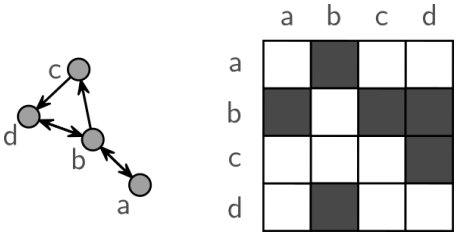


	a	b	c	d
a		■		
b	■		■	■
c				■
d		■		

- Not about individual ties.
- Statistical inference on *motifs* (triangles, dyads, homophily, etc.)

Keck School of Medicine of USC

Social Network



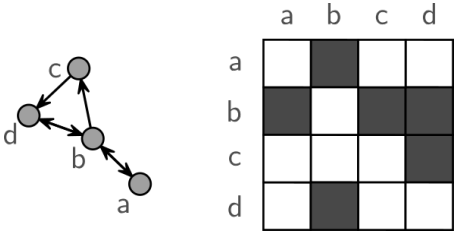
- Not about individual ties.
- Statistical inference on *motifs* (triangles, dyads, homophily, etc.)

Ultimately...

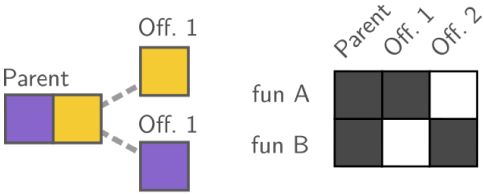
ERGM $\equiv$ Modeling binary arrays

Keck School of Medicine of USC

Social Network



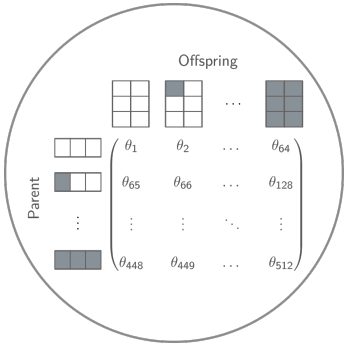
Evolutionary Event



Social Networks are usually represented as **adjacency matrices**, and so can evolutionary events!

If we wanted to build a model with 3 functions, we would need to estimate...

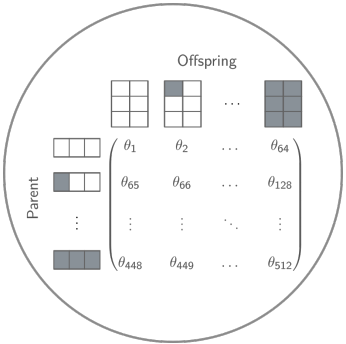
Full Markov Transition Matrix



512 parameters

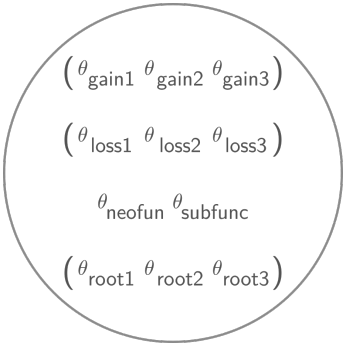
If we wanted to build a model with 3 functions, we would need to estimate...

Full Markov Transition Matrix



512 parameters

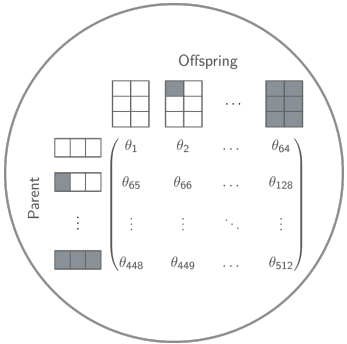
Sufficient statistics



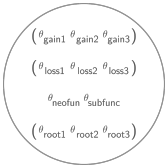
11 parameters (for example)

If we wanted to build a model with 3 functions, we would need to estimate...

Full Markov Transition Matrix



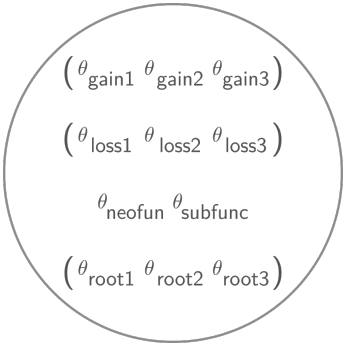
512 parameters



Easier to fit

Easier to interpret

Sufficient statistics



11 parameters (for example)

term examples

Keck School of Medicine of USC

Also known as *dynamic programming* or *postorder tree traversal*

All possible transitions from $\mathbf{x}_n$

Transition Probability (ERGM)

$$\mathbb{P}\left(\tilde{D}_n \,\middle|\, \mathbf{x}_n, \Theta\right) = \sum_{\mathbf{x}} \mathbb{P}\left(\mathbf{x} \mid \mathbf{x}_n\right) \prod_{m \in \mathbf{O}(n)} \mathbb{P}\left(\tilde{D}_m \,\middle|\, \mathbf{x}_m\right)$$

