

SO SÁNH SỰ BIỂU HIỆN GEN LIÊN QUAN ĐẾN TÍNH TRẠNG SINH SẢN CỦA TÔM SÚ (*Penaeus monodon*) BẰNG KỸ THUẬT GIẢI TRÌNH TỰ RNA (RNA-seq) VÀ PCR ĐỊNH LƯỢNG (qPCR)

Trần Thị Hải Yến¹, Nguyễn Minh Thành^{1,*}

¹Trường Đại học Quốc tế, Đại học Quốc gia Thành phố Hồ Chí Minh

*Email: nmthanh@hcmiu.edu.vn

TÓM TẮT

Nghiên cứu này sử dụng kỹ thuật giải trình tự RNA (RNA-seq) và PCR định lượng (qPCR) để khảo sát sự biểu hiện của các gen liên quan đến tính trạng sinh sản ở các mô đích khác nhau của tôm sú cái tự nhiên và gia hóa. Mục tiêu của nghiên cứu là so sánh sự tương đồng trong phân tích sự biểu hiện gen bằng 2 kỹ thuật khác nhau. Đối với mô gan tụy, 6 trong số 10 gen khảo sát có sự biểu hiện tương đồng giữa 2 kỹ thuật sử dụng. Tuy nhiên, gen *cytochrome P450 2L1-like* có biểu hiện giảm khi phân tích bằng kỹ thuật RNA-seq và biểu hiện tăng khi phân tích bằng qPCR. Đối với mô buồng trứng, gen *anti-lipopolysaccharide factor-like* và gen *NADH dehydrogenase subunit 1* có biểu hiện tương đồng khi phân tích bằng cả 2 phương pháp. Các gen biểu hiện ở mô gan tụy cũng được khảo sát ở mô buồng trứng. Các gen này không có biểu hiện hoặc biểu hiện không có ý nghĩa thống kê khi phân tích bằng kỹ thuật RNA-seq hoặc qPCR. Nghiên cứu cũng thảo luận các công cụ tin sinh học ảnh hưởng đến mức độ chính xác trong định lượng và những vấn đề có thể làm sai khác kết quả giữa các kỹ thuật sử dụng để phân tích sự biểu hiện gen.

Từ khóa: Tôm sú, *Penaeus monodon*, sự biểu hiện gen, RNA-seq, qPCR.

1. ĐẶT VẤN ĐỀ

Kỹ thuật giải trình tự RNA (RNA-seq) đã trở thành kỹ thuật phổ biến để nghiên cứu sự biểu hiện gen, thay thế cho kỹ thuật vi mảng (microarray). Kỹ thuật RNA-seq có nhiều ưu điểm so với kỹ thuật vi mảng: (i) không cần thông tin ban đầu của các trình tự gen, (ii) tốc độ giải trình tự cao, tăng đáng kể mức độ bao phủ toàn bộ hệ phiên mã nên có nhiều khả năng phát hiện các đoạn phiên mã hiếm, (iii) khắc phục được nhược điểm thiên lệch của kỹ thuật vi mảng trong quá trình lai các mồi (probe), (iv) chi phí thấp hơn so với kỹ thuật vi mảng, (v) kết nối các đoạn trình tự thành các contig (chuỗi các trình tự ADN liên tục và liền kề, được ghép lại từ các đoạn ADN nhỏ hơn chồng chéo lên nhau mà không có khoảng trống) cho phép nghiên cứu cấu trúc của các gen, (vi) có thể định lượng các gen biểu hiện bằng cách đếm số lượng đoạn phiên mã của mẫu phân tích [1]. Do đó, khi nghiên cứu sự biểu hiện gen đối với đa số loài thủy sản thiếu thông tin di truyền thì RNA-seq là công cụ hữu hiệu nhất. So với kỹ thuật vi mảng đã có qui trình

