

應用奈米孔定序平台進行蝴蝶蘭褐斑病菌基因體 *de novo* 組裝之研究

杜元凱¹ 蔡佳欣² 林彥君³ 歐柏均³ 陳涵葳^{1,*}

摘要

杜元凱、蔡佳欣、林彥君、歐柏均、陳涵葳。2022。應用奈米孔定序平台進行蝴蝶蘭褐斑病菌基因體 *de novo* 組裝之研究。台灣農業研究 71(2):171–183。

蝴蝶蘭出口產值達 40 億元，是台灣外銷花卉之首；高溫多濕的栽培環境，易發生由 *Acidovorax avenae* subsp. *Cattleyae* (AAC) 所引起的蝴蝶蘭褐斑病，影響苗株出口造成產業損失。本研究應用奈米孔定序平台對兩個 AAC 菌株收集系進行基因體解序，AAC_10001 與 AAC_9906R 之長讀序資料量、平均讀序長度、讀序品質分別為 1.49 Gbp、18.7 kb、11.9 與 1.96 Gbp、17.29 kb、12.3，基因體 DNA 具有高度連續性，定序品質佳。對解序資料進行 *de novo* 組裝，AAC_10001 與 AAC_9906R 之連續體數量分別為 6 個與 5 個，最大的連續體長度為 5.7 Mb 與 5.6 Mb，序列 guanine-cytosine (GC) 含量約為 67.4% 與 67.5%，基因體平均定序深度為 230 倍。AAC_10001 與 AAC_9906R 基因體結構比對結果顯示，兩者具有 12 個主要的局部共線性區塊 (locally collinear blocks; LCB)，其中 AAC_10001 菌株 LCB 2 有一段 132 kb 的基因體島遺傳元件插入，包含一個 DNA 修復蛋白、一群參與第四型分泌系統的蛋白、一個整合酶與一個轉位酶，可能影響菌株致病性、共生性、代謝特性、環境適應性或對抗生素產生抗性。本研究使用奈米孔定序平台對蝴蝶蘭 AAC 菌株進行基因體解序，建立細菌基因體長片段讀序生物資訊分析流程，完成 AAC 菌株基因體序列 *de novo* 組裝與初步分析，是國內首例以三代定序技術完成的蝴蝶蘭褐斑病細菌基因體研究，可提供未來準確診斷和長期監測的方向，期望有助於國內蘭園病害防治策略。

關鍵詞：奈米孔定序技術、蝴蝶蘭褐斑病、基因體解序、細菌抗藥性。

前言

蝴蝶蘭褐斑病是由 *Acidovorax avenae* subsp. *Cattleyae* (AAC) 所引起，為短桿型具有鞭毛的革蘭氏陰性菌，好發於高溫多濕環境。AAC 主要分布在菲律賓、義大利、美國、葡萄牙、澳洲及台灣等地 (Ding *et al.* 2010)，其典型病徵為葉片出現水浸狀斑點，隨後轉為暗褐色或黑色不規則凹陷壞疽斑，病徵周圍具有明顯黃暈，病斑持續擴大會導致葉片黃化乾枯，嚴重時植株死亡。依據財政部統計資料顯示，2020 年

蝴蝶蘭出口總值為 40 億元，是台灣外銷花卉之首，蝴蝶蘭多採溫室密集栽培，一旦染病便快速擴散，而潛伏苗株上的 AAC 可能影響商品檢疫，造成產業損失 (Hseu *et al.* 2011)。研究指出 *A. avenae* 可透過基因水平轉移機制獲得嘉賜黴素 (kasugamycin) 抗性基因 (Yoshii *et al.* 2015)，而國內業者常以鏈黴素進行蝴蝶蘭褐斑病防治，長期施用易產生抗藥性變異菌株，造成防治工作越日趨困難 (Huang 2001)。透過細菌基因體解序與分析，有助建立菌株外表型差異與其基因型之關聯，如生長勢、致

投稿日期：2021 年 3 月 15 日；接受日期：2021 年 11 月 18 日。

* 通訊作者：swaychen@tari.gov.tw

¹ 農委會農業試驗所生物技術組助理研究員。台灣 台中市。

² 農委會農業試驗所植物病理組助理研究員。台灣 台中市。

³ 農委會農業試驗所生物技術組計畫助理。台灣 台中市。

病力、移動力、生物膜形成，或是對抗生素的耐受性等；*A. citrulli* M6 菌株基因體序列組裝研究，證實了低拷貝數的 pACM6 質體的存在，且該質體具有 Fic-VbhA 抗毒性模組，可能影響 M6 菌株對抗生素的耐受性 (Yang *et al.* 2019)。此外，比對柑橘潰瘍病之 *Xanthomonas citri* 菌株基因體序列，結果支持質體融合 (fusion) 與重組的發生，且提示傳統栽培管理會提高環境選拔壓力，可能影響致病型 (pathotype) 之進化 (Huang *et al.* 2021)。

次世代定序 (next-generation sequencing; NGS) 技術藉由橋式聚合酶連鎖反應 (solid-phase bridge polymerase chain reaction; solid-phase bridge PCR)，可對小於 450 bp 之短讀序 (short reads) 進行定序，具有高通量、高準確性與低成本之優點，但短讀序無法正確拼接高度重複序列區域，基因體序列組裝時易產生序列間隙 (gaps)。近年快速成長的第三代定序技術，著重於即時、單分子與長片段讀序 (long reads) 技術之發展，如 Oxford Nanopore Technologies (ONT) 所開發的奈米孔定序 (nanopore sequencing) 以及 Pacific Biosciences (PacBio) 研發的單分子即時 (single-molecule real time; SMRT) 定序技術等，其中 PacBio SMRT 讀序長度為 10–15 kb，ONT 奈米孔讀序長度更可達到 200–300 kb，長片段定序可彌補 NGS 序列組裝時，重複序列、二級結構與極端 guanine-cytosine (GC) 含量所造成的影響。目前簡易式奈米孔定序平台 (MinION) 售價僅需 1,000 美元，連結筆記型電腦即可進行定序工作，單個定序晶片 (flow cell) 可產生 30 Gb 定序資料量，對原始 DNA 片段進行讀序，可避免 PCR 擴增的錯誤及鹼基偏好的影響。相對於其他定序平台而言，MinION 具有攜帶方便、超長讀序，即時監控定序情形等優點，廣泛應用於微生物鑑定、微生物抗藥性診斷、微生物基因體組裝與基因水平轉移等研究領域。如溫室番茄細菌性斑點病 (*Pseudomonas syringae* pv. *tomato*; *Pto*) 菌株層級的診斷 (Mechan Llonatop *et al.* 2020)，或黑麥草種子上新型 *Xanthomonas* 屬菌株之鑑定 (Li *et al.* 2020)。本研究使用 MinION 對蝴蝶蘭 *AAC* 菌株進行基因體解序，建立細菌基因體長片段讀序生物資訊分析

