

The Use of SSR and SNP Molecular Markers to Analyze Genetic Diversity Among Wild and Farmed *Rhynchocinetes durbanensis*

Chai-Hsia Gan¹, Chen-Cheng Cheng² and Chia-Hsuan Sung¹

¹Planning and Information Division, Fisheries Research Institute

²Penghu Marine Biology Research Center, Fisheries Research Institute

ABSTRACT

DNA sequence and restriction-site-associated DNA sequencing (RADseq) databases were used to perform *Rhynchocinetes durbanensis* genotyping using a next-generation sequencing (NGS) platform. Ten simple-sequence repeat (SSR) markers and 429,078 single-nucleotide polymorphism (SNP) loci were identified for the analysis of genetic diversity among wild and farmed *R. durbanensis* populations. The genetic differentiation index (F_{st}) for the SSR analysis between the wild and farmed populations was 0.0184, and population structure analysis using Bayesian inference indicated no significant genetic differentiation. A total of 162,109 specific SNP alleles were identified within the farmed population, which were fewer than the number identified in the wild population, which contained 194,579 specific SNP alleles. Homozygosity within the farmed population was higher than that for the wild population. A maximum-likelihood phylogenetic tree separated the farmed and wild samples into 2 groups. However, p-value analysis did not reveal significant differences between farmed and wild samples. The F_{st} value inferred by SNP analysis between the two populations was 0.087, indicating no significant differentiation.

Key words: *Rhynchocinetes durbanensis*, Next-generation sequencing (NGS), single nucleotide polymorphism (SNP), short sequences repeat (SSR), genetic diversity

簡單重複序列與單一核酸多型性分子標誌在野生及養殖 德班氏活額蝦 (*Rhynchocinetes durbanensis*) 基因歧異度之研究

顏綵葭¹·城振誠²·宋嘉軒^{1*}

¹行政院農業委員會水產試驗所企劃資訊組

²行政院農委會水產試驗所澎湖海洋生物研究中心

摘要

本研究為開發德班氏活額蝦 (*Rhynchocinetes durbanensis*) 種原之遺傳資料庫，應用次世代定序平台建構 DNA 核酸序列資料庫與限制酶切位點標定法 (restriction associated DNA sequencing, RAD) 資料庫，並以生物資訊分析從資料庫中篩選出 10 組多型性的簡單重複序列 (simple sequence repeat, SSR) 分子標誌及 429,078 個單一核苷酸多型性 (single nucleotide polymorphism, SNP) 基因序列來檢測水產試驗所澎湖中心種原庫穩定繁殖 F3 種蝦遺傳資訊，並與市售野生族群比較基因歧異度 (genetic diversity)。10 組 SSR 基因座對偶基因頻度分析顯示養殖族群平均異型合子觀測值 H_o 大於期望值 H_e ，表示基因歧異度高；與野生族群比較平均基因變異數為 0.04478，兩族群間基因歧異度為 0.0184，皆無統計顯著差異；貝氏歸群分析法顯示養殖與野生族群在統計上無顯著遺傳分化。以解析度較高之 SNP 序列分析養殖與野生個體之間基因型差異的結果顯示，429,078 個 SNP 位點，養殖個體特有對偶基因數量為 162,109 個，較少於野生個體特有對偶基因數 194,579 個；利用集群親緣分析解析養殖及野生個體之間遺傳距離為 0.087，親緣關係樹狀圖可以區分養殖及野生群。以次世代定序解構之 SSR 及 SNP 分子標記解析養殖三代德班氏活額蝦群內基因歧異度高，沒有顯著基因窄化現象，與野生族群沒有顯著遺傳分化，能區分養殖及野生群的 SNP 基因序列可以用來制定種原維護及育種目標。

關鍵詞：德班氏活額蝦、次世代定序、單一核苷酸多型性、簡單重複序列、基因歧異度

前言

德班氏活額蝦 (*Rhynchocinetes durbanensis*) 中文俗名機械蝦、駱駝蝦，分類學上屬於節肢動物門、甲殼綱、十足目、活額蝦科，廣泛分布於印度洋與太平洋的熱帶、亞熱帶及溫帶海域，主要棲息於珊瑚礁及岩礁的近海地區 (Gordon, 1936)。成蝦體長約 4 cm，具有寶藍色大眼睛，體表呈鮮豔紅白色相間條紋及白點，形態優美，動作曼妙，頗受觀賞水族愛好者喜愛，為具經濟價值的海水觀賞蝦物種 (城與蔡, 2007)。觀賞水族產業具備高技術研發、高附加價值之產業鏈及具全球貿易消費等特性 (徐等, 2015)，根據行政院農業委員會農業貿

易統計資料顯示，2019 年臺灣觀賞蝦活體外銷出口 1.15 億，主要出口國為美國、德國及日本。面對國際競爭市場，臺灣的觀賞蝦活體外銷為具競爭優勢的農漁產品，為了突破種原產地地質的局限，開創更多精緻迷人的觀賞蝦新品系，藉分子生物學與基因體學技術拓展傳統育種研發的深度與廣度，進行性狀基因型鑑定、調控基因與穩定優勢性狀等分子輔助育種是未來研究發展方向 (黃, 2015)。為厚植產業競爭力，未來觀賞水族發展將導向精緻農業模式，結合資訊科技、生物技術與繁養殖技術研發，開創高附加價值新興觀賞水族精品，為臺灣觀賞水族業增加全球競爭優勢的籌碼 (黃等, 2012)。

大多數的觀賞性水族供應來源倚賴野生採捕，為了紓解採集地野生動物資源流失引發的生態壓力，也同時為了因應各國水族領域的動物保護政策緊縮，農業委員會水產試驗所澎湖海洋生物研究中

*通訊作者 / 基隆市和一路 199 號; TEL: (02) 2463-2101 ext. 2509; FAX: (02) 2462-4627; E-mail: chsung@mail.tfrin.gov.tw

心 (以下簡稱澎湖中心) 已研發出低資源消耗、高技術、高效益之德班氏活額蝦完全養殖量產技術 (城與蔡, 2005, 2007), 鼓勵業者人工繁殖蝦苗來提升生產效能及市場流通率, 以替代採捕野生資源, 避免引發產地海域生物多樣性之流失 (城與林, 2014; 城, 2015)。此技術推廣的德班氏活額蝦種原在澎湖中心水產生物種原庫中保種。

