Anaylsis of the Subfunctionalization Model for the Fate of Gene Duplicates

Tristan L. Stark [1]
Malgorzata O'Reilly [1]
Barbara Holland [1]
David Liberles [2]

[1]University of Tasmania

[2]Temple University, Philadelphia

November 18, 2015

Subfunctionaliztion

We consider the evolution of a pair of gene duplicates, each with z regulatory regions and a coding region.

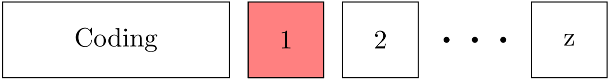
Subfunctionaliztion

We consider the evolution of a pair of gene duplicates, each with z regulatory regions and a coding region.
Under the subfunctionalization model a null mutation can fix

Subfunctionaliztion

We consider the evolution of a pair of gene duplicates, each with z regulatory regions and a coding region.
Under the subfunctionalization model a null mutation can fix

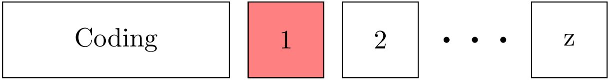
- in any of the z regulatory regions of either copy. We assume this occurs at equal Poisson rate u_r for all $2z$ regions.



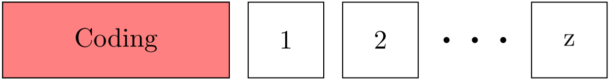
Subfunctionaliztion

We consider the evolution of a pair of gene duplicates, each with z regulatory regions and a coding region.
Under the subfunctionalization model a null mutation can fix

- in any of the z regulatory regions of either copy. We assume this occurs at equal Poisson rate u_r for all $2z$ regions.



- in the coding region of either gene. We assume this occurs at Poisson rate u_c for each gene.



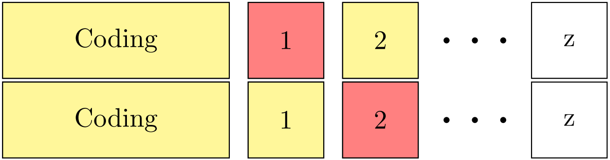
Fates of Duplicates

As mutations build up in the two copies, one of two possible fates eventually occurs

Fates of Duplicates

As mutations build up in the two copies, one of two possible fates eventually occurs

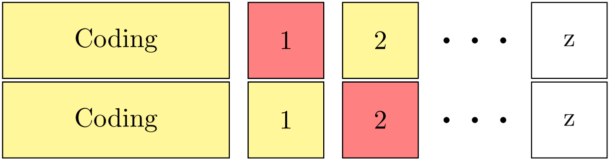
- Subfunctionalization



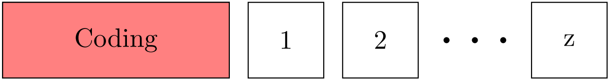
Fates of Duplicates

As mutations build up in the two copies, one of two possible fates eventually occurs

- Subfunctionalization



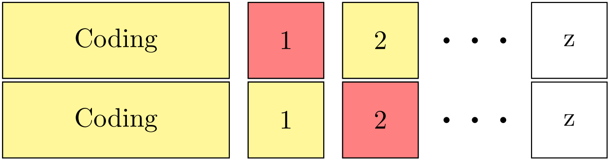
- Pseudogenization, or gene loss



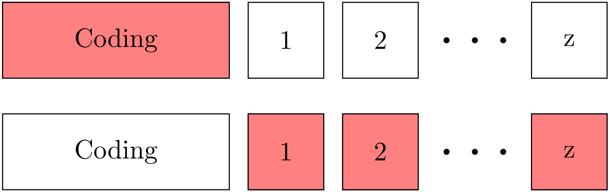
Fates of Duplicates

As mutations build up in the two copies, one of two possible fates eventually occurs