流程，完成 *AAC* 菌株基因體序列 *de novo* 組裝與初步分析，是國內首例以三代定序技術完成的蝴蝶蘭褐斑病細菌基因體研究，可提供未來準確診斷和長期監測的方向，期望有助於國內蘭園病害防治策略。

材料與方法

蝴蝶蘭褐斑病細菌基因體文庫製備與定序

本研究所使用的 *AAC* 菌株收集系由農試所植病組蔡佳欣博士提供，包括 AAC_10001 與 AAC_9906R。*AAC* 菌株接種於 lysogeny broth (LB) 液體培養基，於 37°C 轉速 100 rpm 條件下培養 2 d，並使用 cetyltrimethylammonium bromide (CTAB) extraction buffer 進行 DNA 萃取，接續經 phenol/chloroform/isoamyl alcohol (25 : 24 : 1) 去除蛋白質和有機溶劑，再以異丙醇沉澱 DNA，最後以 Nanodrop (Thermo Fisher Scientific, Wilmington, DE, USA) 進行 DNA 定量與品質檢測。基因體文庫製備使用 Ligation Sequencing Kit (SQK-LSK109, Nanopore, Oxford, UK) 和 Kapa Hyper Prep Kit (KK8502, Roche, Pleasanton, CA, USA)，流程簡述如下：

1. DNA end-repair 和 dA tailing：取 1 μ l 待測基因體 DNA，加入 End Repair & A-Tailing Buffer (Roche) 和 End Repair & A-Tailing Enzyme Mix (Roche)，於 20°C 下反應 60 min 和 65°C 下反應 30 min。
2. Adapter Ligation and Clean-up：完成端點修飾的基因體 DNA 加入 Adapter Mix (AMX, Nanopore)、Ligation Buffer (Roche) 和 DNA Ligase (Roche)，於 20°C 反應 4 h，隨後放入 4°C 冰箱持續進行接合反應。反應完成後使用 Agencourt[®] AMPure XP beads (Beckman Coulter Life Sciences, Indianapolis, IN, USA) 進行樣本純化，並使用 Long Fragment Buffer (LFB, Nanopore) 移除小片段 DNA。
3. Priming and loading the SpotON flow cell：本研究使用 MinION 機型的 R9.4.1 版本 Spot ON 定序晶片 (FLO-MINSP6, Nanopore)，依照原廠指示步驟進行 Priming Step 並將基因

體 DNA 與 Sequencing Buffer (SQB, Nanopore)、Loading Beads (LB, Nanopore) 均勻混合，接續將此混合溶液滴入晶片的樣本槽，最後按照 MinKNOW 軟體 (3.6.0 版本) 指示設定並執行定序程序，總定序時間為 4 h。

蝴蝶蘭褐斑病細菌基因體序列 *de novo* 組裝與分析

MinION 輸出的原始資料檔案格式為 FAST5，定序資料前處理首先以 Guppy 套件 (0.0.7a1 版本) 進行鹼基判定 (basecalling)，再以 Porechop (0.2.4 版本) 套件移除 adapter 序列，最後以 NanoPlot 套件 (1.33.0 版本) 進行序列品質控管 (quality control, QC)，並接續進行基因體序列分析，包括：

1. 基因體組裝：經品質控管後的序列資料使用 Unicycler 套件 (0.4.8 版本) 進行基因體 *de novo* 組裝。以 Bandage 套件 (0.8.1 版本) 呈現視覺化圖形協助組裝判定。
2. 定序深度分析：以 BWA (0.7.17 版本) 套件將品質控管後的序列資料對 AAC 基因體解序資料進行序列比對，再以 Bedtools 套件 (2.28.0 版本) 進行定序深度計算。
3. 基因體 GC 含量分析：AAC 基因體解序資料以 GC Content Plot Online (<http://www.endmemo.com/bio/gcdraw.php>) 功能進行 GC 含量分析與繪圖。

4. 基因體註解：使用 Prokka 套件 (1.14.5 版本) 進行基因註解。
5. 基因體序列結構分析：使用 MUMmer 套件 (4.0 版本) 對基因體解序資料進行比對，比對條件設定連續長度大於 1 kb，且同源性大於 90% 以上，並將結果繪製共線性圖。以 Mauve 套件 (2.4.0 版本) 進行基因體結構與序列重組分析。

結果與討論

本研究利用奈米孔定序平台進行 AAC 菌株基因體定序，AAC_10001 與 AAC_9906R 菌株分別取得 1.49 Gbp 和 1.96 Gbp 的長讀序資料，讀序長度分布如圖 1 所示。AAC_10001 菌株共計有 80,694 條讀序，平均讀序長度為 18.47 kb，讀序長度中位數 (N50) 為 31.47 kb，平均讀序品質為 11.9；而 AAC_9906R 得到 113,120 條讀序，平均長度為 17.29 kb，長度中位數 35.22 kb，讀序品質為 12.3 (表 1)。奈米孔定序技術原理是在薄膜上鑲嵌 α -hemolysin、MspA 或 CsgG 奈米孔洞跨膜蛋白，給予穩定電壓後離子通過奈米孔洞可形成固定電流，待測 DNA 片段與奈米孔洞跨膜蛋白會以主動運輸的方式讓單股 DNA 通過薄膜，此時序列上的鹼基會改變奈米孔洞的電流訊號，記錄電流訊號變化即可得到 DNA 片段序列，

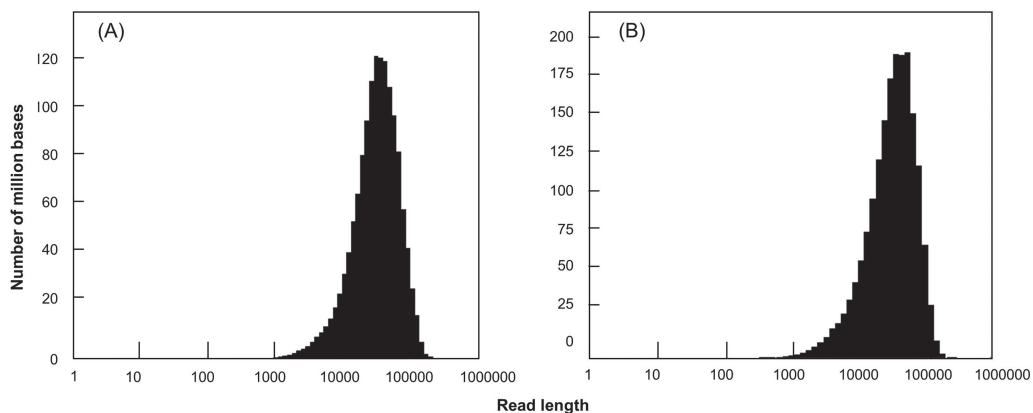


圖 1. *Acidovorax avenae* subsp. *cattleyae* (AAC) 基因體奈米孔定序之長度分布。(A) 菌株 10001；(B) 菌株 9906R。

Fig. 1. Read-length distributions of (A) *Acidovorax avenae* subsp. *cattleyae* (AAC) strain 10001 and (B) strain 9906R. Data obtained by nanopore sequencing.

