

2022 Workshop For Biomedical Big Data

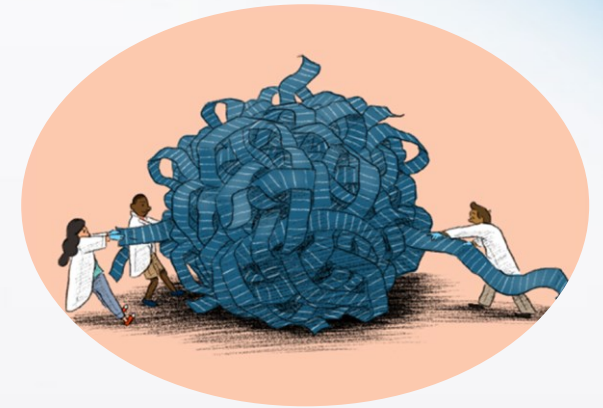


Data Science for High Throughput Biomedical Research

林 仲彥

Chung-Yen Lin

13 Jan 2022



中央研究院
資訊科學研究所
Institute of Information Science
Academia Sinica



LAB OF System Biology & Network Biology
中央研究院資訊科學研究所 @iis, Academia Sinica, TAIWAN
系統生物學與網路生物學實驗室



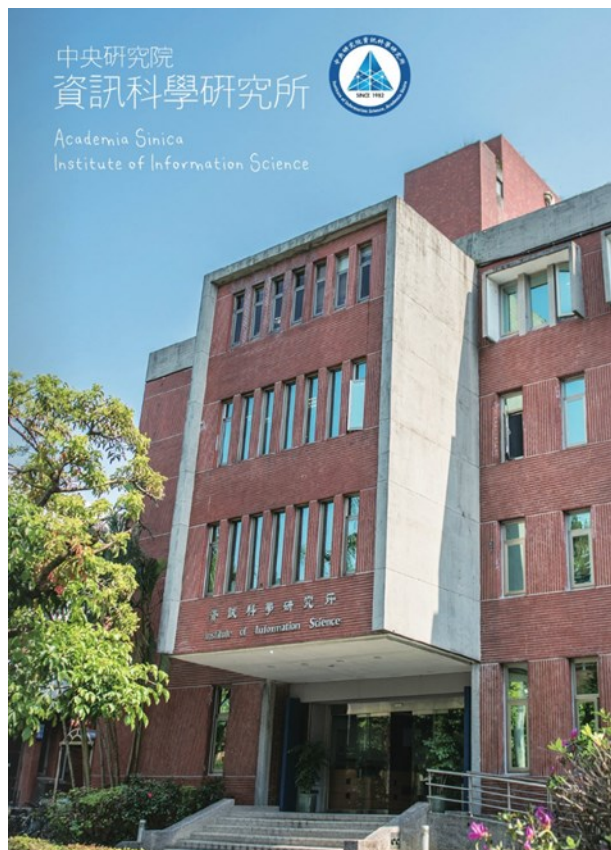
國家衛生研究院
National Health Research Institutes



