

Beyond BLAST

Shuji Shigenobu / 重信秀治

Aim

- BLAST以外の配列解析の手法を概観する。

Advanced BLAST search tools

- ▶ **PSI-BLAST: Position Specific Iterative BLAST**
 - ▶ Automatically generates a position specific score matrix (PSSM)
 - ▶ PSI-BLAST finds sequences significantly similar to the query in a database search and uses the resulting alignments to build a PSSM for the query. With this PSSM the database is scanned again to eventually pull in more significant hits, and further refine the scoring model.
 - ▶ More sensitive than standard BLAST
- ▶ **RPS-BLAST: Reverse Position-Specific BLAST**
 - ▶ RPS-BLAST uses the query sequence to search a database of pre-calculated PSSMs, and report significant hits in a single pass.
 - ▶ Used in CD-search (Conserved Domain search) at NCBI website.
- ▶ **DELTA-BLAST**
 - ▶ DELTA-BLAST searches a protein sequence database using a PSSM constructed from conserved domains matching a query. It first searches the NCBI CDD database to construct the PSSM.

Explore other sequence analysis tools

► What do you need?

► Speed

BLAT, UBLAST

► Consider exon/intron structure

exonerate

► NGS

► short read

bowtie2, bwa, TopHat

► long read

blasr

► large genome vs large genome

lastz

► Multiple alignment

clustalW, muscle, mafft

Exonerate

► Slater GS and Birney E (2005) Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6:31

► <http://www.ebi.ac.uk/about/vertebrate-genomics/software/exonerate>

A generic tool for sequence alignment

Exonerate is a generic tool for pairwise sequence comparison. It allows you to align sequences using a many alignment models, either exhaustive dynamic programming or a variety of heuristics.

Documentation

See the [Exonerate User Guide](#) for examples and tips for how to make the most of this software.

For further details about using exonerate and examples, see the [Exonerate manual](#) and the [Exonerate-server manual](#).

Many of the algorithms in exonerate are described in Slater GS and Birney E (2005) [Automated generation of heuristics for biological sequence comparison](#). BMC Bioinformatics 6:31; doi: 10.1186/1471-2105-6-31

Download

Exonerate is written in C, and currently uses the [glib](#) library for portability. It is portable to all UNIX-like systems, and has been used on various Linux distributions, TRU64, OSX, and BSD.

It is licensed under the [GPL](#).

You can download the source code or a precompiled version.

Exonerate version 2.2 includes fixes for problems with excessive memory consumption when compiled against glib-2, and fixes a bug with using exonerate-server with unmasked sequences.

Source code	exonerate-2.2.0.tar.gz
Linux/i386 binaries	exonerate-2.2.0-i386.tar.gz
Linux/x86_64 binaries	exonerate-2.2.0-x86_64.tar.gz

Exonerate: map cDNA onto genome

Intron/exon構造を考慮してtranscriptをゲノムにマッピングする。
(BLASTでは不可能なdonor/acceptor siteのGU/AGルールを考慮するマッピングソフトウェアが必要)

Exonerate を使う

キイロショウジョウバエのnos遺伝子のORFの配列が手元にある。ゲノムにマッピングせよ。(ex1-1と同じ問題)

- ▶ Transcript: Dmel_nos-PA.nuc.fasta
- ▶ Genome: dmel-all-chromosome-r6.13.fasta

```
exonerate --model est2genome --bestn 1 \  
Dmel_nos-PA.nuc.fasta Dmel_genome.3R.fasta
```

Exonerate: map protein onto genome

Intron/exon構造を考慮してprotein をゲノムにマッピングする。

Exonerate を使う

キイロショウジョウバエのnos遺伝子のタンパク質の配列が手元にある。ゲノムにマッピングせよ。

- ▶ Transcript: Dmel_nos-PA.pep.fasta
- ▶ Genome: dmel-all-chromosome-r6.13.fasta

```
exonerate --model protein2genome --bestn 1 \  
Dmel_nos-PA.nuc.fasta Dmel_genome.3R.fasta
```

Practice

- ▶ Ex8-2: map cDNA onto genome using exonerate
- ▶ Ex8-4: map protein onto genome (cross-species)

Why Multiple Sequence Alignments?

- ▶ Compare multiple sequences
- ▶ Identify conserved regions, patterns, and domains
 - ▶ Predicting function
 - ▶ Predicting structure
 - ▶ Identifying new members of protein families
- ▶ Perform phylogenetic analysis
- ▶ Generate position-specific scoring matrices for profile search

Software for Multiple Alignment

- ▶ Summary of Multiple Sequence Alignment software @EBI
 - ▶ <http://www.ebi.ac.uk/Tools/msa/clustalo/>
- ▶ Clustal Omega
 - ▶ <http://www.clustal.org/omega/>
- ▶ MUSCLE
 - ▶ <http://www.drive5.com/muscle/index.htm>
- ▶ MAFFT
 - ▶ <http://mafft.cbrc.jp/alignment/server/>

Beyond “sequence” search – **profile** search

- ▶ profile search
 - ▶ PSI-BLAST
 - ▶ HMMER
- ▶ profile search (DNA)
 - ▶ MEME toolkit
- ▶ motif db and search tool
 - ▶ InterProScan