表 1. *Acidovorax avenae* subsp. *cattleyae* (AAC) 菌株 10001 與菌株 9906R 基因體之奈米孔定序平台定序結果。

Table 1. Post-sequencing results of the *Acidovorax avenae* subsp. *cattleyae* (AAC) strain 10001 and 9906R genome.

Nanopore statistics	AAC_10001	AAC_9906R
No. of reads	80,694	113,120
Mean read length (bp)	18,474.8	17,291.2
Median read length (bp)	12,835.5	9,661.0
Read length N50 (bp)	31,473.0	35,216.0
Mean read quality ^z	11.9	12.3
Median read quality	12.2	12.7
Total bases (bp)	1,490,806,246	1,955,978,135

^z Phred score.

因此奈米孔定序長度完全由待測 DNA 片段品質決定，10–30 kb 最為常見 (Amarasinghe *et al.* 2020)，最長甚至可達 2.3 Mb (Payne *et al.* 2019)，此次上機結果多數序列超過 10 kb，且集中在 15–35 kb 長度區間，顯示本研究所製備的基因體 DNA 完整性好、品質佳。

Unicycler 套件可選擇 conservative、normal、bold 三種模式進行基因體組裝，其中 conservative 模式會帶入短讀序橋梁 (short read

bridge)，bold 模式則會盡可能的合併多個連續體 (contig)；本試驗選擇 normal 模式進行 AAC 菌株基因體 *de novo* 組裝，該模式在組裝的完整性與正確性間取得一個折中的方案，AAC_10001 與 AAC_9906R 分別得到 6 個與 5 個 contigs，其中最大的 contig 為 5.7 Mb 與 5.6 Mb，利用 Bandage 套件協助定序資料視覺化呈現 (圖 2)，理想的細菌基因體組裝為單個或數個環狀 contigs，代表細菌染色體 DNA 與數個質體 DNA (Wick *et al.* 2015)，5.7 Mb 之 AAC_10001_01 contig 首尾沒有出現重疊序列，因此呈現線性，而 5.6 Mb 之 AAC_9906R_01 可順利組成環狀，後續便以上述 2 個 contigs 作為 AAC_10001 與 AAC_9906R 的染色體 DNA 進行分析。

AAC 菌株基因體 DNA 鹼基組成分析結果如圖 3A、B，AAC_10001 與 AAC_9906R 之 GC 含量，分別為 67.4% 及 67.5%。定序資料經 Bedtools 套件進行定序深度分析，AAC 基因體平均定序深度約為 230 倍，以 AAC_10001 為例 (圖 3C)，定序深度在基因體中分布相當均勻。在 Illumina 短讀序平台下，細菌基因體組裝之正確率，受到分析套件、定序覆蓋深度、品質過濾步驟，以及參考序列選擇等

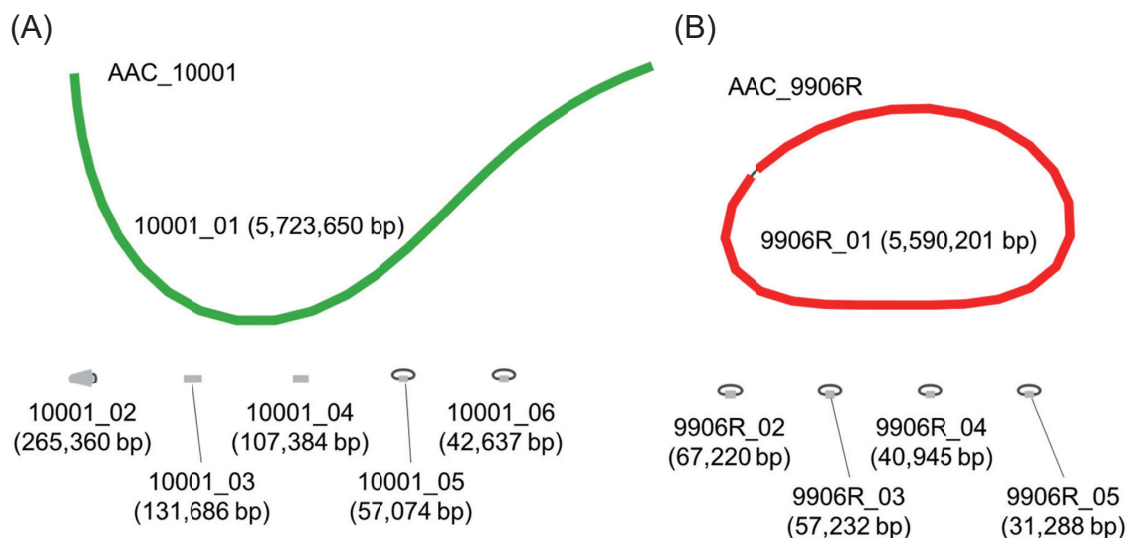


圖 2. *Acidovorax avenae* subsp. *cattleyae* (AAC) 基因體序列 *de novo* 組裝結果。(A) 菌株 10001；(B) 菌株 9906R。
Fig. 2. *De novo* assembly graph of (A) *Acidovorax avenae* subsp. *cattleyae* (AAC) strain 10001 and (B) strain 9906R genome.