中央研究院 資訊科學研究所

Institute of Information Science, Academia Sinica

成立於1977年



研究重點



Bioinformatics in

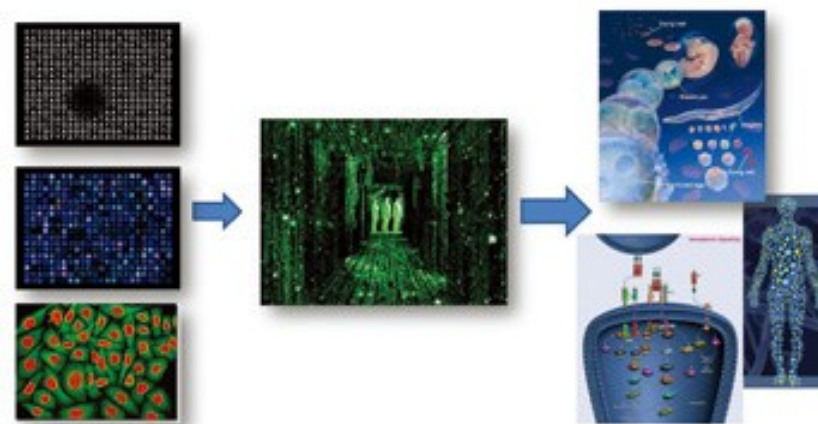


中央研究院
資訊科學研究所
Institute of Information Science
Academia Sinica



- 我們的研究是以資訊技術為主，針對不同生物體學(omics) 上的生物醫農問題進行解析，大致分為：

- ✓ 利用新世代序列定序儀進行基因體和轉錄體註解與分析
- ✓ 基因體/ 轉錄體重組演算法與程式開發
- ✓ 基因調控機制與網路
- ✓ 質譜儀分析的蛋白體學及代謝體學
- ✓ 蛋白質結構與功能預測
- ✓ 蛋白質交互作用網路分析
- ✓ 膜蛋白知識庫平台
- ✓ 高齡健康與台灣人體資料庫
- ✓ 人工智慧與新型藥物開發etc....



LAB OF System Biology & Network Biology

中央研究院資訊科學研究所 @iis, Academia Sinica, TAIWAN

系統生物學與網路生物學實驗室

近五年研究領域

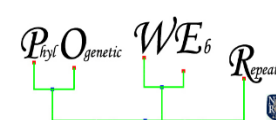
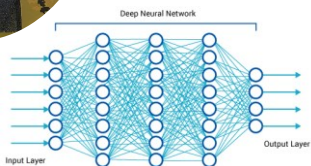
基因體學

深度學習在
病原分類與
新型抗生素之開發

次世代序列
分析與組裝

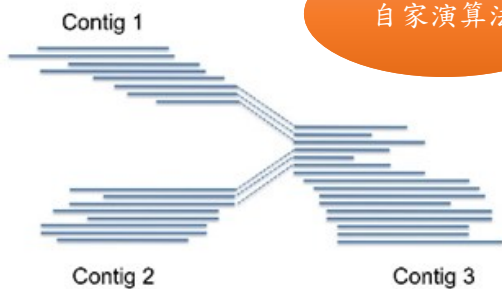
雲端實驗室
電子記事本

生物資訊資料庫
與演算法建置



以原始序列為基礎之染色體層級組裝演算法 到建構線上基因體資料庫分析平台之整合方案

以二代及三代定序平台所組裝之基因體初稿



自家演算法

illumina®

OXFORD NANOPORE Technologies®

PACIFIC BIOSCIENCES®

10X GENOMICS®

染色體層級之基因體組裝

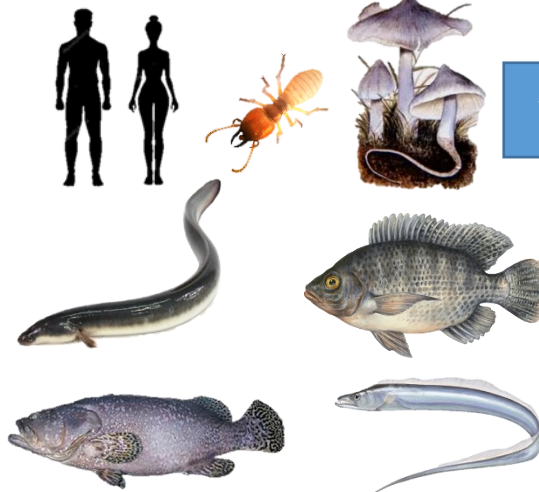


自家演算法

Hi-C



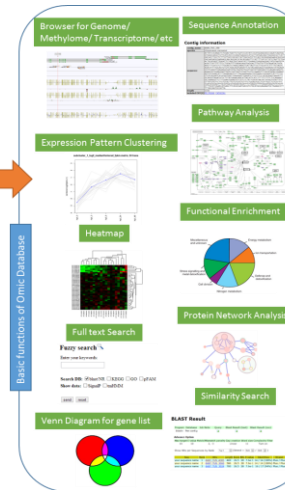
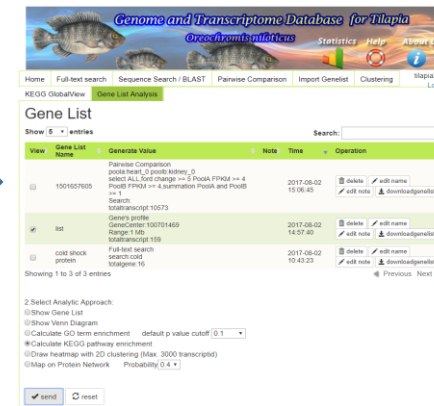
基因體初稿



