

# **Biologia Molecular Computacional**

# **Homologia**

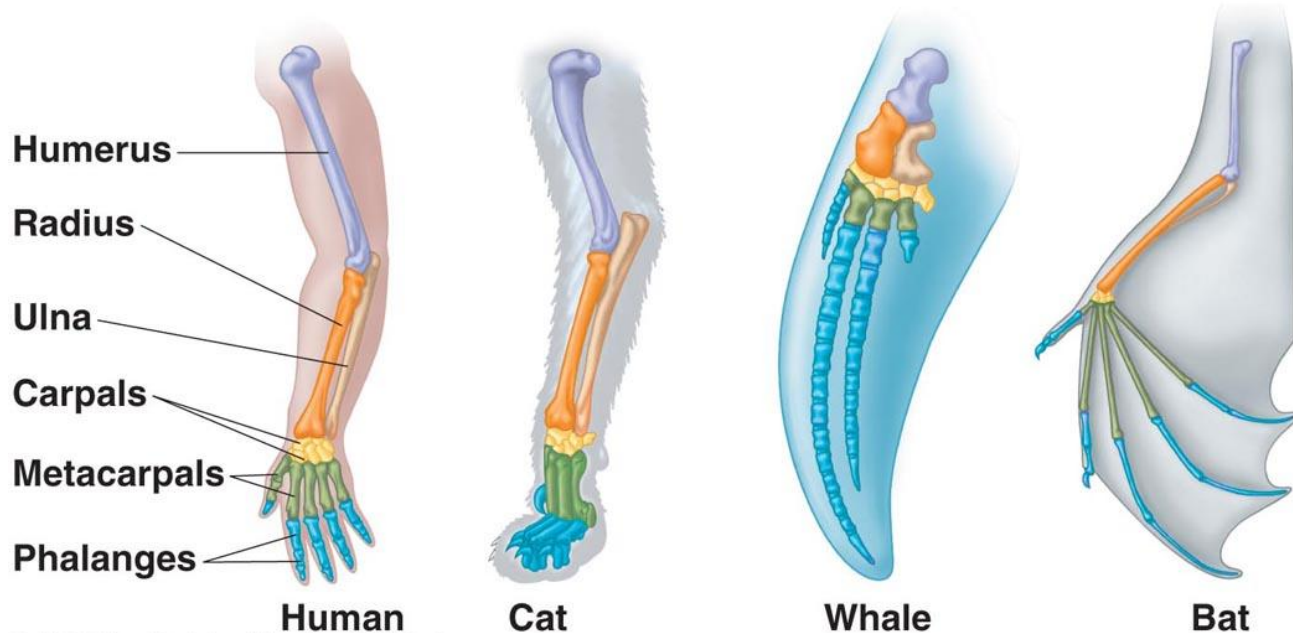
Luiz Thibério Rangel

# O que é homologia?

- Conceito básico para estudos de genômica comparativa;
- Passo inicial para estudos de filogenia(omica);
- Importante para pipelines de anotação funcional;
- Utilizado em processos de modelagem estrutural;

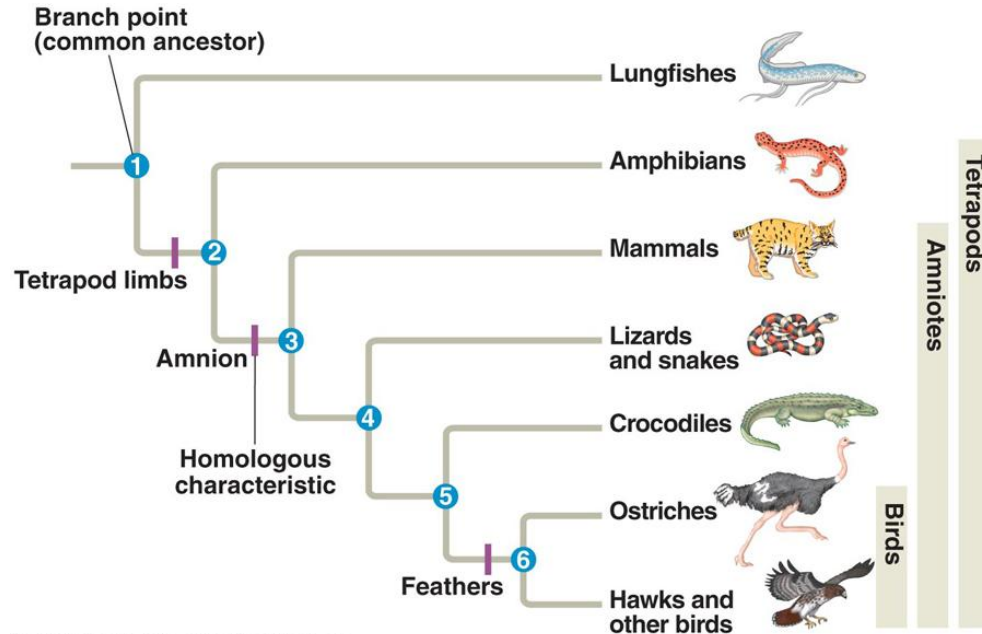
# O que é homologia?

- É o compartilhamento de uma estrutura com **origem comum!**



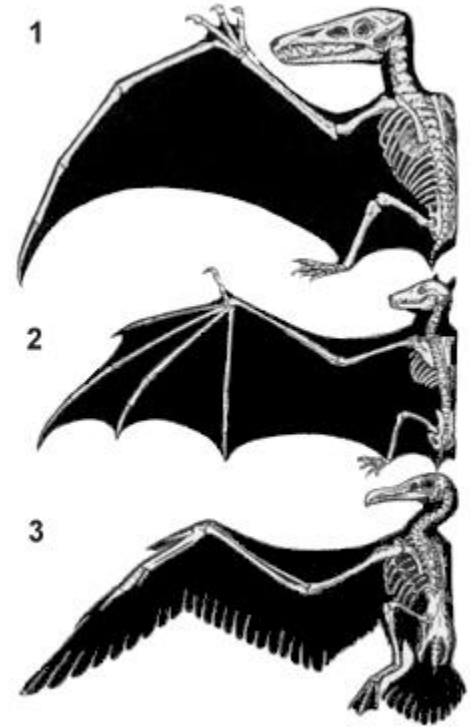
# O que é homologia?

- Essa origem é comum porque foi herdada a partir de um **ancestral comum!**



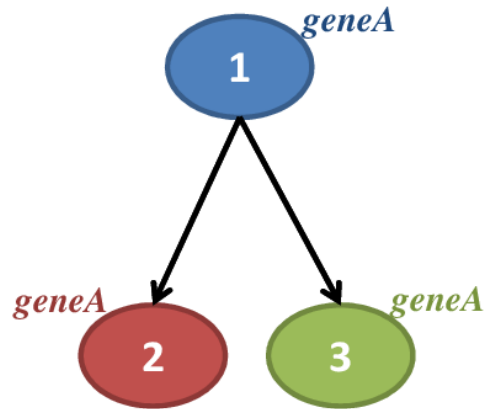
# O que é homologia?

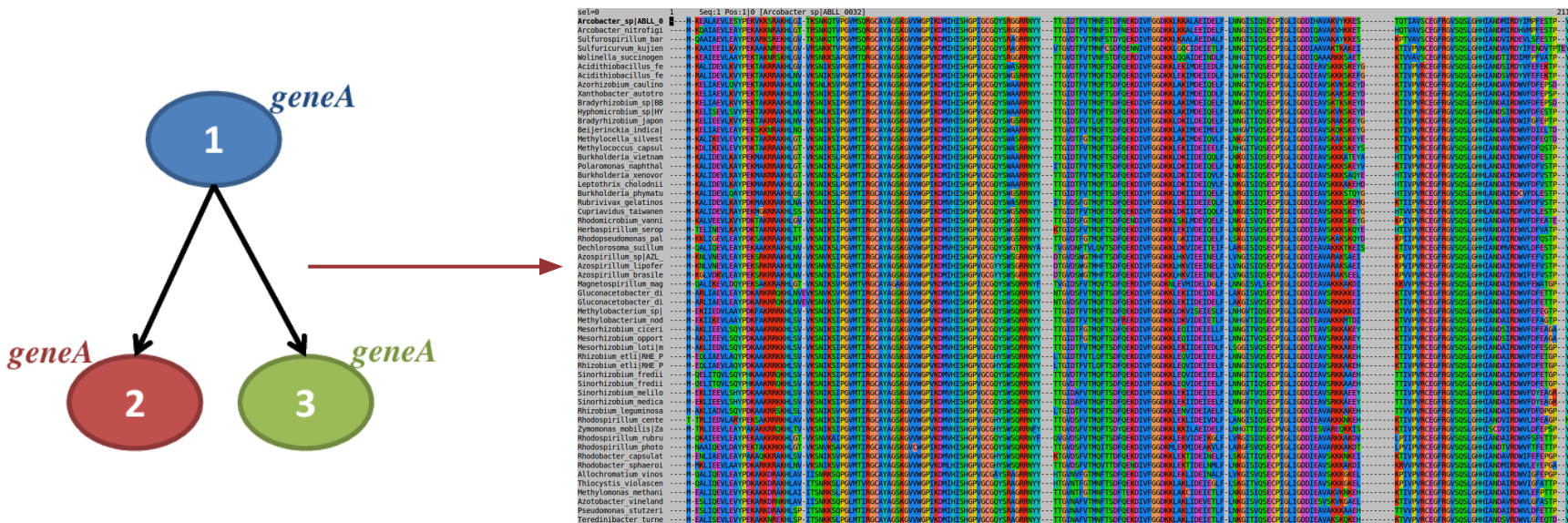
- O conceito de homologia é oposto ao de **Analogia**:
  - caracteres análogos **não possuem origem comum**, mesmo possuindo funções semelhantes!



# Adaptação para biologia molecular

- Genes de organismos distintos herdados a partir do mesmo ancestral são **homólogos**!





# Adaptação para biologia molecular

- Homologia um caracter **booleano!**
  - Verdadeiro ou Falso
- Afirmações como:
  - 30% de homologia;
  - 78% homólogos;
  - Forte homologia!



# Adaptação para biologia molecular

- Homologia um caracter **booleano!**
  - Verdadeiro ou Falso
- Afirmações como:
  - ~~30% de homologia;~~
  - ~~78% homólogos;~~
  - ~~Forte homologia!~~

**ERRADAS!**

# Adaptação para biologia molecular

- Homologia um caracter **booleano!**
  - Verdadeiro ou Falso
- Afirmações como:
  - 30% **similares**;
  - 78% **de similaridade**;
  - **Divergiram recentemente!**
- Ou a origem é comum ou não!

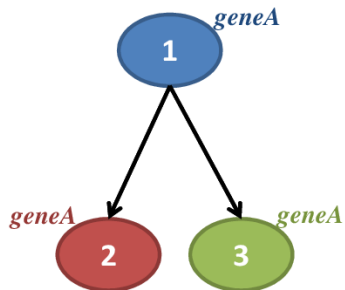
# Tipo de homologia

- Diferentes maneiras de divergir levam a diferentes tipos de homologia:
  - Ortólogos
  - Parálogos
  - Xenólogos

# Tipo de homologia

- Diferentes maneiras de divergir levam a diferentes tipos de homologia:

- Ortólogos
- Parálogos
- Xenólogos



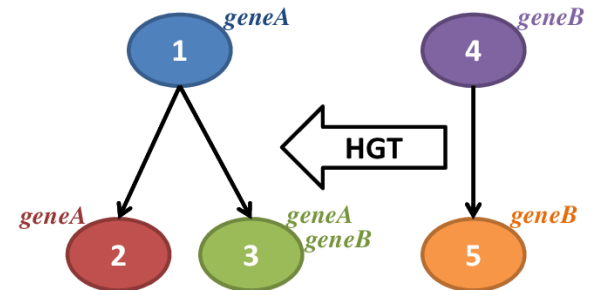
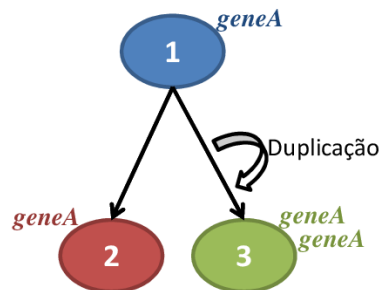
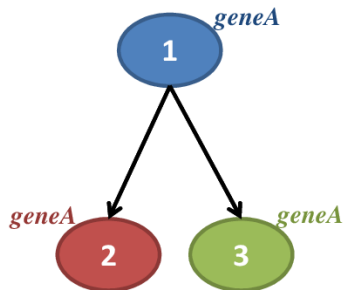
# Tipo de homologia

- Diferentes maneiras de divergir levam a diferentes tipos de homologia:
  - Ortólogos
  - Parálogos
  - Xenólogos



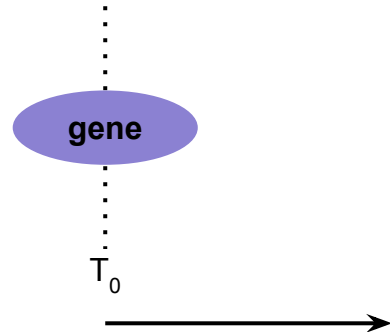
# Tipo de homologia

- Diferentes maneiras de divergir levam a diferentes tipos de homologia:
  - Ortólogos
  - Parálogos
  - Xenólogos