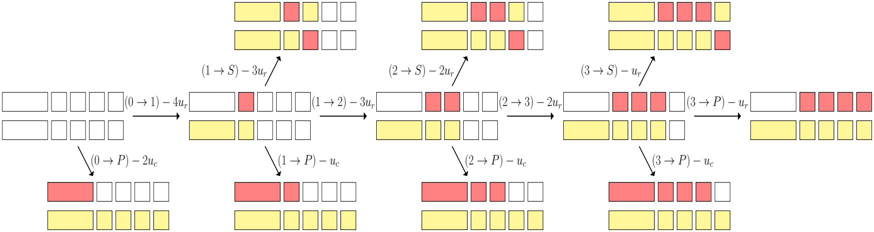
- Subfunctionalization



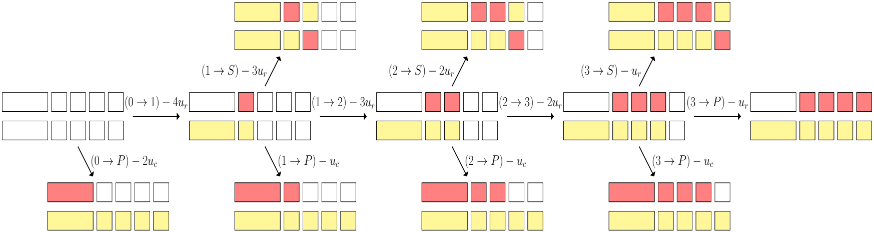
- Pseudogenization, or gene loss



Biological model



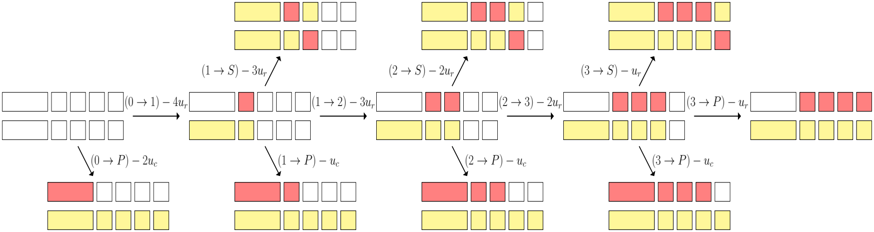
Biological model



Notice

- Initial rate of Pseudogenization is $2u_c$

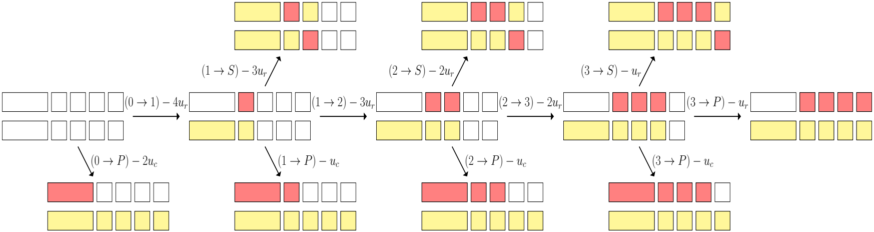
Biological model



Notice

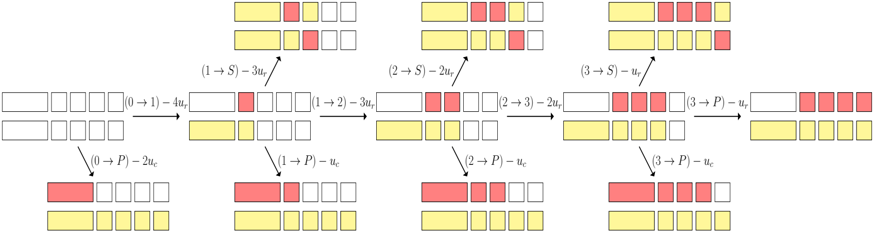
- Initial rate of Pseudogenization is $2u_c$
- After first mutation, drops to u_c

Biological model



Notice

- Initial rate of Pseudogenization is $2u_c$
- After first mutation, drops to u_c
- At final mutation, increases to $u_c + u_r$.



Notice

- Initial rate of Pseudogenization is $2u_c$
- After first mutation, drops to u_c
- At final mutation, increases to $u_c + u_r$.
- Rate of Subfunctionalization equals rate of transition to $i + 1$ equals $(z - i)u_r$.

Model is given by $\mathbf{Q} = [q_{ij}]$ where

Model is given by $\mathbf{Q} = [q_{ij}]$ where

$$q_{ij} = \begin{cases} 2u_c & \text{if } i = 0, j = P \\ 2zu_r & \text{if } i = 0, j = 1 \\ u_c & \text{if } 1 \leq i \leq z-2, j = P \\ (z-i)u_r & \text{if } 1 \leq i \leq z-2, j = i+1 \text{ or } j = S \\ u_r + u_c & \text{if } i = z-1, j = P \\ u_r & \text{if } i = z-1, j = S. \end{cases} \tag{1}$$

Model is given by $\mathbf{Q} = [q_{ij}]$ where

$$q_{ij} = \begin{cases} 2u_c & \text{if } i = 0, j = P \\ 2zu_r & \text{if } i = 0, j = 1 \\ u_c & \text{if } 1 \leq i \leq z-2, j = P \\ (z-i)u_r & \text{if } 1 \leq i \leq z-2, j = i+1 \text{ or } j = S \\ u_r + u_c & \text{if } i = z-1, j = P \\ u_r & \text{if } i = z-1, j = S. \end{cases} \tag{1}$$

The structure of this CTMC is much like those that give rise to the phase-type distribution.

Generator as block matrix

For CTMCs of this structure, it is convenient to write

$$\mathbf{Q} = \left[\begin{array}{c|c} \mathbf{Q}^* & \mathbf{V} \\ \hline \mathbf{O} & \mathbf{O} \end{array}\right], \tag{2}$$

where $\mathbf{Q}^*$ contains the entries corresponding to transitions between transient states, and $\mathbf{V}$ transitions to absorbing states.

Some measures

Its possible to exploit the phase-type-like structure of our chain to derive many measures of interest.

Its possible to exploit the phase-type-like structure of our chain to derive many measures of interest.

Probability density of absorption

$$f_i(t) = \underline{\mathbf{e_i}} e^{\mathbf{Q}^* t} \mathbf{V1} \tag{3}$$

Some measures

Its possible to exploit the phase-type-like structure of our chain to derive many measures of interest.

Probability density of absorption

$$f_i(t) = \underline{\mathbf{e_i}} e^{\mathbf{Q}^* t} \mathbf{V1} \tag{3}$$

Cumulative distribution function

$$\begin{aligned} F_i(t) &= \int_{u=0}^{t} f_i(u) du \\ &= \int_0^t \underline{\mathbf{e_i}} e^{\mathbf{Q}^* u} \mathbf{V}\underline{\mathbf{1}} du \end{aligned}$$

Some measures

Its possible to exploit the phase-type-like structure of our chain to derive many measures of interest.

