

バイオインフォマティクスについて学ぶ

メタデータ	言語: jpn 出版者: 公開日: 2015-02-03 キーワード (Ja): キーワード (En): 作成者: 森内, 良太 メールアドレス: 所属:
URL	https://doi.org/10.14945/00008046

バイオインフォマティクスについて学ぶ

森内良太

技術部 教育研究支援部門

1. 背景と目的

ワトソンとクリックにより DNA の二重らせん構造が提唱されてから 60 年が経過し、その間分子生物学分野の研究はめざましい発展を遂げた。特に近年のコンピューターを使用したバイオインフォマティクス技術の活用により、配列情報解析や発現情報解析、構造情報解析やパスウェイ解析等が行えるようになり、更なる飛躍を遂げている。私は普段、本研究分野の研究支援を行っているが、コンピューターを使用した解析技術、つまり生物学データベースからのデータ取得と解析ツールを使用した解析技術は、必須のスキルとなっている。そこで本研修では、データ取得の方法と解析手法を学び、研究支援に生かせるような能力を身につけることを目的とする。またバイオインフォマティクスの役割と重要性についても学び、知識を広げ、深めることを目的とする。

2. バイオインフォマティクスについて

2.1 バイオインフォマティクスの定義と目的

バイオインフォマティクスの定義について Luscombe らによると、「分子の面で生物学を概念化し、情報を整理し理解することを目的として、情報科学技術へ適用すること」であるという^[1]。つまり生物の高分子に関する情報を、コンピューター技術を使用して整理し理解するということであり、図 1 に示したようなイメージである^[2]。

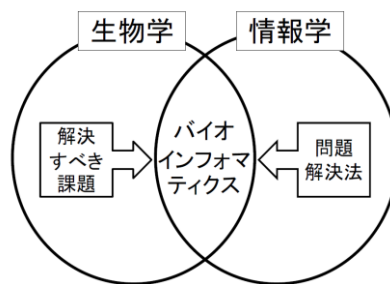


図 1 バイオインフォマティクスの学問的位置付け

生物の生理機能は、遺伝子やアミノ酸などの高分子によってほとんど決定されており、これらの生物学的データは驚くべき速度で生じている。例えば、GenBank（米国生物工学情報センターが提供している世界的な公共塩基配列データベース）の核酸配列エントリー数は 2008 年 12 月の時点で 98,220,409 であったが、2013 年 12 月現在その数は 169,094,459 に上り、わずか 5 年間で 2 倍近くまで増えている^{[3][4]}。生理機能を解明するためにはこれらの情報を活用し、遺伝子発現解析やタンパク質構造の決定、遺伝子やタンパク質間相互作用解析などの様々な解析が必要である。コンピューターを使用することで一度に大量のデータを扱うことができ、自然で観察できる複雑なダイナミクスを容易に調べることができる。

バイオインフォマティクスの目的は、主に以下の3つである^[1]。①研究者が現存する情報にアクセスすることを可能にし、新たな情報を登録できるようにすること、②データ解析に役立つツールと資源を開発すること、③様々なツールを使用してデータを解析し、生物学的に意味を持つように解釈することである。バイオインフォマティクス技術を活用することで大規模解析を行うことができ、一般的原則を明らかにする、あるいは新たな特徴を強調することが可能となる。

2.2 データベースツールと解析ツールの紹介

バイオインフォマティクスの特徴として、多くの実践的なアプリケーションが存在することが挙げられる。本研修では、いくつかの公共データベースツールやフリーの解析ツールの使用方法を学習した。ここでは、研修で扱ったツールの中で使用頻度や重要度が高いウェブツールについて述べる。

2.2.1 各種データベースツール

NCBIは国立生物工学情報センター (National Center for Biotechnology Information)の略称であり、分子生物学やバイオインフォマティクス研究に用いられるデータベースの構築や運営を行っている、アメリカの機関である。NCBIが運営するデータベースには、PubMedやNucleotide、Proteinが存在する。**PubMed**は医学・生物学分野の学術文献検索サービスであり、文献データベースの定番となっている(図2)。著者名、雑誌名、論文タイトル、出版年、PubMed ID、キーワードなどで論文を検索することができ、所属する機関が電子ジャーナルを購入していれば、論文をダウンロードし閲覧することができる。**Nucleotide**と**Protein**はそれぞれ核酸配列、タンパク質関連のデータベースであり、これらもキーワードを入力することで、目的の配列情報を検索できるシステムになっている。**PDB (Protein Data Bank)**は、実験的に決定されているタンパク質と核酸の3次元構造データベースであり、構造バイオインフォマティクス研究共同体RCSB (Research Collaboratory for Structural Bioinformatics)によって管理されている。キーワード検索によって目的のタンパク質や核酸を検索でき、立体構造を閲覧することができる(図3)。