Keck School of Medicine of USC

Also known as *dynamic programming* or *postorder tree traversal*

All possible transitions from $\mathbf{x}_n$

Transition Probability (ERGM)

$$\mathbb{P}\left(\tilde{D}_n \,\middle|\, \mathbf{x}_n, \Theta\right) = \sum_{\mathbf{x}} \mathbb{P}\left(\mathbf{x} \mid \mathbf{x}_n\right) \prod_{m \in \mathbf{O}(n)} \mathbb{P}\left(\tilde{D}_m \,\middle|\, \mathbf{x}_m\right)$$

Model Parameters

Vector of Sufficient Statistics

$$\mathbb{P}\left(\mathbf{x} \mid \mathbf{x}_n\right) = \frac{\exp\left\{\Theta^{\mathsf{t}} s(\mathbf{x}, x_n)\right\}}{\sum_{\mathbf{x}'} \exp\left\{\Theta^{\mathsf{t}} s(\mathbf{x}', x_n)\right\}}$$

Normalizing Constant

the ***lingua franca*** of SNA

Keck School of Medicine of USC

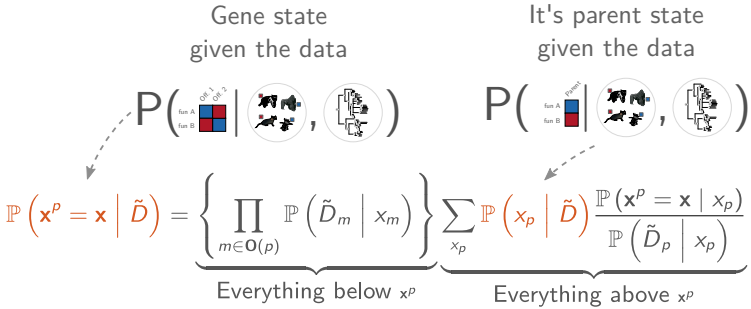
Gene state given the data

It's parent state given the data

Keck School of Medicine of USC

Gene state given the data

It's parent state given the data

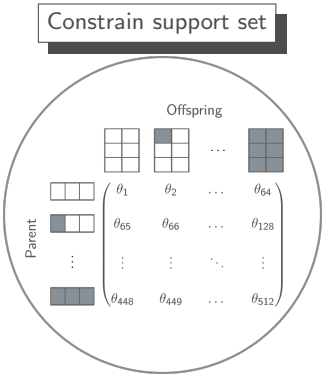


... I implemented this (and more) on **geese**

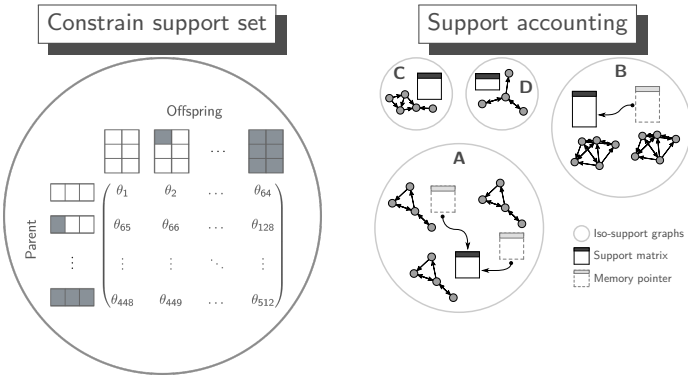
GEne functional **E**volution using **S**ufficiency