Probability density of absorption

$$f_i(t) = \underline{\mathbf{e_i}} e^{\mathbf{Q}^* t} \mathbf{V1} \tag{3}$$

Cumulative distribution function

$$\begin{aligned} F_i(t) &= \int_{u=0}^{t} f_i(u) du \\ &= \int_0^t \underline{\mathbf{e_i}} e^{\mathbf{Q}^* u} \mathbf{V}\underline{\mathbf{1}} du \end{aligned}$$

Using the fact that $\mathbf{Q}^*\underline{\mathbf{1}} + \mathbf{V}\underline{\mathbf{1}} = 0$ its easy to show

Its possible to exploit the phase-type-like structure of our chain to derive many measures of interest.

Probability density of absorption

$$f_i(t) = \underline{\mathbf{e_i}} e^{\mathbf{Q}^* t} \mathbf{V1} \tag{3}$$

Cumulative distribution function

$$\begin{aligned} F_i(t) &= \int_{u=0}^{t} f_i(u) du \\ &= \int_0^t \underline{\mathbf{e_i}} e^{\mathbf{Q}^* u} \mathbf{V}\underline{\mathbf{1}} du \end{aligned}$$

Using the fact that $\mathbf{Q}^*\underline{\mathbf{1}} + \mathbf{V}\underline{\mathbf{1}} = 0$ its easy to show

$$F_i(t) = 1 - \underline{\mathbf{e_i}} e^{\mathbf{Q}^* t} \underline{\mathbf{1}}. \tag{4}$$

With the density and cumulative distribution functions, we're able to derive results for various measures

- Probability of absorption into $j \in \{S, P\}$

$$\begin{aligned} p_{i,j} &= \int_{t=0}^{\infty} \underline{\mathbf{e_i}} e^{\mathbf{Q}^* t} \mathbf{V}_j dt \\ &= -\underline{\mathbf{e_i}} (\mathbf{Q}^*)^{(-1)} \mathbf{V}_j \end{aligned} \tag{5}$$

- The k^{th} moment of time until absorption

$$\begin{aligned} m_i^{(k)} &= \int_{t=0}^{\infty} t^k \underline{\mathbf{e_i}} e^{\mathbf{Q}^* t} \mathbf{V} \underline{\mathbf{1}} dt \\ &= (-1)^k k! \underline{\mathbf{e_i}} (\mathbf{Q}^*)^{(-k)} \underline{\mathbf{1}}, \end{aligned} \tag{6}$$

- Variance of time until absorption

$$var_i = m_i^{(2)} - (m_i)^2. \tag{7}$$

Cause-specific hazard rates

When there are several absorbing states, often interested in the cause-specific hazard rate

UNIVERSITY of TASMANIA

When there are several absorbing states, often interested in the cause-specific hazard rate

Cause-specific hazard rate

$$\lambda_{ij}(t) = \lim_{h \to 0^+} \frac{P(t < T_{\{S,P\}} < t+h, X(T_{\{S,P\}}) = j | T_{\{S,P\}} > t, X(0) = i)}{h}$$
$$= \frac{f_{ij}(t)}{1 - F_i(t)} \quad = \frac{\underline{\mathbf{e_i}} e^{\mathbf{Q}^* t} \mathbf{V}_j}{\underline{\mathbf{e_0}} e^{\mathbf{Q}^* t} \underline{\mathbf{1}}} \tag{8}$$

$$\begin{aligned} f_i(t) &= \sum_{j \in \{S,P\}} f_{ij}(t), \\ \lambda_i(t) &= \sum_{j \in \{S,P\}} \lambda_{ij}(t). \end{aligned} \tag{9}$$

Thinking physically...

The numerical states in our chain are essentially a pure-birth process (recall we track the number of mutations to have occurred).

Thinking physically...

The numerical states in our chain are essentially a pure-birth process (recall we track the number of mutations to have occurred).

Think of the process eventually reaching the state $z - 1$, with no possibility of transition from i to $j < i$.

The numerical states in our chain are essentially a pure-birth process (recall we track the number of mutations to have occurred).

Think of the process eventually reaching the state $z-1$, with no possibility of transition from i to $j<i$.
Recall

- $q_{z-1,P} = u_c + u_r$,

The numerical states in our chain are essentially a pure-birth process (recall we track the number of mutations to have occurred).

Think of the process eventually reaching the state $z-1$, with no possibility of transition from i to $j < i$.
Recall

- $q_{z-1,P} = u_c + u_r$,
- $q_{z-1,S} = u_r$.

The numerical states in our chain are essentially a pure-birth process (recall we track the number of mutations to have occurred).

Think of the process eventually reaching the state $z-1$, with no possibility of transition from i to $j < i$.
Recall

- $q_{z-1,P} = u_c + u_r$,
- $q_{z-1,S} = u_r$.

Since hazard rate assumes process is not absorbed, as t becomes large $X(t)$ is almost certainly $z-1$.

Thinking physically...

The numerical states in our chain are essentially a pure-birth process (recall we track the number of mutations to have occurred).

Think of the process eventually reaching the state $z-1$, with no possibility of transition from i to $j<i$.
Recall

- $q_{z-1,P} = u_c + u_r$,
- $q_{z-1,S} = u_r$.

Since hazard rate assumes process is not absorbed, as t becomes large $X(t)$ is almost certainly $z-1$.

