

実戦演習/Case Study

BLAST等を用いた遺伝子モデルの 機能アノテーション

Aim

- 新規に非モデル生物の網羅的遺伝子モデルを構築し、それらの遺伝子配列が手元に有るとする。これらの機能アノテーションをBLASTによる配列類似性解析によって行いたい、という状況を考える。

https://github.com/nibb-gitc/gitc2018-blast/wiki/case_study-1

ソラマメアブラムシの遺伝子アノ テーション



目的：非モデル昆虫「ソラマメアブラムシ」 *Megoura crassicauda* の網羅的遺伝子レパートリーを構築しそれらの機能アノテーションをしたい。

実験・解析の経過：ソラマメアブラムシから抽出したRNAで、Illumina MiSeqによるRNA-seqを行った。Trinityによるde novo assemblyを行い、その結果 21,730のcontigsが得られた。その中からORFを推定したところ、16,968のcoding genesが推定された。(Shigenobu et al., unpublished)

実戦演習課題：上記の解析で推定されたcoding genesの中から、本コースの練習用に400遺伝子を抽出した。これらをアミノ酸配列に翻訳した配列を

MEGCR_proteins400.pep.fasta

として用意した。それぞれのタンパク質がどのような機能を持っているか、BLAST検索を駆使してアノテーションせよ。

Strategies

- ▶ 1. BLAST search against NCBI nr database
 - ▶ BLAST best hit からの機能推定
 - ▶ ヒット遺伝子のtaxonomy情報を取得

- ▶ 2. 機能アノテーションが充実しているモデル生物を軸にアノテーションを行う。
 - ▶ BLAST best hit からの機能推定、GO term 付与

- ▶ 3. Motif search
 - ▶ モチーフを同定するとともに、GO term 付与

アノテーション戦略 1

NCBI nr (non-redundant protein database) に対してBLASTP検索を行う。

- 1) NCBI nrに対するBLAST検索を行う。ただし、検索には長時間を要するので、今回は計算済みのタブ区切りテキストを提供する。
- 2) description line を取得する。
- 3) トップヒットした配列の生物種とtaxonomy informationを得る。

Workflow

- ▶ 1. BLAST search (今回は計算済み in format7)
- ▶ 2. Retrieve description line
- ▶ 3. Retrieve taxonomy information

アノテーション戦略 2

基本戦略: 機能アノテーションが充実しているモデル生物を軸にアノテーションを行う。

- 1) モデル昆虫 *Drosophila melanogaster* のタンパク質データベースに対するBLAST検索を行う。トップヒットの遺伝子をホモログと見なす。
- 2) *D. melanogaster* ホモログのGene Ontology (GO) termをソラマメアブラムシにアサインする。

Workflow

- ▶ 1. BLAST against Dmel
 - ▶ 1.1 BLAST search against Dmel proteins
 - ▶ output table format
 - ▶ 1.2 Retrieve description line
- ▶ 2. GO term assignment
 - ▶ 2.1 Build FBgn ⇔ FBpp ⇔ GO table
 - ▶ 2.2 Assign GO terms

アノテーション戦略 3

基本戦略: InterProScanでモチーフを同定するとともに、GO termを付与する。

Workflow

- ▶ Run InterProScan