在人工繁養殖的環境下, 維持保種種原有效族群是困難的課題, 往往會面臨因養殖成本及空間的侷限, 族群偏小, 且族群內個體經過數代培育後發生近交機率增加, 族群趨向遺傳相似度非常高的近親個體, 這時族群內個體的基因型趨向純合子 (homozygote) 分布, 促使累積隱性有害純合子, 對偶基因頻度 (allele frequency) 偏移導致基因歧異度流失, 最終導致保種生物生存力、適應性、繁殖力降低等近交衰退 (inbreeding depression) 現象 (曾等, 2013, 2014; 杜等, 2015)。種原庫規畫開發種原之遺傳資料庫, 從中挑選適合的分子標誌來監測種原庫內養殖族群的基因歧異度 (genetic diversity) 參數, 如基因雜合度 (heterozygosity)、有效族群數及近交係數 (inbreeding coefficient) 等基因頻度的分布狀況, 評估種原庫內養殖族群是否面臨近親繁殖壓力而出現近交衰退的趨勢及現象。在評估近親繁殖壓力的指標選擇上, 監測養殖個體的基因型頻度分布會比表現型生長性狀的衰退評估來得敏感, 就能提早進行復育措施, 降低種原基因劣化的風險。

常見用來分析基因歧異度的分子標誌包含 DNA 序列、粒線體 DNA 序列、簡單重複 (simple sequence repeat, SSR) 序列、限制片段長度多型性 (restriction fragment length polymorphism, RFLP)、隨機擴增多型性 DNA (random amplified polymorphic DNA, RAPD)、擴增片段多型性 (amplified fragment length polymorphism, AFLP) 及單一核苷酸多型性 (single nucleotide polymorphism, SNP) 等, 這些分子標誌在因材料方法能獲得的遺傳訊息不一樣, 在多樣性信息含量 (polymorphism information content, PIC) 層面上有解析度的差異。其中在基因體中突變率高、又是應用多基因座 (multilocus) 分析的 SSR 及 SNP 分標記能取得 PIC 大, 是高解析度的分子標記, 適合分析族群之間、甚至族群內個體間的差異, 尤其在分析遺傳相近的小族群的基因雜合度、基因歧異度及近交係數等解析效果很顯著

(Hellberg *et al.*, 2002; Lemopoulos *et al.*, 2019)。

SSR 或稱微衛星 (microsatellite) 是由 Moore 和 Solloterer 在 1991 年同時提出的, 生物基因體中具有大量重複且連續排列組合的 DNA 序列, 重複序列區域的核心序列由 1-6 個核酸組成, 重複排列成總長度約 20 至數百個核酸序列, 其中最常見的是雙核酸重複的組成, 如 $(AT)_n$, 根據重複序列的不同形式可分為: (1) 完全重複序列 (perfect repeat sequence): 為連續單一重複序列組合, 如 $(ATG)_n$; (2) 序列 (imperfect repeat sequence): 重複序列中插入其他核酸, 行成重複不連續組合, 如 $(ATG)_nT(AG)_n$ 以及 (3) 複合重複序列 (compound repeat sequence): 含有兩種以上連續重複序列組合, 如 $(CA)_n(GT)_n$ (Weber, 1990)。而 SSR 具有的高度多型性主要來自於核心序列重複的數量, 此簡單重複序列在 DNA 複製期間, 較容易發生複製錯誤 (replication error, RER), 導致重複序列的增加或缺失 (insertion/deletion, indel), 而且多數 SSR 分佈於基因體非轉錄區位置, DNA 複製容錯率較高, 故比起基因序列突變率更高, 且能遺傳至下一代, 多樣性信息含量大, 是有效的多樣性分析標誌, 尤其 SSR 遺傳多型性在族群之間具有很大差異性, 可應用為鑑別族群間基因歧異度及遺傳距離, 對於數個世代內族群結構發生的變化有較高的解析能力 (Hellberg, 2007)。SSR 分子標記分析的原理是根據 SSR 區域兩端序列 (flanking region) 設計引子, 透過 PCR 反應擴增 SSR 區域片段, 由於核心重複序列的數量不同, 因而可由 PCR 反應擴增出不同長度大小的產物, 再透經膠電泳的區分, 根據片段大小決定基因型並進行相關遺傳分析。

SNP 是指單一核苷酸發生鹼基替代 (base substitution) 或單一鹼基的插入或刪除 (indels) 這種形式的單一位點發生自然變異。鹼基替代包含 transitions 由一個嘌呤 (purine) 轉變成另一個嘌呤, 或是一個嘧啶轉變成另一個嘧啶, A 變成 G 或是 C 變成 T, 反之亦然; 和 transversions 由一個嘌呤轉變成另一個嘧啶, A 或 G 變成 C 或 T, 反之亦然。大多數基因體中序列多型性都是 SNP, SNP 分析多應用在鑑定族群結構、個體間的變異、疾病基因座的定位和數量性狀基因座圖譜 (quantitative trait loci mapping, QTLs mapping)。

在 2005 年次世代定序問世之前, 從基因體中找到具有 SSR 的區域或是找到大量的 SNP 的實驗

步驟是很繁瑣的，而且需耗費大量人力和金錢。首先是利用具有 SSR 核心重複區互補序列的探針 (probe)，從基因體序列中抓出具有 SSR 序列的 DNA 片段，選殖入載體 (vector) 中製備成 genomic DNA library，再利用 DNA 定序儀將每個載體進行定序，得到的序列再進行後續的 SSR 預測分析，這樣連續且重複的過程耗時耗錢。然而，隨著生物科技技術的進步，以往需要龐大經費與時間才能進行基因體研究，現在隨著次世代定序 (next generation sequencing) 技術的發展，突破傳統核酸定序讀取鹼基對 (base pair, bp) 的效率，藉由大量、平行化的短序列片段 (short reads)，同時進行定序作業，又稱為高通量定序 (high throughput sequencing)，能快速且經濟的完成數百萬條的核酸定序，並建構分子生物資訊平台處理序列大數據，有效率進行基因型鑑定並開發數量龐大的分子標誌作為後續基因研究用途 (宋, 2015)。利用次世代高通量定序可以很方便的進行基因組的定序，再加上生物資訊軟體不斷推陳出新，透過軟體組裝解析序列，可以快速建構基因體中的 SSR 及 SNP 資料庫。從次世代定序出來單一條的序列讀度可達到 300 - 700 bp，足夠篩選出 SSR 基因座位置以及從兩端序列 (flanking region) 中設計出可進行 PCR 的引子對 (Kang *et al.*, 2013)，讓開發 SSR 分子標誌的工作變得更簡便，但是後續的驗證仍是相當費時的工作。

