

第5節 データベースおよびウェブツールを用いた

バイオインフォマティクス

バイオインフォマティクスは、生命科学と情報科学が融合した生物学分野であり、「生物情報科学」などと訳される。バイオインフォマティクスは、ヒトゲノムやシロイヌナズナゲノムの解析に伴う大量データを処理するために必然的に生まれたもので、当初は実験プロジェクトをサポートすることが主目的であった。しかし、近年、幾つかの生物のゲノム情報が解読され、転写産物、タンパク質および二次代謝物質などを網羅的に解析しようとするポストゲノム研究においては、バイオインフォマティクスそれ自体が重要な役割を担っている。

バイオインフォマティクスでは、生命現象を理解するために種々の情報を大量に取り扱う。これらの情報は、データベースに収められており、ウェブツールを用いることで必要な情報を迅速に取り出し、解析することができる。本実習では、近年同定された酵母のユビキチン様タンパク質 Urm1 (Furukawa et al. 2000) をモデルに、データベースおよびウェブツールを用いて Urm1 の遺伝子・タンパク質情報の入手方法および、植物における Urm1 ホモログとの系統学的解析法を解説する。

5.1 文献検索

【概要】 『PubMed』は、米国 NLM (National Library of Medicine) の NCBI (National Center for Biotechnology Information) が運営している生物医学関連文献データベース『MEDLINE』に対する検索システムである。1966 年以降の文献が収録されており、文献中の単語や著者名、雑誌名などの項目を検索キーワードとして、目的の文献を探し出すことができる。ここでは、Urm1 を同定した Furukawa らの 2000 年の文献の検索方法を実習する。

【手順】

1. 『PubMed』 (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>) にアクセスする。
2. クエリーボックスに [yeast] を入力し、検索を始める。
3. [Furukawa] を入力し、候補論文を絞込む。
4. [“J Biol Chem”] を入力し、さらに絞り込む。
5. 検索結果から、2000 年に第一著者が [Furukawa] で発表された論文を探し出す。

5.2 キーワード検索

【概要】 『Entrez』は、上記の学術文献データベースと各種の分子生物学データベース (DNA、アミノ酸配列、タンパク質立体構造など) を統合したものであり、米国 NLM の NCBI が運営している。キーワード検索は、遺伝子名やタンパク質名などのキーワードをクエリー (query、問い合わせ) として用い、クエリーに関連する情報を複数のデータベースから一度に引き出すことができる。ここでは、キーワード検索において幾つかのキーワードを組み合わせることで、効率良く Urm1 のデータベース登録情報が得られることを実習する。

【手順】

1. 『Entrez』 (<http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi>) にアクセスする。

2. [Search across databases]に『urml』を入力し、「Go」を押して検索を開始する。
3. [AND yeast]を入力し、候補を絞り込む。
4. [NOT genome]を入力し、さらに候補を絞り込む。
5. [Nucleotide]の中から完全長配列情報(Acc. No. AY558295)を探し、リンクを開く。
6. [FEATURES]の[/translation=]のアミノ酸配列への翻訳をコピーする。

5.3 タンパク質の立体構造データベースの検索

【概要】 タンパク質の分子認識や調節はタンパク質の立体構造に基づいていることから、その機能を分子レベルで理解するためには、タンパク質の立体構造の解析が重要である。立体構造データベースには、立体構造データそのものを登録した一次データベース(Protein Data Base, PDB など)と、それらを比較・分類し、類似したもの同士を階層的にグループ化したデータベース(SCOP、CATH など)がある。ここでは、MMDB に登録されている Urm1 の立体構造モデルを用いて、Urm1 の細胞機能に重要なアミノ酸がどのように配置されているのかを確認する。

【手順】

1. ヘルパーアプリケーション『Cn3D』を準備する。ダウンロードサイト(<http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml>)にアクセスし、OS に合ったプログラムをインストールする。
2. NCBI のトップページ (<http://www.ncbi.nlm.nih.gov/>)の[Search]から[Structure]を選択し、[urml AND yeast]を入力する。もしくは、『Entrez』から[Structure]を選択する。
3. [2AX5]がヒットするので、リンクを開く。
4. [View 3D Structure]をクリックし、3次元構造を表示する。
5. [Sequence/Alignment Viewer]ウィンドでC末端のGly-Glyをクリックし、これらのアミノ酸残基がどこに配置されているか確認する。