基因體完稿

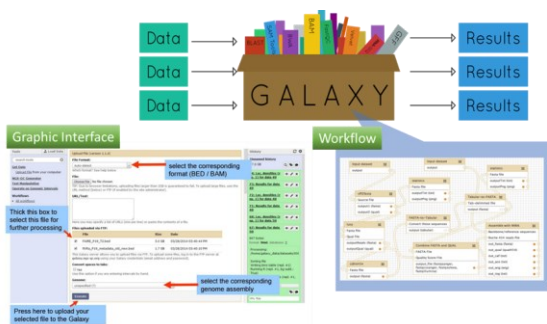
整合資訊技術與資料庫建置視覺化平台

自有網路
平台架構



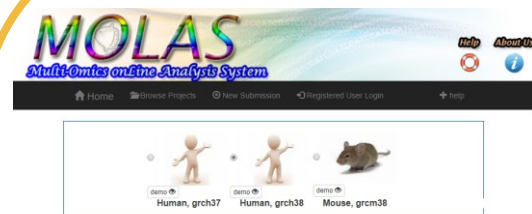
所建置之水產生物基因體資料庫平台

基因體分析應用程式



生物序列線上分析工具表列

基因體線上資料庫與分析平台



*Upload expressed profiling in FPKM in tab file. Example dataset for download: [Fpk-grch37.grch37.grch38](#)