隨著次世代定序技術的發展，衍生許多降低成本的核酸分子標誌分析技術，如 Baird *et al.* (2008) 所提出的限制酶切位點標定法 (restriction associated DNA sequencing, RAD-seq)，利用 Illumina 次世代定序平台，僅對族群中每個個體基因組之特定限制酶切位旁的核酸進行定序，然後透過核酸序列的比對，找出族群內個體間共有的 SNP 位點，藉此不必針對每個個體進行基因體定序，就可以解析出大量的序列數據，可以進行大量的 SNP 位置的預測、篩選及測試，在族群中開發高密度 SNP 分子標誌。利用 NGS 平台開發的分子標誌在許多海洋生物種的研究上得到許多成果 (Saarinen and Austin, 2010; Greenley *et al.*, 2011; Kang *et al.*, 2012)，尤其 NGS 定序適合開發大量的 SSR 及 SNP 分子標誌，已經應用在海洋生態保育 (Novak *et al.*, 2020)、分子遺傳育種 (張等, 2017)、物種鑑別 (Saarinen and Austin, 2010; Greenley *et al.*, 2011)、野生與養殖個體功能性基因表現比對 (黃, 2019)、基因定位

(Valenzuela-Quinonez, 2016)、育種選拔 (Robledo *et al.*, 2018)、系群建構 (Kang *et al.*, 2013) 與漁業行動管理的依據 (Willete *et al.*, 2014) 上。

本研究應用次世代定序平台為澎湖中心水產生物種原庫養殖的德班氏活額蝦建構核酸序列資料庫與限制酶切位點標定法核酸定序 (RAD-Seq) 資料庫，以生物資訊軟體解析出大量 SSR 及 SNP 基因座，篩選具多型性之 SSR 及 SNP 分子標誌來檢測種原庫自群繁殖之 F3 種蝦遺傳結構，並與對照組自觀賞水族市場購得從菲律賓宿霧海域野外採捕的野生族群比較基因歧異度及遺傳距離，來評估種原庫內養殖族群有無顯著近親選拔壓力而發生基因窄化現象，作為種原維護管理及育種策略參考。此外，德班氏活額蝦之核酸序列、SSR、RAD-seq 與 SNP 資料庫及分子標誌可以提供種原利用者在雜交育種、改良或變異品種研發所需的基因型資料，也可在分子指紋鑑定及基因定位辨識作應用。

材料與方法

一、德班氏活額蝦 DNA library 的製備與定序

澎湖中心水產生物種原庫蓄養之德班氏活額蝦種蝦第一代親本 53 尾 (46 雌 7 雄)，第二代、三代親本皆逢機挑選 50 尾 (25 雌 25 雄) 自群逢機繁殖，從群體中挑選 1 隻雌性樣本的肌肉組織作為 DNA library 的定序材料，使用 DNeasy Blood and Tissue kit (Qiagen Ltd.) 萃取基因組 DNA (genomic DNA)，以 DNA 螢光定量儀 Qubit 2.0 檢定基因組 DNA 質量與定量，DNA 濃度總量 > 2 μg ，DNA 質量吸光值 A260/280 比值 = 2；A260/230 比值 > 1.8 的高純度基因組 DNA 送至陽明大學基因體中心製備 DNA library 及核酸定序。DNA library 製備採用 Ovation® Ultralow V2 DNA-Seq library preparation kit (Tecan Genomics, Inc.)。採用次世代定序平台 Illumina MiSeq 核酸定序服務平臺 (Illumina, San Diego, CA, USA) 進行 20X 核酸定序。

二、定序資料處理

定序原始序列 (raw data) 的序列讀長 (read)

為 35 – 301 bp，使用生物資訊分析軟體 CLC Genomics Workbench (Qiagen Ltd.) 進行序列修剪 (trim sequence)，品質值 (quality scores) Q 值設定為 Q20 (定序錯誤率 1%)，trim 完產出 clean data 序列。由於未有德班氏活額蝦基因體序列已解析做為參考序列 (reference sequence)，遂透過生物資訊軟體 CLC biogenomics workbench (Qiagen Ltd.) 及 ABySS (Simpson *et al.*, 2009) 兩種分析分別重新組裝 (*de novo* assembly)，取得重疊群 (contigs) 序列片段清單。

三、SSR 預測與引子設計

經過 CLC Genomics Workbench 及 ABySS 兩種軟體進行 *de novo* assembly，將產出 500bp 以上之 contig，以 MISA v1.0 進行簡單重複序列 (SSR) 的預測分析，預測目標參數設定為雙核酸重複次數 6 次或以上、三核酸重複次數 5 次或以上、四核酸重複次數 5 次或以上、五核酸重複次數 5 次或以上及六核酸重複次數 5 次或以上。將 CLCbio 及 ABySS 組裝 contigs 預測之 SSR 目標區，使用 Bowtie2 v2.1.0 進行比對分析，找出兩種方式皆有的 SSR 目標區。從交集分析的到的目標中，由前後各 100 bp 區域中進行引子設計。針對預測出之 SSR 基因座引子進行多型性測試。初步先利用 SSR 引子對族群和野生族群挑 10 個個體進行 PCR 再現性分析，挑選總共 10 組具有與預測長度相同信號具有可判讀、且具多型性之 SSR 分子標誌進行螢光標定。

四、基因歧異度與族群分化分析

使用 GeneDirtX (GeneDireX, Inc.) 萃取 36 隻種原庫養殖之穩定繁殖 F3 世代德班氏活額蝦與對照組 31 隻由觀賞水族市場購得從菲律賓宿霧海域野外捕撈的野生德班氏活額蝦基因組 DNA，以 10 組多型性 SSR 分子標誌進行 PCR 擴增，以螢光毛細管電泳 ABI 3730XL DNA Analyzer (Thermo Fisher Scientific Inc.) 搭配 GeneMapper v4.0 軟體 (Thermo Fisher Scientific Inc.) 分析 SSR 基因座對偶基因 (allele) 片段大小後，以 Fsat 軟體 (Goudet, 2001) 及 Genepop 雲端軟體

(<https://genepop.curtin.edu.au/>) 計算每組 SSR 基因座之對偶基因數目、對偶基因豐富度 (allele richness, AR)、基因歧異度 (gene diversity, Dk)、PIC 值、異型合子觀察值 (H_o)、異型合子期望值 (H_e)、哈溫平衡 (Hardy-Weinberg equilibrium, HWE) 及與遺傳分化指數 (genetic differentiation, F_{st}) 相關之組間或整體遺傳分化指數 (Raymond and Rousset, 1995)。使用邦弗朗尼校正法 (Bonferroni correction) (Rice, 1989) 校正顯著水準，利用 Arlequin Ver 3.5.2 軟體 (Excoffier and Lischer, 2010) 的 p 值檢定 (p -value < 0.05) 估算 Bootstrap 10,000 次後兩族群之間的基因歧異度相關係數、遺傳分化指數 (F_{st}) 檢定分析。以 STRUCTURE 2.3 軟體 (Pritchard *et al.*, 2000) 提供的貝式歸群分析法檢測兩族群之遺傳結構，以 K 平均數集群分析法模擬 67 個體的分群數 (K) 為 1 – 2，進行 100,000 次的馬爾科夫鏈的演算法 (Markov Chain Monte Carlo, MCMC) 及 10,000 次重複取樣，使用 Structure Harvest (Earl and vonHoldt, 2012) 計算 10 組後驗機率 (posterior probability) 檢測最佳分群數 K 值。

