

次世代定序在白蝦疾病研究之應用

利淑如、葉怡均、蘇義哲、吳豐成
水產試驗所東港生技研究中心



前言

近年來蝦類新興疾病不斷出現，如蝦類早期死亡綜合症 (shrimp early mortality syndrome, EMS) 每年造成數百萬美元經濟損失，導致全球白蝦 (*Litopenaeus vannamei*) 產業接連受到各種疾病之重創 (Yu, et al., 2020)。為徹底解決問題以達成白蝦產業永續經營之目標，需同時進行蝦類疾病管理與完善育種系統。白蝦傳染性疾病是白蝦養殖管理中面臨的重要關鍵問題之一，傳統上，病原體檢測研究依賴於傳統病理檢測及微生物檢測，既費力又費時。

在水產養殖基因體中，因白蝦基因體數據不足而無法完成定序和組裝全基因組，再者全球在蝦基因體解序進度緩慢，主要是因白蝦具有龐大的基因體 (Swathi et al., 2018) 及 DNase 活性強、染色體數量多、雜合性和重複元件水平高，難以分離出高分子量 DNA，使其全基因體定序面臨了重大挑戰 (Abdelrahman et al., 2017)。然而在缺乏蝦全基因體信息的情況下，另有極力開發分子遺傳資源，例如遺傳圖譜 (gene mapping)、表達序列標籤 (shrimp expressed sequence tags)、轉錄體 (transcriptomes)、數量性狀基因座 (quantitative trait loci) 鑑定、微陣列

(microarrays)、差異表達基因 (differential gene) 和細菌人工染色體 (shrimp bacterial artificial chromosome) 等，但仍無法深入探討基因體的開發。

在 2000 年代初期，次世代定序 (next generation sequencing, NGS) 技術的出現，徹底改變了基因體學領域，也為以前無法進行的相關研究領域提供了新的解決方向。透過收集大量細菌的基因體數據，NGS 能夠在一次測序中考慮所有相關訊息。雖然使用 NGS 技術得到的基因體數據成果，但 NGS 技術一開始在蝦病研究中並未得到充分利用，因為它需要高昂的成本 (Nkili-Meyong et al., 2016)。隨著次世代定序技術成熟，定序成本降低，白蝦全基因體已被定序和組裝 (Zhang et al., 2019)。過去幾年白蝦疾病 NGS 相關的研究顯著增加。大幅拓展基因體研究的廣度與深度，為養殖蝦類提供遺傳信息方面取得了巨大進展。

NGS 技術的演變與應用

NGS 定序儀常見的為 Illumina、ABI 和 Roche 平台，其主要步驟相似，其中以 Illumina 的次世代定序 (圖 1) 高通量篩選在過去幾年中發展迅速，在國外已經應用於生

物學和醫學研究領域上，且被證明是發現未知病毒的有用方法，其次世代定序實驗流程包含核酸片段化 (fragmentation)、建構資料庫 (library construction)、高通量定序 (high-throughput sequencing) 和分析 (analysis) 共四個步驟，其原理是利用超聲波將 DNA 打斷成 200–500 bp 的片段大小，然後於片段兩端接上銜接子 (adapter)，再將已接銜接子的 DNA 片段放入到表面帶有互補銜接子序列的晶片 (flowcell)，銜接子互相配對後讓 DNA 片段吸附於 flowcell 上，透過橋式聚合酶連鎖反應合成另一股，進行 DNA 複製，接續使用邊合成邊定序，加入不同已標定特定可移除螢光分子之鹼基 (dNTP) 及合成反應試劑，反覆讓螢光移除並以雷射激發偵測螢光訊號，並將其轉換為一系列 A、T、G 和 C 的輸出，達到高速且大量的 DNA 定序。藉由移除銜接子並低質量讀取，對所有生成的原始數據進行過濾以獲得高質量的干淨讀取並將轉錄體組裝完成，藉