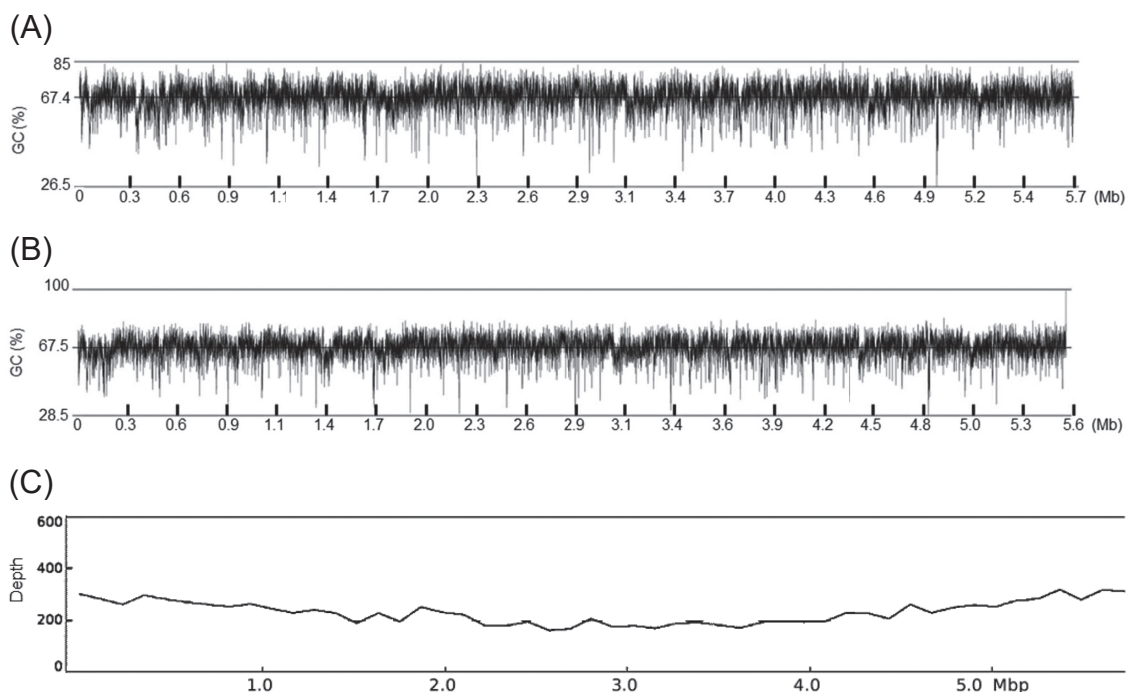


圖 3. *Acidovorax avenae* subsp. *cattleyae* (AAC) 基因體解序 GC 含量分析與定序深度分析。(A) 菌株 10001 與 (B) 菌株 9906R 之 GC 含量分析；(C) AAC_10001 菌株基因組各區域定序深度。

Fig. 3. Results of guanine-cytosine (GC) content and sequencing depth across the assembly of (A) *Acidovorax avenae* subsp. *cattleyae* (AAC) strain 10001 and (B) strain 9906R. (C) Read coverage distribution of AAC_10001.

影響；*Listeria monocytogenes* 解序資料配合參考序列進行組裝，顯示定序深度在 50 倍以上時，可得到 99.9% 組裝正確率 (Pightling *et al.* 2014)；ONT 定序平台需要更高的定序覆蓋深度，根據 Sutton *et al.* (2021) 的研究，*Escherichia coli* 基因體之 ONT 定序資料，以 Canu 或 Flye 套件進行組裝時，定序深度在 62 倍時，才可完成高連續性的基因體組裝，並建議以 ONT 定序平台資料進行基因體組裝時，定序覆蓋深度至少要達到 100 倍才会有可靠的結果。*Pseudomonas aeruginosa* 基因體組裝結果亦顯示，單獨使用 ONT 定序資料進行組裝，在 560 倍的超深覆蓋率情況下其正確率可達 99.8%，但深度增加並不能再提升正確率 (Molina-Mora *et al.* 2020)。

蝴蝶蘭褐斑病 *A. avenae* subsp. *Cattleyae* 命名最早由 Willems *et al.* (1992) 所發表，但近年利用 16S rDNA 進行基因型分析後，則建議將 *A. avenae* subsp. *Cattleyae* 修正為 *A. cat-*

tleyae (Schaad *et al.* 2008)，美國國家生物技術資訊中心 (National center for biotechnology information; NCBI) 發佈之 *A. cattleyae* 參考序列註冊號為 NZ_CP028290，是來自 CAT98_1 菌株的 Illumina MiSeq 定序結果 (Zeng *et al.* 2017)。本研究比較 AAC_10001、AAC_9906R 與 CAT98_1 之序列，探討 ONT 長讀序組裝與 Illumina 短讀序組裝之差異 (表 2)。AAC_10001 與 AAC_9906R 基因體序列長度與 CAT98_1 相當接近，分別為 5.7、5.6 和 5.6 Mb，序列 GC 含量也近似，分別為 65.2、67.4 和 67.5%，但 CAT98_1 為短讀序組裝結果，其 contigs 數量為 428 個，較 AAC_10001 與 AAC_9906R 之 contig 數量多，且 gaps 共計 1.1 Mb，約占 CAT98_1 基因體序列之 20%。CAT98_1 基因體序列經 Prokka 套件註解，共包含 4,564 個編碼序列 (coding DNA sequence; CDS)、3 個 rRNA 序列，以及 53 個 tRNA 序列；本試驗完成之 AAC 基因體解序資料則有 8,630 個以上的 CDS，

表 2. *Acidovorax cattleyae* 菌株 CAT98_1、*Acidovorax avenae* subsp. *cattleyae* (AAC) 菌株 10001 與菌株 9906R 基因體之解序與註解結果比較。

Table 2. Genome statistics of *Acidovorax cattleyae* strain CAT98_1, *Acidovorax avenae* subsp. *cattleyae* (AAC) strain 10001 and 9906R.

Item ^z	Short reads	Long reads	
	CAT98_1	AAC_10001	AAC_9906R
Total length (bp)	5,618,386	5,723,650	5,590,201
Total ungapped length (bp)	4,501,780	5,723,650	5,590,201
Gap ratio (%)	19.6	0	0
No. of contigs	428	1	1
Contig N50 (bp)	23,226	5,723,650	5,590,201
GC content (%)	65.2	67.4	67.5
No. of CDS	4,564	8,689	8,630
No. of rRNA	3	7	7
No. of tmRNA	1	1	1
No. of tRNA	53	36	35

^z GC: guanine-cytosine; CDS: coding DNA sequence.

註解之 rRNA 與 tRNA 序列數量，分別為 7 個與 35 個。Weirather 等人的研究指出，ONT 長讀序組裝時最大的誤差來自鹼基錯配 (mismatch)，通常是鹼基判定異常所致 (Weirather *et al.* 2017)，這些錯配可能造成轉譯框架位移 (frameshift) 或提早出現終止子 (stop codon)，繼而影響基因體解序後的蛋白質預測 (Watson & Warr 2019)。僅使用 ONT (R9.4 版本) 長讀序組裝的情況下，50 倍覆蓋深度的 *E. coli* 基因體解序資料，錯誤率約為 12.5% (Dohm *et al.* 2020)，而 560 倍覆蓋深度的 *P. aeruginosa* 基因體解序資料，錯誤率降至 0.2% (Molina-Mora *et al.* 2020)，本試驗在相同定序平台下取得 230 倍覆蓋深度解序資料，以最高 99.8% 正確率進行估算，若細菌之平均基因長度為 1 kb，則每個基因可能出現 2 次誤配，導致框架位移或提早出現終止子，這可能是本試驗解序所得之 CDS 數量約為 CAT98_1 的 2 倍之原因。此外，短讀序組裝的高碎片化特徵，無法完整重建基因體的重複結構，高度重複出現的 rRNA 數量，與 CDS 數量通常被低估 (Molina-Mora *et al.* 2020)。