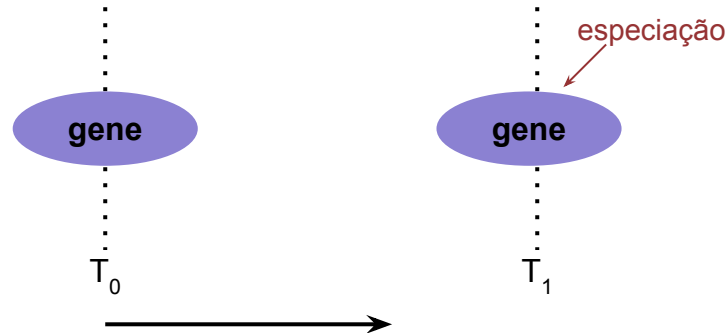
# Ortologia

- **Ortólogos:** genes que divergiram por **especiação**
  - cada descendente possui uma cópia do gene
  - tendência à conservar função:
    - ambas instâncias continuam sendo necessárias



# Ortologia

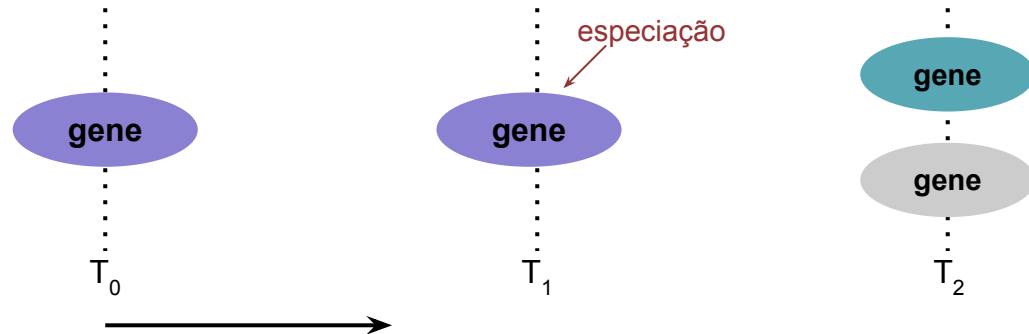
- **Ortólogos:** genes que divergiram por **especiação**
  - cada descendente possui uma cópia do gene
  - tendência à conservar função:
    - ambas instâncias continuam sendo necessárias





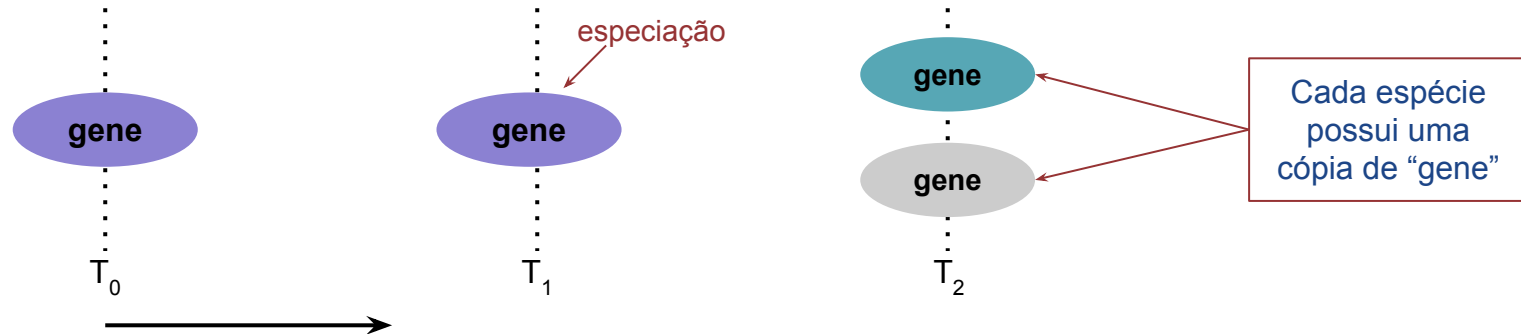
# Ortologia

- **Ortólogos:** genes que divergiram por **especiação**
  - cada descendente possui uma cópia do gene
  - tendência à conservar função:
    - ambas instâncias continuam sendo necessárias



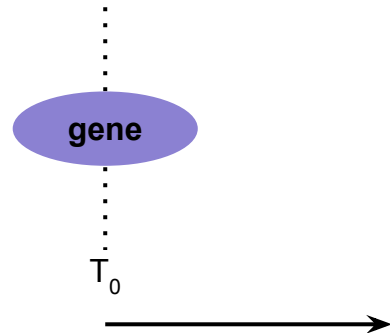
# Ortologia

- **Ortólogos:** genes que divergiram por **especiação**
  - cada descendente possui uma cópia do gene
  - tendência à conservar função:
    - ambas instâncias continuam sendo necessárias



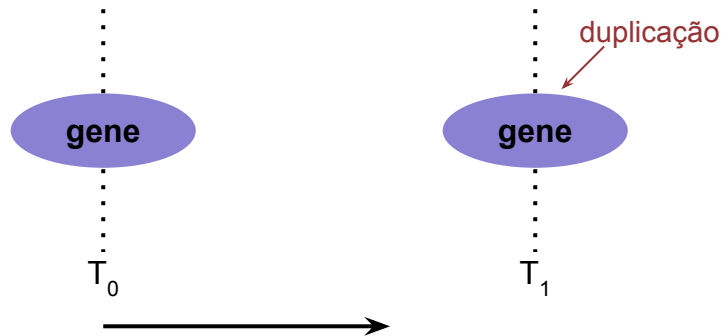
# Paralogia

- **Parálogos:** genes que divergiram por **duplicação!**
  - genes parálogos estão presentes em **mais de uma cópia!**
  - Após a divergência a **função tende a mudar!**



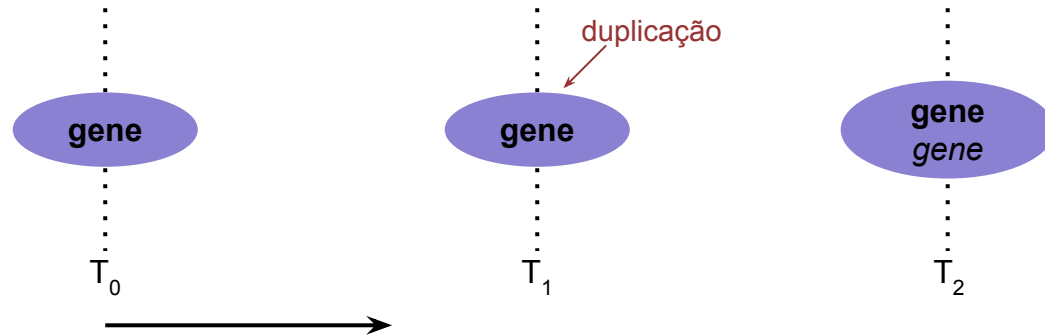
# Paralogia

- **Parálogos:** genes que divergiram por **duplicação!**
  - genes parálogos estão presentes em **mais de uma cópia!**
  - Após a divergência a **função tende a mudar!**



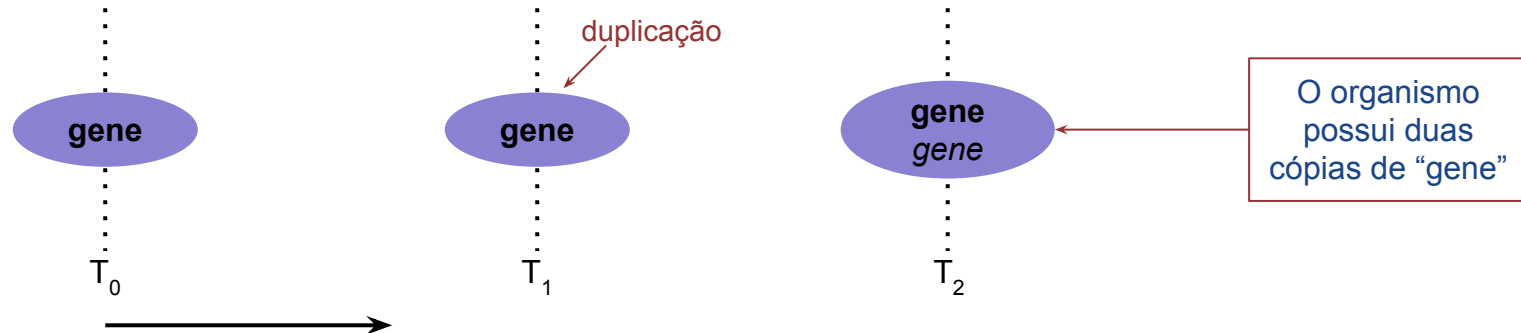
# Paralogia

- **Parálogos:** genes que divergiram por **duplicação!**
  - genes parálogos estão presentes em **mais de uma cópia!**
  - Após a divergência a **função tende a mudar!**



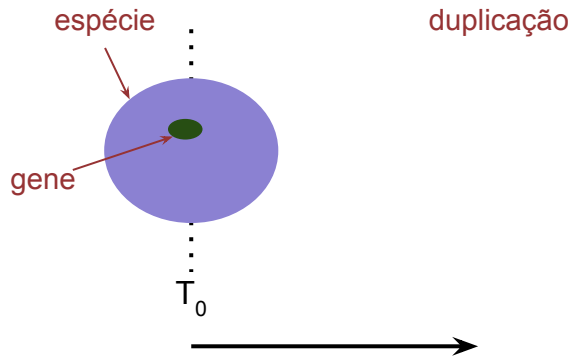
# Paralogia

- **Parálogos:** genes que divergiram por **duplicação!**
  - genes parálogos estão presentes em **mais de uma cópia!**
  - Após a divergência a **função tende a mudar!**



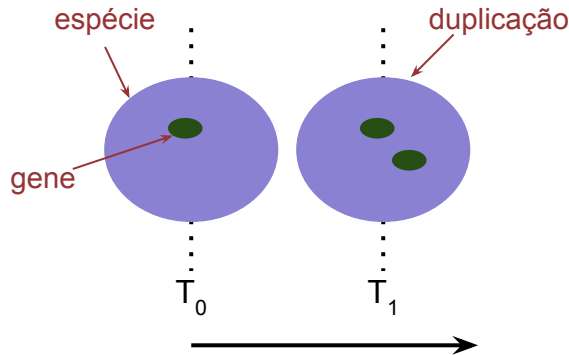
# Paralogia

- **Parálogos:** genes que divergiram por **duplicação!**
  - genes parálogos estão presentes em **mais de uma cópia!**
  - Após a divergência a **função tende a mudar!**



# Paralogia

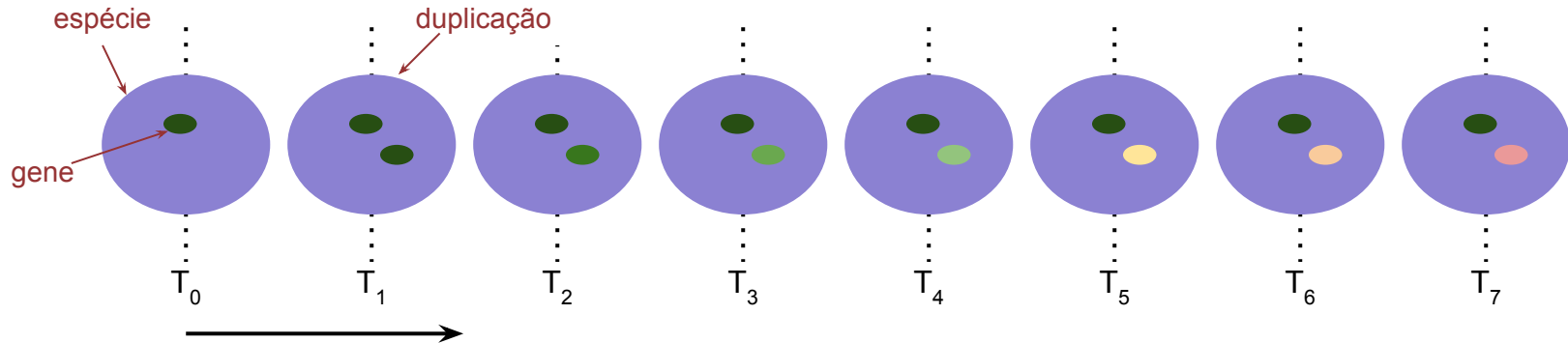
- **Parálogos:** genes que divergiram por **duplicação!**
  - genes parálogos estão presentes em **mais de uma cópia!**
  - Após a divergência a **função tende a mudar!**