由組裝完成後之單基因 (unigene) 再與數據庫 Nr (NCBI non-redundant protein sequences) ; Nt (NCBI non-redundant nucleotide sequences) ; Pfam (蛋白質家族, <http://pfam.xfam.org/search>); KOG/COG (<http://www.ncbi.nlm.nih.gov/COG/>) ; Swiss-Prot (<http://www.ebi.ac.uk/uniprot/>) 和 GO (<http://www.geneontology.org/>) 比對進行基因表現、功能註釋及分子途徑預測分析。註釋其可能具有的功能 (gene ontology, GO) 和所參與的代謝途徑 (Kyoto Encyclopedia of Genes and Genomes, KEGG)。GO 註釋分為三大類：分子功能 (molecular function)、細胞組分 (cellular components) 和生物過程 (biological process)；而 KEGG 是一個整合了基因體、化學和系統功能資訊的綜合資料庫。藉由數據資料的整合，將有助後續對特定基因群進行功能富集分析 (functional enrichment analysis)，並瞭解特定狀況下差異性基因群所代表的生物意義。

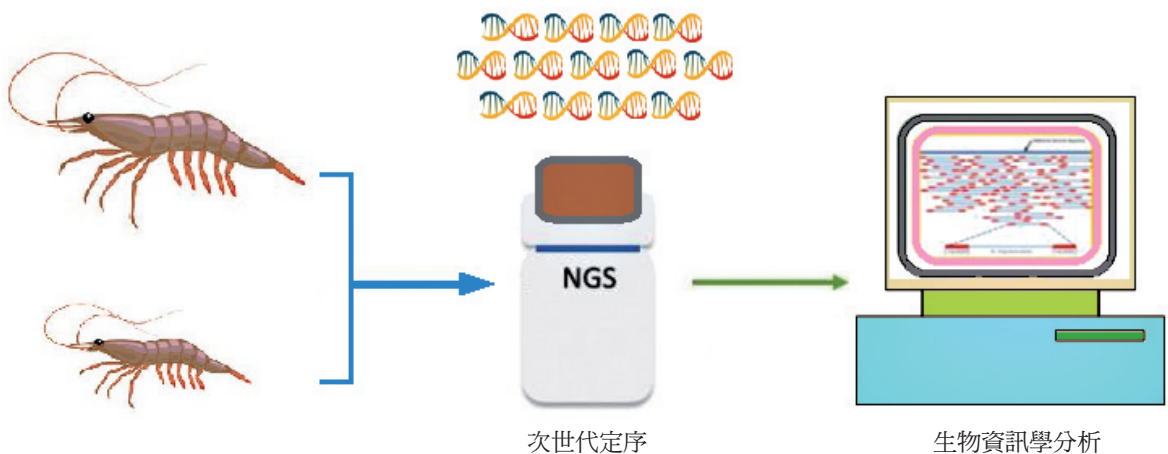


圖 1 Illumina 的次世代定序

NGS 的發展亦可利用物種間的定序資訊比對演化親源性，或是利用外觀差異的取樣定序，尋找生物體內基因表現量的差異與各種表觀基因體變異之遺傳關鍵基因，在現代生物技術工具的應用上，徹底改變了重要的水產養殖物種的基因體學，其重大突破來自報導的南美白對蝦第一個成功的基因體，該基因體使用了定序技術的組合，如 Illumina, PacBio 和 BAC 等 (Zhang et al., 2019)。然而，在蝦類養殖中所發生的疾病，像是世界衛生組織 OIE 表列的白點蝦病毒 (white spot syndrome virus, WSSV)、蝦類早期死亡綜合症 (shrimp early mortality syndrome, EMS)、桃拉症病毒 (Taura syndrome virus, TSV)、傳染性肌肉壞死病毒 (infectious myonecrosis virus, IMNV)、黃頭病毒 (yellow head virus, YHV)、傳染性皮下及造血組織壞死病毒 (infectious hypodermal hematopoietic necrosis, IHNV) 與壞死性肝胰腺炎 (necrotizing hepatopancreatitis, NHP)，以及近年來造成蝦類大量死亡與嚴重減產的幾種新興疾病，如對蝦肝胰腺微孢子蟲症 (enterocytozoon hepatopenaei, EHP)、急性肝胰腺壞死病 (acute hepatopancreas necrosis disease, AHPND)、偷死野田村病毒 (covert mortality nodavirus, CMNV) 與蝦血細胞虹彩病毒 (shrimp hemocyte iridescent virus, SHIV) 等，因對分子生物學和生理功能缺乏了解的狀態下，能夠尋找出可利用的基因體資訊已變得至關重要。