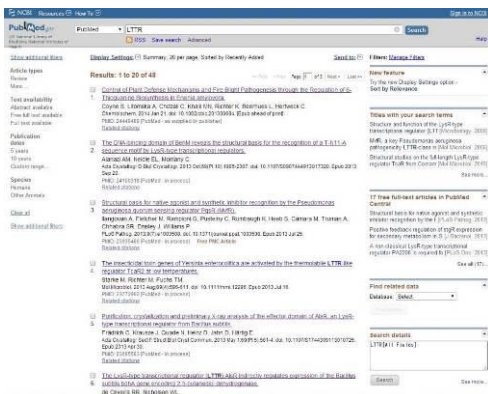


図2 PubMedで論文を検索した結果

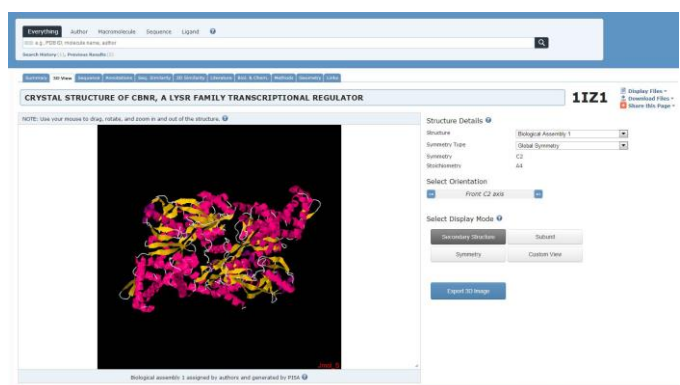


図3 PDBで立体構造を閲覧する様子

2.2.2 塩基配列・アミノ酸配列解析ツール

実際に研究支援を行う場合、様々な場面で塩基配列の解析が要求される。具体的には、制限酵素地図の作成、PCR(ポリメラーゼ連鎖反応)プライマーの設計、タンパク質の分子量予測などである。**EMBOSS GUI**は、配列解析を行うためのソフトウェアである。本ソフトウェアには様々な機能があるが、その一つに制

2.2.4 タンパク質の構造予測

バイオインフォマティクスの目標の1つは、DNAの塩基配列からその配列がコードするタンパク質の機能を知ることであり、現在タンパク質の機能を配列から予測する手がかりとなる方法が開発されている。例えば、**PSORT** というタンパク質の細胞内局在を予測するためのツールや、**SOSUI** という膜タンパク質の判別や膜貫通領域の解析ができるツールがある。PDB データベースではタンパク質の立体構造を簡単に閲覧でき、本研修ではその閲覧方法を学習した (図3)。

2.2.5 タンパク質立体構造の比較

タンパク質において、アミノ酸配列の相同性が高ければ二次構造が似てくるだろうし、二次構造が似ていれば三次構造が似てくるであろう。立体構造が似ていれば機能が予測しやすく、どのアミノ酸残基がどのように重要であるか推測することも可能である。PDBj (PDB の日本語版サイト) に **ASH (Alignment of Structural Homologs)** というツールがあり、名前の通り立体構造を並べて類似度を比較するツールである。PDB に登録されている立体構造のアライメント ID を入力することで、2つの立体構造エントリを比較することができる (図7)。

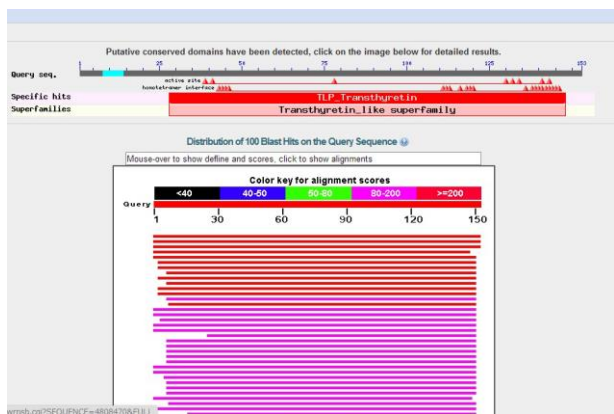


図6 BLAST 解析を行った結果

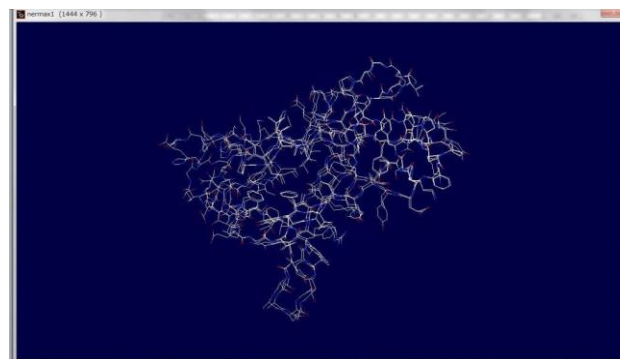


図7 比較した2つの類似なタンパク質

上記で述べたツールについて、以下の表にまとめた。

表 本研修で学習したツール

ツール名	ツールの概要	URL
NCBI PubMed	医学・生物学分野の学術文献検索	http://www.ncbi.nlm.nih.gov/pubmed
NCBI Nucleotide	核酸配列データベース	http://www.ncbi.nlm.nih.gov/nucleotide
NCBI Protein	タンパク質関連データベース	http://www.ncbi.nlm.nih.gov/protein
PDB	タンパク質・核酸構造データベース	http://www.rcsb.org/pdb/home/home.do
EMBOSS GUI	配列解析	http://imed.med.ucm.es/EMBOSS/
Primer 3	プライマー設計支援	http://bioinfo.ut.ee/primer3/
ExPASy	タンパク質解析ツール集	http://www.expasy.org/
BLAST	相同性検索	http://blast.ncbi.nlm.nih.gov/Blast.cgi
PSORT	細胞内局在予測	http://wolfpsort.org/
SOSUI	膜貫通領域の解析	http://bp.nuap.nagoya-u.ac.jp/sosui/
ASH	立体構造の類似度比較	http://pd bj.org/info/ash