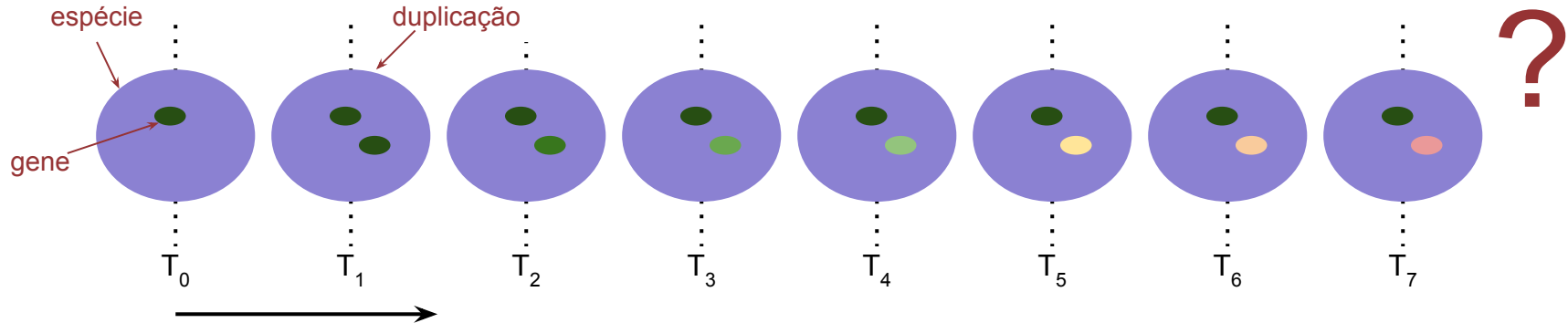
# Paralogia

- **Parálogos:** genes que divergiram por **duplicação!**
  - genes parálogos estão presentes em **mais de uma cópia!**
  - Após a divergência a **função tende a mudar!**



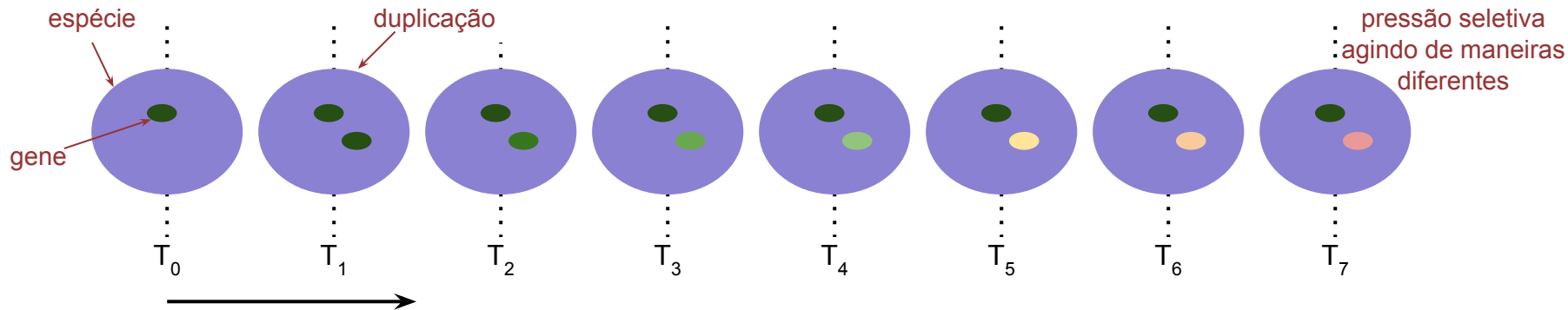
# Paralogia

- **Parálogos:** genes que divergiram por **duplicação!**
  - genes parálogos estão presentes em **mais de uma cópia!**
  - Após a divergência a **função tende a mudar!**



# Paralogia

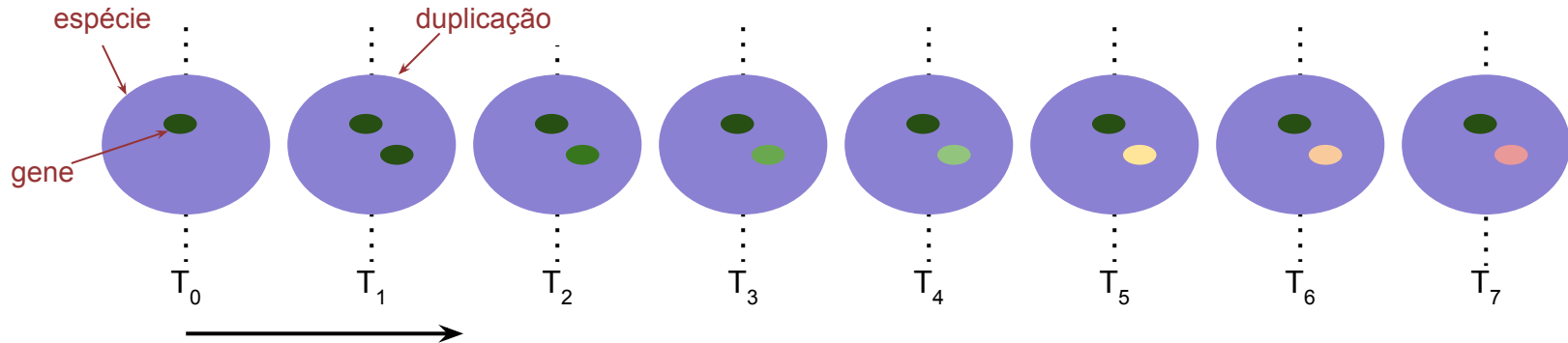
- **Parálogos:** genes que divergiram por **duplicação!**
  - genes parálogos estão presentes em **mais de uma cópia!**
  - Após a divergência a **função tende a mudar!**



# Parálogos: recentes Vs. antigos

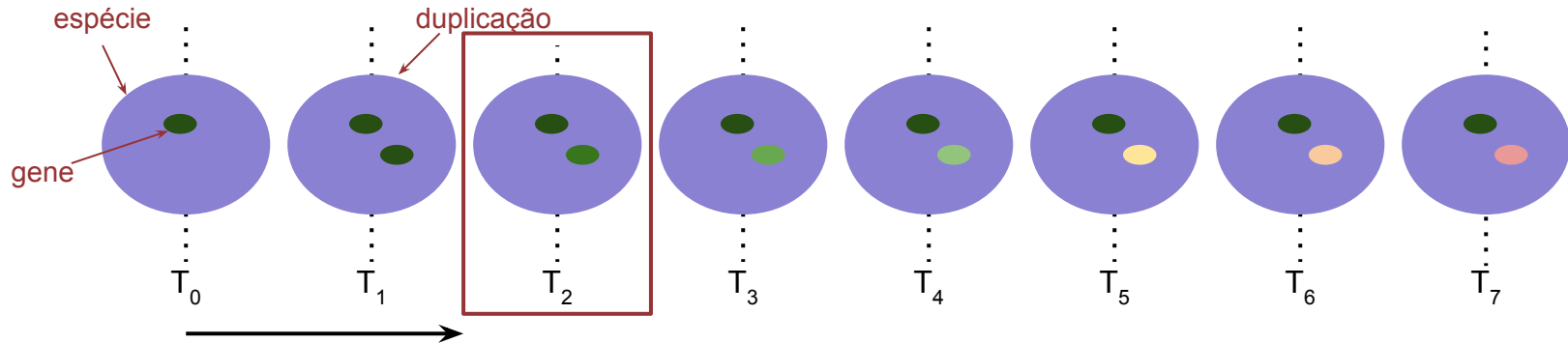
- **Parálogos:**

- duplicação recente: Inparálogos
- duplicação antiga: Outparálogos



# Parálogos: recentes Vs. antigos

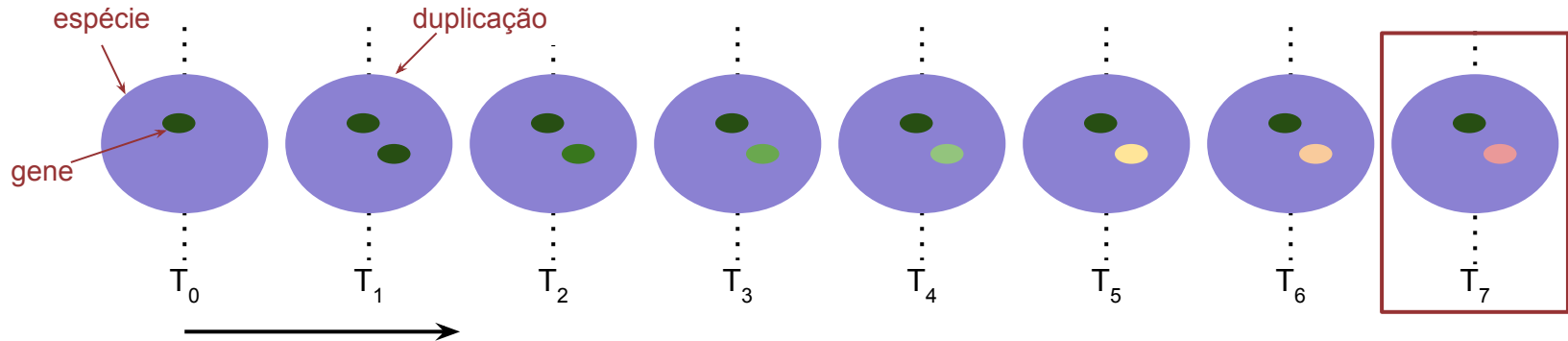
- Parálogos:
  - duplicação recente: **Inparálogos**
  - duplicação antiga: Outparálogos



# Parálogos: recentes Vs. antigos

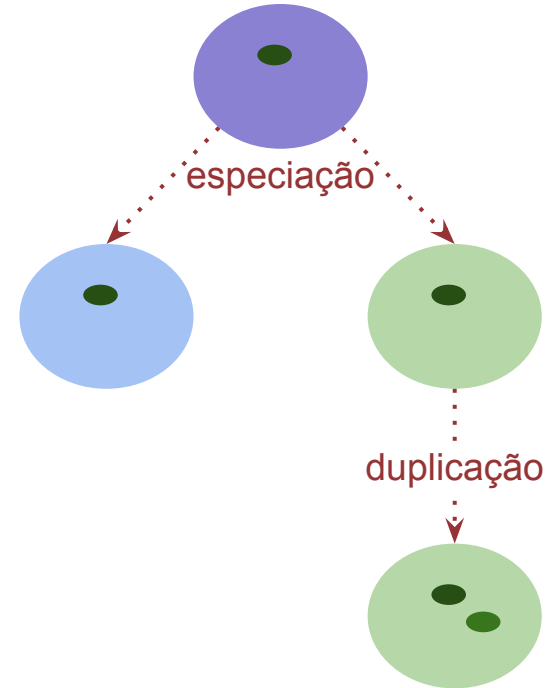
- **Parálogos:**

- duplicação recente: Inparálogos
- duplicação antiga: **Outparálogos**



# Parálogos: recentes Vs. antigos

- Inparálogos
  - Regra do dedão:
    - Duplicação ocorreu **depois** da especiação em questão



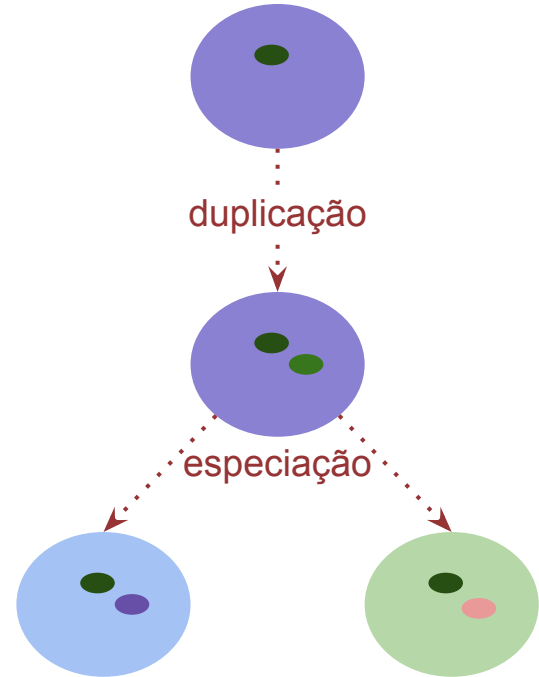
Duplicação recente = Menos tempo para divergir!