... as part of **barry**, your to-go motif accountant

Keck School of Medicine of USC

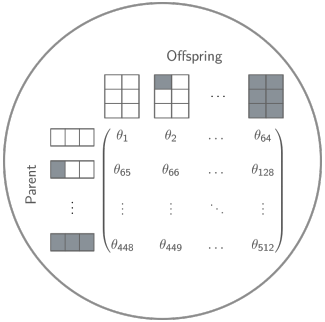


Keck School of Medicine of USC

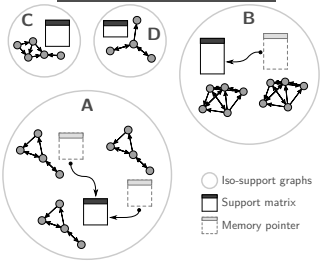


Keck School of Medicine of USC

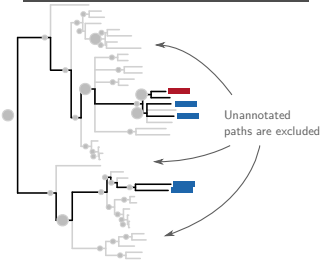
Constrain support set



Support accounting

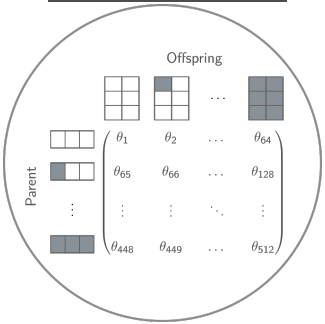


Reduced peeling sequence

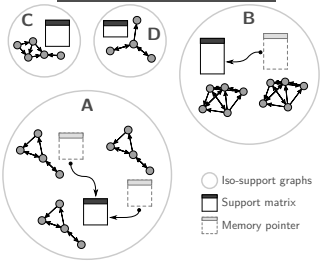


Keck School of Medicine of USC

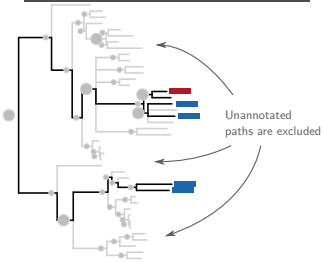
Constrain support set



Support accounting



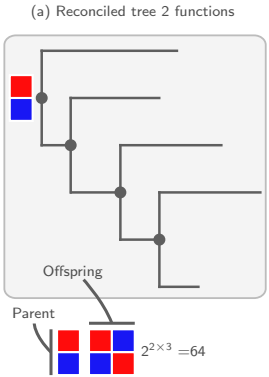
Reduced peeling sequence

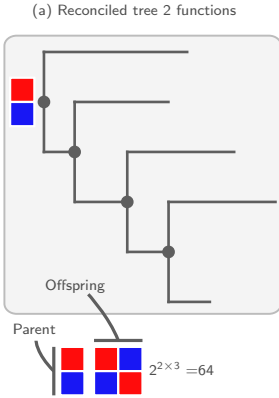


… how big can we go?

Keck School of Medicine of USC

Whether the likelihood is tracktable or not depends on how large is the **"largest event"**





(in practice, arrays up to 32 cells, i.e., 4.3 billion comb., are feasible.)

Analyzing 77 experimentally annotated trees
(Vega Yon, *et al.*, WIP)

Analyzing 77 experimentally annotated trees

Keck School of Medicine of USC

Using data from Vega Yon et al. (2021), we re-estimated the models using **GEESE** and compared the results to those in the **aphylo** paper:

Keck School of Medicine of USC

Using data from Vega Yon et al. (2021), we re-estimated the models using **GEESE** and compared the results to those in the **aphylo** paper:

- 77 experimentally annotated trees

Keck School of Medicine of USC

Using data from Vega Yon et al. (2021), we re-estimated the models using **GEESE** and compared the results to those in the **aphylo** paper:

- 77 experimentally annotated trees
- We only used trees in which

$$2^{(\max \text{Polytomy}+1)\times \text{nfuns}} < 0.5 \times 10^9$$

Keck School of Medicine of USC

Using data from Vega Yon et al. (2021), we re-estimated the models using **GEESE** and compared the results to those in the **aphylo** paper:

- 77 experimentally annotated trees
- We only used trees in which

$$2^{(\max \text{Polytomy}+1)\times \text{nfuns}} < 0.5 \times 10^9$$

  i.e., half billion

Keck School of Medicine of USC

Using data from Vega Yon et al. (2021), we re-estimated the models using **GEESE** and compared the results to those in the **aphylo** paper:

- 77 experimentally annotated trees
- We only used trees in which

$$2^{(\max \text{Polytomy}+1)\times \text{nfuns}} < 0.5 \times 10^9$$

  i.e., half billion

- Both sets of models used "informative" priors.

Keck School of Medicine of USC

Using data from Vega Yon et al. (2021), we re-estimated the models using **GEESE** and compared the results to those in the **aphylo** paper:

- 77 experimentally annotated trees
- We only used trees in which

$$2^{(\max \text{Polytomy}+1)\times \text{nfuns}} < 0.5 \times 10^9$$

  i.e., half billion

- Both sets of models used "informative" priors.
- Both sets of were fitted using Robust Adaptive Metropolis (**fmcmc** R package).

more

Example of code (R)

Keck School of Medicine of USC

After initializing a geese object named `model2fit`:

Example of code (R)