3. 研修内容

3.1 担当講師について

理学研究科生物科学専攻・グリーン科学技術研究所グリーンバイオ研究部門所属の石原顕紀講師に、本研修の講師を担当して頂いた。石原講師の専門分野は分子生物学とバイオインフォマティクスであり、主要な研究テーマは「甲状腺系に及ぼす環境化学物質の影響」である。優秀な若手研究者であり、理学部生物科学科2年生対象の「バイオインフォマティクス演習」を担当しているため、講師を依頼した。

3.2 基礎編「バイオインフォマティクスって何？」

日時：2013年11月21日（木）10:00-11:30

場所：情報基盤センター実習室

参加者：宮澤、早村、高田、井上、森、市川、森内、剣持（敬称略）

内容：分子生物学分野を専門外とする受講者が多数のため、まず遺伝子とは何か、遺伝子が発現するとはどういうことかについて学習した。その後、前半はツールとしてのバイオインフォマティクスに焦点を当て、分子生物学分野の基本的な実験手法と、関連するバイオインフォマティクス技術の利用について学んだ。後半はiPS細胞の研究と講師の先生の研究を例として、バイオインフォマティクス研究がどのような成果をもたらしたのか、解説して頂いた。

3.3 応用編「バイオインフォマティクスツールの使用方法習熟を目指した実習」

日時：2013年11月21日（木）13:00-17:00

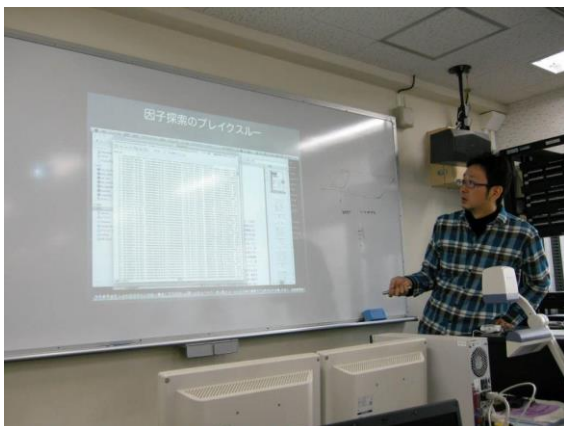
場所：情報基盤センター実習室

参加者：宮澤、市川、森内、剣持（敬称略）

内容：2.2に記されている様々なツールの使用方法について、パソコンを使用しながら研修を行った。まずデータベースツールを使用して、ある遺伝子配列やアミノ酸配列データを取得する方法を学んだ。次に、取得したデータを解析ツールに入力して実際に解析を行い、出力された結果の意味や見方を教えて頂いた。

3.4 研修の風景

研修の風景を以下に示した。





4. 研修の成果と今後に向けて

バイオインフォマティクスとは何か、どのような研究テーマで何を目的として研究を進めているのか、その一端を知ることができ、技術職員としての知識を広げることができたと考える。特に、生物系技術職員だけでなく情報系技術職員も積極的に質問を行い、自らの業務に生かそうとする様子が伺えた。応用編に参加した受講者は、基礎的ではあるが重要なデータベースや解析ツールの使用方法を広く学ぶことができ、身につけることができた。また丁寧に作成されたテキストを頂いたため、参照しながらいつでもツールを使うことができる。

企画として考えた場合、研修を基礎編と応用編に分けたことで、参加者のニーズに合わせた内容の研修を行うことができたと考える。

今後バイオインフォマティクス系の研修を行う場合、マイクロアレイ解析や次世代シーケンサーにより得られる大量データの処理方法など、具体的な事例を取り上げた、一步踏み込んだ内容の研修を行いたいと考えている。

最後に、講師としてご協力頂いた石原顕紀講師、本研修全般にわたって様々な場面でご協力頂いた宮澤俊義技術専門員と高田重利技術専門職員、その他関係者の皆様にこの場をお借りして御礼申し上げます。

参考文献

- [1] Luscombe NM, Greenbaum D and Gerstein M. *Methods Inf Med* 40, 4 (2001)
- [2] 深見-小林薫：「最近流行のバイオインフォマティクスとは何だ」 47, 6 (2001)
- [3] Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. *Nucleic Acids Res* 41,D1 (2013)
- [4] DDBJ Release Note 95.0 (2013)
- [5] 中村保一ら：「バイオデータベースとウェブツールの手とり足とり活用法」 羊土社 (2007)