So $\lim_{t\to\infty} \lambda_{ij}$ is surely $u_c + u_r$ for $j=P$ u_r for $j=S$.

We define the following:

Some events

- The event that processes has not been absorbed by time t, but is absorbed by later time $t + h$
  $A_t^h = \{t < T_{\{S,P\}} < t + h\}$.

We define the following:

Some events

- The event that processes has not been absorbed by time t, but is absorbed by later time $t + h$
  $A_t^h = \{t < T_{\{S,P\}} < t + h\}$.
- The event that process has not been absorbed by time t
  $B_t = \{T_{\{S,P\}} > t\}$.

We define the following:

Some events

- The event that processes has not been absorbed by time t, but is absorbed by later time $t+h$
  $A_t^h = \{t < T_{\{S,P\}} < t+h\}$.
- The event that process has not been absorbed by time t
  $B_t = \{T_{\{S,P\}} > t\}$.
- The event that the process has entered state $z-1$ by time t
  $C_t = \{T_{z-1} \leq t\}$.

We define the following:

Some events

- The event that processes has not been absorbed by time t, but is absorbed by later time $t + h$
  $A_t^h = \{t < T_{\{S,P\}} < t + h\}$.
- The event that process has not been absorbed by time t
  $B_t = \{T_{\{S,P\}} > t\}$.
- The event that the process has entered state $z - 1$ by time t
  $C_t = \{T_{z-1} \leq t\}$.
- The event that the process has not entered state $z - 1$ by time t
  $\overline{C}_t = \{T_{z-1} > t\}$

Proof outline for $\lim_{t \to \infty} \lambda_{ij}$

First, note that

$$\lim_{t \to \infty} P(C_t | B_t, X(0) = i) = 1 \quad \text{and} \quad \lim_{t \to \infty} P(\overline{C}_t | B_t, X(0) = i) = 0,$$

First, note that

$$\lim_{t\to\infty} P(C_t|B_t, X(0)=i) = 1 \quad \text{and} \quad \lim_{t\to\infty} P(\overline{C}_t|B_t, X(0)=i) = 0,$$

Now,

$$\lim_{t\to\infty} \lambda_{ij}(t) = \lim_{t\to\infty} \lim_{h\to 0^+} \frac{P(A_t^h, X(T_{\{S,P\}}) = j|B_t, X(0)=i)}{h}$$

Proof outline for $\lim_{t\to\infty} \lambda_{ij}$

First, note that

$$\lim_{t\to\infty} P(C_t|B_t, X(0)=i) = 1 \quad \text{and} \quad \lim_{t\to\infty} P(\overline{C}_t|B_t, X(0)=i) = 0,$$

Now,

$$\lim_{t\to\infty} \lambda_{ij}(t) = \lim_{t\to\infty} \lim_{h\to 0^+} \frac{P(A_t^h, X(T_{\{S,P\}}) = j|B_t, X(0)=i)}{h}$$

by law of total probability and two limits above this becomes

$$\lim_{t\to\infty} \lim_{h\to 0^+} \frac{P(A_t^h, X(T_{\{S,P\}}) = j|B_t, X(0)=i, C_t)}{h}$$

Proof outline for $\lim_{t\to\infty} \lambda_{ij}$

First, note that

$$\lim_{t\to\infty} P(C_t|B_t, X(0)=i) = 1 \quad \text{and} \quad \lim_{t\to\infty} P(\overline{C}_t|B_t, X(0)=i) = 0,$$

Now,

$$\lim_{t\to\infty} \lambda_{ij}(t) = \lim_{t\to\infty} \lim_{h\to 0^+} \frac{P(A_t^h, X(T_{\{S,P\}}) = j|B_t, X(0)=i)}{h}$$

by law of total probability and two limits above this becomes

$$\lim_{t\to\infty} \lim_{h\to 0^+} \frac{P(A_t^h, X(T_{\{S,P\}}) = j|B_t, X(0)=i, C_t)}{h}$$

Which simplifies to

$$\lim_{t\to\infty} \lim_{h\to 0^+} \frac{P(X(t+h) = j|X(t) = z-1)}{h} \tag{10}$$

Proof outline for $\lim_{t \to \infty} \lambda_{ij}$

First, note that

$$\lim_{t\to\infty} P(C_t | B_t, X(0) = i) = 1 \quad \text{and} \quad \lim_{t\to\infty} P(\overline{C}_t | B_t, X(0) = i) = 0,$$

Now,

$$\lim_{t\to\infty} \lambda_{ij}(t) = \lim_{t\to\infty} \lim_{h\to 0^+} \frac{P(A_t^h, X(T_{\{S,P\}}) = j | B_t, X(0) = i)}{h}$$

