

簡介 DNA 定序技術的演進與第三代定序技術之奈米孔定序法

Brief introduction on the development of DNA sequencing technology and nanopore sequencing technology in third generation sequencing

生物技術課 龔美玲

前言

自從人類發現生命的遺傳物質為一種稱為 DNA (去氧核糖核酸)的長鏈聚合物後，DNA 序列的解析已成為生物遺傳研究的必要工作之一。DNA 透過的每個基本組成單位相接腺嘌呤(A)、胸腺嘧啶(T)、胞嘧啶(C)或鳥糞嘌呤(G)等 4 種鹼基的其中一種，串連組成遺傳密碼(例如 ACGTCTGACTTAA...)引發後續生物各種必需的機能。各種的定序技術應運而生，每一次定序技術的突破與蓬勃發展也造就生命科學研究速度與質量的提升。本文將依序介紹 DNA 定序技術的演進歷程及第三代定序技術—奈米孔定序之運用原理及應用，祈能使未來研究有更好的選擇。

第一代定序技術

第一代定序技術以 1977 年 Frederick Sanger 等人發明的「(雙脫氧)鍊終止法」(dideoxy chain termination) (或簡稱 Sanger 法)為主，透過添加移除 3'端氫氧基的雙去氧核甘酸(ddNTP)，隨機終止 DNA 的聚合反應，再根據膠體電泳的條帶大小讀出序列¹。第一代定序技術應用的高峰即屬 1990 年展開的人類基因體計畫，耗費 30 億美元，於 2003 年完成，總計定序大約 31 億個鹼基對(base pairs) (3.1Gb)，但由於技術上的限制，當時僅能完成 92%的圖譜。後人在 Sanger 法的基礎上進行改良，以螢光標定 ddNTP 與結合毛細管電泳技術，從手工操作發展至自動化核酸定序儀，大幅提高定序效能，最具代表的設備為 2001 年 Applied Biosystems (ABI)公司的 3730 系列定序儀²。目前單條 Sanger 定序的平均長度約 700~800 bp，

¹李思元、莊以光。2010。DNA 定序技術之演進與發展。生物醫學暨檢驗科學雜誌 22(2):49- 58。

²李曜珊。2017。簡介次世代定序技術及美國的法規管理。當代醫藥法規月刊 83:1-12。

最長可達 1,000 bp，準確度 99.99%，最多可同時進行 96 條毛細管的定序反應。

次世代定序技術

通量、費用、耗時上的限制使得第一代定序方法無法被大規模應用，例如個人化基因體定序、基因篩檢、臨床醫學診斷等，直到 2005 年第二代或稱次世代定序(next generation sequencing, NGS)技術問世，各家生物技術公司百花爭鳴，進入了人體基因體定序的降價競賽，從 2003 年的 8 仟萬美元到 2015 年降至 1 仟美元¹。NGS 可做到同時多樣本(pooling samples)、上百萬條 DNA 序列的讀取，結合微製程與光學偵測技術，透過反覆的試劑置換與偵測，讀序量從 Sanger 法的 ~2.6 Mb reads/day，飛躍式提高至 320~900 Gb reads/day，有效縮短定序時間並降低成本，NGS 儀器更在 2013 年首次獲得美國 FDA 核准應用於臨床診斷²。主流的三大 NGS 平台：Roche/454 (2005)、Illumina (2007)及 ABI SOLiD (2007)，雖然各自在採用的定序原理、擴增方法、讀序(read)長度、總讀序量及費用上有所差異，但基本流程皆能分為 4 個階段：樣本文庫(library)製備、樣本文庫擴增、定序反應，以及最後生物資訊資料分析。

NGS 的特點是快速、大量的小片段定序，以 illumina 的 HiSeq 為例，最長讀序為雙端(paired-end, PE) 2 x 150 bp，而 NovaSeq 則可定到 2 x 250 bp，1 Mb 的定序費用僅需要 0.07 塊美金²。一般的 F₂ 分離族群基因定位研究可透過簡化基因體定序技術(genotyping-by-sequencing, GBS)，利用 NGS 同時完成基因型分析與分子標誌開發，同樣以 illumina 平台為例，根據採用購買讀序量(例如 80 Gb)或包下通道(lane)方式的不同，定序費用約落在 5~17 萬元臺幣。但也因為定序片段小，NGS 需要有可參考的基因體序列才能進行較好的拼接組裝(assembly)，尤其基因體中的大量重複序列區域是常見的一大障礙，無法直接僅以 NGS 完成未知物種的全基因體定序。