選擇檔案 | 未選擇任何檔案

Important: Please read this before submission



人類與小鼠



龍膽石斑



日本鰻



草蝦



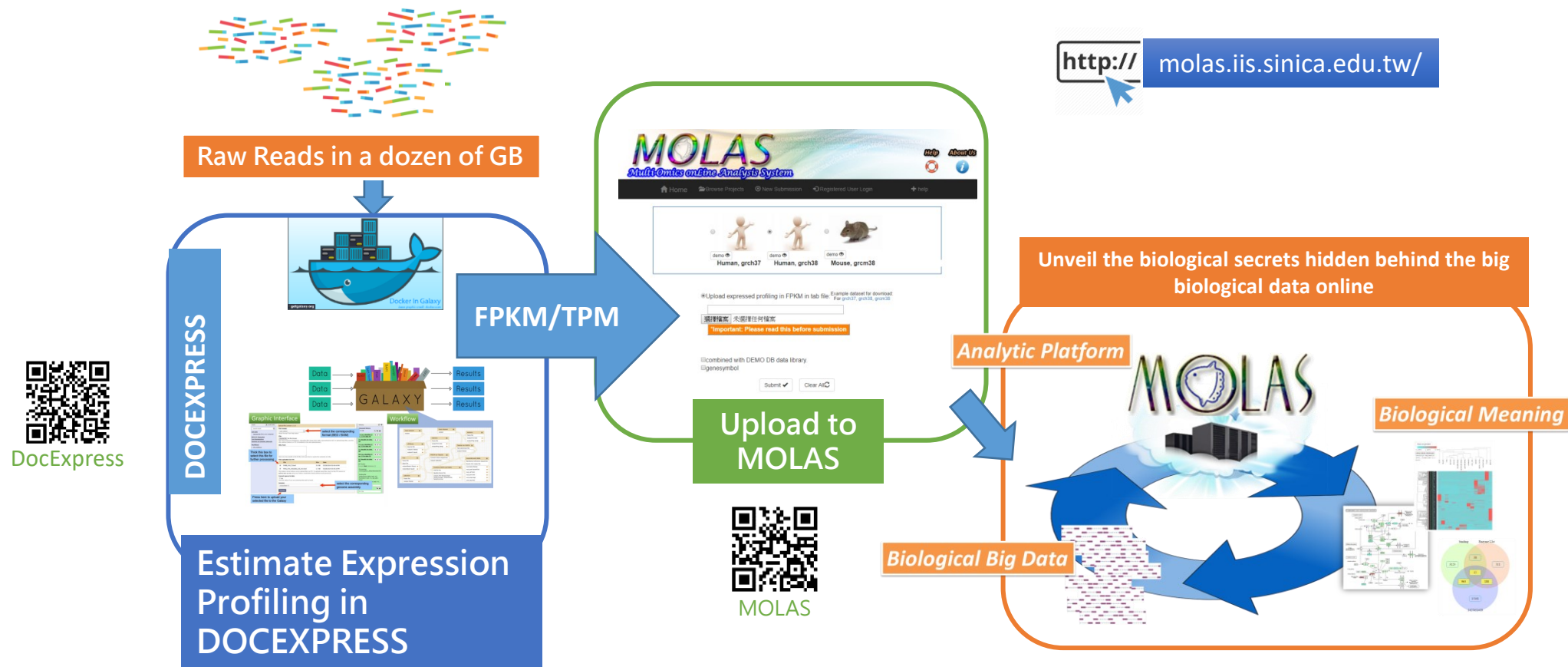
台灣鯛



南美白蝦



由原始數據、基因表現估計到生物意義解析



Technology Transfer to

WELGENE

應用於精準醫學與智慧育種之基因體重組與品質評估整合平台

01 服務/諮詢

結合目前多種最前端的全基因體定序技術平台，利用各平台間的優勢，設計一套基因體全新組裝程式(GABOLA)。GABOLA優勢為可組裝正確且完整的基因體，解析長期難解的重複片段，提供人類基因體中未知片段的遺傳資訊，突破現有基因體組裝軟體限制，及精準組裝個人化基因體進而增加對遺傳變異解釋能力。在非模式物種的日本鰻、台灣鯛、台灣家白蟻及龍膽石斑的組裝上，GABOLA也有相當好的表現，能大幅減少序列中的遺缺，並連結組裝片段，使整個基因體組裝更為完整。

02 台灣產業(或社會經濟)貢獻



提供精準醫療及精準農業育種之藍圖

以個人化基因體組裝更完整呈現疾病相關基因序列

運用於非模式生物智慧育種



增進基因體質與量

解析高重複與低複雜序列區域

提升功能性基因解析度

提高基因體完整性與連續性



可無須參考基因體

可呈現基因體之個體差異



結合不同定序平台優點

10X GENOMICS PACIFIC BIOSCIENCES®
illumina® Oxford Nanopore Technologies
bionano GENOMICS

03 重大成果

本成果可應用的範圍，將包含精準醫療與智慧育種。本研究榮獲2021年科技部未來突破獎與國家生策會學研新創獎。



04 產品/ 成果照片



bioRxiv
THE PREPRINT SERVER FOR BIOLOGY



01 開發尖端技術

「以深度學習機制解析上呼吸道病毒序列資料並建立基因型鑑別機制與網路即時分析平台」為整合深度學習模型及相關自有程式，建構了第一個以上呼吸道病毒基因體大數據為基礎的線上分類網站。

02 台灣產業(或社會經濟)貢獻

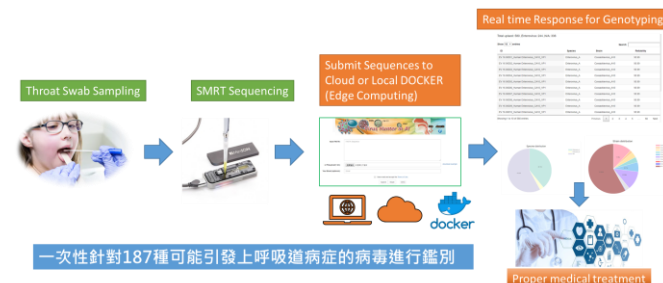
此一上呼吸道病毒基因型分型智慧平台，將能結合第三代小型低價定序儀(如Nanopore)，應用於在上呼吸道病毒感染初期(約三到五天內)的檢測上，應能及早發現所感染的病毒類型，如屬高致病品系，便能提早準備，防止重症的發生，減少寶貴生命受損的機率，同時也降低可能的醫療費用，並有效防禦及控制疫情，且減低醫療和社會經濟成本，進而實現預防醫學的目標。

03 重大成果

本成果可應用的領域業別，將包含第一線診所與醫院，生醫檢驗產業，及病毒學研究團隊與流病監控團隊。本研究榮獲2020年國家新創獎與2021精進獎。



04 產品/ 成果照片



技術創新關鍵優勢利基

結合三代定序與上呼吸道病毒智慧定型的
未來應用場域：虛擬專家到現場快速鑑別

Real time Response for Genotyping

Throat Swab Sampling



SMRT Sequencing



Submit Sequences to
Cloud or Local DOCKER
(Edge Computing)