以 MUMmer 套件分析 AAC_10001、AAC_

9906R 兩菌株基因體序列，與 CAT98_1 之序列差異，根據比對結果繪製共線性圖 (圖 4)，紅色點表示正義股 1 kb 以上之連續片段且具有 90% 之相似性，綠點則表示同源性片段出現在反義股，本研究解序之基因體結構與 CAT98_1 相近，僅少數位置發生倒位 (inversion)，並無大規模的基因重組 (圖 4)。MUMmer 採用修正 Smith-Waterman local alignment 演算法，利用錨點對齊的方式減少隨機序列比對的發生 (Delcher *et al.* 1999)，但 MUMmer 識別高度同源區域 (homologous sites) 設定錨點後，只能進行直系同源序列 (orthologous sequence) 與異源同源序列 (xenologous sequence) 之比對，卻無法進行旁系同源序列 (paralogous sequence) 或多拷貝之直系同源序列 (orthologous sequence) 之比對 (Darling *et al.* 2010)。本試驗另以 Mauve 套件進行 3 個基因體序列比對，Mauve 套件採用錨點對齊的方式提升運算效率，但較 MUMmer 有更寬鬆的比對規則，容許兩個或多個基因體共享同源序列區域，這些區域稱為局部共線區塊 (locally collinear blocks; LCB)，可在兩個或多個序列間進行局部比對，適合分析基因體之重新

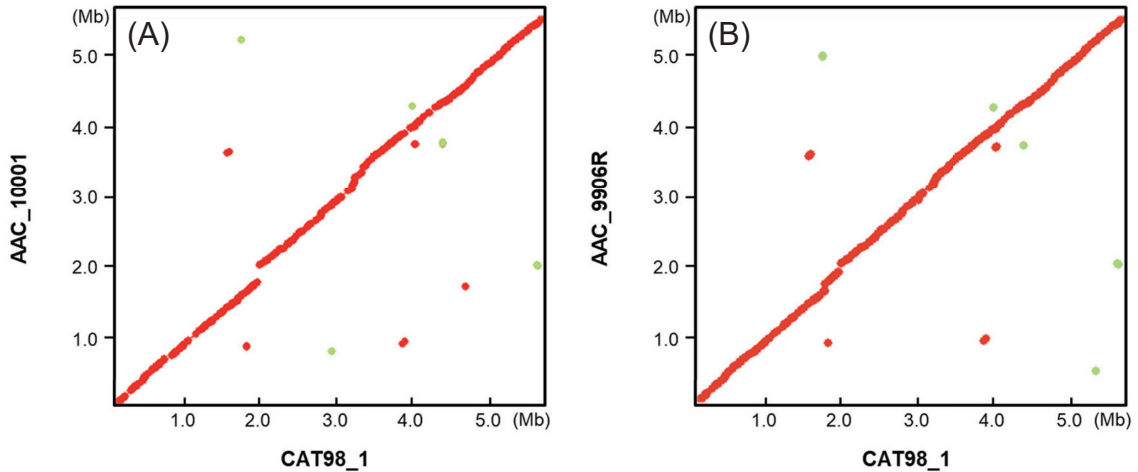


圖 4. *Acidovorax avenae* subsp. *Cattleyae* (AAC) 基因體解序資料與 *Acidovorax cattleyae* 菌株 CAT98_1 基因體序列比對結果。(A) 菌株 10001；(B) 菌株 9906R。

Fig. 4. Dot plot of a nucleotide-based alignment of *Acidovorax cattleyae* strain CAT98_1 chromosomal sequence with (A) *Acidovorax avenae* subsp. *cattleyae* (AAC) strain 10001 and (B) strain 9906R. Aligned segments are represented as dots. Forward matches are plotted in red, reverse matches in green. Figures generated by MUM-mer.

排列、水平轉移或序列發生缺失 (deletion)、重複 (duplication) 現象的探討 (Darling *et al.* 2004)。本研究顯示蝴蝶蘭褐斑病基因體序列共有 12 個主要的 LCB，CAT98_1 缺少 LCB 2，而 AAC_10001 與 AAC_9906R 之 LCB 2 序列長度相差 132 kb，此外，CAT98_1 之 LCB 4 與 LCB 9 與本試驗長讀序組裝位置互換，但 LCB 序列長度一致 (圖 5A)。LCB 7-8 和 LCB 10-11 間，各有一群小於 1 kb 的 LCB 分布，其相對位置如圖 5B、C 所示。差異性片段利用 NCBI blastx 工具與 Refseq_protein 資料庫進行比對 (表 3)，推測 AAC_10001 菌株 LCB 2 可能有一個基因體島 (genomic island; GEI) 的遺傳元件插入，因為該片段長度介於 10-200 kb 間、含有一個 DNA 修復蛋白、一群參與第四型分泌系統 (type IV secretion system; T4SS) 的蛋白、一個整合酶 (integrase) 與一個轉位酶 (transposase)，符合 Juhas *et al.* (2009) 對 GEI 的特徵描述；此外，T4SS 可以協助 GEI 本身、質體 DNA 或是染色體 DNA 進行自我轉移、水平轉移或重組 (Hamilton *et al.* 2005)，可能影

響菌株致病性、共生性、代謝特性、環境適應性或是對抗生素產生抗性。

結論

微生物基因體解序為研究細菌遺傳分類、族群演化和基因水平轉移的最佳策略，各類定序平台的快速進展，使我們可以快速地得到菌株基因體序列，瞭解基因體結構、基因重組、缺失、重複等基因型資料。奈米孔定序平台相較二代短片定序平台，實務上具快速和低成本的優勢，可減少 PCR 擴增錯誤和高度重複區域對於基因體組裝的影響，非常適合進行細菌基因體之解序與組裝。本研究使用奈米孔定序平台進行蝴蝶蘭褐斑病基因體解序，建立細菌基因體長片段讀序生物資訊分析流程，由細菌核酸萃取至資料分析結束僅需 1 wk，並找到一個可能參與致病性或抗抗生素特性的 GEI，是國內蝴蝶蘭褐斑病細菌基因體研究之首例，期望提供國內蘭園準確診斷和長期監測的方向，有助強化國內溫室蝴蝶蘭病害防治工作。