五、RAD-Seq library 製備

挑選種原庫穩定繁殖之 F3 養殖蝦與對照組菲律賓宿霧海域野外捕撈的野生蝦各 5 隻個體，利用肌肉組織以 GeneDirtX (GeneDireX, Inc.) 萃取基因組 DNA，再以 DNeasy Blood and Tissue kit (Qiagen) 對粗萃 DNA 再次進行純化，以達成 Qubit 螢光定量儀定量質量 (QC) 檢驗，DNA 總量 > 2 μ g，無雜質污染，質量達 A260/280 值約 2.0；A260/230 值 > 1.8 的基因組 DNA。兩族群高純度之 1 μ g 基因組 DNA 以 1 unit EcoRI (New England BioLabs Inc.) 限制酶置於 37°C 培養箱中切割 16 hr，隨後利用 Illumina DNA Prep kit 連接 P1 adapter (Illumina Inc.) 連接酶，以超音波震盪器 Covaris E210 (Covaris, Inc.) 進行 DNA 隨機斷裂，以 agarose 膠篩選 300 – 500 bp DNA 片段及使用 Illumina DNA Prep kit 萃取 agarose DNA，以微流體電泳分析儀 Agilent Technologies 2100 bioanalyzer (Agilent Technologies, Inc.) 測定 DNA 片段 QC 檢驗，質量須達 A260/280 值約 2.0；A260/230 值 > 1.8 的 DNA 片段，以 ABI

StepOnePlus RealTime PCR 儀隨後合上 A-tailing 及 P2 adapter 來製備 RAD library，委託源資國際生物科技股份有限公司採用高通量定序系統 Illumina HiSeq 4000 次世代核酸定序服務平台，進行 Illumina Paired-End 101bp 的 RAD-seq 定序，定序量約為 20G。

六、RAD-Seq 資料處理與 SNP 預測

將產出的 RAD-Seq 定序原始數據利用國家實驗研究院國家高速網路與計算中心的台灣杉 TAIWANIA 高速計算主機上進行生物資訊分析。以巨量核酸與蛋白質結構高效能分析平台上的生醫軟體 STACK v2.4 (Catchen *et. al.*, 2013) 進行 *de novo* pipeline 解析 SNP 基因座並進行族群分析。先以 Process_radtags 指令對 Paired-End raw data 序列修剪，切割 EcoRI 限制酶、barcode 序列，長度設定 120 bp。然後以 ustacks 指令將樣本 clean data 定序資料整理成 10 個個體基因座資訊，序列差異門檻參數設定 M 為 5，最小定序深度 m 為 5；以 cstacks 彙整所有樣本基因座資料建立完整族群基因座資訊目錄，平均覆蓋率參數設定 n 為 5；再以 sstacks 判讀個體 SNP 基因型；以 tsv2bam 指令將個體 SNP 基因型轉換為以 SNP 基因座排序輸出；指定最小變異頻率 (minor allele frequency, MAF) 在所有樣本要大於為 5%，為後續族群分析篩選變異點用。最後以 population 指令進行族群基因歧異度分析，輸出所有 haplotype 資訊，設定最小定序深度 m 為 7；r 參數設定 0.7 指定輸出的基因座至少出現在 70% 樣本中；structure 的參數設定 p 值檢定小於 0.05。

利用 CLC Genomics Workbench (Qiagen Ltd.) 將輸出 10 個體 SNP 對偶基因序列以歐式距離 (Euclidean distance) 運算個體之間遺傳距離，進行集群分析 (cluster analysis)，製作親緣關係樹狀圖。

結果與討論

本研究針對種原庫持續培育三代的德班氏活額蝦進行基礎基因體研究，以次世代定序平台 Illumina MiSeq 定序產出基因組核苷酸序列總數據

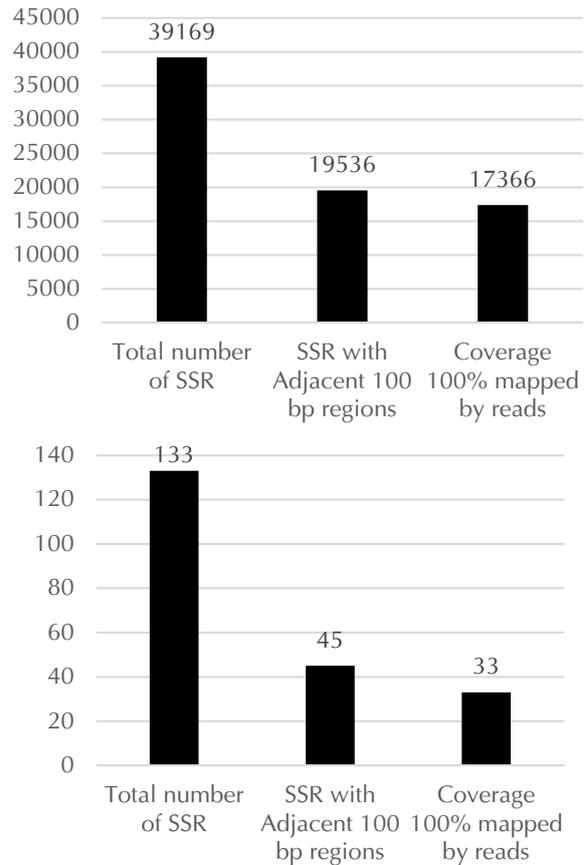


Fig. 1 Summary of candidate SSRs in assembly contigs (A) assembled by CLCbio and (B) assembled by ABySS.