一次可對187種可能引發上呼吸道病症的病毒進行鑑別

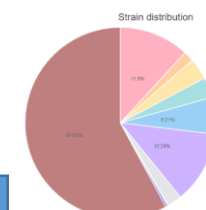
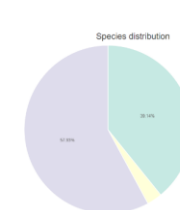
Total upload: 580, Enterovirus: 244, N/A: 336

Show 10 entries

ID	Species	Strain	Reliability
EV 16 00001_Human Enterovirus_CA10_VP1	Enterovirus_A	Coxsackievirus_A10	95.99
EV 16 00002_Human Enterovirus_CA10_VP1	Enterovirus_A	Coxsackievirus_A10	95.99
EV 16 00003_Human Enterovirus_CA10_VP1	Enterovirus_A	Coxsackievirus_A10	95.99
EV 16 00004_Human Enterovirus_CA10_VP1	Enterovirus_A	Coxsackievirus_A10	95.99
EV 16 00005_Human Enterovirus_CA10_VP1	Enterovirus_A	Coxsackievirus_A10	95.99
EV 16 00006_Human Enterovirus_CA10_VP1	Enterovirus_A	Coxsackievirus_A10	95.99
EV 16 00007_Human Enterovirus_CA10_VP1	Enterovirus_A	Coxsackievirus_A10	95.99
EV 16 00008_Human Enterovirus_CA10_VP1	Enterovirus_A	Coxsackievirus_A10	95.99
EV 16 00009_Human Enterovirus_CA10_VP1	Enterovirus_A	Coxsackievirus_A10	95.99
EV 16 00010_Human Enterovirus_CA10_VP1	Enterovirus_A	Coxsackievirus_A10	95.99

Showing 1 to 10 of 580 entries

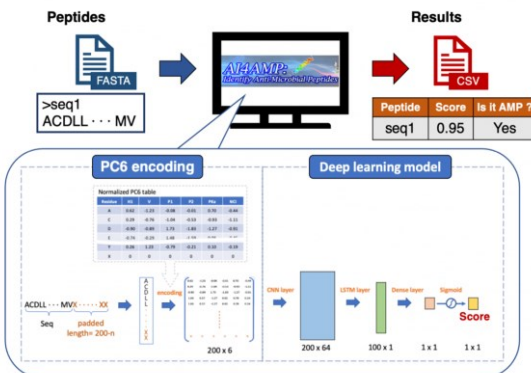
Previous 1 2 3 4 5 ... 58 Next



Proper medical treatment

01 開發尖端技術

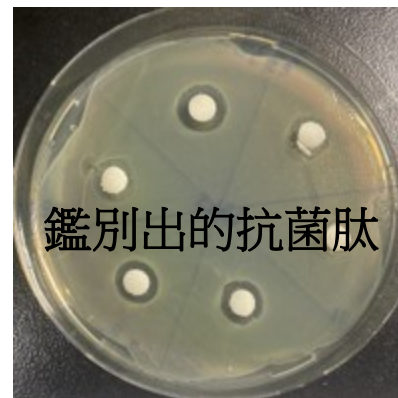
本團隊利用蛋白質物化特性設計序列編碼模型，並結合深度學習卷積神經網路，建構蛋白質序列抗菌肽活性預測模型，並透過生成對抗網路模型，製造出與真實抗菌肽高度相似的序列，**加速抗菌藥物的新藥開發與找尋。**



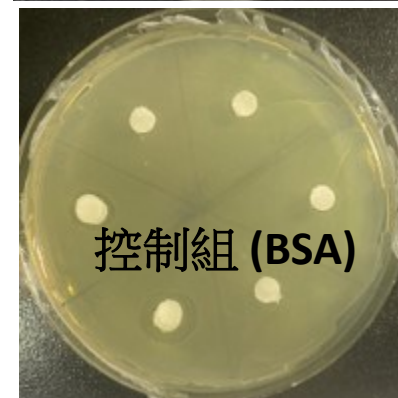
02 台灣產業(或社會經濟)貢獻

藥廠及生技公司可透過本團隊開發的抗菌肽生成器，快速設計出可能具抗菌效果的蛋白序列，並經由抗菌肽預測系統平台，來預測序列是否具抗菌效果，再進行實驗驗證，如此可以大幅減低抗菌肽藥物的開發時間及成本，同時因其誘發的**抗藥性低**，將變成**未來新一代抗生素**，可**減少後續醫療相關的費用支出。**

03 重大成果



鑑別出的抗菌肽



控制組 (BSA)

04 產品/ 成果照片

YouTube



5分鐘簡介

GitHub



程式模型

www.



