

Bioinformática - Práctica 5

Barbed Pérez, Óscar León (666137)

Clemente Villafranca, Alba (652204)

R_AxML

Contenidos

- Introducción
- Instalación
- Modelos evolutivos
- Resultados



Contenidos

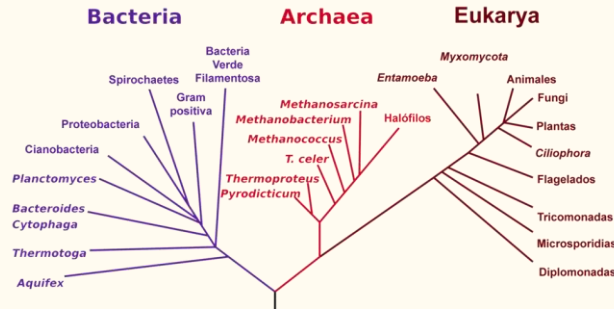
- Introducción
- Instalación
- Modelos evolutivos
- Resultados



Introducción

- Herramienta creada por The Exelixis Lab
- Crear árboles filogenéticos basados en máxima verosimilitud.
- Método basado en secuencias tal que pretende encontrar el valor de theta tal que $\Pr(x | \theta)$ se maximiza, siendo x los datos observados.

Árbol Filogenético de la Vida



Contenidos

- Introducción
- Instalación
- Modelos evolutivos
- Resultados



Instalación

- Repositorio de GitHub: <https://github.com/stamatak/standard-RAxML>
- Para Linux, compilar
 - Instrucciones de compilación: standard-RAxML/README
- Para Windows, utilizar los ejecutables precompilados
 - Ejecutables: standard-RAxML\WindowsExecutables_v8.2.10\
- Hay diferentes variantes (optimizaciones) del programa
 - Estándar: raxmlHPC
 - Multithread: raxmlHPC-PTHREADS
 - Para procesadores de hace 6 ó 7 años: [...] -SSE3
 - Para procesadores de hace 4 ó 5 años: [...] -AVX y [...] -AVX2

Instalación

- Online:
 - <https://embnet.vital-it.ch/raxml-bb/>
 - <http://www.trex.uqam.ca/index.php?action=raxml> ---> Solo admite formato PHYLIP. Puede visualizar el árbol.

Conversor a PHYLIP:

http://sequenceconversion.bugaco.com/converter/biology/sequences/fasta_to_phylip.php

Contenidos

- Introducción
- Instalación
- Modelos evolutivos
- Resultados



Modelos evolutivos según el tipo de secuencia:

```
For BINARY data use:  BINCAT[X]                or BINGAMMA[X]                or
                     BINCATI[X]               or BINGAMMAI[X]               or
                     ASC_BINGAMMA[X]          or ASC_BINCAT[X]
For DNA data use:    GTRCAT[X]                or GTRGAMMA[X]                or
                     GTRCATI[X]              or GTRGAMMAI[X]              or
                     ASC_GTRGAMMA[X]         or ASC_GTRCAT[X]
For Multi-state data: MULTICAT[X]            or MULTIGAMMA[X]            or
                     MULTICATI[X]           or MULTIGAMMAI[X]           or
                     ASC_MULTIGAMMA[X]      or ASC_MULTICAT[X]
For AA data use:     PROTCATmatrixName[F|X] or PROTGAMMAmatrixName[F|X] or
                     PROTCATImatrixName[F|X] or PROTGAMMAImatrixName[F|X] or
                     ASC_PROTGAMMAmatrixName[X] or ASC_PROTCATmatrixName[X]
The AA substitution matrix can be one of the following:
DAYHOFF, DCMUT, JTT, MTREV, WAG, RTREV, CPREV, VT,
BLOSUM62, MTMAM, LG, MTART, MTZOA, PMB, HIVB, HIVW,
JTTDCMUT, FLU, STMTREV, DUMMY, DUMMY2, AUTO, LG4M, LG4X,
PROT_FILE, GTR_UNLINKED, GTR
With the optional "F" appendix you can specify if you want to use empirical base frequencies.
With the optional "X" appendix you can specify that you want to do a ML estimate of base frequencies.
```

Hablaremos sólo de los de nucleótidos.