為 20,965,331 bp，序列條數 (read) 約 4.1 百萬，read 長度為 301 bp。由於德班氏活額蝦非模式物種，無已發表的基因體序列做為參考序列可協助進行組裝拼接，遂透過生物資訊軟體重新組裝，直接對序列進行拼接、組裝分析取得較長、較完整、覆蓋率高的 contigs 序列片段，以獲得較長基因組核苷酸序列。利用 CLC Genomics Workbench 對 raw data 進行重新組裝拼接，組裝後大於 500bp 的 contig 數為 76,394，最大長度為 16,557 bp，總核苷酸數為 66,090,000 bp；利用 ABySS 軟體對 CLC 組裝一次的 contig 再進行一次重新組裝，組裝後 clean data 大於 500 bp 的 contig 數為 548，最大長度為 17,053 bp，總核苷酸數為 503,952 bp。組裝出的 contig 序列片段以 misa v1.0 進行 SSR 的預測分析，以 CLC 組裝的 contig 預測的 SSR 總數為 39,169，在 SSR 區域兩端具有 100 bp 可設計 primer 的數量為 19,536，可被 raw reads 覆蓋率達 100% 的數量有 17,366 (Fig. 1A)。以 ABySS 組裝的 contig 預測的 SSR 總數為 133，在 SSR 區域兩端具有 100 bp 可

Table 1 Summary table of genetic variability between farmed (RPH) and wild (RO) *Rhynchocinetes durbanensis* populations

Loci	A		A _R		D _k		PIC		H _o		H _e		F _{is}		
	RPH	RO	RPH	RO	RPH	RO	RPH	RO	RPH	RO	RPH	RO	RPH	RO	
P1	1	1	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	NA	NA
P2	2	2	2.000	2.000	0.475	0.503	0.355	0.368	0.000	0.000	0.468	0.495	1.000**	1.000**	
P3	2	3	2.000	3.000	0.457**	0.572**	0.346	0.456	0.000	0.000	0.451	0.563	1.000**	1.000**	
P4	2	2	2.000	2.000	0.500	0.500	0.375	0.375	1.000	1.000	0.507	0.508	-1.000	-1.000	
P5	3	3	3.000	3.000	0.514	0.241	0.449	0.220	0.000	0.000	0.507	0.237	1.000**	1.000**	
P6	1	1	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	NA	NA	
P7	1	1	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	NA	NA	
P8	2	2	2.000	2.000	0.437	0.452	0.334	0.342	0.000	0.000	0.430	0.444	1.000**	1.000**	
P9	2	2	2.000	2.000	0.500	0.500	0.375	0.375	1.000	1.000	0.507	0.508	-1.000	-1.000	
P14	1	1	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	NA	NA	

A: number of alleles, AR: allele richness, D_k: gene diversity, PIC: polymorphism information content. H_o: observed heterozygosity, H_e: expected heterozygosity, F_{is}: inbreeding coefficient value. Symbol * indicates significant difference at $p < 0.05$; ** indicates significant difference at $p < 0.001$.

設計 primer 的數量為 45，可被 raw reads 覆蓋率達 100% 的數量有 33 (Fig. 1B)。利用 Bowtie2 軟體進行兩種組裝方式預測 SSR 的交集分析，交集分析產生的 SSR 基因座目標數為 16。

以 10 組 SSR 引子對 36 隻澎湖中心養殖族群 F3 世代與對照組 31 隻由觀賞水族市場購得從菲律賓宿霧海域野外捕撈的野生德班氏活額蝦進行基因歧異度分析。由以上 10 組 SSR 引子進行種族群和野生族群之基因歧異度分析 (Table 1)，結果顯示養殖群對偶基因數目及對偶基因豐富度與野生族群比較皆沒有顯著差異；10 組 SSR 分子標記 PIC 值在養殖族群和野生族群之間介於 0.334 (P8) 至 0.456 (P3) 左右。養殖群在大部分 SSR 基因座分析皆顯示平均 H_o 大於平均 H_e，表示養殖群內異型合子個體在族群中所佔之比例大於根據哈溫定律所估算之期望值，養殖群內遺複雜度高。測試異質度缺失 (heterozygote deficit) HWE 顯示 10 組基因座皆顯著，F_{is} 近交係數的 p 值檢定 (p -value < 0.05) 也半數是顯著的，顯示 SSR 對偶基因異型合子不足。兩族群 F_{st} 為 0.0184 (Table 2)， p 值檢定無顯著，顯示在澎湖中心繁殖族群與庫外之野生

Table 2 Pairwise estimate of genetic diversity F_{st} (below) and Delta μ^2 distance (above) between farmed (RPH) and wild (RO) *Rhynchocinetes durbanensis* populations

Population	RPH	RO
RPH		0.04478
RO	0.01841	

No significantly difference ($p > 0.05$) between populations

族群之間分化 (genetic differentiation) 相當低，兩個族群之間的平均基因變異數為 0.04478，無顯著差異，表示 2 族群非常相似。利用貝氏演算法分析種族群和野生族群個體的 K 平均數集成分成 2 群的圖形 (Fig. 2)，各柱體分別代表每隻德班氏活額蝦樣本，相同顏色代表具有近似的基因型，後驗機率 (posterior probability) 檢測兩個族群無明顯的遺傳結構，兩個族群 10 組 SSR 對偶基因頻度分布無顯著差異存在。本研究 SSR 分析結果得知自群穩定繁殖三代的德班氏活額蝦族群基因歧異度仍很高，與市售野生群無顯著遺傳差異性及遺傳分

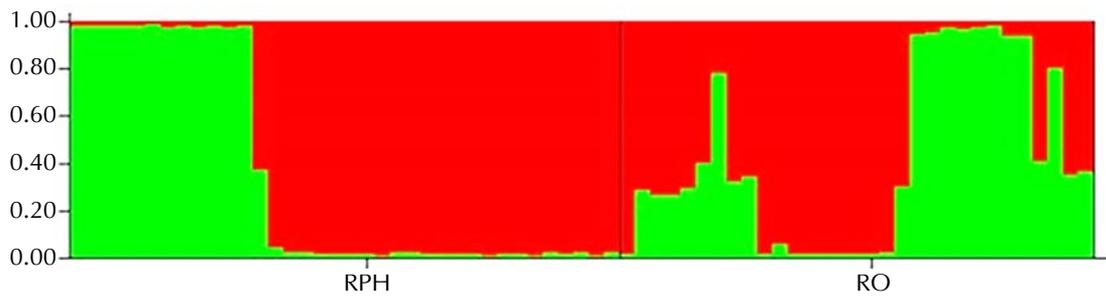


Fig. 2 Genetic clustering patterns of *R. durbanensis* for different sampled sites as inferred using Bayesian methods implemented in STRUCTURE. Each bar represents an individual within the farmed (RPH) and wild (RO) *Rhynchocinetes durbanensis* populations.