第三代定序技術

為了持續突破定序成本，以及追求更簡化的定序方式，第三代定序融合了奈米製程技術，最大特點為單條長片段 DNA 的即時(real-time)定序；第三代定序又

可稱為「長讀取定序技術」(long-read sequencing)，目前平均的序列讀取長度可大於 1 萬個鹼基(10 kb)，能直接定序原始 DNA 樣本，省去前代定序技術需要的 PCR 增幅過程，也就可避開 PCR 擴增的錯誤率及偏好性問題，甚至能分析一部分 DNA 甲基化修飾。提到第三代定序，目前較廣為人知的為 Pacific BioSciences (PacBio)公司 2010 年推出的「單分子即時定序」(single-molecule real-time sequencing, SMRT)，以及 Oxford Nanopore Technologies (ONT)公司於 2012 年底推出的「奈米孔定序」(nanopore-based platform)，隨著不斷提升定序通量與降低錯誤率，使得 ONT 公司已成為近年歐洲最大的生技獨角獸之一，故以下僅針對「奈米孔(nanopore)定序技術」做更進一步的介紹。

(一)技術原理

nanopore 定序的技術核心為一片具有上千個奈米孔的電流感測晶片，所謂奈米孔是將經過修飾的蛋白質鑲嵌於人工合成的高電阻脂質雙層薄膜，結構仿造生物細胞膜上的跨膜蛋白，讓小分子仍可以通過膜進入細胞內；當施加電位差於薄膜兩端時，會驅動 DNA 分子藉由馬達蛋白(motor protein)靠近奈米孔表面並被解開雙股結構，經設計過的奈米孔只容許一條單股 DNA 通過，當帶有不同鹼基的核苷酸通過奈米孔時，會引起不同程度的電流強度改變，即時偵測每個奈米孔的特定電流變化，進而回推通過的 DNA 鹼基序列為何^{1,3}。

(二)設備優勢

nanopore 與其他定序技術的不同之處為不需要使用螢光分子標定與千萬元臺幣的精密光學量測定序儀，每秒可偵測 400~450 個鹼基，定序設備成本大幅降低。當 ONT 公司於 2014 年推出定價 1,000 美金起價、重量不到 100 克的可攜式 MinION 時最備受矚目，因其輕巧及操作的簡便性，使得定序研究工作不再被侷限於實驗室，不只田間、野外或海洋探測船上，甚至在南極大陸、國際太空站等極限環境都已有使用案例，也利於深入傳染病疫區，在極短時間內確認病毒等

³高唯真。2018。第三代定序—Oxford nanopore 技術及原理。臺灣生物科學研發策進社群 <https://investigator.tw/>。

病原的序列。

nanopore 定序設備的入手價格親民，具有易整合且滿足不同情境的低、中、高通量快速定序系列機型供使用者選擇，持續優化操作與 QC 分析軟體介面，加上 ONT 彈性、經濟的包裹式銷售方式，以及搭上 Covid-19 疫情期間快速鑑別、追蹤傳播鏈及病毒序列變異之檢測需求⁴；極具競爭力的檢測成本，使得 ONT 的定序設備迅速拓展規模與搶攻市占率，在植物基因體應用相關的發表也開始以倍數成長⁵。隨者奈米孔材質及數據分析演算法等相關定序技術越趨成熟，目前所宣稱準確度已可達 99% 以上。

(三) 硬體需求與定序樣品準備

然而採用 nanopore 定序仍需考量感測晶片(定價約 5 萬元臺幣)與試劑的成本，另外軟體操作需要之筆記型電腦需配備至少 16GB RAM、高階中央處理器(CPU)(例如至少 Intel i7 以上、規格至少達 4 cores/8 threads)、至少 1TB 的內接式固態硬碟(SSD)，以及為了分析電流物理圖形訊號的高階繪圖晶片(GPU)(或稱圖形處理器)(例如 NVIDIA GPU RTX 2060 SUPER 或更高階，並搭配至少 8GB 的 GPU memory)，才能做到即時序列讀取(base-calling)⁵。

雖然使用第三代定序通常有著追求長片段讀序的期待，但超長的 DNA 片段容易塞住奈米孔洞而影響定序效能，30~50 kb 為實驗室最常使用的定序片段長度，並需盡可能移除 10 kb 以下的片段，讓定序資源利用最佳化。物種本身的基因體大小與特性、樣品 DNA 萃取與純度品質、建庫與上機的熟練度，皆會影響平均讀序長度、最終總讀序量與單次分析的成本，而一般來說，動物的讀序量產出會優於植物。

(四) Nanopore 之應用潛力

由於 nanopore 的長讀序可以降低植物基因體組裝的難度，除了大基因體的

⁴財團法人生技醫療科技政策研究中心。2021。【獨角獸】英國 ONT 奈米孔單分子定序 點燃第三代基因定序戰火。新創幫 vol.021。

⁵ Oxford Nanopore Technologies 官網。https://nanoporetech.com/及 https://nanoporetech.com/resource-centre。

