生物資訊
Bioinformatics

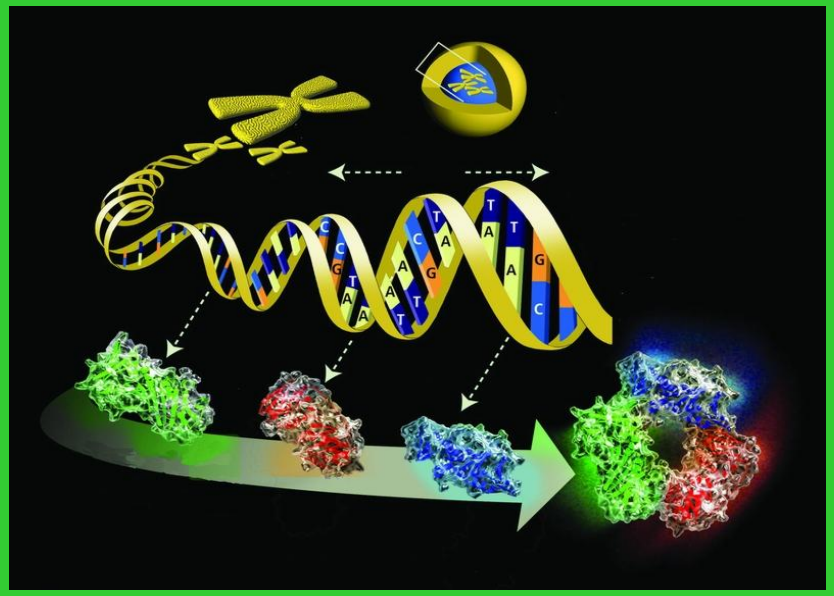
生物資訊

簡單介紹

生物資訊學（Bioinformatics）

利用**應用數學、資訊學、統計學**和**計算機科學**
的方法研究生物學的問題。

目前的**生物資訊學**基本上只是分子生物學與資訊技術
（尤其是**網際網路技術**）的結合體



生物資訊在現代醫學上之研究、
教學、實際應用等各方面的意義

生物學技術往往生成大量的嘈雜數據。

與數據挖掘類似，生物資訊學利用**數學工具**從大量數據中提取有用的生物學資訊。

生物資訊學所要處理的典型問題包括：

重新組裝在霰彈槍定序法測序過程中被打散的DNA序列
→
從蛋白質的胺基酸序列預測蛋白質結構
→
利用mRNA微陣列或質譜儀的數據檢驗基因調控的假說

某些人將計算生物學作爲生物資訊學的同義詞處理，
在英語維基百科中就是如此。

計算生物學 = 生物資訊學？

但是另外一些人認爲計算生物學和生物資訊學
應當被當作不同的條目處理。

➔因爲・・・・・・

生物資訊學更重於生物學領域中計算方法的使用、發展

而**計算生物學**強調應用資訊學技術對生物學領域中的假說進行檢驗，並嘗試發展新的理論。

生物資訊學的研究材料和結果就是各種各樣的生物學數據，其研究工具是**計算機**。

●研究方法：

1. 利用「計算、模擬」
2. 處理「編輯整理、管理和顯示」
3. **對生物學數據的搜索「收集和篩選」**

●目前主要的**研究方向**有：

1.序列比對
2.基因重組
3.基因表達
4.建立進化論的模型
5.基因識別
6.蛋白質結構預測
7.蛋白質反應的預測

生物資訊

研究方向

1. 序列分析

1977年噬菌體Φ-X174
成爲第一個被完整測定基因組順序的生物體。

自此以後，越來越多生物體的DNA序列被人類測定。

通過對這些序列的分析，
人們希望獲知其中對應 蛋白質編碼的基因和 基因調控序列。

不同物種間的基因比對能夠→
1. 解釋和預測他們蛋白質的功能的相似性
2. 揭示不同物種間的聯繫

基因組中並不是所有的核苷酸都構成基因

所以序列分析的另一個研究課題
是對 基因組 中的基因和 基因調控序列進行自動識別。

●在較高等的生物體中，DNA序列的大部分並沒有明顯的作用。

●但是，這些所謂的「 垃圾DNA」卻可能具有未被識別的功能。



Macvector

是一套生物資訊和序列分析軟體→

2.基因組註釋

在 基因組學中，
對基因和其他生物特徵的標註稱爲基因組註釋。

1995年，Owen White和他的組員
在1995年完成的人類首次爲獨立的生物體**「流感嗜血桿菌」**測序的工作中
設計出了第一套基因組註釋軟件系統。

該系統能夠自動識別基因、轉錄基因和其他生物學特徵，
並能夠初步的分析它們的功能。

大多數現今的註釋系統的與之類似，
但用以分析DNA序列的軟件在不斷在進化。

在尋找基因和致力於發現新蛋白的努力中，
人們習慣於把新的序列同已知功能的蛋白序列作比對。

由於這些比對通常都希望能夠推測新蛋白的功能，
不管它們是雙重比對還是多序列比對，
都可以回答大量的其它的生物學問題。

舉例來說，
面對一堆蒐集的比對序列，
人們會研究隱含於蛋白之中的系統發生的關係，
以便於更好地理解蛋白的進化。

人們並不只是著眼於某一個蛋白，
而是研究一個家族中的相關蛋白，
看看進化壓力和生物秩序如何結合起來創造出新的具有雖然不同但是功能相關的蛋白。

研究完多序列比對中的高度保守區域，
我們可以對蛋白質的整個結構進行預測，
並且猜測這些保守區域對於維持三維結構的重要性。

顯然，

分析一群相關蛋白質時，很有必要了解比對的正確構成。