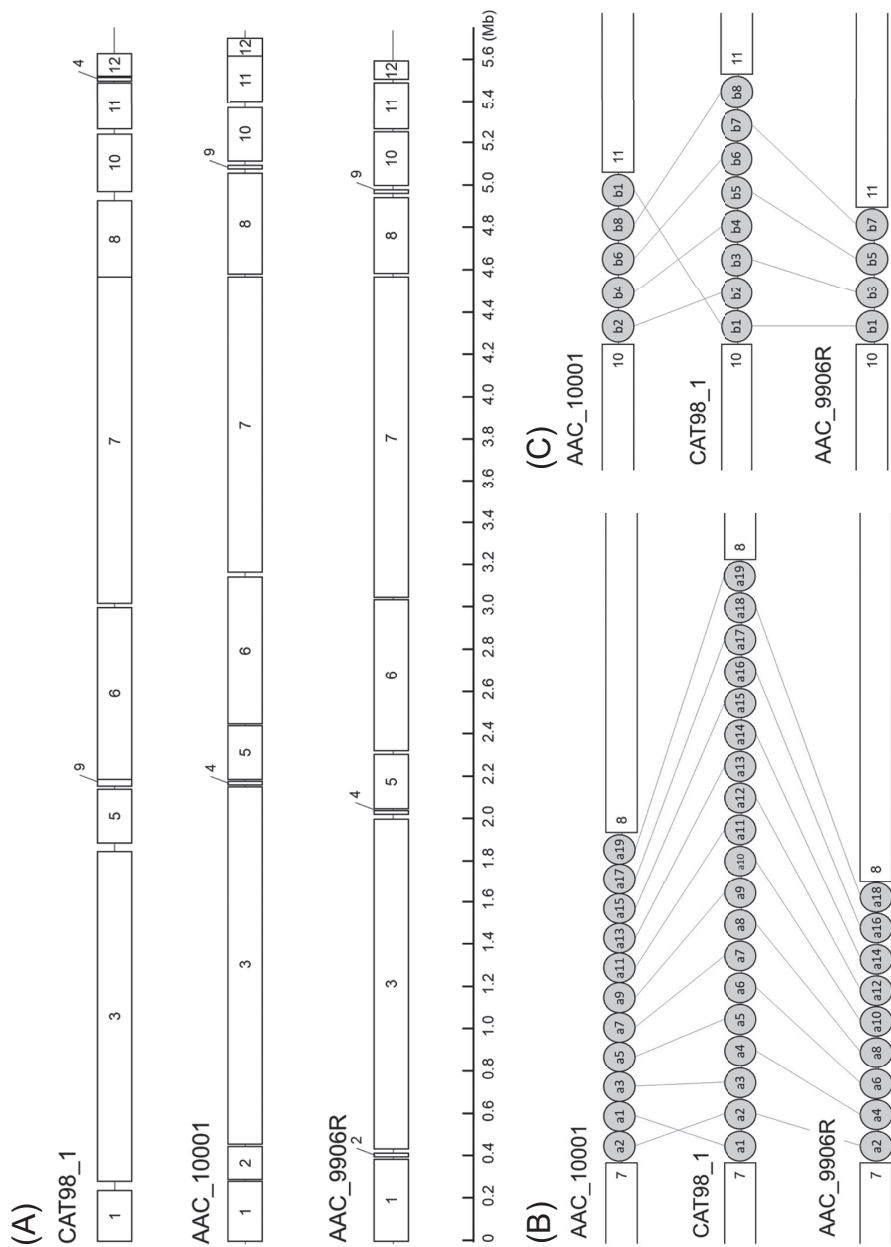


圖 5. *Acidovorax cattleyae* 菌株 CAT98_1、*Acidovorax avenae* subsp. *cattleyae* (AAC) 菌株 10001 與菌株 9906R 基因體序列比對分析。(A) 以 Mauve 軟體分析共有 12 個局部共線性區塊 (LCBs)，另有兩群小於 1 kb 的 LCBs，分布於 (B) LCB 7-8 以及 (C) LCB 10-11 之間。
 Fig. 5. Visualized genome alignment of the genome of *Acidovorax cattleyae* strain CAT98_1, *Acidovorax avenae* subsp. *cattleyae* (AAC) strain 10001 and strain 9906R. (A) In total, 12 locally collinear blocks (LCBs) were identified by Mauve and labelled in. There were two set of small LCBs (< 1 kb) located between (B) LCB 7-8 and (C) LCB 10-11.

表 3. *Acidovorax avenae* subsp. *cattleyae* (AAC) 菌株 10001 與菌株 9906R 基因體註解結果。
Table 3. Annotation results of *Acidovorax avenae* subsp. *cattleyae* (AAC) strain 10001 and 9906R.

Location	AAC_10001				AAC_9906R			
	Accession no.	Length	E-value	Description	Accession no.	Length	E-value	Description
LCB 2	WP_139160155.1	Full	8e-158	PLxREG domain-containing protein	WP_218297416.1	Partial	2e-69	PLxRFG domain-containing protein
	WP_017438281.1	Full	2e-36	Endolysin	WP_017438281.1	Full	8e-37	Endolysin
	WP_011794861.1	Full	6e-30	DUF4376 domain-containing protein				
	WP_160853018.1	Partial	1e-47	DUF6361 family protein				
	WP_114865077.1	Partial	5e-36	Helicase				
	WP_157768779.1	Full	2e-21	SMI1/KNR4 family protein				
	WP_202907192.1	Partial	7e-17	DNA repair protein RadC				
	WP_111877013.1	Full	2e-120	Conjugal transfer protein TraH				
	WP_011804808.1	Full	9e-98	Type IV secretion system protein TraC				
	WP_011804801.1	Full	4e-81	Type-F conjugative transfer system mat-ing-pair stabilization protein TraN				
	WP_202907158.1	Full	4e-58	Conjugal transfer protein TraG N-terminal domain-containing protein				
	WP_162017004.1	Full	2e-53	Conjugal transfer protein TraF				
	WP_055397146.1	Full	8e-52	TraJ family protein				
	WP_059400278.1	Full	1e-95	Conjugative transfer system coupling protein TraD				
	WP_063460687.1	Full	2e-60	TraI domain-containing protein				
	WP_011804827.1	Full	1e-35	DUF932 domain-containing protein				
	WP_005792891.1	Full	1e-30	AcaB family transcriptional regulator				
	WP_107134185.1	Full	1e-27	SEC-C domain-containing protein				
	WP_107204843.1	Partial	4e-155	type VI secretion system tip protein VgrG				
	WP_059400388.1	Full	5e-54	ATP-dependent DNA helicase DinG				
	WP_011804837.1	Partial	6e-39	DUF1173 domain-containing protein				
	WP_055399654.1	Partial	4e-26	Fic family protein				
	WP_092939581.1	Full	2e-21	GNAT family N-acetyltransferase				
	WP_218300121.1	Partial	2e-15	HNH endonuclease				

表 3. *Acidovorax avenae* subsp. *cattleyae* (AAC) 菌株 10001 與菌株 9906R 基因體註解結果。(續)
 Table 3. Annotation results of *Acidovorax avenae* subsp. *cattleyae* (AAC) strain 10001 and 9906R. (continued)

