

FAST TREE

Alfonso Delgado
Miguel Ángel Cosculluela
Miguel Crespo

Instalación

- Windows y Linux: Descargar ejecutable
- Mac OS X: Descargar fichero fuente [FastTree.c](#) y compilarlo con
gcc -O3 -finline-functions -funroll-loops -Wall -o FastTree FastTree.c -lm
necesario tener xcode

Ejecución

Forma de ejecutarlo:

- `FastTree -nt alignment_file > tree_file`

Parámetros diferentes

- `-nt` -> Si son nucleótidos
- `-quote` -> Para obtener el nombre entero de cada secuencia
- `-gamma` -> Reescala las ramas del árbol y calcula la probabilidad gamma20
- `-quiet` -> Silencia los mensajes de salida

Métodos para inferir el árbol:

- Por defecto Jukes-Cantor + CAT
- `-gtr` -> GTR+CAT (Generalized time-reversible model, solo nucleótidos)
- `-wag` -> WAG+CAT (Modelo Whelan-And-Goldman 2001, solo proteínas)
- `-lg` -> LG+CAT (Modelo Le-Gascuel 2008, solo proteínas)

Modelos

- Nucleótidos
 - Generalized time-reversible (**GTR**)
- Aminoácidos
 - Jones-Taylor-Thornton 1992 (**JTT**)
 - Whelan & Goldman 2001 (**WAG**)
 - Le and Gascuel 2008 (**LG**)

GTR : https://en.wikipedia.org/wiki/Substitution_model#GTR: Generalised time reversible

JTT : <https://www.ncbi.nlm.nih.gov/pubmed/1633570>

WAG : <http://www.ncbi.nlm.nih.gov/pubmed/11319253>

LG : <http://mbe.oxfordjournals.org/content/25/7/1307.short>

Fases del proceso

1. Heurística de fusión de vecinos
2. Reducción de la longitud del árbol
3. Maximizar la probabilidad del árbol
4. Estimación de la confianza de cada rama

Fases del proceso

1. **Heurística de fusión de vecinos**
2. Reducción de la longitud del árbol
3. Maximizar la probabilidad del árbol
4. Estimación de la confianza de cada rama

Utiliza tres heurísticas:

1. Recuerda la mejor unión entre nodos
2. De una unión candidata realiza una optimización Hill-climbing
3. Descarta casos a comprobar mediante *“Top hits”*

Fases del proceso

1. Heurística de fusión de vecinos
2. **Reducción de la longitud del árbol**
3. Maximizar la probabilidad del árbol
4. Estimación de la confianza de cada rama

Utiliza una mezcla de dos estrategias:

- nearest-neighbor interchanges (**NNIs**)
 - Trabaja sobre árboles sin raíz, generando una relación e intercambiando sus sub-árboles
- subtree-prune-regraft moves (**SPRs**)
 - Selecciona un sub-árbol del principal y lo introduce en otro sitio generando un nuevo nodo

Fases del proceso

1. Heurística de fusión de vecinos
2. Reducción de la longitud del árbol
3. **Maximizar la probabilidad del árbol**
4. Estimación de la confianza de cada rama

- Utiliza uno de los modelos anteriores
- Clasifica posiciones en categorías
 - Rango de 0.05 a 20
 - Divisiones alejadas geoméricamente
- Clasificación mediante Bayes con prioridad *Gamma*

Fases del proceso

1. Heurística de fusión de vecinos
2. Reducción de la longitud del árbol
3. Maximizar la probabilidad del árbol
4. **Estimación de la confianza de cada rama**

- Estimación rápida de la confianza
- Aplica la prueba **Shimodaira-Hasegawa**
 - Dados $(A,B),(C,D)$
 - Usa la prueba para compararlos
 - Y alterar la topología a $(A,C),(B,D)$ o $(A,D),(B,C)$

Resultados

- Árboles en formato Newick.
- Confianza de cada rama entre 0 y 1.
- Estructura del árbol:
 - Raíz 3 hijos.
 - Nodos internos 2 hijos
- Secuencias iguales => nodo internos más de 2 hijos



Resultados

-Muestra las opciones y procesos del árbol a calcular.

-Se incluyen tiempos de procesos.

-Muestra caracteres no reconocidos.

-Cadenas únicas.

```
C:\Users\Hifonso\Downloads\F practica4>Fasttree.exe -nt alineamientoFelinos.fasta > arbolFelinosFI.nwk
Fasttree Version 2.1.10 No SSE3
Alignment: alineamientoFelinos.fasta
Nucleotide distances: Jukes-Cantor Joins: balanced Support: SH-like 1000
Search: Normal +NNI +SPR (2 rounds range 10) +ML-NNI opt-each=1
TopHits: 1.00*sqrtN close=default refresh=0.80
ML Model: Jukes-Cantor, CAT approximation with 20 rate categories
Ignored unknown character m (seen 1 times)
Ignored unknown character r (seen 2 times)
Initial topology in 0.29 seconds
Refining topology: 21 rounds ME-NNIs, 2 rounds ME-SPRs, 11 rounds ML-NNIs
Total branch-length 1.698 after 3.41 sec1, 1 of 38 splits
ML-NNI round 1: LogLk = -188496.756 NNIs 3 max delta 34.50 Time 5.27
Switched to using 20 rate categories (CAT approximation)16 of 20
Rate categories were divided by 0.825 so that average rate = 1.0
CAT-based log-likelihoods may not be comparable across runs
Use -gamma for approximate but comparable Gamma(20) log-likelihoods
ML-NNI round 2: LogLk = -167540.294 NNIs 0 max delta 0.00 Time 6.05
Turning off heuristics for final round of ML NNIs (converged)
ML-NNI round 3: LogLk = -167499.665 NNIs 0 max delta 0.00 Time 7.39 (final)
Optimize all lengths: LogLk = -167499.650 Time 7.81
Total time: 11.88 seconds Unique: 40/40 Bad splits: 0/37
```