Keck School of Medicine of USC

After initializing a geese object named `model2fit`:

```
# For later use (see last two lines)
term_overall_changes(model2fit, duplication = TRUE)
term_overall_changes(model2fit, duplication = FALSE)

# Couting how many genes change
term_genes_changing(model2fit, duplication = TRUE)

# Gain at duplication
term_gains(model2fit, funs = 0:nfuns, duplication = TRUE)

# Gain and loss at speciation
term_gains(model2fit, funs = 0:nfuns, duplication = FALSE)
term_loss(model2fit, funs = 0:nfuns, duplication = FALSE)

# Constraining the support set
rule_limit_changes(model2fit, id = 0, lb = 0, ub = 4, duplication = TRUE)
rule_limit_changes(model2fit, id = 1, lb = 0, ub = 4, duplication = FALSE)
```

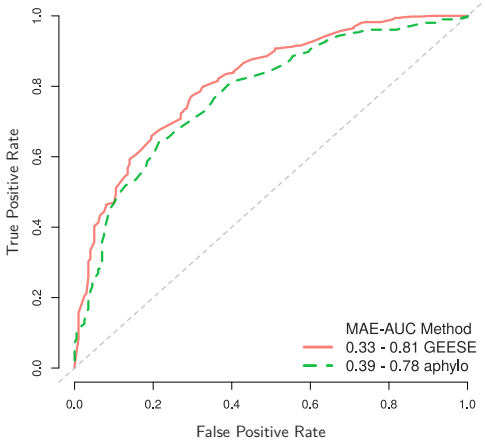


Figure 1 *Receiver Operating Characteristic Curve*. Comparing each set of genes from the 77 trees totaling 709 present/absent annotations in 77 phylogenetic trees. **GEESE** outperforms **aphylo** in both MAE and AUC.

Tapping into computational scalability

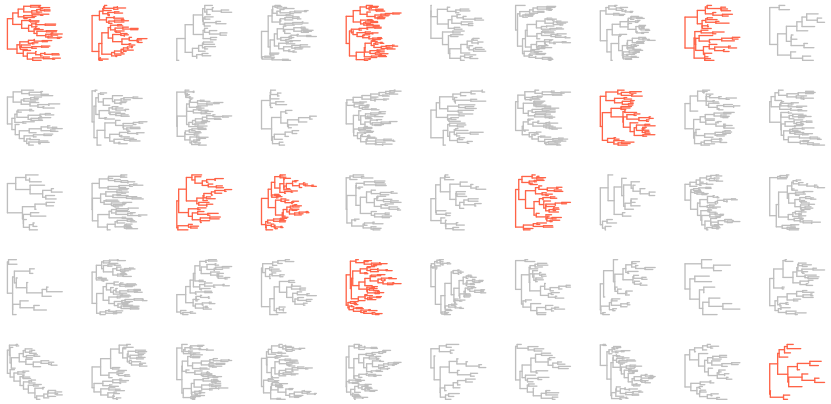


Figure 2 A dramatization of how a group of GEESE, i.e., a flock, looks like.

Keck School of Medicine of USC

Same as before, with the exception of:

Analyzing 70 experimentally annotated trees (Flock)

Keck School of Medicine of USC

Same as before, with the exception of:

- 70 experimentally annotated trees

Analyzing 70 experimentally annotated trees (Flock)

Keck School of Medicine of USC

Same as before, with the exception of:

- 70 experimentally annotated trees
- Reduced the set to trees with a single function

Analyzing 70 experimentally annotated trees (Flock)

Keck School of Medicine of USC

Same as before, with the exception of:

- 70 experimentally annotated trees
- Reduced the set to trees with a single function
- A single pooled-model with $\approx$**34,000 arrays**.

Analyzing 70 experimentally annotated trees (Flock)

Keck School of Medicine of USC

Same as before, with the exception of:

- 70 experimentally annotated trees
- Reduced the set to trees with a single function
- A single pooled-model with $\approx$**34,000 arrays**.
- Largest polytomy 27, meaning about **250,000,000 combinations**.

Analyzing 70 experimentally annotated trees (Flock)

Keck School of Medicine of USC

Same as before, with the exception of:

- 70 experimentally annotated trees
- Reduced the set to trees with a single function
- A single pooled-model with $\approx$**34,000 arrays**.
- Largest polytomy 27, meaning about **250,000,000 combinations**.
- Fitted within 15 minutes

Analyzing 70 experimentally annotated trees (Flock)

Keck School of Medicine of USC

Same as before, with the exception of:

- 70 experimentally annotated trees
- Reduced the set to trees with a single function
- A single pooled-model with ≈**34,000 arrays**.
- Largest polytomy 27, meaning about **250,000,000 combinations**.
- Fitted within 15 minutes

