FastTree

Guillermo Reloba

Sandra Navarro

Instalación

- Se accede a: http://www.microbesonline.org/fasttree/#Install
- Windows y Linux: descargar ejecutable.
- Mac OS X: descargar fichero fuente FastTree.c (es necesario tener instalado xcode.gcc) y compilar con

gcc -03 –finline-functions –funroll-loops –Wall –o FastTree FastTree.c -lm

FastTree

Also see Fast Tree-Comparison Tools

FastTree infers approximately-maximum-likelihood phylogenetic trees from alignments of nucleotide or protein sequences. FastTree can handle alignments with up to a million of sequences in a reasonable amount of time and memory. For large alignments, FastTree is 100-1,000 times faster than PhyML 3.0 or RAxML 7. FastTree is open-source software -- you can download the code below.

FastTree is more accurate than PhyML 3 with default settings, and much more accurate than the distance-matrix methods that are traditionally used for large alignments. FastTree uses the Jukes-Cantor or generalized time-reversible (GTR) models of nucleotide evolution and the JTT (Jones-Taylor-Thornton 1992), WAG (Whelan & Goldman 2001), or LG (Le and Gascuel 2008) models of amino acid evolution. To account for the varying rates of evolution across sites, FastTree uses a single rate for each site (the "CAT" approximation). To quickly estimate the reliability of each split in the tree, FastTree computes local support values with the Shimodaira-Hasegawa test (these are the same as PhyML 3's "SH-like local supports").

More information:

Downloading and Installing FastTree

FastTree 2.1.10 is available as

- Linux 64-bit executable (+SSE)
  - Double-precision executable for nearly-identical sequences (see explanation)
  - Multi-threaded executable (+SSE +OpenMP) (see usage guide)
- Windows 32-bit command-line executable (no SSE)
- C code

If you use a Mac or other platform not included above, download FastTree.c and run

```
gcc -O3 -finline-functions -funroll-loops -Wall -o FastTree FastTree.c -lm
```

Ejecución

- Se ejecuta con: FastTree [parámetros] archivo alineado > archivo de salida
- Acepta ficheros en formato FASTA o PHYLIP.
- Parámetros:
  - -**nt**: para nucleótidos.
  - -**quiet**: suprime mensajes de salida.
  - -**log**: guarda árboles intermedios y detalles del modelo.
  - -**n <number>:** analiza múltiples alineamientos (solo formato PHYLIP).
  - -**intrees in_tree**: establece el árbol de partida.
  - -**quote**: obtiene número entero de cada secuencia.
  - -**gamma**: reescala ramas del árbol y calcula la probabilidad gamma20.
  - -**expert**: muestra más información.

```
Usage for FastTree version 2.1.10 No SSE3:
  FastTree protein_alignment > tree
  FastTree < protein_alignment > tree
  FastTree -out tree protein_alignment
  FastTree -nt nucleotide_alignment > tree
  FastTree -nt -gtr < nucleotide_alignment > tree
  FastTree < nucleotide_alignment > tree
FastTree accepts alignments in fasta or phylip interleaved formats

Common options (must be before the alignment file):
  -quiet to suppress reporting information
  -nopr to suppress progress indicator
  -log logfile -- save intermediate trees, settings, and model details
  -fastest -- speed up the neighbor joining phase & reduce memory usage
       (recommended for >50,000 sequences)
  -n <number> to analyze multiple alignments (phylip format only)
       (use for global bootstrap, with seqboot and CompareToBootstrap.pl)
  -nosupport to not compute support values
  -intree newick_file to set the starting tree(s)
  -intree1 newick_file to use this starting tree for all the alignments
       (for faster global bootstrap on huge alignments)
  -pseudo to use pseudocounts (recommended for highly gapped sequences)
  -gtr -- generalized time-reversible model (nucleotide alignments only)
  -lg -- Le-Gascuel 2008 model (amino acid alignments only)
  -wag -- Whelan-And-Goldman 2001 model (amino acid alignments only)
  -quote -- allow spaces and other restricted characters (but not ' ) in
         sequence names and quote names in the output tree (fasta input only;
         FastTree will not be able to read these trees back in)
  -noml to turn off maximum-likelihood
  -nome to turn off minimum-evolution NNIs and SPRs
       (recommended if running additional ML NNIs with -intree)
  -nome -mllen with -intree to optimize branch lengths for a fixed topology
  -cat # to specify the number of rate categories of sites (default 20)
     or -nocat to use constant rates
  -gamma -- after optimizing the tree under the CAT approximation,
     rescale the lengths to optimize the Gamma20 likelihood
  -constraints constraintAlignment to constrain the topology search
      constraintAlignment should have 1s or 0s to indicates splits
  -expert -- see more options
For more information, see http://www.microbesonline.org/fasttree/
Windows users: Please remember to run this inside a command shell
Hit return to continue
```