# Parálogos: recentes Vs. antigos

- Outparálogos

- Regra do dedão:
  - Duplicação ocorreu **antes** da especiação em questão

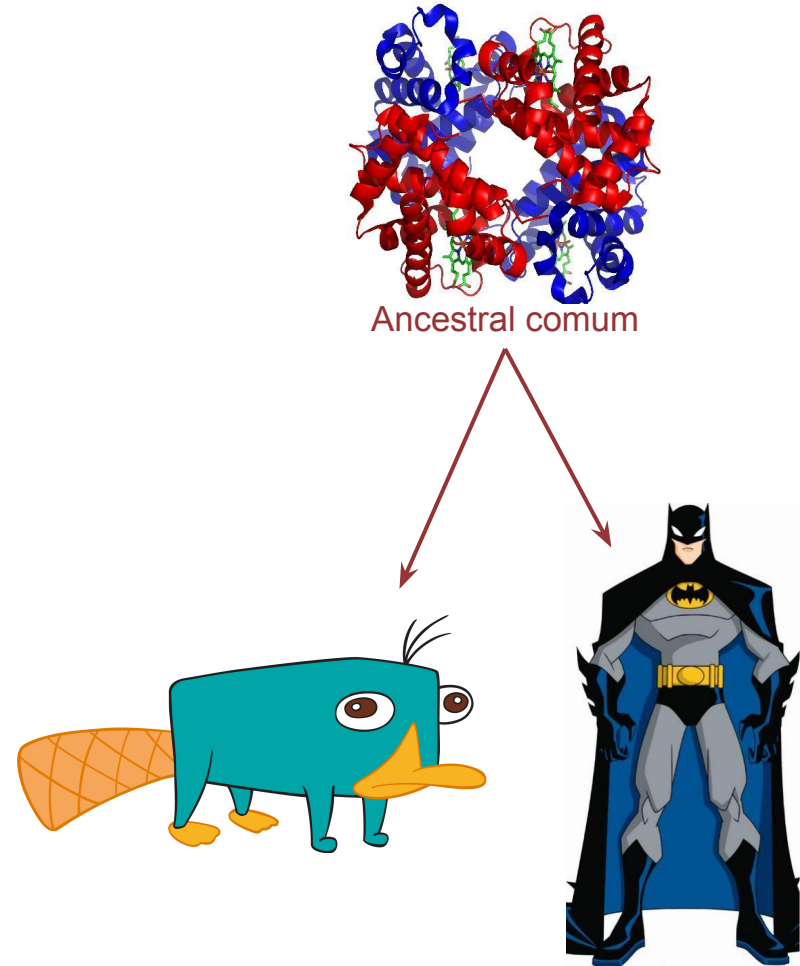
Duplicação antiga = Mais tempo para divergir!





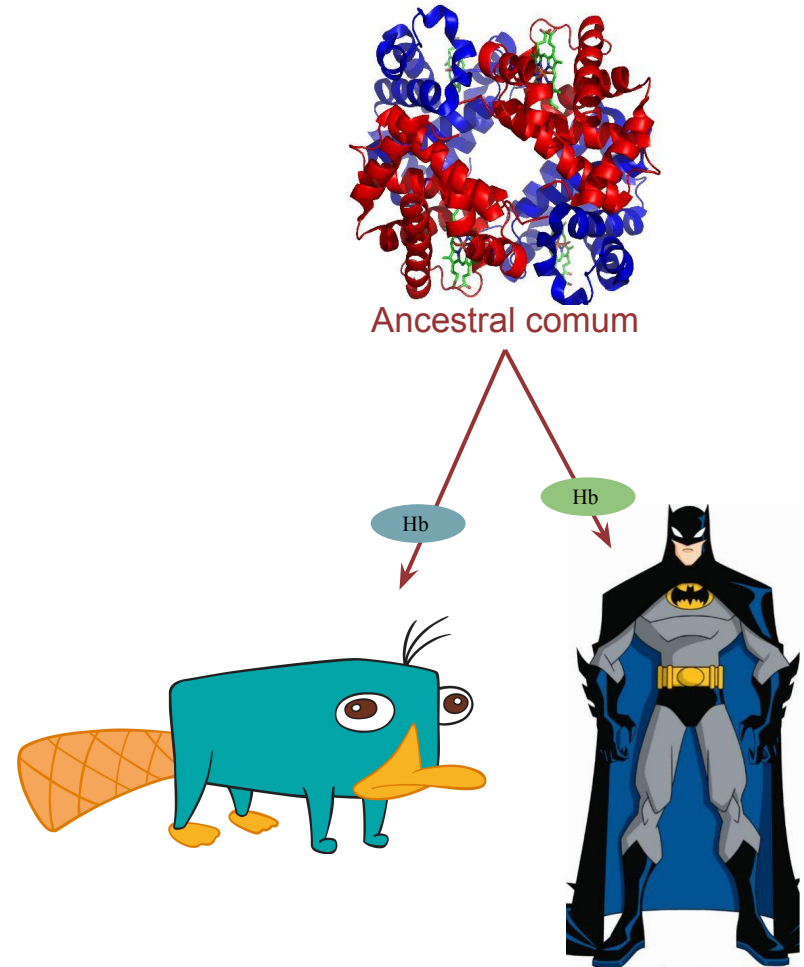
# Exemplo: Ortologia

- Hemoglobinas  $\alpha$  e  $\beta$  são conservadas em praticamente todos vertebrados!
  - portanto são as hemoglobinas de humano e ornitorrinco são homólogas!



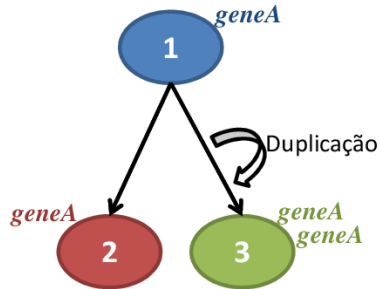
# Exemplo: Ortologia

- Hemoglobinas  $\alpha$  e  $\beta$  são conservadas em praticamente todos vertebrados!
  - portanto são as hemoglobinas de humano e ornitorrinco são homólogas!



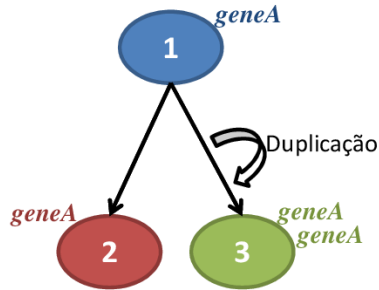
# Exemplo: Paralogia

- Agnatha possuem hemoglobinas “homotetraméricas”
  - apenas um gene codifica as 4 subunidades;

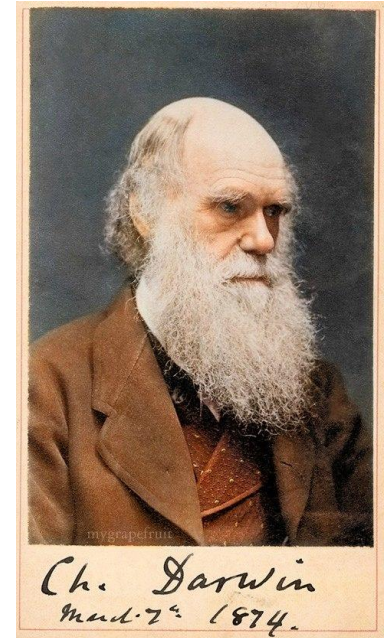


# Exemplo: Paralogia

- Agnatha possuem hemoglobinas “homotetraméricas”
  - apenas um gene codifica as 4 subunidades;
  - as subunidades  $\alpha$  e  $\beta$  surgiram a partir da **duplicação do gene ancestral!**

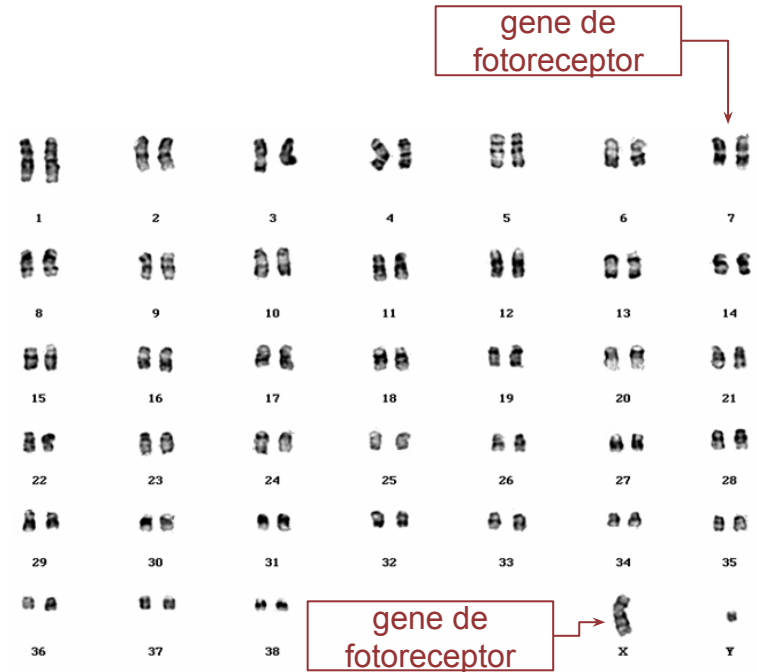


# Exemplo: Paralogia



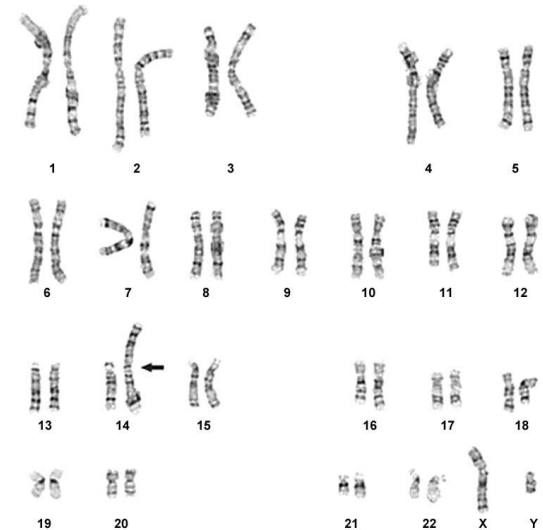
# Exemplo: Paralogia

- Cachorro (e.g.) só possui 2 fotorreceptores:
  - enxerga em PB!



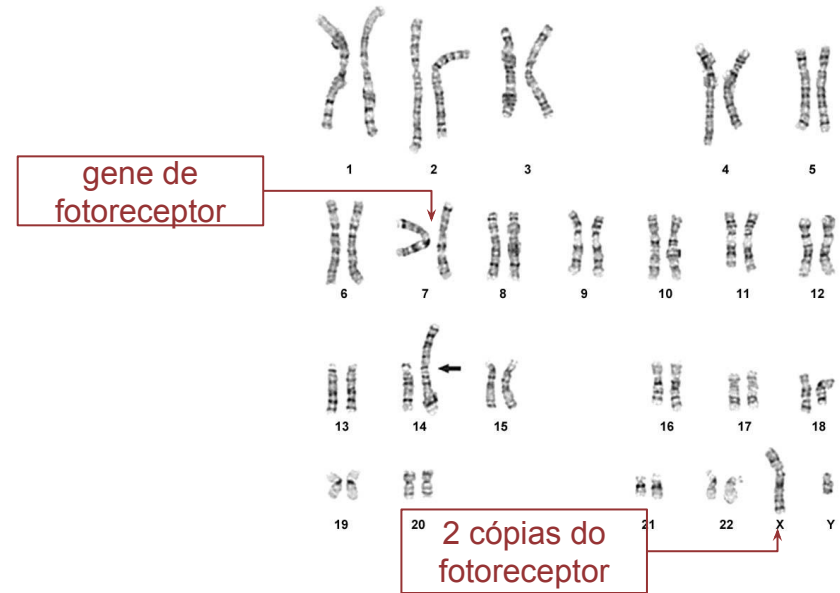
# Exemplo: Paralogia

- Primatas possuem 3 fotorreceptores!
  - como?!?!?



# Exemplo: Paralogia

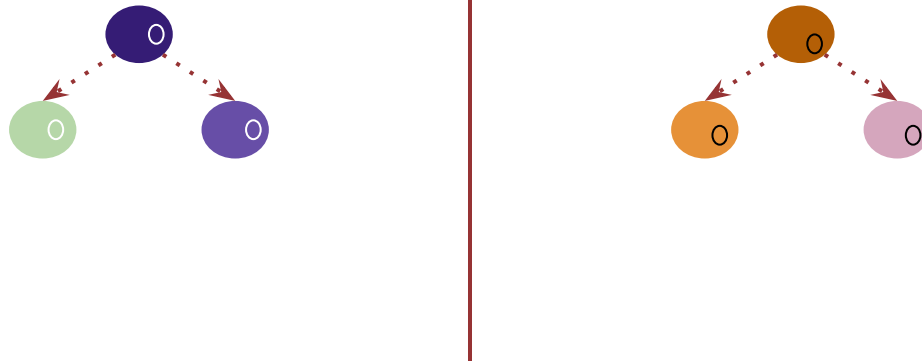
- Primatas possuem 3 fotorreceptores!
  - como?!?!?
  - duplicação do gene presente no cromossomo X!
- Por isso diptonismo é mais comum em homens!