	MCMC est.	
# of genes changing at dupl	-1.85	(0.05)
Gains at dupl.	-1.14	(4.30)
Gains at spec.	-2.50	(0.09)
Loss at spec.	-3.70	(0.07)
Root has the fun.	5.83	(3.78)

Keck School of Medicine of USC

Transition Probabilities

Model Parameters

Vector of Sufficient Statistics

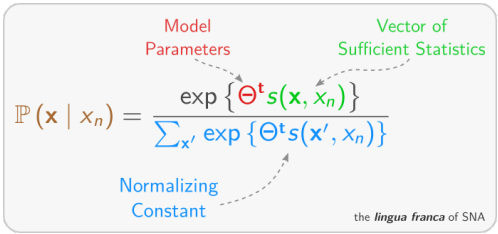
$$\mathbb{P}(\mathbf{x} \mid x_n) = \frac{\exp\left\{\Theta^t s(\mathbf{x}, x_n)\right\}}{\sum_{\mathbf{x}'} \exp\left\{\Theta^t s(\mathbf{x}', x_n)\right\}}$$

Normalizing Constant

the ***lingua franca*** of SNA

Keck School of Medicine of USC

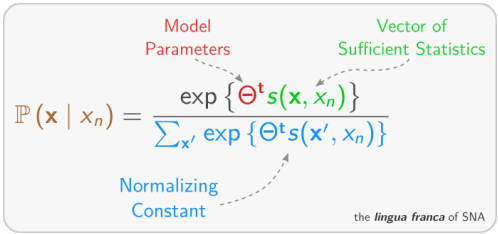
Transition Probabilities



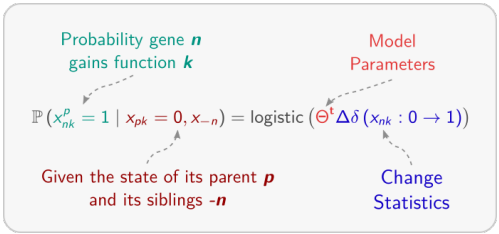
Event type	Duplication	Speciation
Both gain	0.02	0.01
Both lose	0.02	$<$0.01

Keck School of Medicine of USC

Transition Probabilities



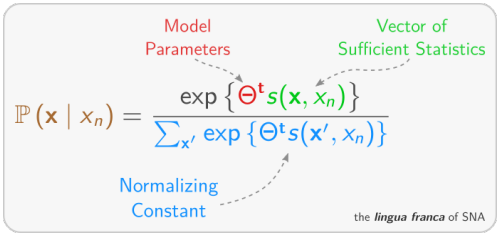
Conditional Probabilities ("Gibbs")



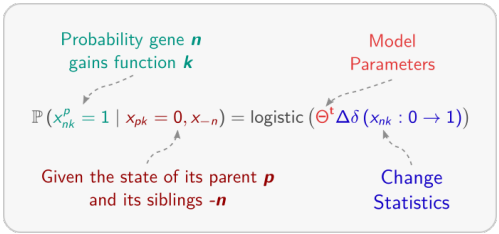
Event type	Duplication	Speciation
Both gain	0.02	0.01
Both lose	0.02	<0.01

Keck School of Medicine of USC

Transition Probabilities



Conditional Probabilities ("Gibbs")



Event type	Duplication	Speciation
Both gain	0.02	0.01
Both lose	0.02	<0.01

Event type	Duplication	Speciation
Preserve 0	0.85	0.90
Preserve 1	0.86	0.97

Take home

Keck School of Medicine of USC

Goal

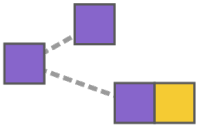
- We are in a race for uncovering **what genes do**.
- **Automatic algorithms** provide a way.

Keck School of Medicine of USC

Goal

- We are in a race for uncovering **what genes do**.
- **Automatic algorithms** provide a way.

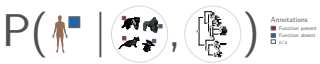
Today



- Many alternatives... many unrealistic **assumptions**.
- **geese** (ERGMs): **GE**ne function **E**volution using **S**ufficiency.

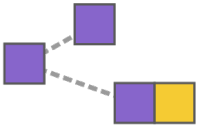
Keck School of Medicine of USC

Goal



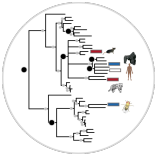
- We are in a race for uncovering **what genes do**.
- **Automatic algorithms** provide a way.

Today



- Many alternatives... many unrealistic **assumptions**.
- **geese** (ERGMs): **GE**ne function **E**volution using **S**ufficiI**E**ncy.

Next steps



- Further study its properties (bias, power, accuracy).
- Find applications for this ~~model~~ **modeling framework**.