NGS 在白蝦疾病研究的應用探討

Fang 等 (2019) 採取生長 100 天且體長僅 2–3 cm 的發育不良的南美白對蝦和無特定病原 (SPF) 健康幼蝦進行總 RNA 提取，cDNA 文庫構建和利用 NGS 的方法將轉錄體定序 (transcriptome sequencing)，並使用 Illumina 系統完成測序，其結果得到 108221 個單基因，將單基因進行數據庫的比對並註釋功能，在 NR、NT、KO、KOG、SwissProt、PFAM 和 GO 中註釋的單基因分別為 26005 (24.02%)、12468 (11.52%)、12194 (11.26%)、16806 (15.52%)、23050 (21.29%)、29084 (26.87%) 和 29911 (27.63%)。根據 GO 分析，總共 29912 個單基因被分為三個主要功能類別，南美白對蝦的細胞生物過程中涉及 1235 種的單基因，其次是負責合成細胞成分有 369 種單基因，173 種單基因具有一定的分子功能，如催化活性，分子功能調節劑，轉運蛋白活性和結合能力，為了確定發育不良和健康的南美白對蝦中的生物學途徑，所有單基因都被映射到 KEGG 數據庫中的標準途徑，總共有 15205 個單基因被分配到 229 條 KEGG 途徑，其中，前 10 條 KEGG 途徑是：信號轉導 (1526)、轉譯 (1438)、運輸和分解代謝 (1194)、內分泌系統 (1051)、蛋白質摺疊、分類和降解 (939)、碳水化合物代謝 (780)、免疫系統 (703)、氨基酸代謝 (615)、細胞通訊 (589) 和消化系統 (529)。而在差異表達基因上，發育不良的南美白對蝦中發現了 13 個與病毒相關的上調基因 (表 1)，而在健康蝦類中則未發現，其中發現有 5 個與微小核糖核酸病毒相關的基因片段，例如 RNA 解旋酶 (RNA helicase)，RNA 依賴性 RNA 聚合酶 (RNA-directed RNA polymerase)

和微小 RNA 病毒衣殼蛋白 (Picornavirus capsid protein) 等 (表 2)，微小核糖核酸病毒是小的無包膜 RNA 病毒，25–30 nm 二十面

體，它們可以感染並在細胞的細胞質中複製，且與南美白對蝦生長遲緩的發病機理相關，亦可在人類和動物中引起多種疾病。此

表 1 在發育不良的南美白對蝦中 13 個與病毒相關的上調基因 (Fang et al., 2019)

基因 ID	發育不良白蝦序列數量	健康白蝦序列數量	Swiss-Prot 資料庫描述	蛋白質家族描述	細胞組分描述
c75311_g1	352.52	0	複製酶多聚蛋白 OS=蟋蟀麻痺病毒	RNA 解旋酶	RNA 依賴性 RNA 聚合酶複合物
c78682_g2	351.11	0		馬動脈病毒 Nsp2 型半胱氨酸蛋白酶	細胞外區域/膜
c76407_g1	63.83	0		腺病毒 IVa2 蛋白質	
c73240_g1	46.94	0			病毒衣殼
c56446_g1	73.69	0		病毒 RNA 解旋酶/病毒甲基轉移酶	RNA 依賴性 RNA 聚合酶複合物
c65243_g1	23.47	0		桿狀病毒多面體包膜蛋白	病毒衣殼/病毒被膜
c68785_g1	21.592	0		微小 RNA 病毒衣殼蛋白	病毒衣殼
c69554_g1	60.55	0		微小 RNA 病毒衣殼蛋白	病毒衣殼
c20203_g1	29.57	0		皰疹病毒主要外層包膜糖蛋白 (BLLF1)	病毒被膜
c79968_g1	122.51	0		非洲豬瘟病毒多基因家族 360 蛋白	
c57532_g1	86.83	0		N 端病毒多聚蛋白	病毒被膜
c79999_g1	3179.27	0	複製酶多聚蛋白 OS=蟋蟀麻痺病毒	微小 RNA 病毒衣殼蛋白	RNA 依賴性 RNA 聚合酶/病毒衣殼
c79905_g1	180.25	0		輪狀病毒非結構蛋白 NSP3	膜

