

生物研究コンピューター時代

バイオインフォマティクス

バイオ(生命科学)とインフォマティクス(情報科学)を融合した研究分野



DNA配列の個人差、病気情報など
体

コンピューターの中の生物学の例

- 遺伝子の働きを推定
- ゲノムの比較で進化を探る
- アミノ酸配列から、たんぱく質の立体構造を予測

AMINOSACCHIDE
VLESLNRVHLL
LATNEMWRSQSEI
SLRDVAYQVYKKGSR
IATTHADNIFLSDQTKI
LLGRVGDQDPLRQVE
TVYQLGDVSAKTTI
VLEKIDYGE

たんぱく質の立体構造

たんぱく質と反応する化合物を探す



グラフィック・山田 英利子 / The Asahi Shimbun

がんの原因、薬の候補を解析

「ゴオオオ……」。東京大学医科学研究所ヒトゲノム解析センターのスーパーコンピューター室は大音量に包まれていた。ジェット機のエンジンの近くにいるかのような。パソコン6144台分にあたる機器が24時間、動いている。

「ゴオオオ……」。ここでは、肝臓がん患者のゲノム(全遺伝情報)を解析して、がんの原因を探っている。1人のゲノムの解析量は約4テラバイト。市販の外付けハードディスク3台分、標準的なDVD(4.7ギバイト)に保存するなら870枚が必要だ。共同研究を

コンピューターの中で生物進化の道筋を解明し、新薬の候補となる物質も探す。生物に関する膨大なデータをコンピューターで解析したり、そのための計算法を開発したりする研究分野が発展している。生命科学(情報科学)の融合分野「バイオインフォマティクス(生物情報科学)」だ。(瀬川茂子、杉本崇)

ゲノム計画で発展

90年代以降、ヒトゲノムをはじめ、さまざまな生物のゲノムの解析計画を通じて発展した。大量のデータを処理する方法が開発され、データを処理して生物学的な意味を探る研究も進展した。

ゲノムは、ATCGで表される4種類の「塩基」という物質がつながってできている。解析装置が読める長さは限られているので、短く切ったゲノムの断片を読んでいく。ATCGの4文字で書かれた暗号文のような多数の断片をコンピューターの中でつなぎあわせるのは、大変な作業だ。

現在、森下教授は、神経の病気の患者のゲノムを解析中だ。解析法はメタカルの時と違う。30億塩基対からなる人類の標準的なゲノムはすでに公開されているので、新たに読んだゲノム断片が、標準的なものどこに相当するか、コンピューターで探す。断片の配列を読むのは、次世代シーケンサーと呼ばれる解析装置だ。

ヒトゲノム解析の主な出来事

年	出来事	1人分のゲノムを読む費用と時間
86	米エネルギー省がヒトゲノム研究を開始	
90	国際ヒトゲノム計画がスタート	
03	国際コンソーシアムが複数の人をつないだヒトゲノムを解析	2700億円 13年
07	セラ社社長のベンター博士が自分のゲノムを公開	100億円 4年
	ワトソン博士が自分のゲノムを公開。個人では世界初	1億5千万円 4.5カ月
08	米企業や中国の研究機関がアフリカ・中国人のゲノム解析	2500万円 8週間
10	ゲノム解析の低コスト化が進む	100万円 2週間

ゲノム配列だけでなく、遺伝子の働きなども網羅的に調べる時代になってきた。「以前は病気がかわる一つの遺伝子に注目して、診断や治療に役立てられなかった調べる時代が多数の遺伝子を一度に調べる時代になった」と京都大学の辻本豪三教授(薬学)は話す。

創薬分野でも、バイオインフォマティクスが使われている。病気がかわるたんぱく質が見つかると、そのたんぱく質の働きを促したり、止めたりする物質は薬の候補になる。たんぱく質の立体構造から、反応する化合物を探してシミュレーションすることで薬の候補になる物質を探す方法は、すでに広く使われている。

◆「アスパラマリン」(<http://aspara.asahi.com/>)の「アサロン」・秋手面(かひい)にもアップ記事を掲載していきます。

「メタカルの頃は1台の解析装置が1日に200万文字読んだ。今は250億文字。解析速度は1万倍以上になった」と森下教授。解析装置の高速化はさらに進み、コストも下がっている。

バイオインフォマティクスの課題は人材育成だ。若い人に興味をもってもらうと、理化学研究所の生命情報基盤研究部門は「空気中から、(有害化学物質の)ホルムアルデヒドを除く植物を作る」というプロジェクトを実施中だ。

現在、森下教授は、神経の病気の患者のゲノムを解析中だ。解析法はメタカルの時と違う。30億塩基対からなる人類の標準的なゲノムはすでに公開されているので、新たに読んだゲノム断片が、標準的なものどこに相当するか、コンピューターで探す。断片の配列を読むのは、次世代シーケンサーと呼ばれる解析装置だ。

実際に遺伝子組み換え植物を作るのではなく、ウェブ上の遺伝子配列などの情報を利用して「ゲノム設計」を競ってもらう。後に専門家が「設計配列」を植物に組み込んで審査する。安全で合理的な設計を競う。「ロボコン(ロボットコンテスト)」のように若者に夢中になってほしいと「ゲノコン」と名付けた。

東京大学の森下真一教授(情報生命科学)らはメタカゲノムを解析し、2007年に発表した。1500万個のゲノム断片をつないで7億塩基対を解読するのに3年かかったという。さらに2年かけて、メタカゲノムをヒト、ニワトリ、ホヤのゲノムと比較した。進化の過程でゲノム全体が2倍になる「遺伝子重複」が繰り返して起こったことを検証した。

は、さらに効率よく薬の候補になる物質を探そうと、たんぱく質を構成するアミノ酸配列の文字列から、反応する化合物を推定する方法を開発した。答えがわかっているアミノ酸配列と化合物の組み合わせをコンピューターに学習させ、未知の組み合わせでも答えられるようにする。情報科学でよく知られている計算法を応用した。

将来は設計士も?