Modelos evolutivos

Modelos:

- GAMMA: más costoso en memoria y tiempo
- CAT: $\frac{1}{4}$ del coste computacional de gamma

CAT es recomendable para aproximaciones en modelos evolutivos de 50 secuencias o más. En muestras más pequeñas no hay suficientes datos para calcular una predicción de manera fiable.

Paper de un experimento que compara el uso de ambos en RAxML:

<http://w.hicomb.org/papers/HICOMB2006-05.pdf>

Modelos evolutivos

Variaciones de los modelos:

- GAMMA:
 - GTRGAMMA
 - GTRGAMMAI: estima la proporción de porciones invariables, en vez de utilizar las de por defecto.
 - ASC_GTRGAMMA: corrige el sesgo producido en el proceso de evaluación. Necesario especificar la corrección con el parámetro `--asc-corr` (stamatakis|felsenstein|lewis).
- CAT:
 - GTRCAT
 - GTRCATI: el árbol final es evaluado con GTRGAMMAI, en vez de GTRGAMMA.
 - ASC_GTRCAT: corrige el sesgo producido en el proceso de evaluación. Necesario especificar la corrección con el parámetro `--asc-corr` (stamatakis|felsenstein|lewis).

Modelos evolutivos

Parámetros importantes:

- -m: modelo evolutivo
- -n: nombre del resultado
- -f: algoritmo de búsqueda. Por defecto Hill-Climbing.
- -s: fichero de alineamiento
- -p: semilla para procesos que precisen de aleatoriedad

Modelos evolutivos

Otros parámetros:

Consultar el manual ---> raxmlHPC -h

```
raxmlHPC      -s sequenceFileName -n outputFileName -m substitutionModel
               [-a weightFileName] [-A secondaryStructureSubstModel]
               [-b bootstrapRandomNumberSeed] [-B wcCriterionThreshold]
               [-c numberOfCategories] [-C] [-d] [-D]
               [-e likelihoodEpsilon] [-E excludeFileName]
               [-f a|A|b|B|c|C|d|D|e|E|F|G|H|h|I|i|J|K|k|m|n|N|o|p|q|r|R|s|S|t|T|u|
               v|V|w|W|x|Y]
               [-F]
               [-g groupingFileName] [-G placementThreshold] [-h] [-H]
               [-i initialRearrangementSetting]
               [-I autoFC|autoMR|autoMRE|autoMRE_IGN]
               [-j] [-J MR|MR_DROP|MRE|STRICT|STRICT_DROP|T_<PERCENT>] [-k] [-K]
               [-L MR|MRE|T_<PERCENT>] [-M]
               [-o outGroupName1[,outGroupName2[,...]]] [-O]
               [-p parsimonyRandomSeed] [-P proteinModel]
               [-q multipleModelFileName] [-r binaryConstraintTree]
               [-R binaryModelParamFile] [-S secondaryStructureFile]
               [-t userStartingTree]
```



Contenidos

- Introducción
- Instalación
- Modelos evolutivos
- **Resultados**

Resultados

La ejecución del programa produce varios archivos como resultado en el directorio en el que se ejecuta. Los más básicos son (puede haber más según las opciones)

Resultados

- Ficheros con árboles en formato Newick:
 - RAxML_parsimonyTree.<nombre> : Árbol de parsimonia aleatorio inicial utilizado para comenzar el proceso.
 - RAxML_result.<nombre> : Árbol tras finalizar el proceso iterativo.
 - RAxML_bestTree.<nombre> : Árbol tras las últimas optimizaciones.

Resultados

- Ficheros de texto:
 - RAxML_info.<nombre> : contiene información sobre los pasos seguidos por el programa. Principal fuente de información.
 - RAxML_log.<nombre> : Por cada línea, contiene un *timestamp* y el valor de *likelihood* del árbol en ese momento (una línea por iteración).
 - Contienen puntuaciones (*score*, *likelihood*) del árbol en distintos instantes, que miden su calidad en función del modelo.

Dudas / otras opciones:

- [standard-RAxML/manual/NewManual.pdf](#)
- Preguntadnos



Gracias por vuestra atención