Triads, Dyads, and Gene Functions
When Social Network Analysis meets Phylogenetics

George G Vega Yon

https://ggvy.cl

vegayon@usc.edu

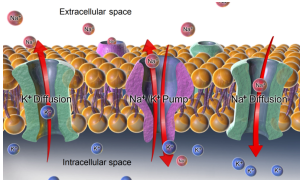
Thank you!

Keck School of Medicine of USC

Gene functions can be classified in three types:

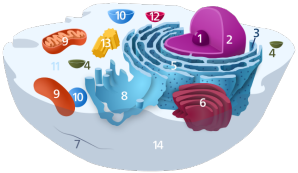
Molecular function

Active transport GO:0005215



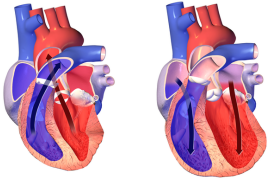
Cellular component

Mitochondria GO:0004016



Biological process

Heart contraction GO:0060047



go back

Keck School of Medicine of USC

Example of GO term

Accession	GO:0060047
Name	heart contraction
Ontology	biological_process
Synonyms	heart beating, cardiac contraction, hemolymph circulation
Alternate	IDs None
Definition	The multicellular organismal process in which the heart decreases in volume in a characteristic way to propel blood through the body. Source: GOC:dph

Table 1 Heart Contraction Function. source: amigo.geneontology.org

go back

Example: Molecular function in family PTHR1128

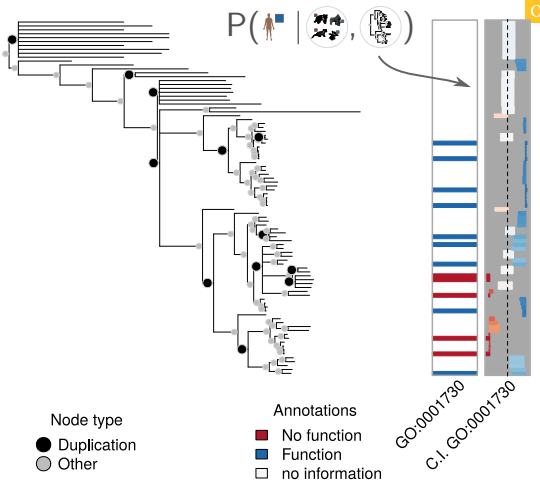
Keck School of Medicine of USC

Name: 2'-5'-oligoadenylate synthetase activity

Desc: GO:0001730 involved in the process of cellular antiviral activity (wiki on interferon).

MAE: 0.34

AUC: 0.91



Note: Prediction made using **aphylo** (Vega Yon and *et al*, PLOS Comp. Bio 2021)

go back

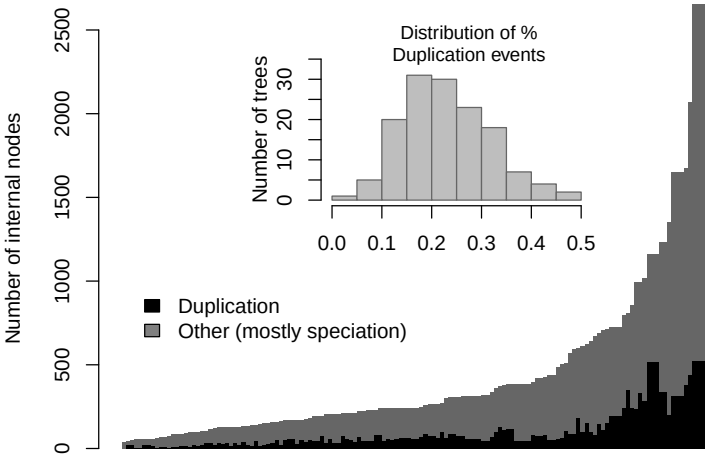
Keck School of Medicine of USC

Sample of annotations (first 10 in a single tree, Phosphoserine Phosphatase [PTHR10000])

Internal id	Branch Length	type	ancestor
AN0		S	LUCA
AN1	0.06	S	Archaea-Eukaryota
AN2	0.24	S	Eukaryota
AN3	0.44	S	Unikonts
AN4	0.42	S	Opisthokonts
AN6	0.68	D	
AN9	0.79	S	Amoebozoa
AN10	0.18	D	
AN15	0.57	S	Dictyostelium
AN18	0.52	S	Alveolata-Stramenopiles

go back

Data: Node type (events)