by law of total probability and two limits above this becomes

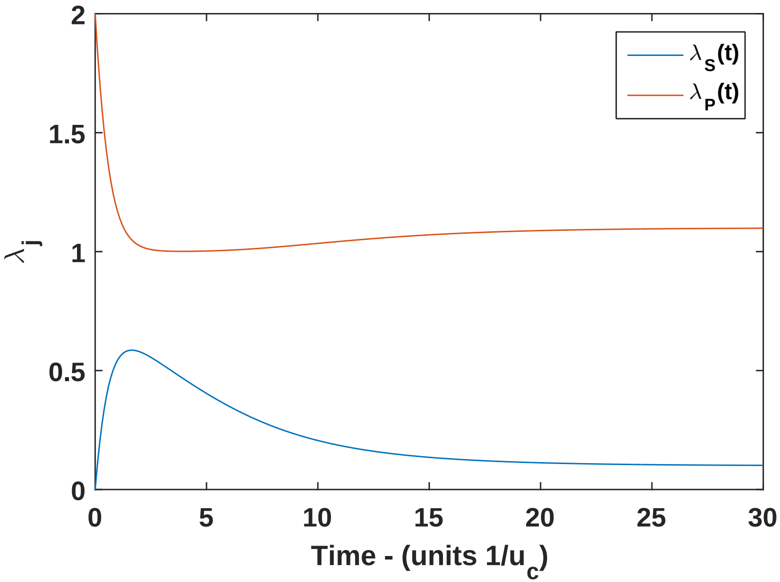
$$\lim_{t\to\infty} \lim_{h\to 0^+} \frac{P(A_t^h, X(T_{\{S,P\}}) = j | B_t, X(0) = i, C_t)}{h}$$

Which simplifies to

$$\lim_{t\to\infty} \lim_{h\to 0^+} \frac{P(X(t+h) = j | X(t) = z-1)}{h} \tag{10}$$

By Markov Property we can drop the t's, and we're left with $q_{z-1,j}$

Cause-specific hazard rates



Something slightly different

We might be interested in the rate of absorption into state P at time t conditional only on not having been absorbed into P.

Something slightly different

We might be interested in the rate of absorption into state P at time t conditional only on not having been absorbed into P.
We define the following rate

Pseudogenization rate

$$
\begin{aligned}
h_P^z(t) &= \lim_{h\to 0^+} \frac{P(t < T_P < t+h | T_P > t, X(0)=0)}{h} \\
&= \frac{f(t)}{1-F(t)} \\
&= \frac{\underline{\mathbf{e}}_0 e^{\mathbf{Q}^* t}\mathbf{V}_P}{1-\int_{u=0}^{t} \underline{\mathbf{e}}_0 e^{\mathbf{Q}^* u}\mathbf{V}_P du} \\
&= \frac{\underline{\mathbf{e}}_0 e^{\mathbf{Q}^* t}\underline{\mathbf{V}}_P}{1-\underline{\mathbf{e}}_0 \left(e^{\mathbf{Q}^* t}-\mathbf{I}\right)(\mathbf{Q}^*)^{(-1)}\mathbf{V}_P}.
\end{aligned} \tag{11}
$$

Here T_P is RV tracking time to absorption into P, and could be infinity.

Some limit analysis

Intuitively, expect pseudogenization rate to go to 0 as $t \to \infty$.

Intuitively, expect pseudogenization rate to go to 0 as $t \to \infty$.
This is easily proved using fact that

$$lim_{t \to \infty} e^{\mathbf{Q}^* t} = 0, \tag{12}$$

Some limit analysis

Intuitively, expect pseudogenization rate to go to 0 as $t \to \infty$.
This is easily proved using fact that

$$lim_{t\to\infty} e^{\mathbf{Q}^* t} = 0, \tag{12}$$

Notice

$$\lim_{t\to\infty} h_P(t) = \frac{\lim_{t\to\infty} \underline{\mathbf{e}}_0 e^{\mathbf{Q}^* t} \underline{\mathbf{v}}_P}{1 - \lim_{t\to\infty} \underline{\mathbf{e}}_0 \left(e^{\mathbf{Q}^* t} - \mathbf{I}\right) (\mathbf{Q}^*)^{(-1)} \mathbf{v}_P}$$

Intuitively, expect pseudogenization rate to go to 0 as $t \to \infty$.
This is easily proved using fact that

$$lim_{t\to\infty} e^{\mathbf{Q}^* t} = 0, \tag{12}$$

Notice

$$\begin{aligned}\lim_{t\to\infty} h_P(t) &= \frac{\lim_{t\to\infty} \underline{\mathbf{e}}_0 e^{\mathbf{Q}^* t} \underline{\mathbf{v}}_P}{1 - \lim_{t\to\infty} \underline{\mathbf{e}}_0 \left(e^{\mathbf{Q}^* t} - \mathbf{I}\right) (\mathbf{Q}^*)^{(-1)} \mathbf{v}_P} \\ &= \frac{0}{1 + \underline{\mathbf{e}}_0 (\mathbf{Q}^*)^{(-1)} \mathbf{v}_P}\end{aligned}$$

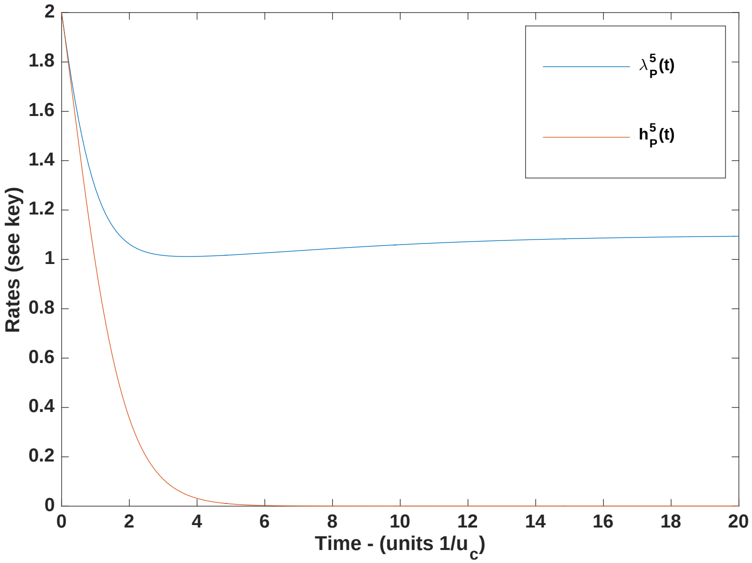
Intuitively, expect pseudogenization rate to go to 0 as $t \to \infty$.
This is easily proved using fact that