網站平台

抗菌肽線上鑑別網站平台

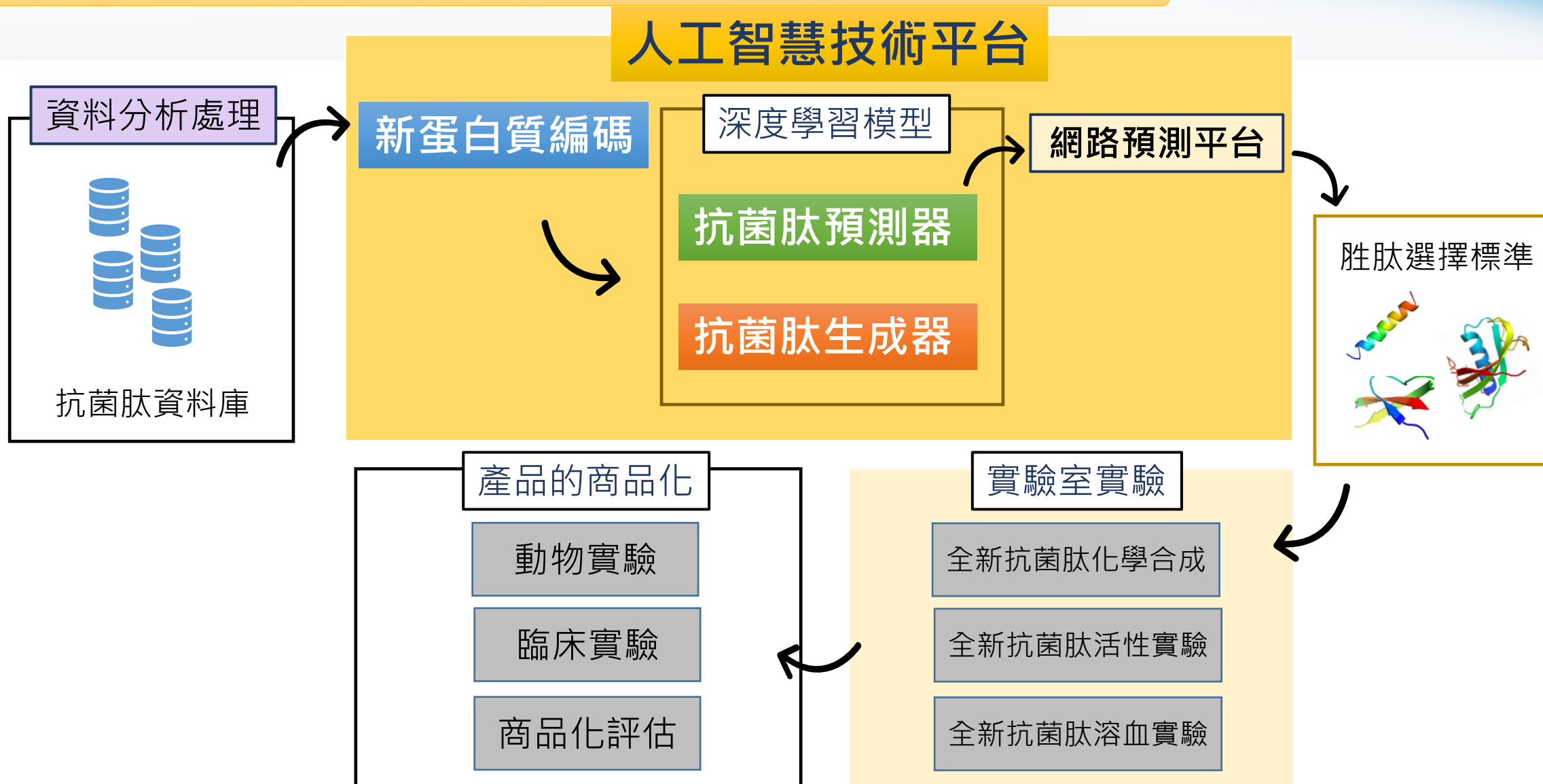
已發表於