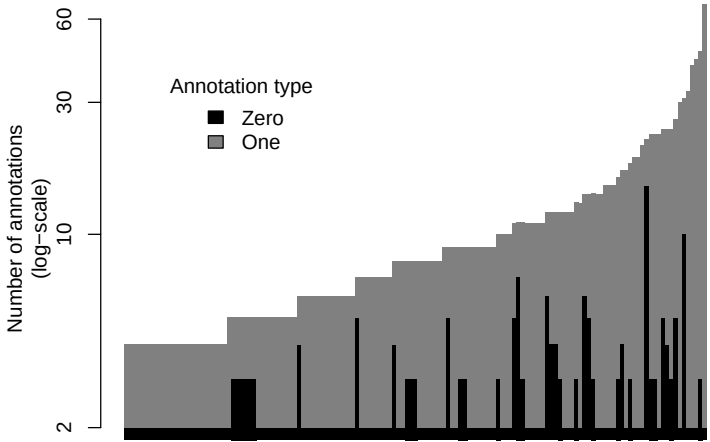


Keck School of Medicine of USC

This is the first 10 of $\sim$ 400,000 experimental annotations used:

	Family	Id	GO term	Qualifier
1	PTHR12345	HUMAN|HGNC=15756|UniProtKB=Q9H190	GO:0005546	
2	PTHR11361	HUMAN|HGNC=7325|UniProtKB=P43246	GO:0016887	CONTRIBUTES_TO
3	PTHR10782	MOUSE|MGI=MGI=3040693|UniProtKB=Q6P1E1	GO:0045582	
4	PTHR23086	ARATH|TAIR=AT3G09920|UniProtKB=Q8L850	GO:0006520	
5	PTHR32061	RAT|RGD=619819|UniProtKB=Q9EPI6	GO:0043197	
6	PTHR46870	ARATH|TAIR=AT3G46870|UniProtKB=Q9STF9	GO:1990825	
7	PTHR15204	MOUSE|MGI=MGI=1919439|UniProtKB=Q9Z1R2	GO:0045861	
8	PTHR22928	DROME|FlyBase=FBgn0050085|UniProtKB=Q9XZ34	GO:0030174	
9	PTHR35972	HUMAN|HGNC=34401|UniProtKB=A2RU48	GO:0005515	
10	PTHR10133	DROME|FlyBase=FBgn0002905|UniProtKB=O18475	GO:0097681	

go back


Number of annotations
(log−scale)
2
10
30
60
Annotation type
Zero
One

go back

Overview of Prediction Results: aphylo

Keck School of Medicine of USC

	Pooled	Type of Annotation: Molecular Function	Type of Annotation: Biological Process	Type of Annotation: Cellular Comp.
Mislabeling				
ψ_{01}	0.23	0.18	0.09	
ψ_{10}	0.01	0.01	0.01	
Duplication Events				
μ_{d01}	0.97	0.97	0.10	
μ_{d10}	0.52	0.51	0.03	
Speciation Events				?
μ_{s01}	0.05	0.05	0.05	
μ_{s10}	0.01	0.01	0.02	
Root node				
π	0.79	0.71	0.88	
Trees	141	74	45	22
Accuracy under the by-aspect model				
AUC	-	0.77	0.83	
MAE	-	0.34	0.26	
Accuracy under the pooled-data model				
AUC	-	0.77	0.75	
MAE	-	0.35	0.34	

Previously, joint estimates out-performed one-at-a-time

- **Molecular Function** No change.
- **Biological Process** Significantly better.
- **Cellular Component** Does not converge.

Molecular Function $\neq$ Biological Process ? Cellular Component

data

go back

Keck School of Medicine USC

- Sparse matrix represented using double hashmaps (fast row/column access).
- Template implementation for flexible weights and metadata.
- Fast counting using change statistics (Ch. 4).
- Calculation of support for sufficient stats.

`https://USCbiostats.github.io/binaryarrays`

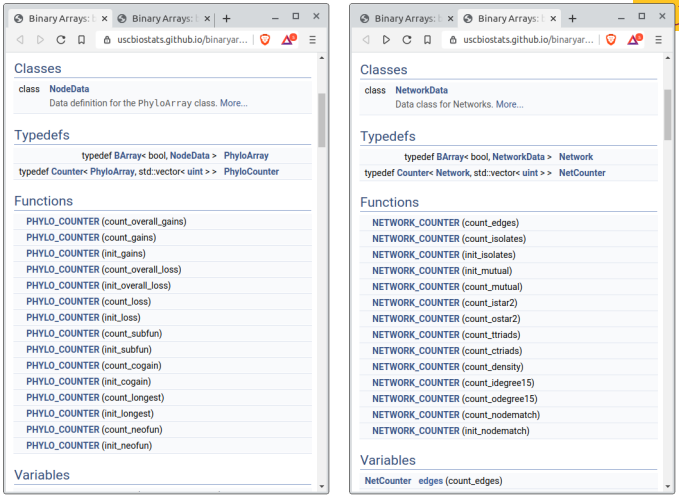


Figure 3 Screenshots from the project's website on GitHub.