$$lim_{t\to\infty} e^{\mathbf{Q}^* t} = 0, \tag{12}$$

Notice

$$\begin{aligned} \lim_{t\to\infty} h_P(t) &= \frac{\lim_{t\to\infty} \underline{\mathbf{e}}_0 e^{\mathbf{Q}^* t} \underline{\mathbf{v}}_P}{1 - \lim_{t\to\infty} \underline{\mathbf{e}}_0 \left(e^{\mathbf{Q}^* t} - \mathbf{I}\right) \left(\mathbf{Q}^*\right)^{(-1)} \mathbf{v}_P} \\ &= \frac{0}{1 + \underline{\mathbf{e}}_0 (\mathbf{Q}^*)^{(-1)} \mathbf{v}_P} \\ &= 0. \end{aligned} \tag{13}$$

Wake up! A picture.



A quick (historical) aside

Some work has been done in the past on modelling this subfunctionalization process.

A quick (historical) aside

Some work has been done in the past on modelling this subfunctionalization process.

- Hughes and Liberles (2007,2008) did approximate mechanistic analysis implicitly based on embedded DTMC of process considered here.

A quick (historical) aside

Some work has been done in the past on modelling this subfunctionalization process.

- Hughes and Liberles (2007,2008) did approximate mechanistic analysis implicitly based on embedded DTMC of process considered here.
- Later, phenomenological approximations have been used, informed by analysis of Hughes and Liberles (Notable Konrad (2011), Tuefel (2014)).

A quick (historical) aside

Some work has been done in the past on modelling this subfunctionalization process.

- Hughes and Liberles (2007,2008) did approximate mechanistic analysis implicitly based on embedded DTMC of process considered here.
- Later, phenomenological approximations have been used, informed by analysis of Hughes and Liberles (Notable Konrad (2011), Tuefel (2014)).
- Sigmoid functions have been successful in fitting to real data.

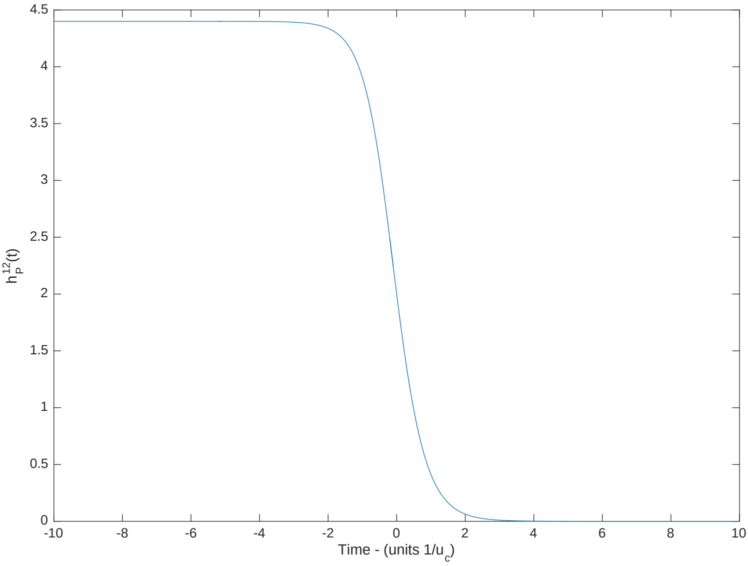
A quick (historical) aside

Some work has been done in the past on modelling this subfunctionalization process.

- Hughes and Liberles (2007,2008) did approximate mechanistic analysis implicitly based on embedded DTMC of process considered here.
- Later, phenomenological approximations have been used, informed by analysis of Hughes and Liberles (Notable Konrad (2011), Tuefel (2014)).
- Sigmoid functions have been successful in fitting to real data.

This motivates us to look into the behaviour of the model in negative time!

Sigmoid Function



Limit as $t \to -\infty$

Usually interested in physical time. Analysis of negative limit is novel.

Limit as $t \to -\infty$

Usually interested in physical time. Analysis of negative limit is novel. We considered a general CTMC with initial distribtuion $\underline{\alpha}$ and structure

$$\mathbf{Q} = \left[\begin{array}{c|c} \mathbf{Q}^* & \mathbf{V} \\ \hline \mathbf{O} & \mathbf{O} \end{array}\right], \tag{14}$$

Limit as $t \to -\infty$

Usually interested in physical time. Analysis of negative limit is novel. We considered a general CTMC with initial distribtuion $\underline{\alpha}$ and structure

$$\mathbf{Q} = \left[\begin{array}{c|c} \mathbf{Q}^* & \mathbf{V} \\ \hline \mathbf{O} & \mathbf{O} \end{array}\right], \tag{14}$$

$h_P(t; \alpha)$ is an obvious generalization of our earlier function (rate of transition into absorbing state P assuming not already in P).