# Xenologia

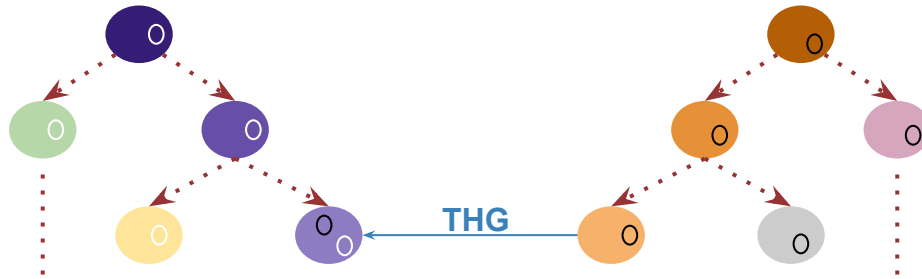
- **Xenólogos**
  - Divergiram por **Transferência Horizontal de Genes**
    - Aquisição de novas funções



# Xenologia

- Xenólogos

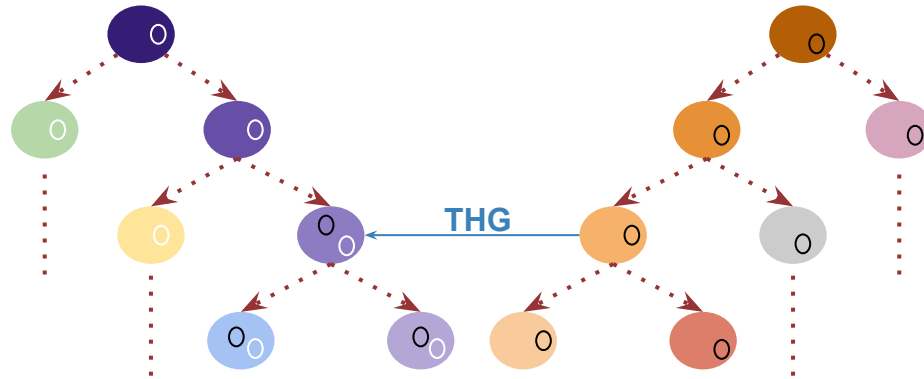
- Divergiram por **Transferência Horizontal de Genes**
  - Aquisição de novas funções



# Xenologia

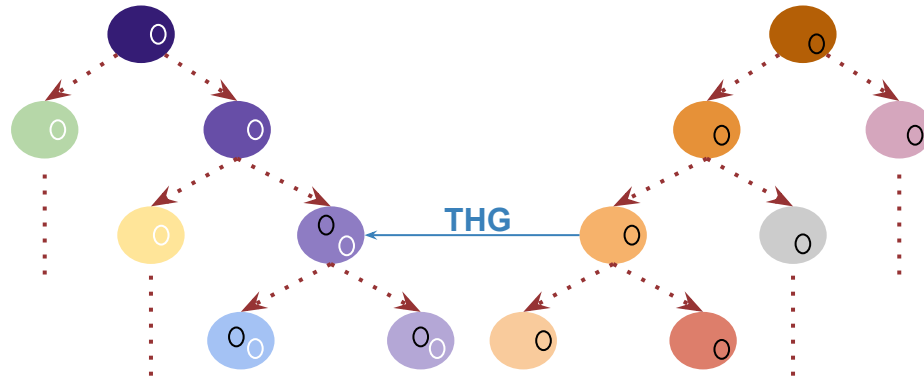
- Xenólogos

- Divergiram por **Transferência Horizontal de Genes**
  - Aquisição de novas funções



# Xenologia

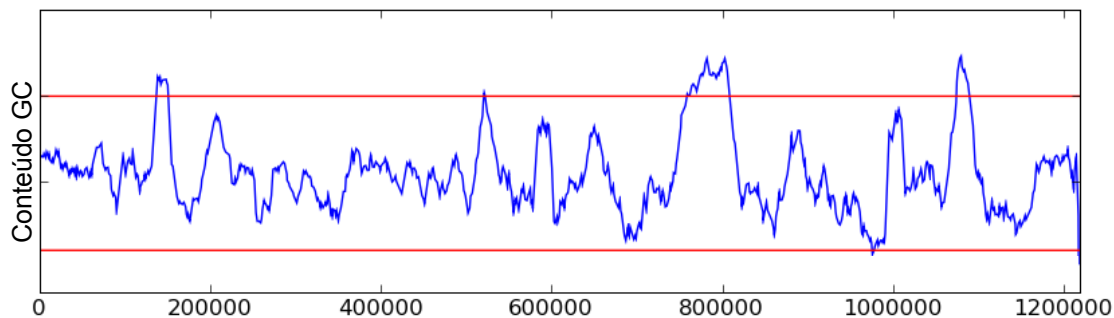
- **Xenólogos**
  - Divergiram por **Transferência Horizontal de Genes**
    - Aquisição de novas funções



# Xenologia

- Xenólogos

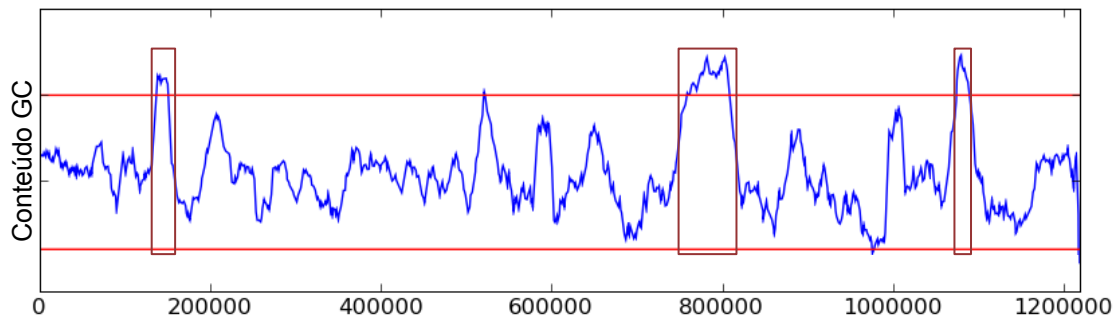
- Divergiram por **Transferência Horizontal de Genes**
  - Leva a “desvios composicionais” no genoma



# Xenologia

- Xenólogos

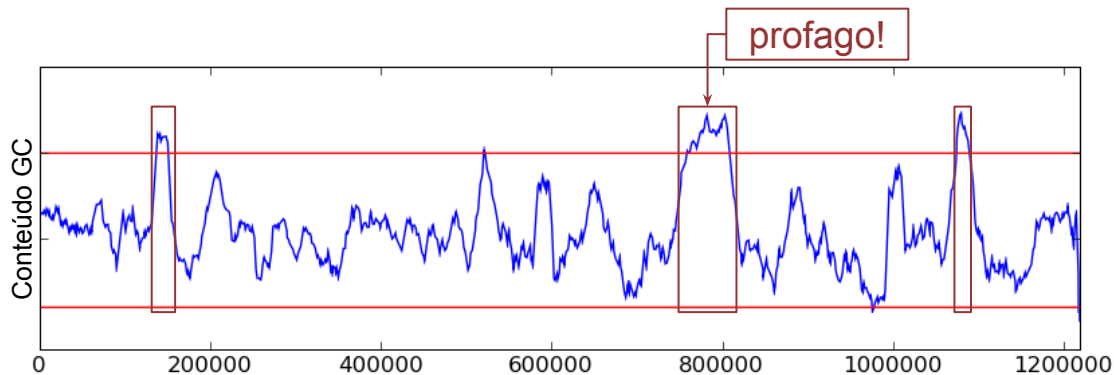
- Divergiram por **Transferência Horizontal de Genes**
  - Leva a “desvios composicionais” no genoma



# Xenologia

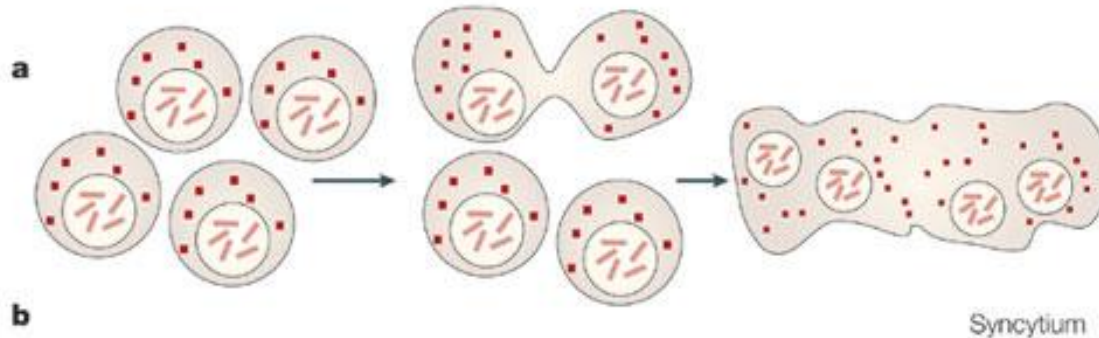
- Xenólogos

- Divergiram por **Transferência Horizontal de Genes**
  - Leva a “desvios composicionais” no genoma



# Exemplo: Xenologia

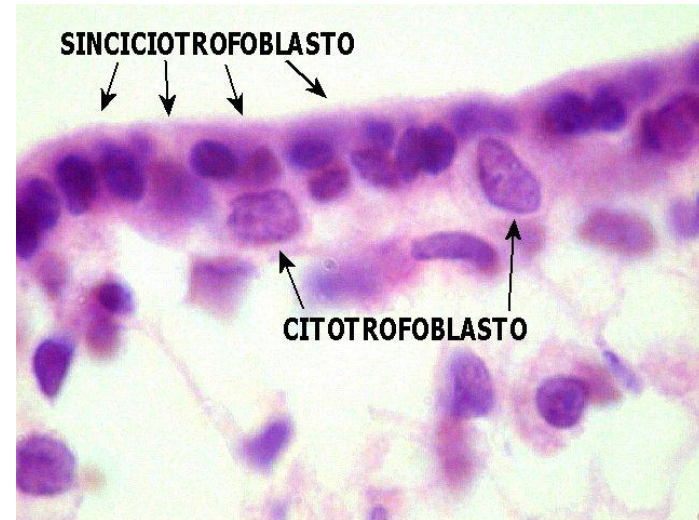
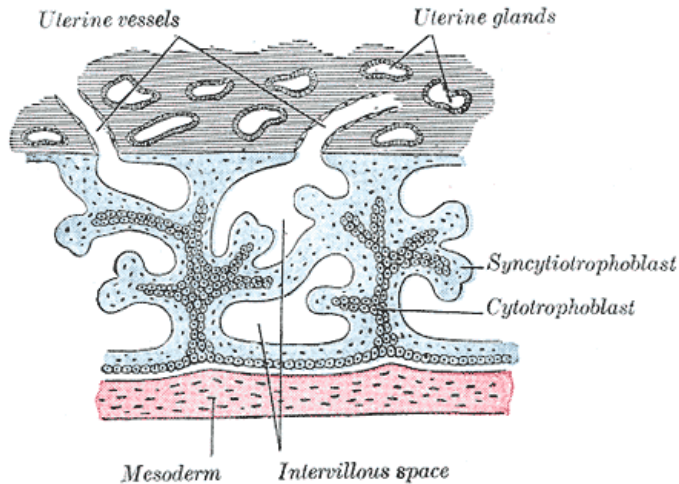
- Induzir a fusão de células é um mecanismo utilizado por vírus para expandir;
  - formação de um sincício;





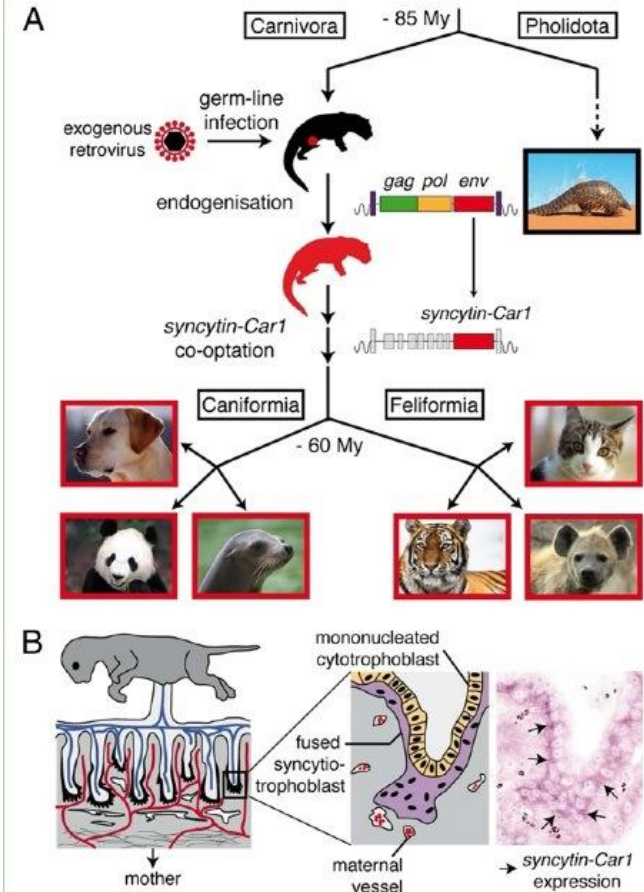
# Exemplo: Xenologia

- Muito semelhante ao sinciciotrofoblasto de mamíferos placentários!