Table 3 Molecular characteristics of SNP allele information between farmed (RPH) and wild (RO) *Rhynchocinetes durbanensis* populations

Population	Private alleles	Num.	Homo.	Het.	Hom. rate (%)	Het. rate (%)	Nucleotide Diversity, π
RPH	194,579	5	156,533	38,046	80.45	19.55	0.32 \pm 0.02
RO	162,109	5	137,622	24,487	84.89	15.11	0.27 \pm 0.02

Private alleles: the number of private alleles in each population; Num.: the number of individuals per locus in each population; Homo.: homozygote alleles found within each population; Het.: heterozygote alleles found within each population; Hom. rate (%): homozygosity rate found within each population; Het. rate (%): heterozygosity rate found within each population.

化。單一核苷酸多型性 SNP 分子標誌解析度比 SSR 更高 (Toonen *et al.*, 2013; Lemopoulos *et al.*, 2019)，可以區別親緣相近的育種系群遺傳差異 (Iheshiulor *et al.*, 2016; Robledo *et al.*, 2018) 與判別親緣相近的地域適應性(local adaptation) 族群 (Perrier *et al.*, 2017; Vendrami *et al.*, 2019)，可以應用於檢測只有自群繁殖 F3 種蝦基因型序列和野生個體基因序列差異性。從澎湖中心繁殖族群及野生族群個體建構限制酶切位點標定法 RAD-seq 定序資料庫，獲得定序品質值 Q20 錯誤率 1% 的 clean data 數據量有 138.49 Gb，辨識出變異點共 4,382,312 位點，contig 長度平均 427.98 bp。以生物資訊族群分析養殖與野生族群個體之間 SNP 位點共 429,078 個，養殖族群特有對偶基因數量為 162,109 個，較少於野生族群特有對偶基因數量為 162,109 個 (Table 3)；養殖族群平均同型合子 (homozygote) 有 137,622 個 (佔 SNP 總數 84.89%)、平均異型合子為 24,487 個 (佔 SNP 總數 15.11%)、核酸多樣性值 (π , nucleotide diversity) 為 0.32 \pm 0.03；野生族群平均同型合子有 156,533 個 (佔 SNP 總數 80.45%)、平均異型合子為 38,046 個

(佔 SNP 總數 19.55%)、 π 值 0.26 \pm 0.02；以上數據 p 值檢定結果 p -value > 0.05 無顯著差異，顯示養殖族群基因歧異度狀況與野生族群無顯著差異。以歐式距離運算個體之間遺傳距離的親緣集群分析，顯示兩族群間遺傳距離 0.087， p 值檢定無顯著遺傳分化。親緣遺傳距離樹狀圖顯示養殖個體與野生個體分成兩群 (Fig. 3)，表示養殖隨機交配 F3 世代種蝦和野生族群基因型已有差異，SNP 分子標誌在基因差異度檢測比 SSR 分子標誌的解析度更高。種原庫自群繁殖 3 世代已漸漸和市售野生族群有些微基因型差異，仍應持續追蹤累積更多子代數固定性狀及基因型，可得顯著遺傳分化的養殖品系/品種。另一方面，SNP 分析樣本數應增加至 30 隻，在族群統計上才不易有偏差，且種原庫外也應從多個區域/國家進行採樣，增加族群地理的遺傳多樣性分析。

推廣人工繁養殖產技術來降低消費市場對野生資源的依賴，兼具生態保育意義及成本優勢。經由澎湖中心水產生物種原庫自群隨機交配三代種苗，可以持續並穩定供應人工繁養殖產技術轉移的蝦苗。根據近交係數的計算，3 代全同胞兄

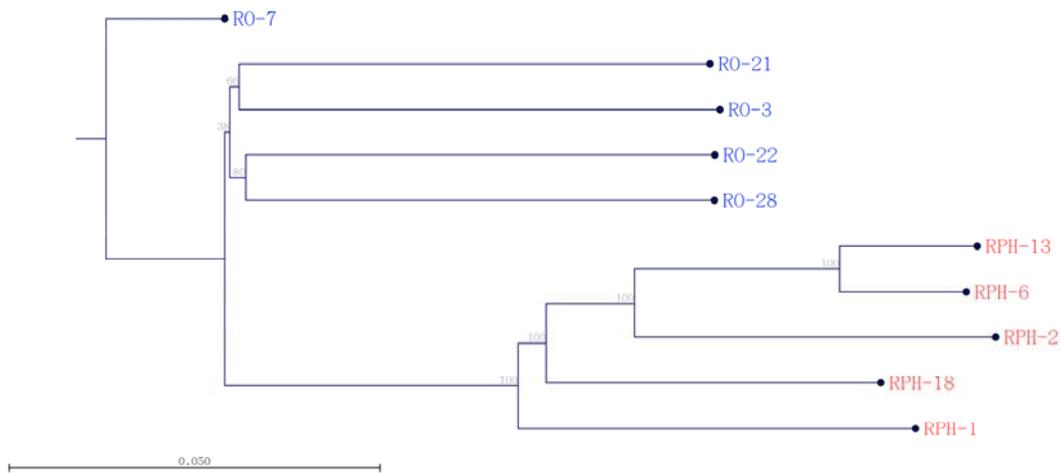


Fig. 3 Cluster analysis genetic distance phylogeny tree of *Rhynchocinetes durbanensis* generated from farmed (RPH) and wild (RO) samples based on 429,078 SNP sequences. Branch numbers indicate bootstrap support values. Scale bar indicates substitutions per site.

妹交配後代的近交係數高達 50%，平均衰退 36.5%，雖然自群繁殖為較緩慢近交（近交率低於 2.5%），平均衰退低於 10%（劉, 2007a, b），然而，為了確保種蝦的遺傳質量，避免因有效族群數量少或近交頻繁累積有害基因，隨機交配的種蝦個體基因型頻度調查對小族群的表現形質保持、基因漂變偵測與復育策略和未來開發新品系的育種策略的規劃是必要的基礎研究（James, 1962）。曾等（2014）對紅色吳郭魚進行近親交配對生長影響研究指出近親退化現象始於 F3 世代，在體長及死亡率有顯著影響，種原庫亦有關注到養殖種蝦 F3 世代有體長變小及死亡率增加現象；故本研究應用次世代定序及生物資訊解析基因序列，已有許多魚類基因型育種研究在小樣本數、節省實驗時間及經費上皆能有效分析養殖種蝦族群狀態（population status）如對偶基因頻率分布、基因歧異度及育種系群之間遺傳距離（Reitzel *et al.*, 2013; Toonen *et al.*, 2013）。SSR 對偶基因頻度分布可以區別近代培育出的養殖系群與野生群的遺傳差異（Kohlmann *et al.*, 2005），SSR 分析能為鯰魚（Nakron and Moeikum, 2009）及鮭魚（Norris *et al.*, 1999）種魚育種策略管理提供有效的數據驗證。本研究使用的 10 組 SSR 基因座在 2 族群內對偶基因數皆不多，大部分未能符合異質度缺失哈溫平衡，推測原因是本研究針對 SSR 的篩選參數設定嚴格及經過 2 個軟體篩選 2 次，雖然可取得可信