5.4 ホモロジー検索

【概要】 ホモロジー(相同性)検索とは、ある一つの配列(たとえば遺伝子A)と類似する配列が、データベース中にあるかどうかを調べることである。遺伝子Aと高い相同性を示す遺伝子Bが得られた場合、これらは共通の祖先遺伝子から由来している可能性が高い。また、遺伝子Aの機能が未知であっても、遺伝子Bの機能が既知であれば、遺伝子Aの機能を推定することができる。このように、ホモロジー検索は、単に類似する配列の検索のみならず、遺伝子の進化・系統学的解析やタンパク質の機能予測など様々な解析に用いられる。ここでは、ホモロジー検索ツール『BLAST』を用いて、植物における Urm1 のホモログ(相同性を示す遺伝子・タンパク質)の同定を試みる。

【イネにおける Urm1 ホモログ(RURM1)の検索】

【手順】

1. 『NCBI-BLAST』 (<http://www.ncbi.nlm.nih.gov/BLAST/>)にアクセスする。
2. 『blastp』 (Protein - protein BLAST)を開く。
3. [search]に Urm1 のアミノ酸配列もしくは Protein_id を入力する。

4. Options の[Limit by entrez query]の select from から [Oryza sativa] を選択する。
5. [BLAST!] を押して、検索を始める。
6. [Format!] を押して、結果を表示する。

【植物における Urm1 ホモログの検索】

【手 順】

1. DNA Data Bank of Japan (DDBJ, <http://www.ddbj.nig.ac.jp/Welcome-j.html>) にアクセスし、『BLAST』を開く。
2. [プログラム] から 『pblastn』 (アミノ酸配列クエリー × 塩基配列データベース [アミノ酸配列に翻訳]) を選択する。
3. RUM1 のアミノ酸配列を入力する。
4. [検索対象データベース] の中から、植物のみを選択する。
5. [入力内容の送信] を押して、検索を始める。
6. [View Result] を押して、結果を表示する。

5.5 マルチプルアライメントおよび系統樹作成

【概 要】 マルチプルアライメントとは、一度に複数の配列を比較し、類似する部分を縦に揃えて並び合わせる操作である。マルチプルアライメントは、配列中の生物学的に意味がある部位の予測や、配列間の進化的解析に用いられる。ここでは、上記で同定した Urm1 ホモログをマルチプルアライメントツール『ClustalW』を用いてアライメントし、それらの系統学的解析を行う。

【アミノ酸配列への変換】

【手 順】

1. 『pblastn』の検索結果ページで各植物の[E Value]が一番高い配列情報を開く。
2. [Accession No.] もしくは塩基配列情報をコピーする。
3. 別のウィンドで『NCBI Home Page』にアクセスし、『ORF Finder』を開く。
4. コピーした[Accession No.] もしくは塩基配列を貼り付ける。
5. [OrfFind] を押して、ORF の候補を表示する。
6. 候補の中から Urm1 と同じ特徴(C 末端 Gly-Gly) をもつ ORF を探し出す。

【アライメント解析および系統樹作成】

【手 順】

1. 『ORF Finder』の配列表示画面で、[Accept] を押して次の画面へ進む。
2. 左側のプルダウンメニューから、[3 Fasta protein] を選択する。
3. [View] を押して、FASTA 形式で表示された配列情報をコピーする。
4. 別のウィンドで『DDBJ』にアクセスし、『ClustalW』を開く。
5. [解析配列データ] にコピーした配列情報を貼り付ける。
6. [>]以降を生物名に書き直す。
7. この操作を『pblastn』で選択した生物全てに対して繰り返す。
8. Urm1 および Rum1 のアミノ酸情報も FASTA 形式で入力する。

9. [TYPE]を[PROTEIN]に設定し、[SHOW ALIGNMENT SCORE]にチェックをつける。
10. [入力内容の送信]を押して、解析を始める。
11. [View Result]を押して、結果を表示する。
12. 系統樹表示アプリケーション『TreeView』のダウンロードサイト (<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>)にアクセスし、『TreeView』をインストールする。
13. [query.ph]を押して、系統樹を表示する。