# Exemplo: Xenologia

- Os genes que possibilitam a formação do sinciciotrofoblasto são os syncytin A e B:
  - genes de origem **VIRAL!**
  - provável infecção em um óvulo ou espermatozoide



**Intervalo?!?!?**

# Aplicações: alinhamentos globais

- Objetivo de alinhamentos globais é sobrepor “**sítios homólogos**” das sequências alinhadas:

```
AAB24882      TYHMCQFHCERYVNNHSGEKLYECNERSKAFSCPSHLQCHKRRQIGEKTHEHNQCGKAFPT 60
AAB24881      -----YECNQCCKAFAQHSSLKCHYRTHIGEKPYECNQCCKAFSK 40
                ****: .***: * *:*** * :****.:* *****..
```

```
AAB24882      PSHLQYHERTHTGEKPYECHQCGQAFKKCSLLQRHKRTHTGEKPYE-CNQCCKAFAQ- 116
AAB24881      HSHLQCHKRTHTGEKPYECNQCCKAFSQHGLLQRHKRTHTGEKPYMNVINMVKPLHNS 98
                **** *:*****:***:***.: .*****: : *.: :
```

# Aplicações: alinhamentos globais

- Objetivo de alinhamentos globais é sobrepor “**sítios homólogos**” das sequências alinhadas:
  - **lacunas**: são adicionadas quando há uma **inserção/deleção**, e portanto não há equivalente na outra sequência

```
AAB24882      TYHMCQFHC RYVNNHSGEKL YECNERSKAFSCPSHLQCHKRRQIG EKT HEHNQCGKAFPT 60
AAB24881      -----YECNQCGKAFAQHSS LKCHYRTHIG EKP YECNQCGKAFSK 40
                ****: .***: * *:** * :****.:* *****..
```

```
AAB24882      PSHLQYHERTHTG EKP YECHQCGQAFKKCSLLQRHKRTHTG EKP YECNQCGKAFAQ- 116
AAB24881      HSHLQCHKRTHTG EKP YECNQCGKAFSQHGLLQRHKRTHTG EKP YMNVINMVKPLHNS 98
                **** *:*****:***:*. : .*****: : *.:
```

# Aplicações: alinhamentos globais

- Objetivo de alinhamentos globais é sobrepor “**sítios homólogos**” das sequências alinhadas:
  - **pareamentos e não-pareamentos**: essas posições vão ser sobrepostas de acordo com a inserção das lacunas

```
AAB24882      TYHMCQFHCRCYVNNHSGEKLYECNERSKAFSCPSHLQCHKRRQIGEKTHEHNQCGKAFPT 60
AAB24881      -----YECNQCCKAFAQHSSLKCHYRTHIGEKPYECNQCCKAFSK 40
                ****: .***: * *:*** * :****.:* *****..
```

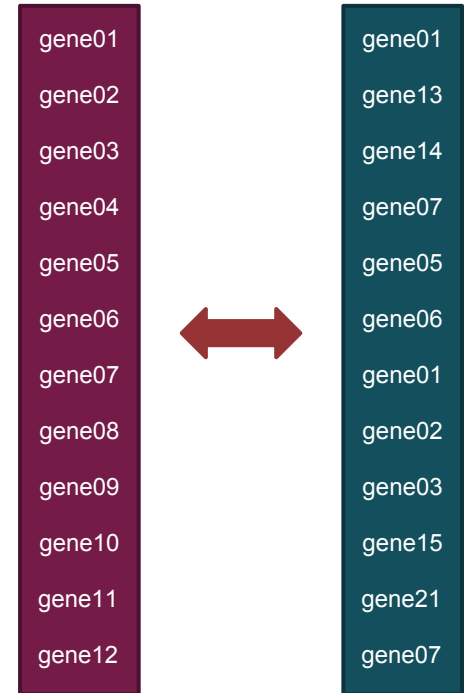
```
AAB24882      PSHLQYHERTHTGEKPYECHQCGQAFKKCSLLQRHKRTHTGEKPYE-CNQCGKAFQA- 116
AAB24881      HSHLQCHKRTHTGEKPYECNQCCKAFSQHGLLQRHKRTHTGEKPYMNVINMVKPLHNS 98
                **** *:*****:***:*. : .*****: : *.:
```

# Agrupamento de homólogos

- Agrupar genes com origens comuns;
- Geralmente a partir de **sequências protéicas**;
- Objetivos:
  - **anotação funcional**;
    - homologia  $\neq$  função
    - ortólogos e inparálogos tendem a conservar função;
    - genes com funções equivalentes;
  - **filogenia**;
  - **genômica comparativa**;

# Algoritmo de melhor hit Bidirecional:

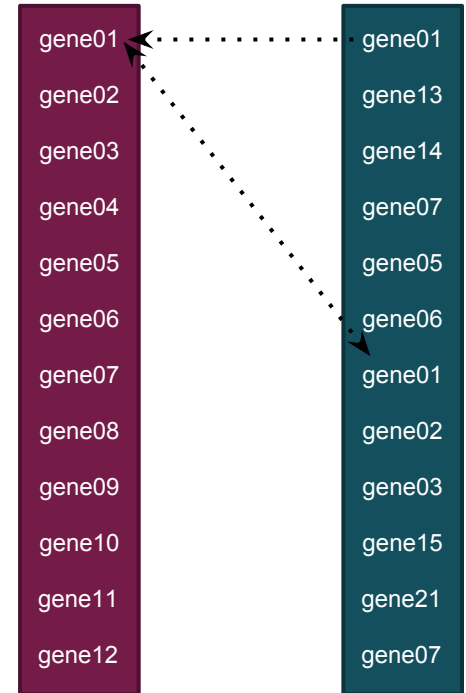
## 1. Comparações **all-vs-all** (BLAST)





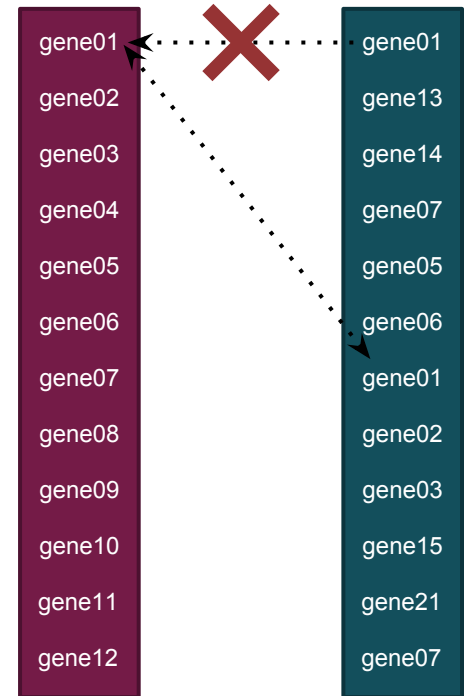
# Algoritmo de melhor hit Bidirecional:

1. Comparações **all-vs-all** (BLAST)
2. Buscar melhores hits **recíprocos**
  - a. recíproco?!?!



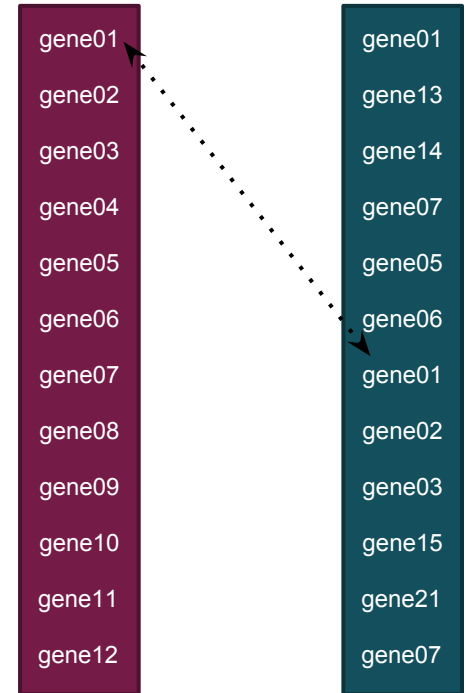
# Algoritmo de melhor hit Bidirecional:

1. Comparações **all-vs-all** (BLAST)
2. Buscar melhores hits **recíprocos**
  - a. recíproco?!?!



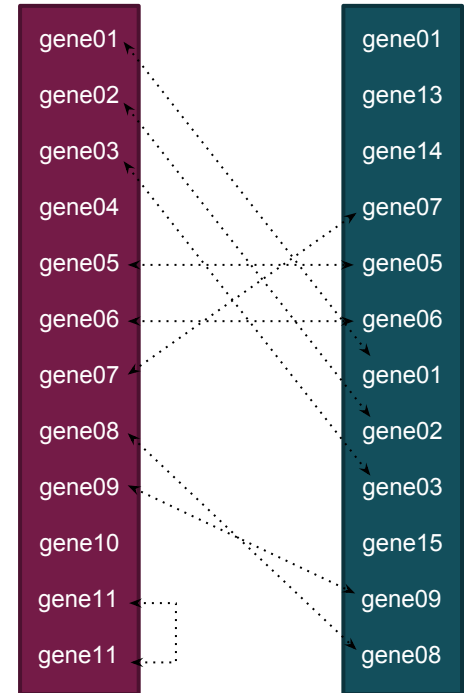
# Algoritmo de melhor hit Bidirecional:

1. Comparações **all-vs-all** (BLAST)
2. Buscar melhores hits **recíprocos**
  - a. recíproco?!?!?



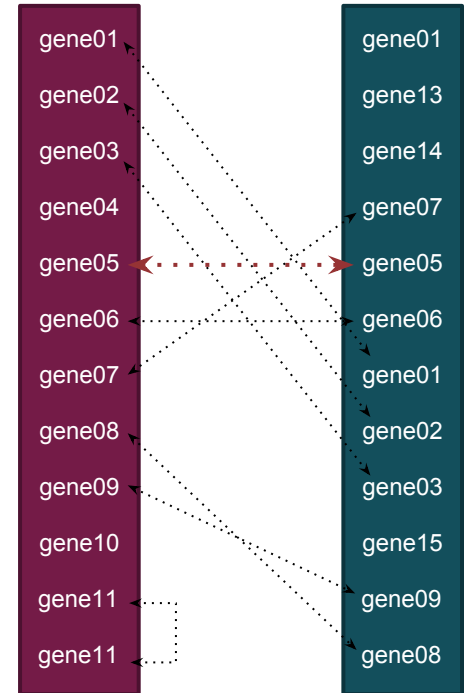
# Algoritmo de melhor hit Bidirecional:

1. Comparações **all-vs-all** (BLAST)
2. Buscar melhores hits **recíprocos**



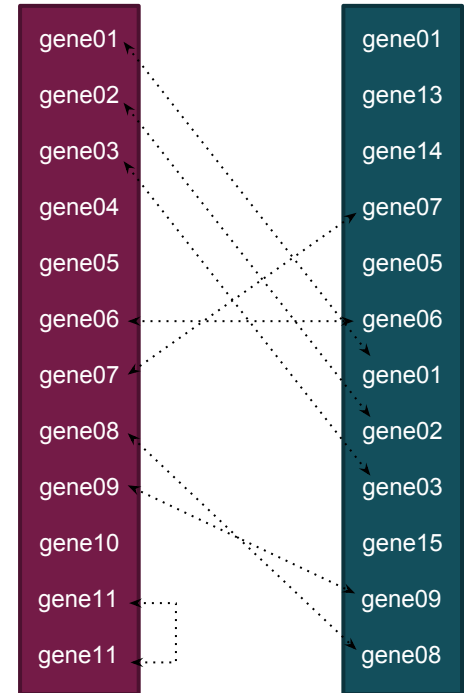
# Algoritmo de melhor hit Bidirecional:

1. Comparações **all-vs-all** (BLAST)
2. Buscar melhores hits **recíprocos**
  - a. remover hits **não significantes!**
    - i. **Identidade:** 30%
    - ii. **cobertura:** 50%
    - iii. **e-value:**  $10^{-10}$



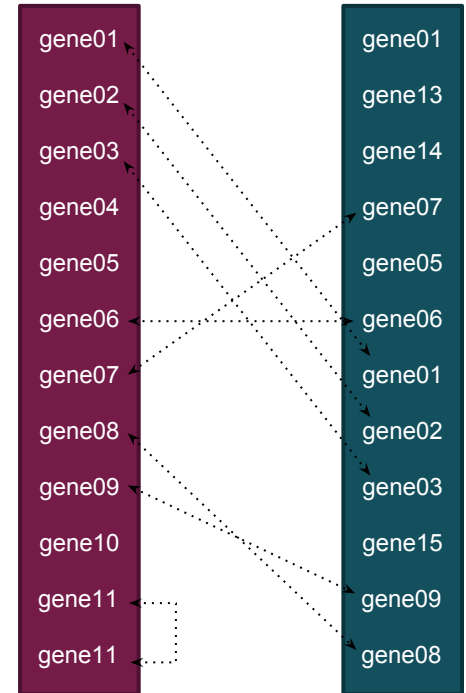
# Algoritmo de melhor hit Bidirecional:

1. Comparações **all-vs-all** (BLAST)
2. Buscar melhores hits **recíprocos**
  - a. remover hits **não significantes!**
    - i. **Identidade:** 30%
    - ii. **cobertura:** 50%
    - iii. **e-value:**  $10^{-10}$