Location	AAC_10001				AAC_9906R			
	Accession no.	Length	E-value	Description	Accession no.	Length	E-value	Description
	WP_059400427.1	Partial	4e-61	ParB N-terminal domain-containing pro-teín	WP_092840623.1	Full	2e-36	IPTL-CTERM sorting domain-containing protein
	WP_059400429.1	Partial	2e-29	Integrase domain-containing protein	WP_056669722.1	Full	4e-38	Tripartite tricarboxylate transporter per-mease
	WP_059400431.1	Full	3e-54	PRTRC system protein D	WP_196289847.1	Full	3e-47	Type 1 glutamine amidotransferase
	WP_108625398.1	Full	4e-77	IS5 family transposase	WP_092839864.1	Partial	1e-78	Methyl-accepting chemotaxis protein
	WP_092840623.1	Full	3e-37	IPTL-CTERM sorting domain-containing protein	WP_196289847.1	Full	3e-64	Type 1 glutamine amidotransferase
	WP_092839864.1	Partial	1e-78	Methyl-accepting chemotaxis protein	WP_196289848.1	Partial	3e-48	Polyhydroxyalkanoate synthesis re-pressor PhaR
	WP_196289847.1	Full	3e-64	Type 1 glutamine amidotransferase	WP_092745669.1	Full	1e-24	Chemotaxis response regulator CheY
	WP_196289848.1	Partial	3e-48	Polyhydroxyalkanoate synthesis re-pressor PhaR	WP_092835372.1	Full	3e-61	EscU/YscU/HrcU family type III secretion system export apparatus switch protein
LCB 7-8	WP_092745669.1	Full	1e-24	Chemotaxis response regulator CheY	WP_092835370.1	Full	1e-17	Protein phosphatase CheZ
	WP_092835372.1	Full	3e-61	EscU/YscU/HrcU family type III secretion system export apparatus switch protein	WP_218296960.1	Full	4e-41	Flagellar type III secretion system protein FlhB
	WP_196290539.1	Partial	6e-111	Flagellar biosynthesis protein FlhA	WP_196290539.1	Partial	5e-111	Flagellar biosynthesis protein FlhA
LCB 10-11	WP_092840645.1	Full	6e-75	DUF6402 family protein	WP_107204716.1	Full	8e-06	ArgE/DapE family deacylase
	WP_107204716.1	Full	3e-17	ArgE/DapE family deacylase	WP_175503168.1	Full	2e-05	M20/M25/M40 family metallo-hydro-lase
	WP_175503168.1	Full	2e-14	M20/M25/M40 family metallo-hydro-lase	WP_092840645.1	Full	8e-35	DUF6402 family protein

引用文獻

- Amarasinghe, S. L., S. Su, X. Dong, L. Zappia, M. E. Ritchie, and Q. Gouil. 2020. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* 21:30. doi:10.1186/s13059-020-1935-5
- Darling, A. C. E., B. Mau, F. R. Blattner, and N. T. Perna. 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14:1394–1403. doi:10.1101/gr.2289704
- Darling, A. E., B. Mau, and N. T. Perna. 2010. ProgressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. doi:10.1371/journal.pone.0011147
- Delcher, A. L., S. Kasif, R. D. Fleischmann, J. Peterson, O. White, and S. L. Salzberg. 1999. Alignment of whole genomes. *Nucleic Acids Res.* 27:2369–2376. doi:10.1093/nar/27.11.2369
- Ding, C. Z., W. J. Zhao, D. Y. Cun, H. Y. Chen, and S. F. Zhu. 2010. Detection of *Acidovorax avenae* subsp. *cattleyae* by real-time fluorescent PCR. *Acta Phytopathol. Sin.* 40:235–241. (in Chinese with English abstract)
- Dohm, J. C., P. Peters, N. Stralis-Pavese, and H. Himmelbauer. 2020. Benchmarking of long-read correction methods. *NAR Genom. Bioinform.* 2(2):lqaa037. doi:10.1093/nargab/lqaa037
- Hamilton, H. L., N. M. Domínguez, K. J. Schwartz, K. T. Hackett, and J. P. Dillard. 2005. *Neisseria gonorrhoeae* secretes chromosomal DNA via a novel type IV secretion system. *Mol. Microbiol.* 55:1704–1721. doi:10.1111/j.1365-2958.2005.04521.x
- Hseu, S. H., C. J. Sung, and H. Shentu. 2011. A specific primer pair for diagnosis and detection of bacterial brown spot pathogen: *Acidovorax avenae* subsp. *Cattleyae* on *Phalaenopsis* orchid. *Plant Pathol. Bull.* 20:52–64. (in Chinese with English abstract) doi:10.6649/PPB.201106_20(1_2).0006
- Huang, C. J., T. L. Wu, P. X. Zheng, J. Y. Ou, H. F. Ni, and Y. C. Lin. 2021. Comparative genomic analysis uncovered evolution of pathogenicity factors, horizontal gene transfer events, and heavy metal resistance traits in citrus canker bacterium *Xanthomonas citri* subsp. *citri*. *Front. Microbiol.* 12:731711. doi:10.3389/fmicb.2021.731711
- Huang, T. C. 2001. Brown spot. p.51–56. *in*: *Plant Protection Illustrated Handbook 6: Orchid Protection*. (Bureau of Animal and Plant Health Inspection and Quarantine, ed.) Bureau of Animal and Plant Health Inspection and Quarantine. Taipei, Taiwan. 127 pp. (in Chinese)
- Juhas, M., J. R. van der Meer, M. Gaillard, R. M. Harding, D. W. Hood, and D. W. Crook. 2009. Genomic islands: Tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.* 33:376–393. doi:10.1111/j.1574-6976.2008.00136.x
- Li, T., R. Mann, T. Sawbridge, J. Kaur, D. Auer, and G. Spangenberg. 2020. Novel *Xanthomonas* species from the perennial ryegrass seed microbiome—assessing the bioprotection activity of non-pathogenic relatives of pathogens. *Front. Microbiol.* 11:1991. doi:10.3389/fmicb.2020.01991
- Mechan Llontop, M. E., P. Sharma, M. Aguilera Flores, S. Yang, J. Pollok, L. Tian, C. Huang, S. Rideout, L. S. Heath, S. Li, and B. A. Vinatzer. 2020. Strain-level identification of bacterial tomato pathogens directly from metagenomic sequences. *Phytopathology* 110:768–779. doi:10.1094/PHYTO-09-19-0351-R
- Molina-Mora, J. A., R. Campos-Sánchez, C. Rodríguez, L. Shi, and F. García. 2020. High quality 3C *de novo* assembly and annotation of a multidrug resistant ST-111 *Pseudomonas aeruginosa* genome: Benchmark of hybrid and non-hybrid assemblers. *Sci. Rep.* 10:1392. doi:10.1038/s41598-020-58319-6
- Payne, A., N. Holmes, V. Rakyán, and M. Loose. 2019. BulkVis: A graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics* 35:2193–2198. doi:10.1093/bioinformatics/bty841
- Pightling, A. W., N. Petronella, and F. Pagotto. 2014. Choice of reference sequence and assembler for alignment of *Listeria monocytogenes* short-read sequence data greatly influences rates of error in SNP analyses. *PLoS One* 9:e104579. doi:10.1371/journal.pone.0104579
- Schaad, N. W., E. Postnikova, A. Sechler, L. E. Claffin, A. K. Vidaver, J. B. Jones, I. Agarkova, A. Ignatov, E. Dickstein, and B. A. Ramundo. 2008. Reclassification of subspecies of *Acidovorax avenae* as *A. Avenae* (Manns 1905) emend., *A. cattleyae* (Pavarino, 1911) comb. nov., *A. citrulli* (Schaad *et al.*, 1978) comb. nov., and proposal of *A. oryzae* sp. nov. *Syst. Appl. Microbiol.* 31:434–446. doi:10.1016/j.syapm.2008.09.003
- Sutton, J. M., J. D. Millwood, A. Case McCormack, and J. L. Fierst. 2021. Optimizing experimental design for genome sequencing and assembly with Oxford Nanopore Technologies. *Gigabyte* 2021:1–26. doi:10.46471/gigabyte.27
- Watson, M. and A. Warr. 2019. Errors in long-read assemblies can critically affect protein prediction. *Nat. Biotechnol.* 37:124–126. doi:10.1038/s41587-018-0004-z
- Weirather, J. L., M. de Cesare, Y. Wang, P. Piazza, V. Sebastiano, X. J. Wang, D. Buck, and K. F. Au. 2017. Comprehensive comparison of Pacific Biosciences