發展用於多序列比對的程序是一個很有活力的研究領域，

絕大多數方法都是基於**漸進比對**（progressive alignment）的概念。

漸進比對的思想依賴於→

使用者用作比對的蛋白質序列之間確實存在的生物學上

或→

更準確地說是系統發生學上　　的相互關聯。

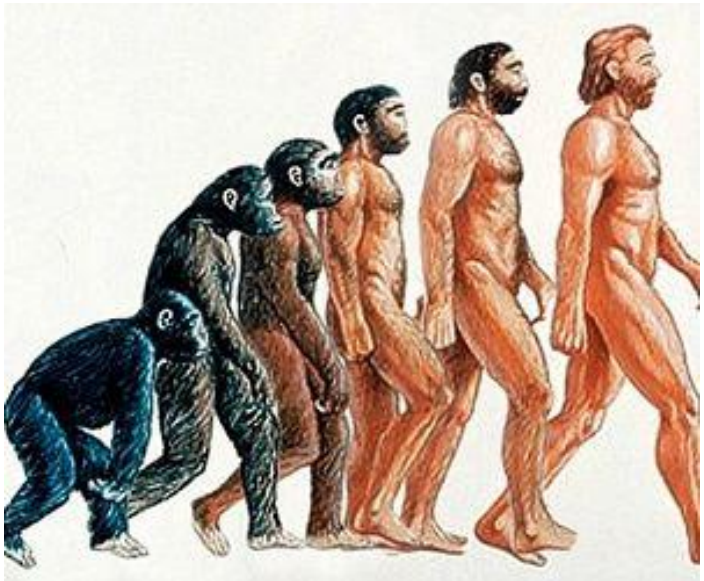
不同算法從不同方面解決這一問題，
但是當比對的序列大大地超過兩個時（雙重比對），
對於計算的挑戰就會**很難**。

在實際操作中，算法會在計算速度和獲得最佳比對之間尋求平衡，常常會接受足夠相近的比對。

不管最終使用的是什麼方法，
使用者都必須審視結果的比對，
因爲再次基礎上作一些手工修改是十分必要的，尤其是對保守的區域。

3.計算進化生物學

未來的研究工作包括重建業已相當複雜的 進化樹。
計算進化生物學 常常與用遺傳算法的計算機科學相混淆。
後者受到生物進化原理的啓發，
發展出一套軟件用於改進配方、算法、集成電路設計等等。



進化生物學研究物種的起源和演化。
引入信息學到進化生物學中，使得研究者能夠：

〈一〉通過度量DNA序列的改變研究眾多生物體間的進化關係
（超越了以前基於身體和生理特徵觀察的研究方法）

〈二〉通過整個 基因組的比對，研究更爲複雜的進化論課題，
如 基因複制 基因橫向遷移

〈三〉爲種群進化建立復雜的計算模型，預測種群隨時間的演化

〈四〉保存大量物種的 遺傳信息

4.生物多樣性的度量

該領域的一個重要前景是爲瀕危物種建立基因銀行，即將各物種的基因組信息保存下來。 這樣即便在將來這些物種滅絕了，人類也可能利用它們的基因組信息重新創造出它們。

5.蛋白質結構預測

是生物資訊學的重要應用。蛋白質的氨基酸序列可以容易的由它的基因編碼序列獲得。
在絕大多數情況下，在蛋白質的原生環境中，其結構由它的氨基酸序列唯一的決定。
蛋白質的結構對於理解蛋白質的功能十分重要。
這些結構信息通常被稱爲二級、三級、四級結構。 目前尚沒有普遍可行方案實現這些結構的準確預測；大多數方案爲啓發式的。

6.蛋白質表達分析

對生物標本進行測量所獲得的數據中，包含有**大量生物標本內蛋白質的資訊。 生物資訊學被廣泛的應用於這些數據的分析。**

對於前者，生物資訊學所面臨的問題與 RNA微陣列數據分析中遇到的問題相似

對於後者，生物信息學將所獲得的大量質譜數據與通過已知蛋白質數據庫預測的數據進行比較，並使用複雜的統計學方法進一步分析。

7.生物系統模擬

系統生物學通過對細胞子系統的計算機模擬，分析這些細胞過程間的複雜聯繫，並將分析結果可視化。 例如，**人工生命就通過計算機模擬簡單的生物形式，以幫助人類了解進化過程**。

8.在藥物方面研發應用

生物資訊學可用於**藥物靶標基因的發現和驗證**。

有許多數據庫可用來獲得在不同組織在正常／疾病狀態下基因表達的差異，通過搜索這些數據庫，可以得到候選基因作爲藥物靶標，特異性地針對某一種疾病。

另外，還可**根據蛋白質功能區和三維結構的預測來對藥物靶標進行鑑定，以便早期了解所研究蛋白的屬性，預測它是否適用於藥物作用**。

資料來源：

維基百科/Google

http://zh.wikipedia.org/zh/%E7%94%9F%E7%89%A9%E4%BF%A1%E6%81%AF%E5%AD%A6

http://www.google.com.tw/images?hl=zh-tw&rlz=1T4SKPB_zh-TWTW331TW331&q=%E7%94%9F%E7%89%A9%E8%B3%87%E8%A8%8A&um=1&ie=UTF-8&source=og&sa=N&tab=wi&biw=1003&bih=440