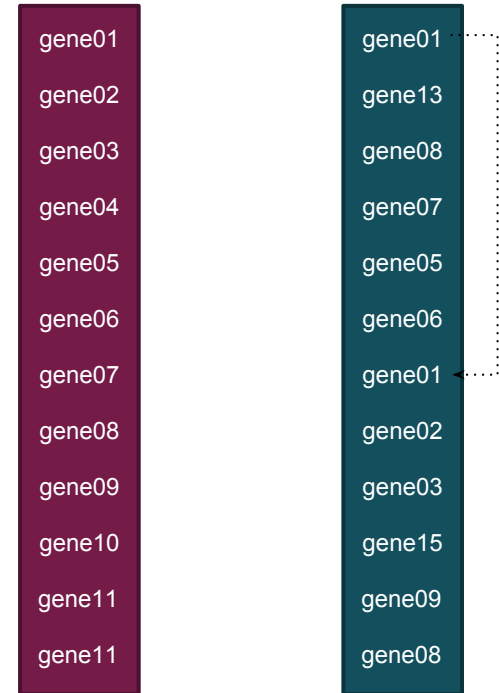
# Algoritmo de melhor hit Bidirecional:

1. Comparações **all-vs-all** (BLAST)
2. Buscar melhores hits **recíprocos**
  - a. remover hits **não significantes!**
  - b. montagem dos pares “semente”



# Algoritmo de melhor hit Bidirecional:

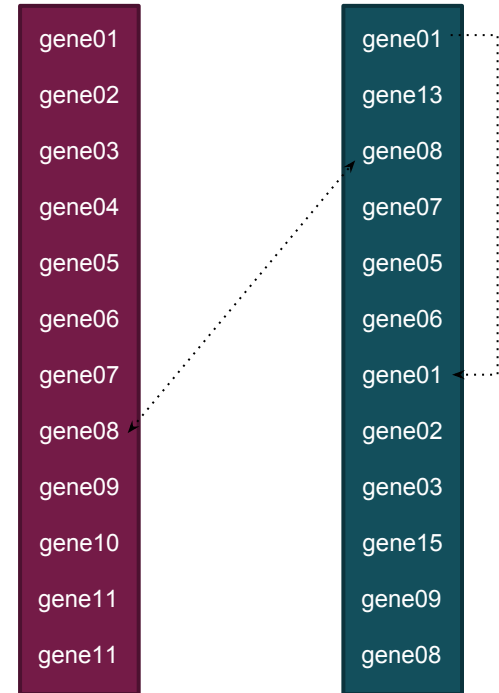
1. Comparações **all-vs-all** (BLAST)
2. Buscar melhores hits **recíprocos**
  - a. remover hits **não significantes!**
  - b. montagem dos pares “semente”
3. Adição de **Inparálogos**
  - a. hit significativo no mesmo genoma





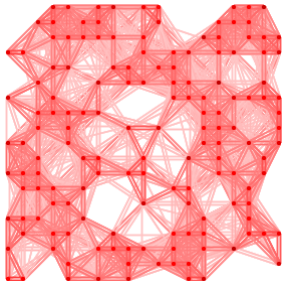
# Algoritmo de melhor hit Bidirecional:

1. Comparações **all-vs-all** (BLAST)
2. Buscar melhores hits **recíprocos**
  - a. remover hits **não significantes!**
  - b. montagem dos pares “semente”
3. Adição de **Inparálogos**
  - a. hit significativo no mesmo genoma
  - b. hit significativo no outro genoma



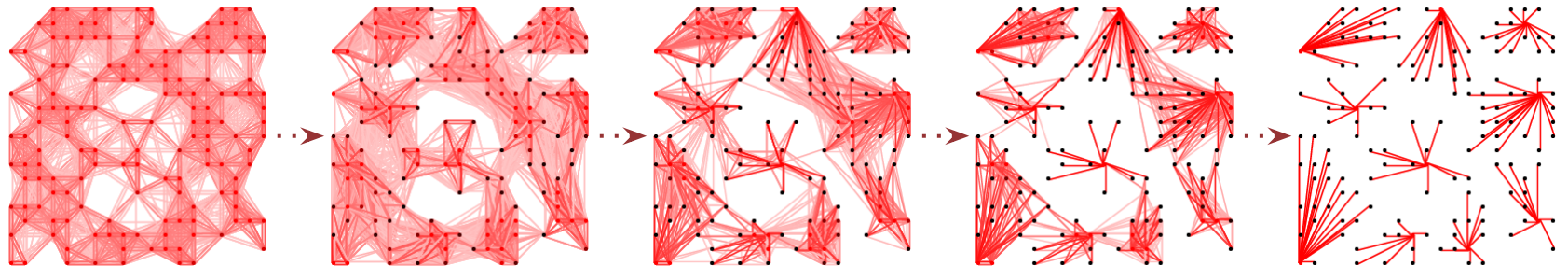
## Algoritmo de melhor hit Bidirecional: **extras**

- Algoritmo de Clusterização de Markov (MCL):
  - método de clusterização de grafos (redes);
    - refinar resultado do BBH (granularidade);
    - Parâmetros básicos: **expansão** e **inflação**



# Algoritmo de melhor hit Bidirecional: **extras**

- Algoritmo de Clusterização de Markov (MCL):
  - método de clusterização de grafos (redes);
    - refinar resultado do BBH (granularidade);
    - Parâmetros básicos: **expansão** e **inflação**

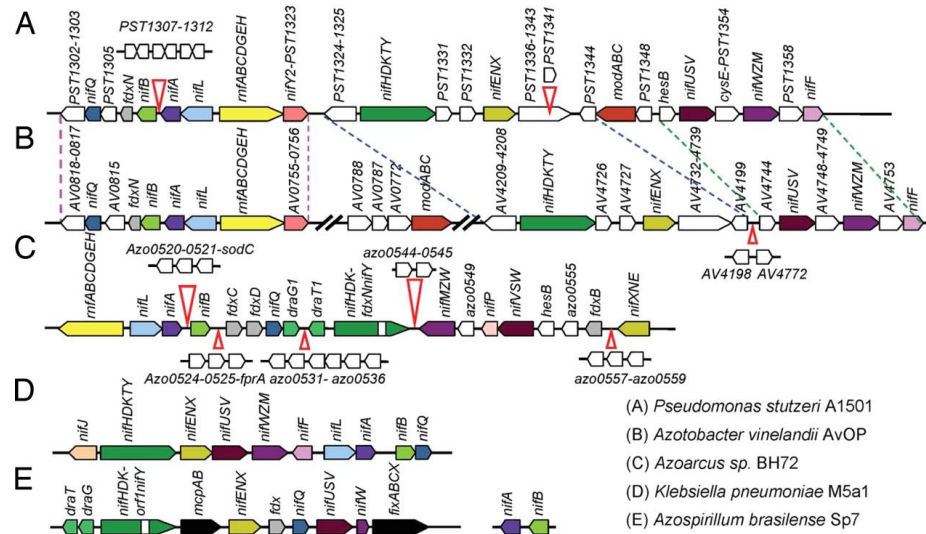


## Algoritmo de melhor hit Bidirecional: **extras**

- Sintenia:
  - **suporte** para hipótese de homologia;
  - identificar o “**verdadeiro ortologo**”;

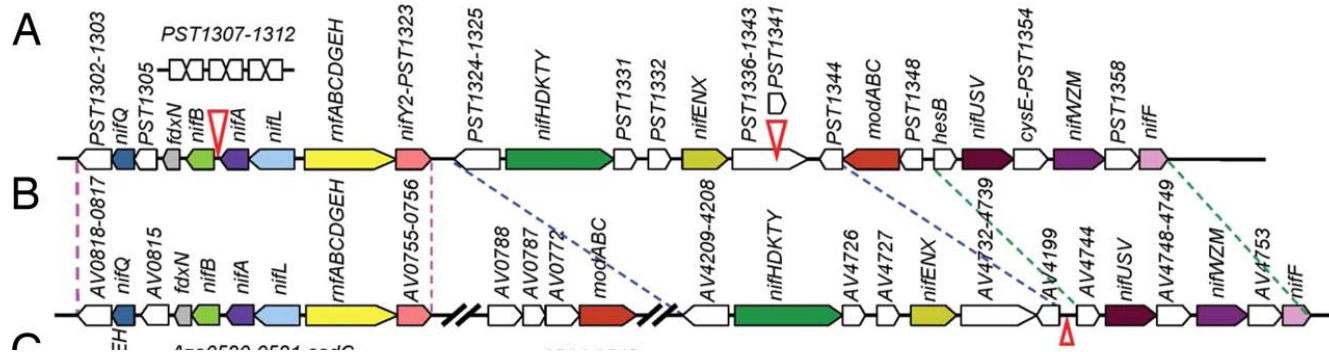
## Algoritmo de melhor hit Bidirecional: extras

- Sintenia:
  - **suporte** para hipótese de homologia;
  - identificar o “**verdadeiro ortólogo**”;



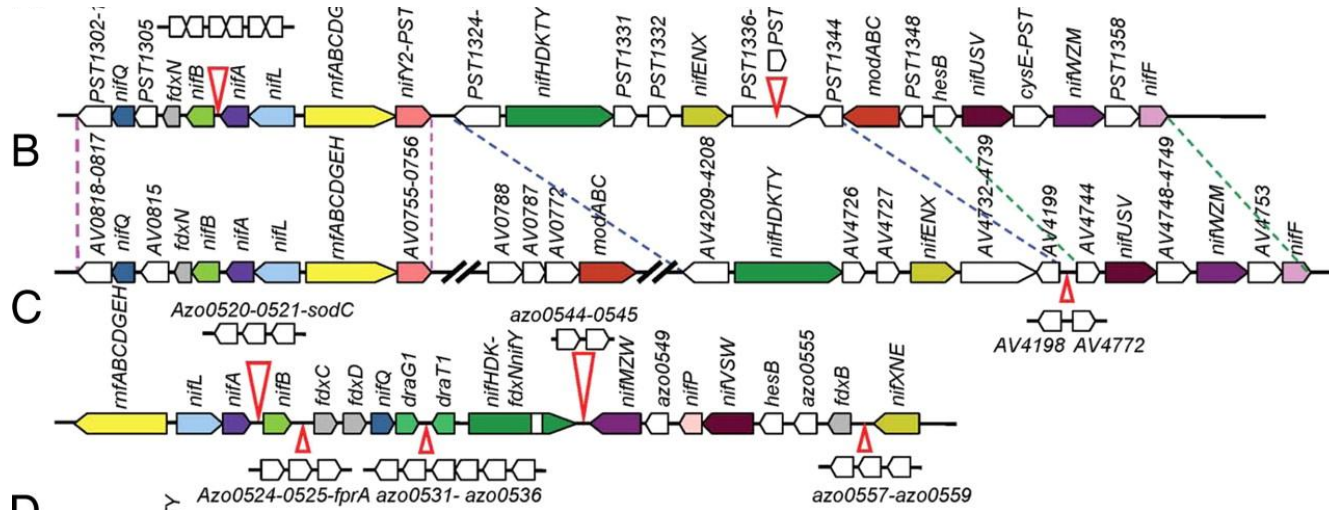
# Algoritmo de melhor hit Bidirecional: **extras**

- Sintenia:
  - **suporte** para hipótese de homologia;
  - identificar o “**verdadeiro ortólogo**”;



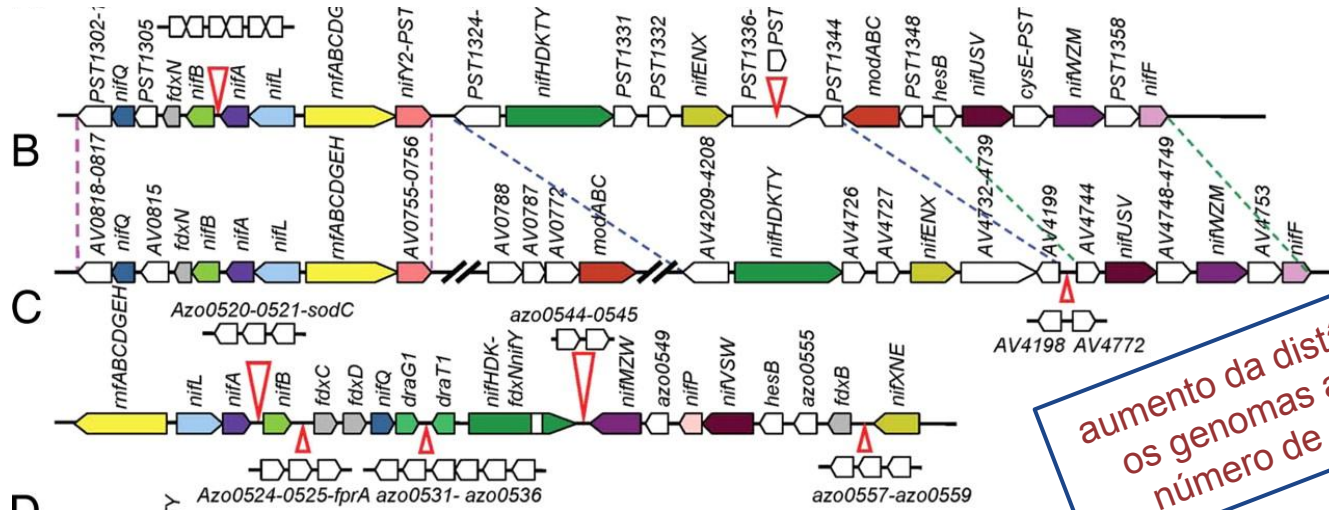
# Algoritmo de melhor hit Bidirecional: **extras**

