

特集 2

生物資訊的技術動向

資訊通訊單位 客座研究員 乃木篤*、香月祥太郎

1. 前言

人類基因體計畫的進展速度，不但超過國際間的預期，且在 2000 年 6 月，初稿序列 (draft sequence) 也宣告完成，目前正處於完全解讀基因體的最後階段。雖然依據原先的預定規模，日本提供 6~7% 的貢獻，但由於缺乏策略，導致進度落後，引發各方的指摘，還留下許多未完的課題。不過，在 2002 年 8 月，以物理化學研究所為中心，將老鼠全長 cDNA^(註) 與海鞘基因體解讀完畢；在 12 月，日美雙方共同宣布完成海鞘基因體解讀等，替基因體科學的新進展，跨出了一大步。隨著人類基因體或老鼠 cDNA 序列的公開，可看見今後的焦點，將從序列的分析，轉移到包羅萬象的系統性功能分析。而日本擁有水稻基因體和全長 cDNA library 等基本技術的優勢，所以不管在學術上還是在應用領域，都掌握了引領全世界的良機。

近年來，隨著分析設備的高速化和自動化，實驗資料的數量急速膨大，因而使資訊技術的角色更形重要。而且不只是一要提高資料處理的效率，為了活用基因體，解決分析上包羅萬象的複雜性，還必須有相當的數學與資訊科學知識。為了讓生物資訊朝此方向有系統的邁進，在生物科學和資訊科學的範圍內誕生了全新的學術領域。

本期刊在 2001 年 12 月已提出生物資訊特集，從生命科學的觀點，概述其動向。本文將涵蓋之後的動向，為了使大家對基因體與生命產生系統性的理解，將逐一討論，並說明當前的課題與人才培育的策略。

(註) 全長 cDNA：cDNA 是指去除基因體 DNA 中不必要的序列，只保留蛋白質的基因序列，利用經處理的訊息 RNA，也就是 mRNA (messenger RNA)，作為模版，所做出的 DNA。全長 cDNA 與 cDNA 片段不同，全長 cDNA 具有蛋白質合成時所需的所有資訊，因此可合成蛋白質。為了有效合成全長 cDNA，需具備頂尖的技术，我國在這方面領先全球。

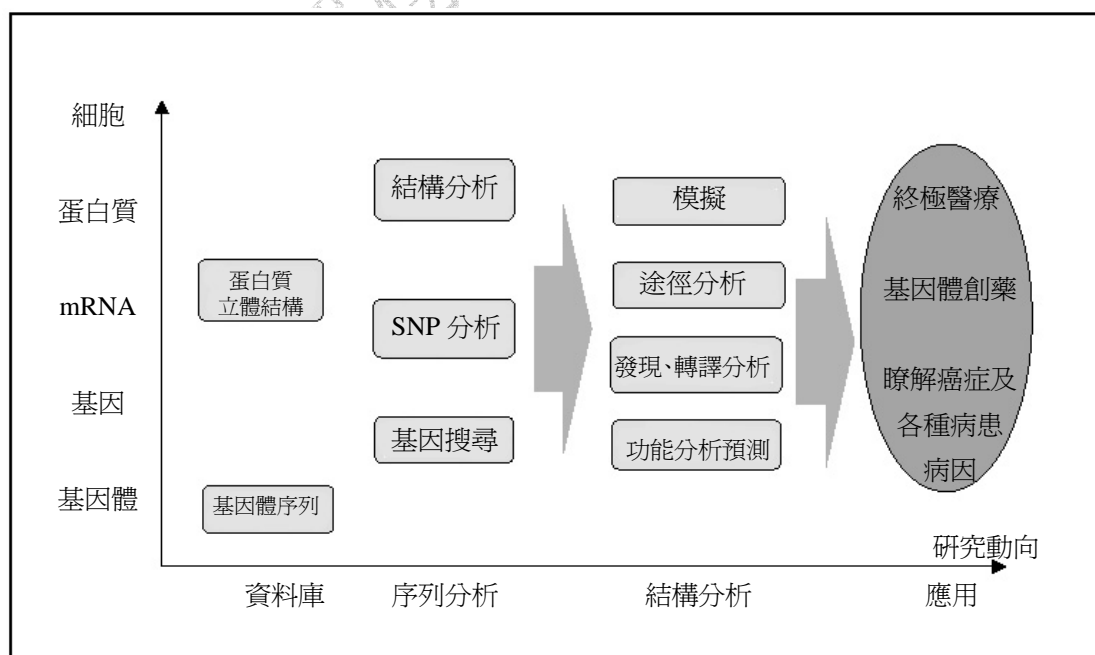
2. 生物資訊的資訊科學技術

若將基因體研究的種種問題，當作一般的資訊科學處理，將會發生很大的困難，遠超過現在大多數電腦的負荷能力。這是因為基因體序列非常長，又必須配合分析對象的特性，故需要近似解法和啟發式的技術。因此，開發實用的演算法，已成為生物資訊的主要研究領域，而其成果對於基因體定序也帶來很大的貢獻。美國的 Celera Genomics 公司，利用全基因體隨機定序（whole-genome shotgun sequencing）法作為定序的技術，將數千萬隨機的序列片段，像是玩拼圖一樣進行組合。當初大家還對此項技術抱持疑慮，但該公司利用高性能的電腦，與獨特的演算法，以令人驚異的速度完成定序，讓大家對資訊技術的威力和效果，留下深刻的印象。

由於近似解法未必能引導出最適的結果，所以必須改進分析結果的生物學評價與基本演算法或參數。生物資訊的特徵在於，儘可能地將實驗無限寬廣的搜尋空間縮小，同時進行包羅萬象的系統性分析。也可說是將各別基因分析，無法看到的生命現象全貌，完整地展現出來。

到目前為止，為了建構基因體資料庫和進行資料的分析，生物資訊以工具的開發及使用的技術為中心。今後將會針對基因體研究的基因發現和功能預測，更加活用資訊處理的能力。從解讀基因體開始的研究，如圖表 1 所示，從 DNA，延伸到基因、蛋白質，甚至以生物個體為對象，大大增加了複雜性和多樣性，而分析的層次也從序列，延伸到功能、作用機制，研究焦點明顯地轉移。因此，生物資訊也必須遵循此趨勢進行研究開發。

圖表 1 生物資訊研究的動向



(科學技術動向中心製作)