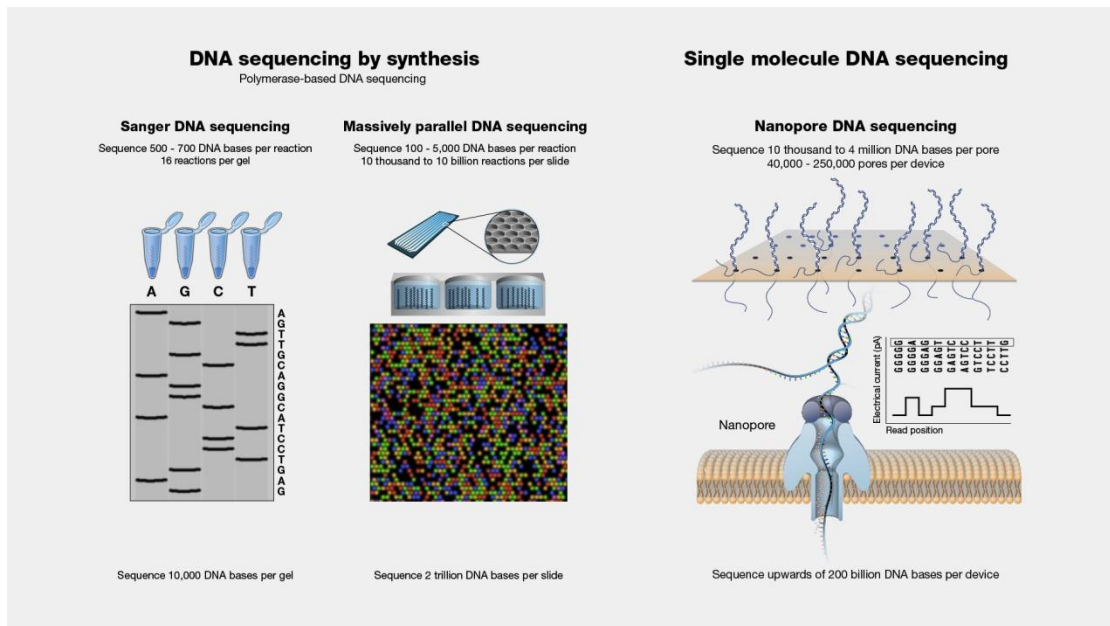
物種或多倍體植物的定序，可預見對植物物種內多品系的泛基因組(Pan-genome)或核心種原研究會更為容易，且能更準確的分析染色體結構變異，找出與外表性狀的關聯性，提供作物育種更多寶貴的訊息。

另一項應用為利用定序設備的小型化優勢，建立植物病毒基因組的快速檢測技術及流程，順應氣候變遷之下植物病毒檢測日漸成長的需求。例如法國農業研究發展國際合作中心(CIRAD)研究重要經濟作物—參薯(water yam, *Dioscorea alata*)，利用 Oxford Nanopore 的裝置 MinION 檢測病毒序列，與傳統定序結果比較，相似程度高達 99.8%。高準確度意味著能於作物栽種現場，特別是交通不便的地區，即時進行有效的病害診斷、監控與管理，省去採樣送驗的時間，避免病毒短時間大爆發的風險⁶。

結語

次世代定序克服第一代定序的通量限制與成本障礙，帶來龐大的序列資訊，而第三代定序又破解了 NGS 技術之短讀序難以解決的重複序列問題，讓科學家能截長補短，以大塊拼圖更容易去拼湊出基因體的全貌。原先人類基因體計畫中那未能解碼的 8% (大約 2 億個鹼基)區域，包含端粒、中節、核糖體 DNA 序列，也是藉由新世代的定序技術，終於在 2021 年完成史上第一個完整的人類基因體 (T2T-CHM13)，並發表在今(2022)年的《科學》(Science)期刊上。技術的革新能為人類帶來福祉，隨著第三代定序技術不斷提升準確度與通量，結合小型化、長讀序等優勢，以及持續開發更成熟且容易操作的組裝分析軟體，相信未來能以更低的價格實現定序普及化與擴大應用領域，讓定序技術更貼近人們的生活。

⁶ 農業科技決策資訊平台。2019。 <https://agritech-foresight.atri.org.tw/article/contents/1688>。



圖一、定序技術的演進從第一代 Sanger 定序到次世代定序(NGS)，再到第三代 Nanopore 定序之技術原理示意圖。(圖片來源：

<https://www.genome.gov/genetics-glossary/DNA-Sequencing>)