- Sintenia:
  - **suporte** para hipótese de homologia;
  - identificar o “**verdadeiro ortólogo**”;



# Algoritmo de melhor hit Bidirecional: **extras**

- Sintenia:
  - **suporte** para hipótese de homologia;
  - identificar o “**verdadeiro ortólogo**”;



aumento da distância entre os genomas aumenta o número de rearranjos



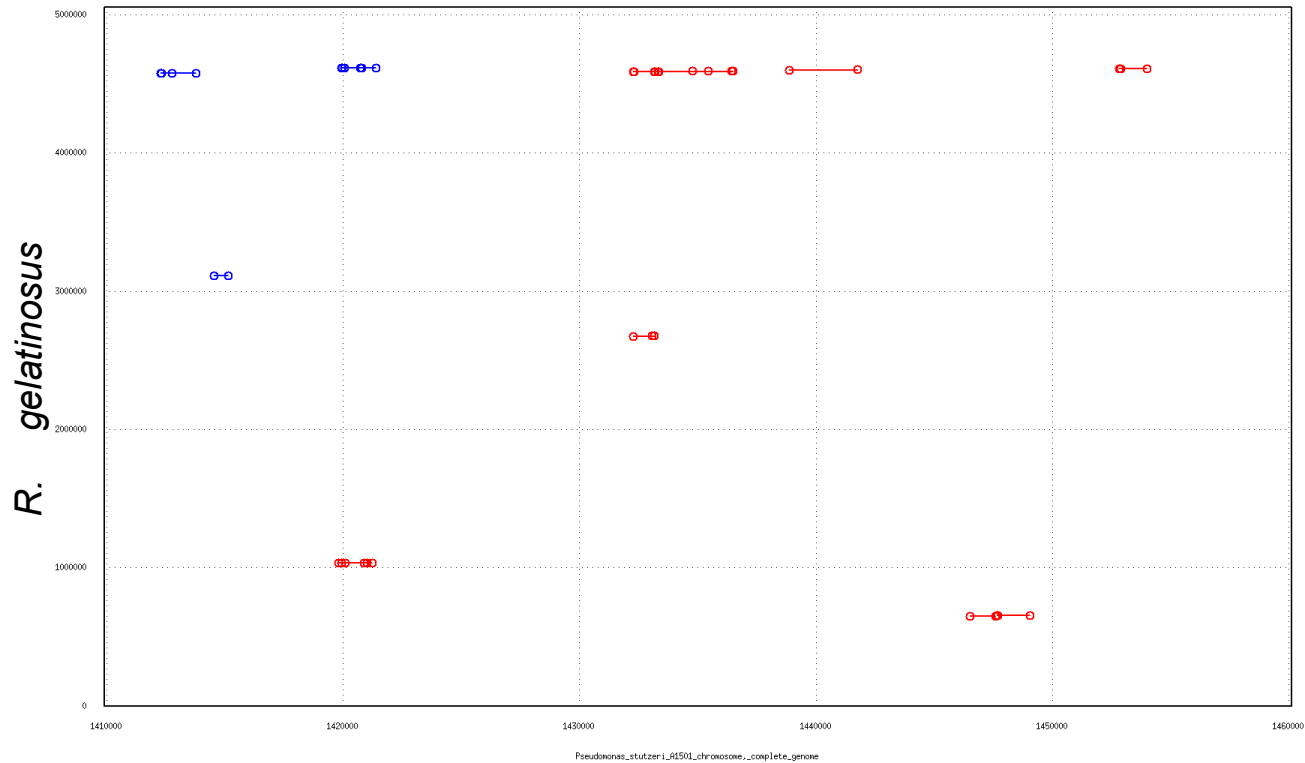


# Exemplo: sintenia

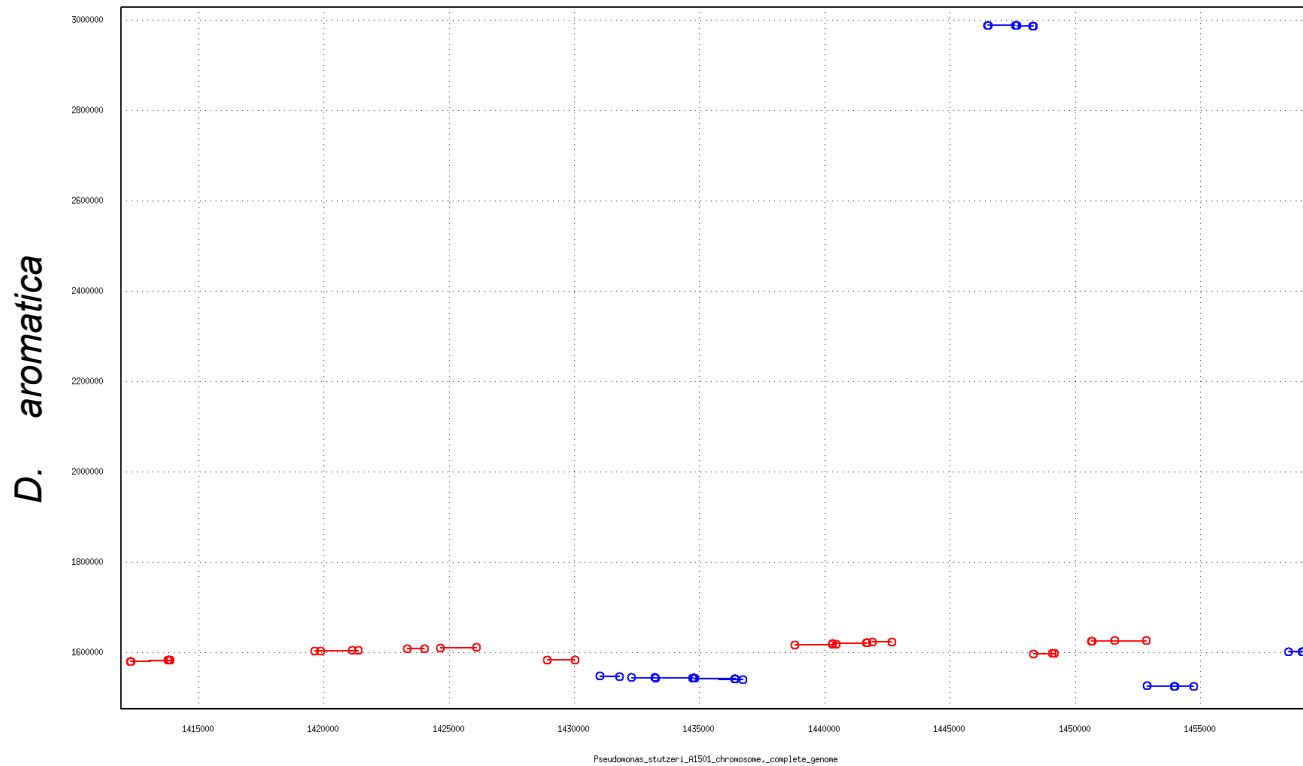
- Identificar transferência horizontal!
  - comparar com *P. stutzeri*:
    - *R. gelatinosus*
    - *D. aromatica*
  - apresentam sintenia?



# Exemplo: sintenia

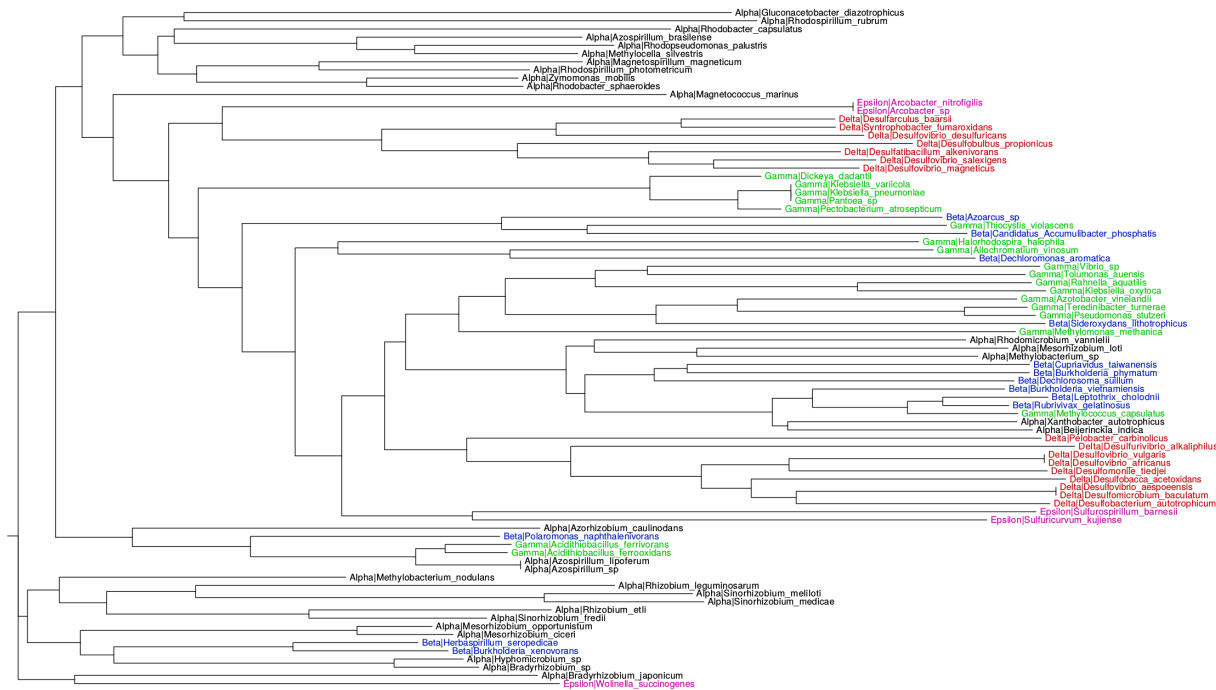


# Exemplo: sintenia



# Exemplo: sintenia

- Diferentes abordagens para avaliar



# Aplicações: agrupamento de proteínas

- Exemplos de ferramentas
  - InParanoid/Multiparanoid
    - apenas BBH;
    - 2 processos separados:
      - formação de pares;
      - sobreposição de pares de genomas;

# Aplicações: agrupamento de proteínas

- Exemplos de ferramentas
  - orthoMCL
    - BBH e ajuste com o MCL;
    - suporta maior quantidade de genomas;
    - mais usado;

# Aplicações: bases de dados

- Se você quer apenas saber o homólogo de um gene específico, o processo é bem mais simples!
  - comparar sua sequência com um conjunto já agrupado;
    - diversas BDs com focos diferentes;



# Aplicações: bases de dados

- COG:
  - 1º BD de grupos homólogos;
  - considerado por muitos o “**gold standard**”;
    - curados manualmente;
    - poucos genomas;
  - desativado em 2012

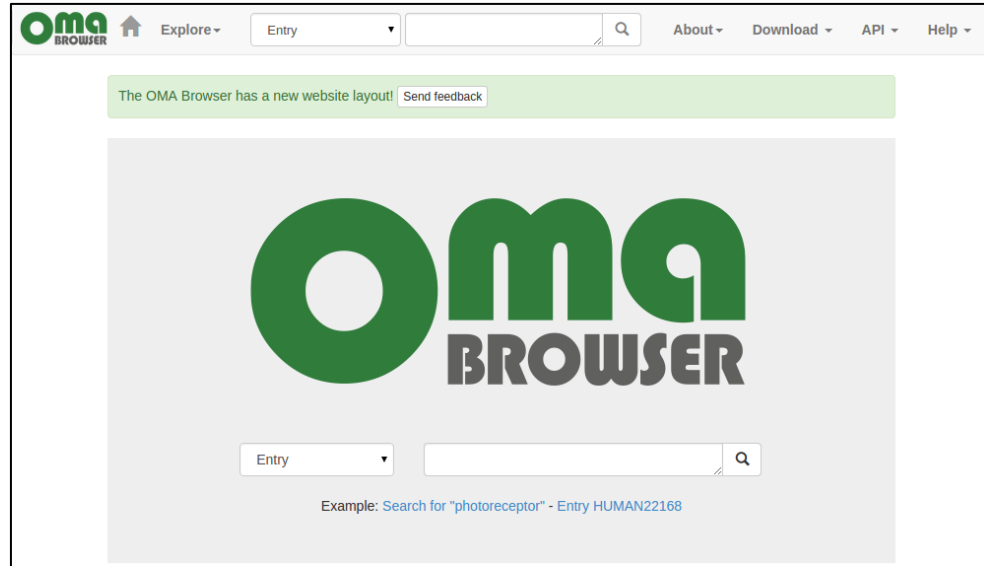
# Aplicações: bases de dados

- eggNOG
  - **baseado no COG;**
  - o mais abrangente;
  - baseado em funções
  - diversas formas de busca;
  - grupos hierárquicos;



# Aplicações: bases de dados

- OMA
  - diversas formas de busca;
  - baseado em funções
  - grupos hierárquicos;



# Aplicações: filogen(ia)ômica)

1. Agrupamento de hómologos
2. Genes comuns às espécies de interesse
3. Descartar grupos contendo parálogos
  - a. porque?
4. Alinhamento e árvore