Usually interested in physical time. Analysis of negative limit is novel. We considered a general CTMC with initial distribtuion $\underline{\alpha}$ and structure

$$\mathbf{Q} = \left[\begin{array}{c|c} \mathbf{Q}^* & \mathbf{V} \\ \hline \mathbf{O} & \mathbf{O} \end{array}\right], \tag{14}$$

$h_P(t; \alpha)$ is an obvious generalization of our earlier function (rate of transition into absorbing state P assuming not already in P).

Using l'Hôpital's rule we get

$$\lim_{t \to -\infty} h_P(t; \underline{\alpha}) \quad = \quad \lim_{t \to -\infty} \frac{\underline{\alpha} e^{\mathbf{Q}^*} \mathbf{Q}^* \underline{\mathbf{V}}_P}{-\underline{\alpha} e^{\mathbf{Q}^* t} \mathbf{v}_P}. \tag{15}$$

Limit as $t \to -\infty$

We diagonalize $\mathbf{Q}^*$

$$\mathbf{Q}^* = \mathbf{A}^{-1}\mathbf{\Lambda}\mathbf{A}$$

We diagonalize $\mathbf{Q}^*$

$$\mathbf{Q}^* = \mathbf{A}^{-1}\mathbf{\Lambda}\mathbf{A}$$

Then we do some algebra to get

$$\lim_{t\to-\infty} h_P(t;\underline{\alpha}) = \lim_{t\to-\infty} \frac{\sum_k [\alpha\mathbf{A}^{-1}]_k e^{\lambda_k t}\lambda_k(\mathbf{AV}_p)}{-\sum_l [\alpha\mathbf{A}^{-1}]_l e^{\lambda_l t}(\mathbf{AV}_p).} \tag{16}$$

We diagonalize $\mathbf{Q}^*$

$$\mathbf{Q}^* = \mathbf{A}^{-1}\mathbf{\Lambda}\mathbf{A}$$

Then we do some algebra to get

$$\lim_{t\to-\infty} h_P(t;\underline{\alpha}) = \lim_{t\to-\infty} \frac{\sum_k [\alpha\mathbf{A}^{-1}]_k e^{\lambda_k t}\lambda_k(\mathbf{AV}_p)}{-\sum_l [\alpha\mathbf{A}^{-1}]_l e^{\lambda_l t}(\mathbf{AV}_p).} \tag{16}$$

Letting λ_m be the eigenvalue of maximum absolute real part of $\mathbf{Q}^*$ the numerator is

Limit as $t \to -\infty$

We diagonalize $\mathbf{Q}^*$

$$\mathbf{Q}^* = \mathbf{A}^{-1}\mathbf{\Lambda}\mathbf{A}$$

Then we do some algebra to get

$$\lim_{t \to -\infty} h_P(t; \underline{\alpha}) = \lim_{t \to -\infty} \frac{\sum_k [\alpha \mathbf{A}^{-1}]_k e^{\lambda_k t} \lambda_k (\mathbf{AV}_p)}{-\sum_l [\alpha \mathbf{A}^{-1}]_l e^{\lambda_l t} (\mathbf{AV}_p).} \tag{16}$$

Letting λ_m be the eigenvalue of maximum absolute real part of $\mathbf{Q}^*$ the numerator is

Numerator

$$\sum_k [\alpha \mathbf{A}^{-1}]_k e^{(\lambda_k - \lambda_\mathrm{m})t} \lambda_k (\mathbf{Av}_p)$$

Which in the limit is just λ_m

Limit as $t \to -\infty$

Similarly, the denominator is

Denominator

$$-\sum_{l}[\alpha \mathbf{A}^{-1}]_l e^{(\lambda_l - \lambda_\mathrm{m})t} \lambda_l (\mathbf{A}\mathbf{V}_p)$$

Similarly, the denominator is

Denominator

$$-\sum_{l}[\alpha \mathbf{A}^{-1}]_l e^{(\lambda_l - \lambda_{\mathrm{m}})t} \lambda_l (\mathbf{AV}_p)$$

Which in the limit is just 1.

Similarly, the denominator is

Denominator

$$-\sum_{l}[\alpha \mathbf{A}^{-1}]_l e^{(\lambda_l - \lambda_m)t} \lambda_l (\mathbf{AV}_p)$$

Which in the limit is just 1. So, it turns out that

Result

$$\lim_{t \to -\infty} h_P(t) = -\lambda_m = Sp(\mathbf{Q}) \tag{17}$$

Limit as $t \to -\infty$

Similarly, the denominator is

Denominator

$$-\sum_l [\alpha \mathbf{A}^{-1}]_l e^{(\lambda_l - \lambda_\mathrm{m})t} \lambda_l(\mathbf{AV}_p)$$

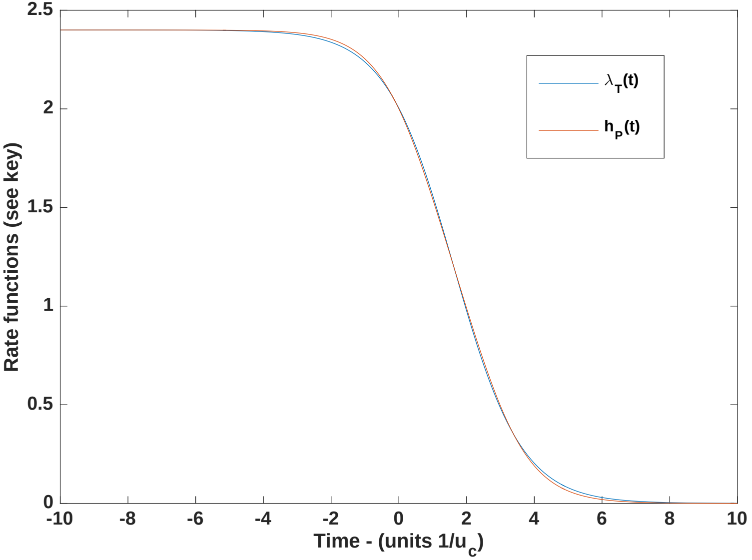
Which in the limit is just 1. So, it turns out that