表 2 五個病毒相關基因片段的蛋白質家族資料庫 (PFAM) 分析僅出現在生長不良的南美白對蝦中 (Fang et al., 2019)

片段編碼	基因長度	蛋白質家族預測
F1	6868	RNA 解旋酶/RNA 依賴性 RNA 聚合酶
F2	4707	病毒甲基轉移酶/RNA 解旋酶/RNA 依賴性 RNA 聚合酶
F3	1518	RNA 解旋酶/RNA 依賴性 RNA 聚合酶/微小 RNA 病毒衣殼蛋白
F4	4392	微小 RNA 病毒衣殼蛋白/二順反子病毒科 VP4/蟋蟀麻痺病毒 (CRPV) 衣殼
F5	9609	RNA 解旋酶/RNA 依賴性 RNA 聚合酶

外，被病毒感染的南美白對蝦不能生長到正常大小的原因，可能是由於被病毒感染的蝦不能有效代謝蛋白質和糖（澱粉和蔗糖）的緣故，在植物上，病毒感染和蛋白質和糖的代謝關係已得到很好的研究。但是，沒有關於動物的任何信息。無論如何，蛋白質和糖代謝的下降無疑會抑制植物或動物細胞的生長。

另一位學者 Liao 等 (2020) 為了研究十足類虹彩病毒 1 (Decapod iridescent Virus 1, DIV1) 的感染機制，比較南美白對蝦血球細胞感染和未感染 DIV1 的轉錄組分析，發現 DIV1 感染後，有 1112 個差異表達基因 (differential gene expression)，其中鑑定出 889 個基因被上調，而 223 個基因被下調。這些基因主要參與果糖、甘露糖、碳和磷酸肌醇代謝。在這些代謝途徑中，磷酸三糖異構酶 (triose-phosphate isomerase, TPI) 家族是最引人注目的差異表達基因，因為它參與了多個代謝過程。TPI 的三種類型為 LvTPI-like，LvTPI-Blike 和 LvTPI-Blike1，通過 RNA 干擾獲得了基因沉默，LvTPI-like 和 LvTPI-Blike1 沉默導致南美白對蝦的高死亡率。然而，在南美白對蝦感染 DIV1 時，LvTPI-like 和 LvTPI-Blike 的沉默減少了 DIV1 的複製，所有結果顯示 TPI-like 基因在此過程中起著重要作用。

透過 NGS 比較不同來源的基因體序列來識別細菌菌株的遺傳變異性，快速診斷技術對於 AHPND 傳播的防止至關重要。藉由 NGS 的應用，全基因體定序 (whole genome sequencing, WGS) 和靶向擴增子定序 (targeted amplicon sequencing, TAS) 是細菌

病原體研究中常用的兩種方法 (Lefterova et al., 2015)。WGS 是對細菌的整個基因體 DNA 序列的分析，序列經軟體比對後，發現了一個大型染色體外質粒 (~69 kbp)，pVA 攜帶兩個毒素基因，基因體研究在很大程度上有助於更深入地了解 VP_{AHPND} 菌株的基因體，提供重要的遺傳信息，允許針對疾病設計干預措施，特別是在基於病原體研究的 PCR 方法開發以及引起白蝦 AHPND 的細菌檢測。NGS 將基因體學整合到水產養殖研究中，加重了基因體學對水生疾病研究的重要性 (Yu et al., 2020)。

結語

白蝦在病毒感染機制探討上，隨著白蝦疾病的前人研究，根據次世代高通量定序所得到的訊息，結合生物晶片，觀察不同狀況時序下的病毒基因表現變化，與宿主之間的交互影響。在對蝦類的蛋白質體研究上，則以基因體表現概況作為基礎，藉以推論其中可能存在的生理調控機制與關連代謝路徑。這些研究成果有利於我們更加瞭解病原體感染宿主細胞的過程，了解免疫反應與病原複雜的感染機制，找出調控病原複製與感染細胞的關鍵蛋白，進而探尋可能的致病蛋白，有效地抑制病原體的感染與蔓延，減少對養殖產業所帶來的衝擊。在白蝦養殖遺傳學研究上，對於欲選育性狀的蝦可根據次世代高通量定序所得到的訊息，結合生物晶片，作為未來育種的參考依據，有助於抗病品系的南美白對蝦的生產。