Modelos evolutivos

- Nucleótidos
  - Jukes-Cantor
  - GTR: usando -gtr
- Aminoácidos:
  - JTT
  - WAG: usando -wag
  - LG: usando -lg

Resultados

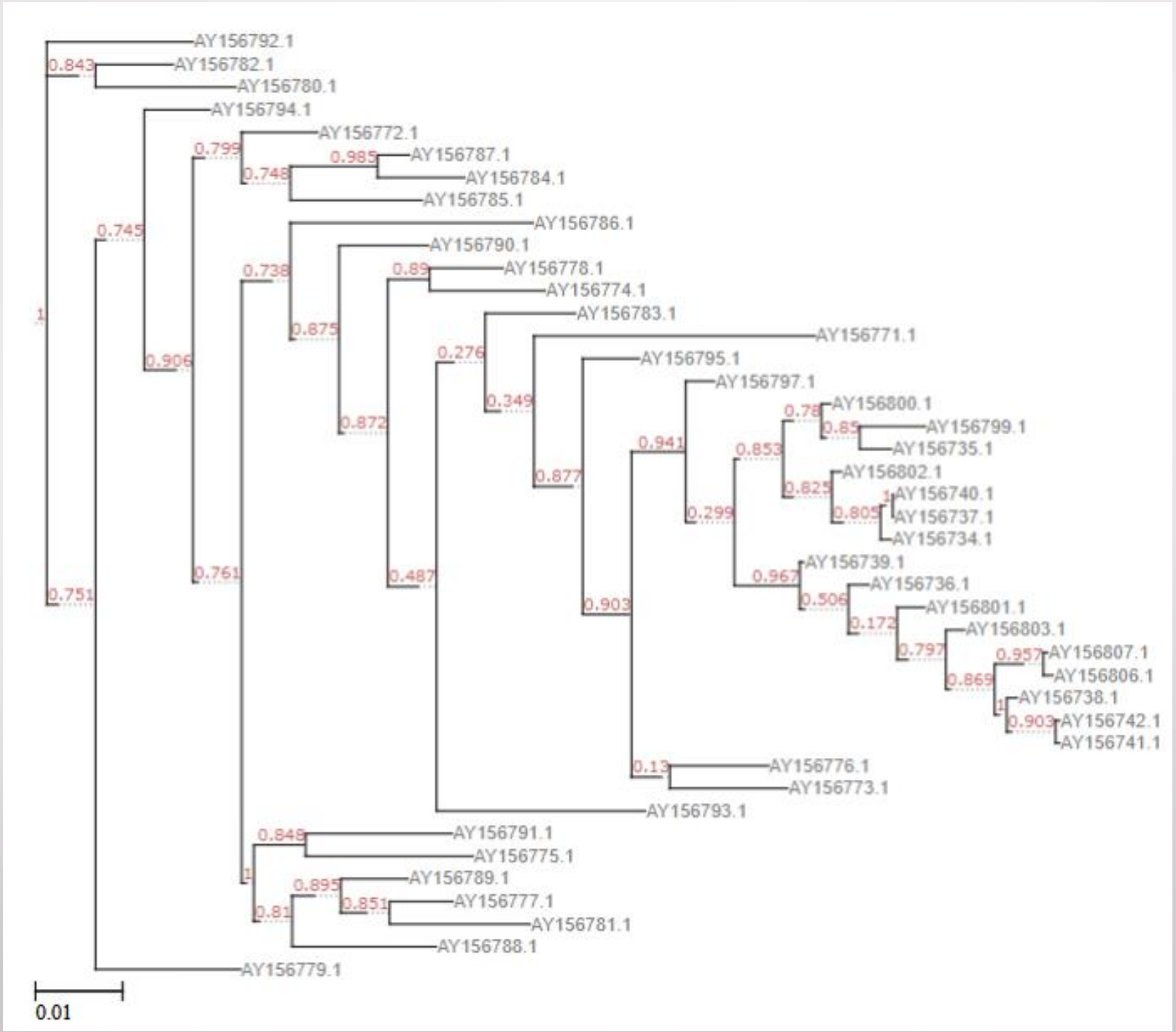
- Ignora las letras desconocidas e informa
- Muestra el número de cadenas únicas
- Muestra el tiempo consumido

