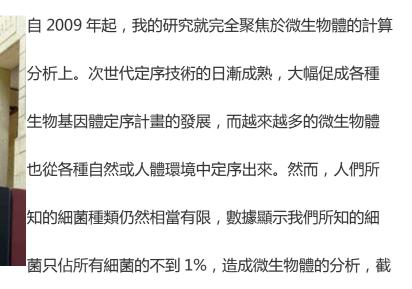


吳育瑋老師學術分享:微生物體基因序列分析

吳育瑋助理教授於 2012 年 8 月從美國印第安那大學布魯明頓分校取得生物資訊博士學位,並在加州柏克萊美國國家實驗室進行為期 4 年的博士後研究工作後,於 2016 年 10 月至本校任教,現為醫學科技學院醫學資訊研究所助理教授。(編按)



至目前為止仍然是件相當困難的工作。

微生物體的分析難度主要來自於各種細菌基因序列的混雜。試著想像一盒高難度的 拼圖:一般人或許需要花上一、兩個禮拜甚至一個月去拼湊這幅拼圖,再將幾百盒 完全不同的高難度拼圖碎片全部混在一起;而將這幾百盒完全不同的拼圖全部拆解 並拼湊出來的複雜度,就是我們在分析微生體時遇到的難度。【圖:醫學資訊研究 所吳育瑋助理教授】



為了解決這個問題,我開發出了一種分群演算法「MaxBin」,運用機率分析與機器學習演算法將微生物體基因序列進行分類,以簡化後續的各種分析與處理。以拼圖來說,就是將幾百盒混在一起的拼圖分類為幾百堆,每堆只包含來自一種拼圖的碎片。這種方法能夠大幅降低後續的分析複雜度,並藉由分群後的個別基因體促成我們對未知物種的認知。

MaxBin 已於 2014 年發表在《微生物期刊》上(Microbiome 2:26, 2014);第二版 MaxBin 2.0 也在 2016 年發表於《生物信息學期刊》(Bioinformatics 32(4):605-607)。截至 2017 月 4 月 27 日止,MaxBin 已被引用 164 次(Google Scholar; MaxBin 108 次;MaxBin 2.0 56 次),顯示其被廣泛地運用在各種微生體研究上。其準確性更受到國際專家學者的肯定:一項國際性的軟體評比結果指出 MaxBin 在分類個別細菌上「有著最高的準確率」,而且「能夠最精準地分類細菌」(Nature Methods 14: 1063-1071, 2017)。

我也運用自己開發的軟體分析了數個微生物體,並將研究成果發表在各國際期刊上(BioEnergy Res 8(3): 1031-1035, 2015;AEM 82(1): 255-267, 2016; ISME 10:833-845, 2016; mBio 7(4): e01106-16, 2016; mSystems 1(6): e00120-16, 2016; Nature Microbiology: in press)。未來我計畫繼續開發與維護這套軟體,讓它持續在微生物體的計算分析上扮演著不可或缺的角色。(文/吳育瑋,醫學科技學院醫學資訊研究所助理教授)