- and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research* 6:100. doi:10.12688/f1000research.10571.2
- Wick, R. R., M. B. Schultz, J. Zobel, and K. E. Holt. 2015. Bandage: Interactive visualization of *de novo* genome assemblies. *Bioinformatics* 31:3350–3352. doi:10.1093/bioinformatics/btv383
- Willems, A., M. Goor, S. Thielemans, M. Gillis, K. Kersters, and J. De Ley. 1992. Transfer of several phytopathogenic *Pseudomonas* species to *Acidovorax* as *Acidovorax avenae* subsp. *avenae* subsp. nov., comb. nov., *Acidovorax avenae* subsp. *citrulli*, *Acidovorax avenae* subsp. *cattleyae*, and *Acidovorax konjaci*. *Int. J. Syst. Bacteriol.* 42:107–119. doi:10.1099/00207713-42-1-107
- Yang, R., D. Santos Garcia, F. Pérez Montaña, G. M. da Silva, M. Zhao, I. Jiménez Guerrero, T. Rosenberg, G. Chen, I. Plaschkes, S. Morin, R. Walcott, and S. Burdman. 2019. Complete assembly of the genome of an *Acidovorax citrulli* strain reveals a naturally occurring plasmid in this species. *Front. Microbiol.* 10:1400. doi:10.3389/fmicb.2019.01400
- Yoshii, A., T. Omatsu, Y. Katayama, S. Koyama, T. Mizutani, H. Moriyama, and T. Fukuhara. 2015. Two types of genetic carrier, the IncP genomic island and the novel IncP-1 β plasmid, for the *aac(2')-IIa* gene that confers kasugamycin resistance in *Acidovorax avenae* ssp. *avenae*. *Mol. Plant Pathol.* 16:288–300. doi:10.1111/mpp.12182
- Zeng, Q., J. Wang, F. Bertels, P. R. Giordano, M. I. Chilver, R. B. Huntley, J. M. Vargas, G. W. Sundin, J. L. Jacobs, and C. H. Yang. 2017. Recombination of virulence genes in divergent *Acidovorax avenae* strains that infect a common host. *Mol. Plant Microbe Interact.* 30:813–828. doi:10.1094/mpmi-06-17-0151-r

***De novo* Genome Assembly of the *Acidovorax avenae* subsp. *cattleyae* by Using the Nanopore-Based DNA Sequencing Technology**

Yuan-Kai Tu¹, Chia-Hsin Tsai², Yen-Chun Lin³, Po-Chun Ou³, and Han-Wei Chen^{1*}

Abstract

Tu, Y. K., C. H. Tsai, Y. C. Lin, P. C. Ou, and H. W. Chen. 2022. *De novo* genome assembly of the *Acidovorax avenae* subsp. *cattleyae* by using the nanopore-based DNA sequencing technology. *J. Taiwan Agric. Res.* 71(2):171–183.

In Taiwan, orchids takes the lead in floral exports, with the value reaching NTD 4 billion in 2020. However, the orchid bacterial brown spot disease caused by *Acidovorax avenae* subsp. *cattleyae* (*AAC*) is commonly observed in greenhouses. The combination of warm temperature and high humidity provides the optimum conditions for brown spot disease outbreak, resulting in the low-export quality of orchids. In this study, two *AAC* strains were sequenced using nanopore-based sequencing technology. Overall, the average read length and read quality of AAC_10001 and AAC_9906R were 1.49 Gb, 18.7 kb, 11.9 and 1.96 Gb, 17.29 kb, 12.3, respectively. The results showed high contiguity of DNA quality. After *de novo* genome assembly, 6 and 5 contigs were obtained for AAC_10001 and AAC_9906R, in which the largest contig was 5.7 Mb and 5.6 Mb with 67.4% and 67.5% guanine-cytosine (GC) content, and the average sequencing depths for both largest contigs were 230×. Genome structure comparison revealed 12 locally collinear blocks (LCBs) in AAC_10001 and AAC_9906R. We found a putative 132 kb genomic island (GEI) in LCB2 of AAC_10001, harboring a DNA repair protein, a cluster type IV secretory system proteins, an integrase and a transposase. GEI plays a critical role in the bacterial evolution of pathogenicity, symbiosis, metabolism, environmental adaptation, and antibiotics resistance. In this research, we first present a reliable long-read *de novo* assembly pipeline for *AAC* genome study by using nanopore sequencing data only. This could provide fundamental knowledge for strain detection, identification or long-term monitoring and give us a perspective disease prevention strategy to adopt.

Key words: Nanopore sequencing, *Acidovorax avenae* subsp. *cattleyae*, Genome assembly, Bacterial antibiotic resistance.

Received: March 15, 2021; Accepted: November 18, 2021.

* Corresponding author, e-mail: swaychen@tari.gov.tw

¹ Assistant Research Fellows, Biotechnology Division, Taiwan Agricultural Research Institute, Taichung City, Taiwan, ROC.

² Assistant Research Fellow, Plant Pathology Division, Taiwan Agricultural Research Institute, Taichung City, Taiwan, ROC.

³ Project Assistants, Biotechnology Division, Taiwan Agricultural Research Institute, Taichung City, Taiwan, ROC.