- Muestra la plausibilidad (likelihood) de cada rama que calcula. Para ser comparada es necesario usar la opción -gamma para que la plausibilidad que devuelve sea Gamma20.
- El resultado es un árbol en formato Netwick
- La raíz del árbol es asignada arbitrariamente

Ejemplo

```
pendiente@DESKTOP-OQOS191:/mnt/c/Users/Guillermo/Documentos/Trabajo/Bioinformatica/P5$ ./FastTree -nt alineado.fasta>t.nwk
FastTree Version 2.1.10 SSE3
Alignment: alineado.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 k (seen 12 times)
Ignored unknown character m (seen 12 times)
Ignored unknown character r (seen 85 times)
Ignored unknown character s (seen 2 times)
Ignored unknown character w (seen 14 times)
Ignored unknown character y (seen 54 times)
Initial topology in 0.03 seconds
Refining topology: 21 rounds ME-NNIs, 2 rounds ME-SPRs, 11 rounds ML-NNIs
Total branch-length 0.488 after 0.31 sec1, 1 of 39 splits
ML-NNI round 1: LogLk = -3496.923 NNIs 9 max delta 4.31 Time 0.53
Switched to using 20 rate categories (CAT approximation)1 of 20
Rate categories were divided by 0.748 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 = -3283.944 NNIs 5 max delta 0.45 Time 0.75
ML-NNI round 3: LogLk = -3283.909 NNIs 0 max delta 0.00 Time 0.86
Turning off heuristics for final round of ML NNIs (converged)
ML-NNI round 4: LogLk = -3283.855 NNIs 1 max delta 0.05 Time 1.02 (final)
Optimize all lengths: LogLk = -3283.855 Time 1.07
Total time: 1.30 seconds Unique: 41/42 Bad splits: 0/38
pendiente@DESKTOP-OQOS191:/mnt/c/Users/Guillermo/Documentos/Trabajo/Bioinformatica/P5$
```

Salida


AY156792.1
AY156782.1
AY156780.1
AY156794.1
AY156772.1
AY156787.1
AY156784.1
AY156785.1
AY156786.1
AY156790.1
AY156778.1
AY156774.1
AY156783.1
AY156771.1
AY156795.1
AY156797.1
AY156800.1
AY156799.1
AY156735.1
AY156802.1
AY156740.1
AY156737.1
AY156734.1
AY156739.1
AY156736.1
AY156801.1
AY156803.1
AY156807.1
AY156806.1
AY156738.1
AY156742.1
AY156741.1
AY156776.1
AY156773.1
AY156793.1
AY156791.1
AY156775.1
AY156789.1
AY156777.1
AY156781.1
AY156788.1
AY156779.1
0.01