mSystems

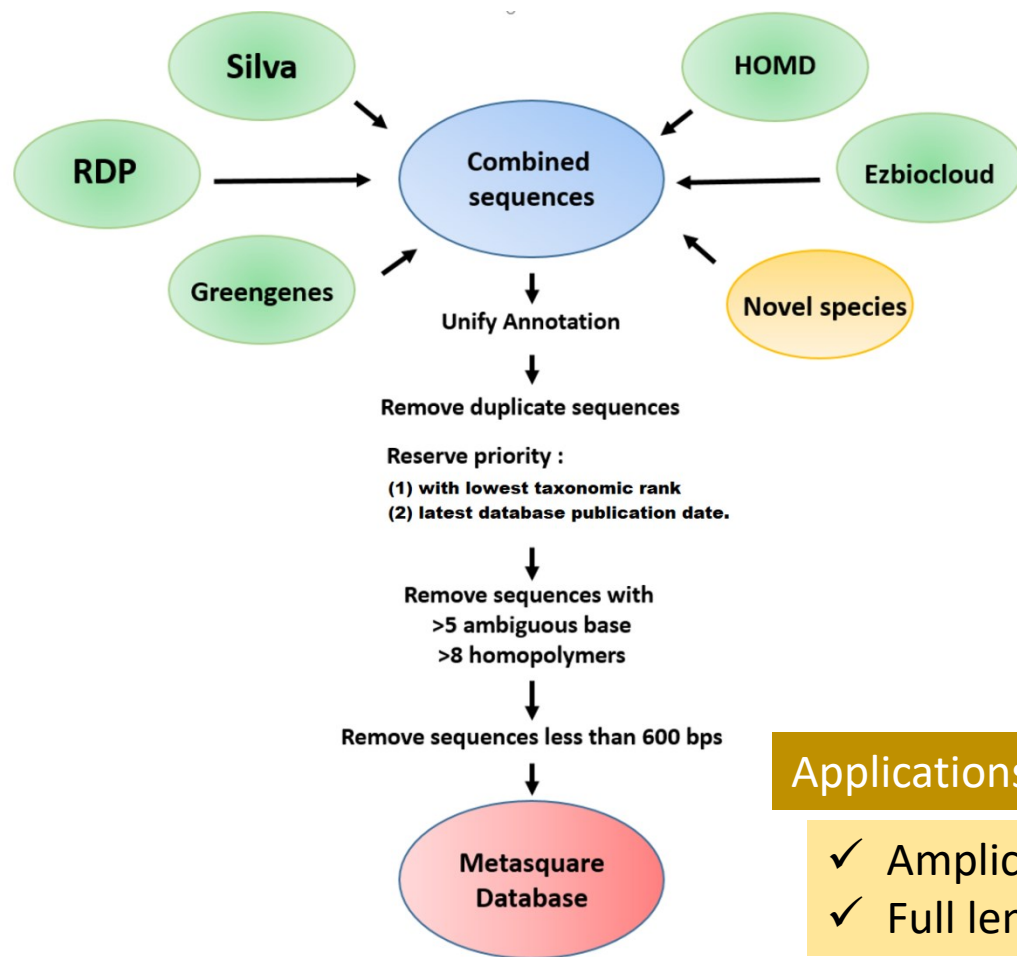
(IF=6.63)