Keck School of Medicine of USC

Imagine that we have 3 functions (rows) and that each node has 2 siblings (columns)

			Transitions to	
			Case 1	Case 2
Parent	A	$\begin{bmatrix} 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
	B	$\begin{bmatrix} 1 \end{bmatrix}$		
	C	$\begin{bmatrix} 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$
Sufficient statistics				
# Gains			1	1
Only one offspring changes (yes/no)			1	0
# Changes (gain+loss)			2	3
Subfunctionalizations (yes/no)			0	1

return

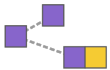
Representation	Description	Definition
	Gain of function	$(1 - x_p) \sum_{n:n \in Off} x_n$
	Loss of function	$x_p \sum_{n:n \in Off} (1 - x_n)$
	Subfunctionalization	$x_p^k x_p^j \sum_{n \neq m} x_n^k (1 - x_n^j)(1 - x_m^k) x_m^j$
	Neofunctionalization	$x_p^k (1 - x_p^j) \sum_{n \neq m} x_n^k (1 - x_n^j)(1 - x_m^k) x_m^j$
	Longest branch gains	$(1 - x_p^k)\mathbf{1}\left(x_m^k : m = \text{argmax}_n \text{blength}_n\right)$

Table 2 Example of sufficient statistics for evolutionary transitions.

go back

Example: Simple model with two functions

Keck School of Medicine of USC

To illustrate, we will **simulate** and then **estimate** the parameters for the following process:

1. 100 genes on a simulated phylogenetic tree.
2. Two functions, **0** and **1**,
3. **Function 0** is likely to be gain **gained at a dupl. event**,
4. **Function 1** is gained as **neofunctionalization** (**from 0**) **at a dupl. event**,
5. There is a higher chance of **changes at duplication** (explicit).
6. Root node starts off without either function (i.e. prob $\to$ 0).

We will fit the model using Robust Adaptive Metropolis with a logistic prior centered at 0 with scale 2.

go back

Keck School of Medicine of USC

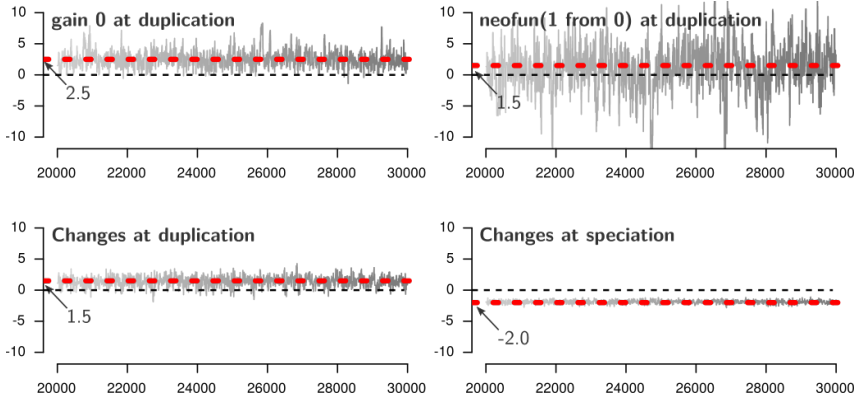


Figure 4 MCMC Trace of the functional gain of 0, neofunctionalization (1 from 0), and change rate (by event type).

Keck School of Medicine of USC

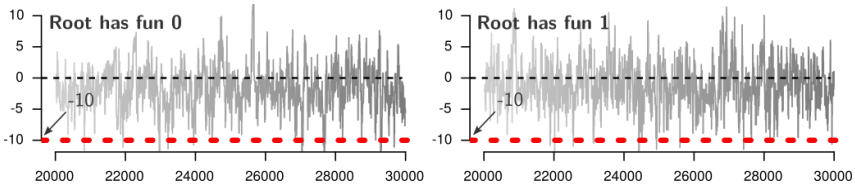


Figure 5 MCMC Trace of root parameters. The true population parameters are $(\theta_{root0}, \theta_{root1}) = (-10.0, -10.0)$. Root node probabilities are always hard to get.

Keck School of Medicine of USC

Repeated this experiment 5,000 times:

- MCMC for fitting.
- RAM kernel.
- Logistic prior at zero with scale two.
- Each tree took $<$ 1min estimation.

Figure 6 Distribution of parameter estimates from 5,000 phylo trees w/ 100 leafs.