Result

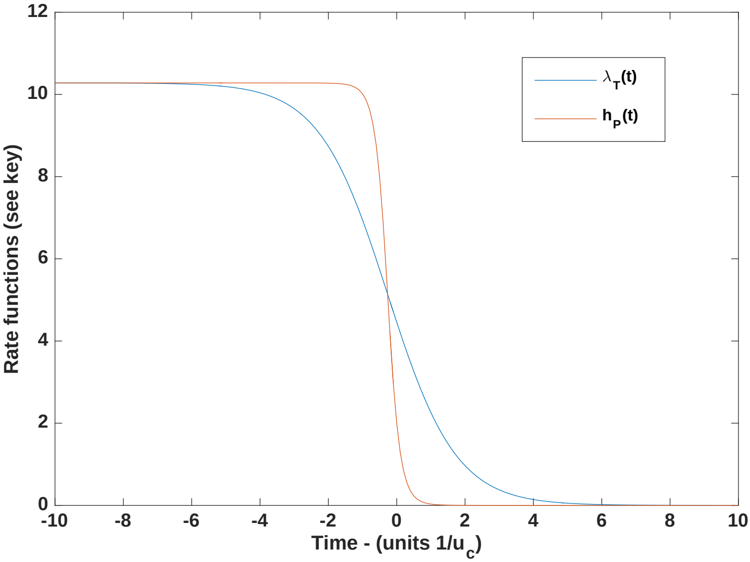
$$\lim_{t \to -\infty} h_P(t) = -\lambda_m = Sp(\mathbf{Q}) \tag{17}$$

We can use this to fit the phenomenological approximation of Tuefel et al (2014) to our exact mechanistically function.

Phenom. vs Exact



Phenom. vs Exact



Using theory of phase-type distributions we are able to derive and analyze mathematical model for the biological subfunctionalization model assuming only that

Summary

Using theory of phase-type distributions we are able to derive and analyze mathematical model for the biological subfunctionalization model assuming only that

- The rate of null mutations in regulatory regions is Poisson u_r

Using theory of phase-type distributions we are able to derive and analyze mathematical model for the biological subfunctionalization model assuming only that

- The rate of null mutations in regulatory regions is Poisson u_r
- The rate of null mutations in the coding region is u_c

Summary

Using theory of phase-type distributions we are able to derive and analyze mathematical model for the biological subfunctionalization model assuming only that

- The rate of null mutations in regulatory regions is Poisson u_r
- The rate of null mutations in the coding region is u_c

Pseudogenization rate implied by this model turns out to have the same behaviour as existing phenomenological approximations

Summary

Using theory of phase-type distributions we are able to derive and analyze mathematical model for the biological subfunctionalization model assuming only that

- The rate of null mutations in regulatory regions is Poisson u_r
- The rate of null mutations in the coding region is u_c

Pseudogenization rate implied by this model turns out to have the same behaviour as existing phenomenological approximations

- We have provided a means to translate between the exact function and the popular approximation

Summary

Using theory of phase-type distributions we are able to derive and analyze mathematical model for the biological subfunctionalization model assuming only that

- The rate of null mutations in regulatory regions is Poisson u_r
- The rate of null mutations in the coding region is u_c

Pseudogenization rate implied by this model turns out to have the same behaviour as existing phenomenological approximations

- We have provided a means to translate between the exact function and the popular approximation
- As well as deriving a host of performance measures

Summary

Using theory of phase-type distributions we are able to derive and analyze mathematical model for the biological subfunctionalization model assuming only that

- The rate of null mutations in regulatory regions is Poisson u_r
- The rate of null mutations in the coding region is u_c

Pseudogenization rate implied by this model turns out to have the same behaviour as existing phenomenological approximations

- We have provided a means to translate between the exact function and the popular approximation
- As well as deriving a host of performance measures

This work provides a mathematically rigorous, mechanistically motivated and exact analysis for the fate of gene duplicates.

Summary

Using theory of phase-type distributions we are able to derive and analyze mathematical model for the biological subfunctionalization model assuming only that

- The rate of null mutations in regulatory regions is Poisson u_r
- The rate of null mutations in the coding region is u_c

Pseudogenization rate implied by this model turns out to have the same behaviour as existing phenomenological approximations

- We have provided a means to translate between the exact function and the popular approximation
- As well as deriving a host of performance measures

This work provides a mathematically rigorous, mechanistically motivated and exact analysis for the fate of gene duplicates.

Future work

Future work will move in two directions

Future work

Future work will move in two directions

- Allowing for multiple duplication events to analyze the fates of whole gene families.

Future work

Future work will move in two directions

- Allowing for multiple duplication events to analyze the fates of whole gene families.
- Expanding the model to allow for a mixture of sub- and neofunctionalization.