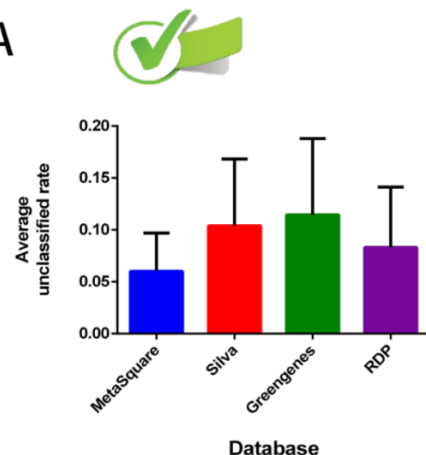
全新抗菌肽開發流程



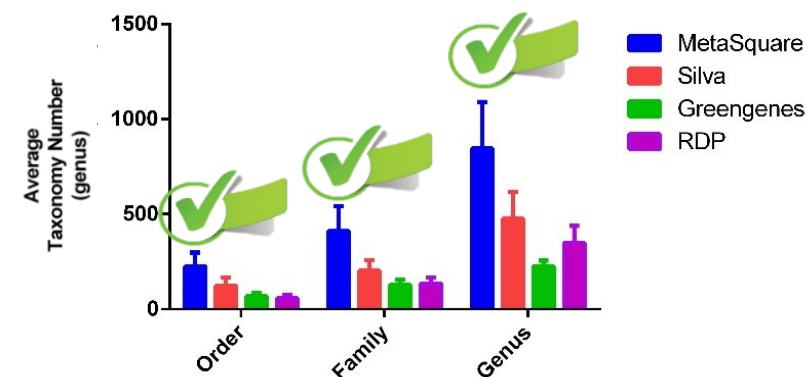
MetaSquare: Integrated 16s rRNA Database



A



B



Applications:

- ✓ Amplicon of 16s rRNA
- ✓ Full length 16s rRNA





線上資料庫 Web Databases



阿拉伯芥基因甲基化線上分析資料庫(*BMC Genomics*, 2016) · <http://tea.iis.sinica.edu.tw>



細枝鹿角珊瑚與共生藻基因資料庫(*Molecular Ecology*, 2016), <http://ips.iis.sinica.edu.tw/coral/>



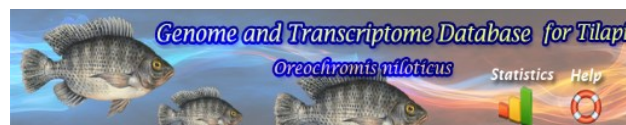
半索動物轉錄體資料庫(*Marine Genomics*, 2014), <http://molas.iis.sinica.edu.tw/hemichordate/>



龍膽石斑基因體與轉錄體資料庫, <http://molas.iis.sinica.edu.tw/grouper2015/>



日本鰻基因資料庫 (*PLoS ONE* 2015) , <http://molas.iis.sinica.edu.tw/ipeel2016>



台灣鯛基因體資料庫 <http://molas.iis.sinica.edu.tw/tilapia2017/>



沉香基因表現資料庫 (*BMC Plant Science*, 2015) , <http://molas.iis.sinica.edu.tw/agarwood>



腸病毒基因分型與重組分析(*BMC Genomics*, 2015) <http://symbiont.iis.sinica.edu.tw/evidence>



人類幽門桿菌蛋白質交互作用資料庫(*Bioinformatics*, 2005), <http://dpi.nhri.org.tw/hp>



線上分析工具平台 Web Applications

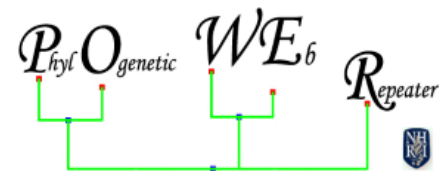
PDA: Primer design Assistant for PCR primer design, *NAR* 2003, <http://dbb.nhri.org.tw/primer>, 聚合酶反應核酸引子線上設計平台



myBLAST: customized databases and results management on your Desktop 在您的電腦上的高速序列比對平行化平台 · web version: <http://mybioweb.nhri.org.tw/myblast>, Windows/ MAC/ Linux version download :<http://eln.iis.sinica.edu.tw>



POWER: Phylogenetic web repeater (*NAR*, 2005), <http://power.nhri.org.tw> 線上生物巨分子親緣關係分析平台



MySort: Estimate Immune Cell Composition (*BMC Bioinfo*, 2018) <http://symbiosis.iis.sinica.edu.tw/mySORT/> 解析浸潤於癌組織中的免疫細胞組合



UPS 2.0: Unique Probe Selector (*BMC Genomic* 2010) <http://array.iis.sinica.edu.tw/ups> 快速病原檢測開發平台/檢測晶片與基因晶片線上設計系統



PALM: Phylogenetic reconstruction by Automatic likelihood Model Selector. *PLoS ONE* 2009, <http://palm.iis.sinica.edu.tw> 最佳化親源分析 高速分析平台



Enterovirus Genotyping in AI/ 腸病毒基因型線上AI智慧定型 <http://symbiosis.iis.sinica.edu.tw/Enterovirus>



Resource