度高，短時間可應用之引子，然而 PIC 含量不高，未來要在從德班氏活額蝦之 SSR 資料庫中篩選更高 PIC 含量、更多對偶基因數目、符合哈溫平衡的 SSR 引子對，作為系群之間基因歧異度分析。

本研究 SSR 及 SNP 對偶基因頻度分析結果顯示養殖族群 SNP 對偶基因同型結合子比率較高，但是與野生族群相比未達顯著差異性，且與野生族群相比基因歧異度並沒有顯著差異，推測可能族群內遺傳變異量（variation）大導致族群之間差異性不顯著。另外，SNP 基因座頻度分析的參數如基因型缺值 n 、最小定序深度 m 、較少基因型讀取數目 MAF、輸出基因座在樣本比率 r 等設定會影響基因歧異度分析結果（Wang, 2016），須得多次分析求最適合參數，以達正確研究成果（Wellmann and Bennewitz, 2019）。

本研究在 SNP 試驗所使用之樣本數僅從養殖及野生群中隨機採樣 5 隻個體進行親緣遺傳距離分析已能明顯區分出養殖個體與野生個體 2 群，表示養殖族群基因型已固定，繼續培育解析基因型的養殖品系，在育種上可以利用 SNP 分子標誌為工具來研究及定位性狀基因型，進行精確的基因調控、加速功能性基因的選拔或穩定優良種原的基因優勢等分子標誌輔助育種（marker-assisted selection, MAS）研究。MAS 廣泛應用在水產品品系改良與育種效率提升上（曾與林, 2011、張等, 2017），如米蝦體色開發（朱, 2016）、白蝦抗病基因

篩選 (李, 2013)、鰻魚消化基因定位 (查, 2014) 及龍膽石班的肌肉增生基因定位 (林等, 2014) 等, 分子標誌輔助育種研究為水產繁養殖產業帶來突破性的發展。

結 論

本研究應用次世代定序平台建構德班氏活額蝦核酸序列資料庫與 RAD-Seq 資料庫;再以生物資訊軟體解析出大量 SSR 及 SNP 分子標誌, 記錄種原庫的保種生物遺傳資源 (genetic resource), 對種原的遺傳提升、育種工作及未來種原的基因體資訊應用非常有幫助。本研究 SSR 對偶基因頻度分析結果顯示在澎湖中心水產生物種原庫所繁殖德班氏活額蝦與庫外之菲律賓野生族群兩個族群之間基因歧異度無顯著分化, 養殖族群多樣性高無顯著近交選拔壓力;高解析度的 SNP 分析以基因序列譜系分析可以區分養殖與野生族群個體。

謝 辭

本研究得以順利完成, 承蒙本所澎湖海洋生物研究中心提供樣本材料, 特此致謝。

參考文獻

- 朱珮云 (2016) 建立彩色米蝦轉錄體基因庫與顯色性狀功能性基因標記. 國立臺灣海洋大學水產養殖學系碩士論文, 79 pp.
- 宋嘉軒, 城振誠 (2015) ISSR 應用在德班氏活額蝦 (*Rhynchocinetes durbanensis*) 遺傳分析的初步研究. 103 年度臺灣水產學會會員大會暨學術論文發表會, 臺灣大學, 臺北, 臺灣, 457-458.
- 李昆霖 (2013) 以次世代定序資料重組南美白蝦的轉錄基因體並探討其基因表現. 國立臺灣大學漁業科學研究所 碩士論文, 60 pp.
- 杜金蓮, 曾福生, 王姿文, 張錦宜 (2015) 九孔基因歧異度對幼苗活存率及幼貝成長表現之影響. 水產研究, 23 (1): 67-75.
- 林峰右, 吳育甄, 葉信利 (2014) 龍膽石班肌肉增生相關基因之初步探討. 水試專訊, 47: 19-23.
- 查昱如 (2014) 以次世代定序技術解析臺灣常見三種養殖鰻魚 (日本鰻、鱸鰻、太平洋雙色鰻) 之轉錄體及探討其基因表現概況. 國立臺灣大學漁業科學研究

- 所 碩士論文, 81 pp.
- 城振誠 (2015) 我國海水觀賞蝦培育技術發展概況. 農業生技產業季刊, 43: 56-65.
- 城振誠, 林金榮 (2014) 海水觀賞蝦繁養殖. 水產試驗所技術手冊6, 行政院農業委員會水產試驗所編印, 基隆, 臺灣, 57 pp.
- 城振誠, 蔡萬生 (2005) 餌料、投餵策略及溫度對德班氏活額蝦初期蝦苗成長之影響. 水產研究, 13 (1): 45-52.
- 城振誠, 蔡萬生 (2007) 德班氏活額蝦幼苗之發育研究. 水產研究, 15 (1): 13-35.
- 徐元華, 林彥宏, 余祁暉 (2015) 觀賞魚及周邊產業趨勢與發展策略. 農業生技產業季刊, 43: 1-13.
- 陳彥群 (2012) 探討次世代定序技術的定序偏差對全新基因體組裝的影響. 國立成功大學工程科學系 碩士論文, 134 pp.
- 張格銓, 黃瀛生, 張湧泉, 劉富光 (2012) 臺灣地區日本鰻鰻線之年內族群遺傳結構. 水產研究, 20 (1): 51-59.
- 張格銓, 黃瀛生, 陳建彰, 陳榮華 (2017) 制定水產分子遺傳育種的研究策略. 水試專訊, 59: 36-39.
- 曾福生, 林金榮 (2011) 應用分子標誌輔助選拔育種問題的探討. 水試專訊, 34: 32-35.
- 曾福生, 杜金蓮, 朱惠真, 謝立偉, 林金榮 (2013) 臺灣及日本野生九孔粒線體細胞色素c氧化酶基因序列變異及遺傳結構之分析. 水產研究, 21 (1): 35-42.
- 曾福生, 杜金蓮, 謝立偉, 王姿文, 張錦宜 (2014) 近親交配對紅色吳郭魚成長與死亡率之影響. 水產研究, 22 (1): 63-70.
- 黃之暘, 何源興, 陳文義, 郭慶老 (2012) 觀賞水族產業發展與管理. 水產試驗所特刊第15號, 行政院農業委員會水產試驗所編印, 基隆, 臺灣, 160 pp.
- 黃章文 (2015) 臺灣觀賞米蝦種原系開發與分子選育應用. 農業生技產業季刊, 43: 20-25.
- 黃慶輝 (2019) 次世代定序及基因體研究應用於水產科技之淺說. 水試專訊, 65: 44-47.
- 劉富光 (2007a) 近親交配與繁殖育種 (上). 水試專訊, 17: 31-34.
- 劉富光 (2007b) 近親交配與繁殖育種 (下). 水試專訊, 18: 12-15.
- Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresk and E. A. Johnson (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLoS ONE, 3 (10): e3376.
- Earl, D. A. and B. M. vonHoldt (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour., 4(2) : 359-361.
- Excoffier, L. and H. E. L. Lischer (2010) Arlequin ver 3.5:

- A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.*, 10: 564-567.
- Greenley, A. P., A. Muguia-Vega, A. Saenz-Arroyo and F. Micheli (2011) New tetranucleotide microsatellite loci in pink abalone (*Haliotis corrugata*) isolated via 454 pyrosequencing. *Conserv. Genet. Resour.*, 4(2):265-268.
- Gordon, I. (1936) On the macruran genus *Rhynchocinetes*, with description of a new species. *Proc. Zool. Soc. Lond.*, 106(1): 75-88
- Goudet, J. (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). <http://www2.unil.ch/popgen/softwares/fstat.htm>.
- Hellberg, M. E., R. S. Burton, J. E. Neigel and S. R. Palumbi (2002) Genetic assessment of connectivity among marine populations. *Bull. Mar. Sci.*, 70 (1): 273-290.
- Hellberg, M. E. (2007) Footprints on water: The genetic wake of dispersal among reefs. *Coral Reefs*, 26 (3): 463-473.
- Iheshiulor, O. O. M., J. A. Woolliams, X. Yu, R. Wellmann, T. H. E. Meuwissen (2016) Within- and across-breed genomic prediction using whole-genome sequence and single nucleotide polymorphism panels. *Genet. Sel. Evol.*, 48: 15.
- James, J. (1962). The spread of genes in random mating control populations. *Genet. Res.*, 3(1): 1-10.
- Jombart T. and I. Ahmed (2011) ADEGENET 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21): 3070–3071.
- Kang, J. H., J. Y. Park and H. S. Jo (2012) Rapid development of microsatellite markers with 454 pyrosequencing in a vulnerable fish, the mottled skate, *raja pulchra*. *Int. J. Mol. Sci.*, 13(6): 7199-7211.
- Kang, J. H., S. G. Yang, T. S. Moon, J. Y. Park and T. J. Choi (2013) Development of microsatellite markers for the kelp grouper *epinephelus bruneus* by 454 pyrosequencing and transfer to related species. *Genet. Mol. Res.*, 12 (4): 5485- 5493.
- Kohlmann, K., P. Kersten and M. Flajshans (2005) Microsatellite-based genetic variability and differentiation of domesticated, wild and feral common carp (*cyprinus carpio* L) populations. *Aquac.*, 247: 253-266.
- Lemopoulos, A., J. Prokkola, S. Uusi-Heikkilä, A. Vasemägi, A. Huusko, P. Hyvaerinen, M. Koljonen, J. Koskiniemi and A. Vainikka (2019) Comparing RADseq and microsatellites for estimating genetic diversity and relatedness- Implications for brown trout conservation. *Mol. Evol.*, 9 (4): 2106-2120.
- Moore, S. S., L. L. Sargeant, T. J. King, J. S. Mattick, M. Georges and D. J. Hetzel (1991) The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. *Genomics*, 10: 654-660.
- Na-nakorn U. and T. Moeikum (2009) Genetic diversity of domesticated stocks of striped catfish, *Pangasianodon hypophthalmus* (Sauvage 1878), in Thailand: Relevance to broodstock management regimes. *Aquac.* 297: 70-77.
- Novak B. J., Fraser D. and T. H. Maloney (2020) Transforming ocean conservation: applying the genetic rescue toolkit. *Genes*, 11 (2): 209.
- Norris, A. T., D. G. Bradley and E. P. Cunningham (1999) Microsatellite genetic variation between and within farmed and wild atlantic salmon (*salmo salar*) populations. *Aquac.*, 180 (3): 247-264.
- Perrier C., A-L. Ferchaud, P. Sirois, I. Thibault and L. Bernatchez (2017) Do genetic drift and accumulation of deleterious mutations preclude adaptation? Empirical investigation using RADseq in a northern lacustrine fish. *Mol. Ecol.*, 26: 6317–6335.
- Pritchard, J. K., M. Stephens, and P. Donnelly (2000) Inference of population structure using multilocus genotype data. *Genet.*, 155: 945–959.
- Reitzel A., Herrera S., Layden M., Martindale M. and T. Shank (2013) Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. *Mol. Ecol.*, 22(11):2953–2970.
- Robledo D., C. Palaiokostas, L. Bargelloni, P. Martínez and R. Houston (2018) Applications of genotyping by sequencing in aquaculture breeding and genetics. *Rev. Aquac.*, 10 (3): 670-682.
- Saarinen, E. V. and J. D. Austin (2010) When technology meets conservation: Increased microsatellite marker production using 454 genome sequencing on the endangered okaloosa darter (*etheostoma okaloosae*). *J. Hered.*, 101 (6):784-788.
- Simpson J.T., K. Wong, S. D. Jackman, J. E. Schein, S. J. Jones, I. Birol (2009) ABySS: A parallel assembler for short read sequence data. *Genome Res.*, 19 (6): 1117–1123.
- Toonen, R. J., J. B. Puritz, Z. H. Forsman, J. L. Whitney, I. Fernandez-Silva, K. R. Andrews and C. E. Bird (2013) ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ* 1,

- e203.
- Tishkoff, S. A. and B. C. Verrelli (2003) Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu. Rev. Genom. Hum. Genet.*, 4, 293–340.
- Valenzuela-Quiñonez F. (2016) How fisheries management can benefit from genomics? *Brief. Funct. Genom.*, 15 (5): 352-357.
- Vendrami, D. L. J., M. De Noia, L. Telesca, W. Handal, G. Charrier, P. Boudry, L. Eberhart-Phillips and J. Hoffman (2019) RAD sequencing sheds new light on the genetic structure and local adaptation of European scallops and resolves their demographic histories. *Sci. Rep.*, 9: 7455.
- Wang, J. (2016) Individual identification from genetic marker data: developments and 2 accuracy comparisons of methods. *Mol. Ecol. Res.*, 16: 163-175.
- Weber, J. L. (1990) Informativeness of human (dC-dA)n.(dG-dT)n polymorphisms. *Genomics*, 7 (4): 524-530.
- Wellmann, R. and J. Bennewitz (2019) Key Genetic Parameters for Population Management. *Front. Genet.*, 10: 667.