phân tích chuẩn, lĩnh vực nghiên cứu sử dụng kỹ thuật RNA-seq còn nhiều tồn tại trong xử lý và phân tích số liệu nên vẫn chưa có qui trình phân tích chuẩn được chấp nhận rộng rãi [2]. Hiện tại nhiều công cụ tin sinh học có sẵn có thể xử lý số liệu RNA-seq và có thể phân loại thành 2 phương pháp phân tích: (i) phương pháp định vị các đoạn trình tự (read) trực tiếp vào hệ gen hoặc hệ phiên mã tham chiếu, tiếp theo là định lượng các đoạn trình tự đã được định vị [3], [4]; (ii) phương pháp gọi là định vị giả định (pseudoalignment) chia nhỏ các đoạn trình tự thành k-mer trước khi định vị chúng vào các gen [5], [6]. Một đoạn trình tự ngắn có thể có nhiều kết quả định vị khác nhau, bao gồm không được định vị vào bất kỳ vị trí của hệ gen (0), hoặc 1 vị trí (1), hoặc nhiều vị trí của hệ gen (> 1, gọi là định vị đa vị trí (multiread)). Sau quá trình định vị, công cụ tin sinh học tiếp tục chuyển đổi kết quả định vị để định lượng sự biểu hiện gen. Mỗi phần mềm xử lý các đoạn trình tự có kết quả định vị đa vị trí theo phương pháp khác nhau (loại trừ hoặc giữ lại) và đây là nguồn gốc sai số trong định lượng biểu hiện gen

của kỹ thuật RNA-seq. Vì vậy, mức độ chính xác của định lượng biểu hiện gen hoàn toàn tùy thuộc vào phần mềm tin sinh học được lựa chọn và các thuật toán mặc định của từng công cụ.

Tôm sú (*Penaeus monodon*) là đối tượng thủy sản chủ lực ở Việt Nam và mang lại giá trị kinh tế cao cho ngành thủy sản. Năm 2024, diện tích nuôi tôm sú đạt 662.000 ha và sản lượng tôm sú đạt 284.000 tấn, chiếm 11,5% cơ cấu các sản phẩm tôm xuất khẩu với giá trị xuất khẩu là 447,4 triệu USD [7]. Kỹ thuật RNA-seq đã được ứng dụng rộng rãi trong nghiên cứu hệ gen phiên mã của tôm sú liên quan đến tính trạng tăng trưởng, dinh dưỡng, sinh sản, kháng bệnh và tính trạng chống chịu với điều kiện môi trường bất thuận. Hiện tại, số lượng hệ gen phiên mã của tôm sú lưu trữ trên cơ sở dữ liệu SRA-NCBI (Sequence Read Archive Database of the National Center for Biotechnology Information) là 587 bộ dữ liệu. Thời gian gần đây, Nguyen và cs (2025) [8] công bố kết quả nghiên cứu so sánh hệ gen phiên mã của tôm sú cái tự nhiên và gia hóa ở giai đoạn tiền thành thực và thành thực sinh sản. Nghiên cứu sử dụng kỹ thuật RNA-seq để xác định sự biểu hiện gen của các gen tiềm năng liên quan đến tính trạng sinh sản của tôm sú và sử dụng kỹ thuật qPCR để đánh giá lại kết quả gen biểu hiện từ kỹ thuật RNA-seq. Tuy nhiên, nghiên cứu chưa phân tích và so sánh chi tiết sự biểu hiện gen giữa phương pháp RNA-seq và qPCR. Vì vậy, nghiên cứu hiện tại tiếp tục phân tích sự biểu hiện gen của một số gen tiềm năng liên quan đến tính trạng sinh sản ở tôm sú cái bằng kỹ thuật RNA-seq và qPCR. Mục tiêu của nghiên cứu này là đánh giá mức độ tương đồng của định lượng gen biểu hiện từ 2 kỹ thuật phân tích khác nhau, từ đó định hướng kỹ thuật phân tích phù hợp trong nghiên cứu sự biểu hiện gen ở các đối tượng thủy sản.

2. VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Mẫu tôm thí nghiệm

Nghiên cứu sử dụng mẫu tôm sú cái tự nhiên và gia hóa ở giai đoạn tiền thành thực (giai đoạn 1) và giai đoạn thành thực sinh sản (giai đoạn 4). Số lượng mẫu tôm tự nhiên và gia hóa là 3 mẫu/giai đoạn. Mô buồng trứng và mô gan tụy là cơ quan đích của nghiên cứu sự biểu hiện gen bằng kỹ thuật RNA-seq và qPCR.

2.2. Phân tích sự biểu hiện gen bằng kỹ thuật RNA-seq

Kỹ thuật RNA-seq phân tích sự biểu hiện gen bao gồm các bước cơ bản sau: Xử lý dữ liệu thô, định vị các đoạn trình tự lên thư viện gen (hệ gen hoặc hệ phiên mã), ước tính số lượng đoạn trình tự được định vị và chuẩn hóa kết quả ước tính, phân tích biểu hiện gen sai khác giữa các điều kiện thí nghiệm khác nhau. Các đoạn trình tự ngắn của từng mẫu mô được định vị vào thư viện hệ phiên mã của tôm sú đã được công bố [8] bằng công cụ Bowtie 2 [9] và tiếp tục được sắp xếp theo thứ tự bằng Samtools. Công cụ RSEM [10] định lượng đoạn trình tự ngắn của từng mẫu mô định vị lên mỗi gen chú giải. Kết thúc bước này tạo ra các tập số liệu định lượng các đoạn trình tự ngắn cho mỗi gen chú giải. Phân tích so sánh sự biểu hiện gen khác biệt giữa giai đoạn 1 và giai đoạn 4 ở các mẫu bằng công cụ DESeq2 phiên bản 3.5.3 [11] trên nền tảng R. Những gen chú giải có tổng số các đoạn trình tự ngắn từ tất cả mẫu nhỏ hơn 10 bị loại bỏ. Một gen được xác định là có biểu hiện gen khác biệt giữa giai đoạn 1 và 4 dựa vào các giá trị: $|\log_2(\text{fold change})| > 2$, giá trị p điều chỉnh ($p\text{-adjusted (p-adj)}$) $< 0,05$.

2.3. Phân tích sự biểu hiện gen bằng kỹ thuật qPCR

Phần mềm Primer 3 được sử dụng để thiết kế các cặp primer cho phản ứng qPCR dựa vào trình tự lắp ráp của các gen bằng phương pháp lắp ráp không có hệ gen tham chiếu [8]. Phản ứng qPCR bao gồm 2x GoTaq® qPCR Master Mix (Promega), 200 nM primer (Bảng 1), 1 µl cDNA và thêm nước cất (Ultrapure™ DNase/RNase-Free Distilled Water (Invitrogen™)) để đạt thể tích phản ứng là 20 µl. Phản ứng PCR được thực hiện trên máy luân nhiệt Eppendorf Realplex 4. Gen *elongation factor-1α (ef-1α)* được lựa chọn là gen đối chứng của phản ứng qPCR. Thí nghiệm thiết kế lặp lại 3 phản ứng qPCR cho mỗi mẫu. Giá trị chu kỳ ngưỡng (Ct) của mỗi mẫu được tự động tính toán bằng phần mềm được tích hợp trên máy luân nhiệt Eppendorf Realplex 4. Từ giá trị chu trình ngưỡng, phân tích sự biểu hiện gen được thực hiện theo phương pháp $\Delta\Delta C_t$, từ đó so sánh mức độ biểu hiện tương đối của các gen được biểu diễn theo số lần thay đổi $\log_2(\text{fold change})$ ($\log_2 FC$). Kết quả qPCR được phân tích thống kê (kiểm định Student's t-test) và xác định sai khác có ý nghĩa thống kê dựa vào giá trị $p < 0,05$.

Bảng 1. Trình tự primer của các gen sử dụng cho phản ứng qPCR

STT	Tên gen	Trình tự primer (5'-3')
1	<i>Anti-lipopolysaccharide factor-like</i>	F: TTGACGTGTATCCGAAGTTGG R: GCTTTCATTGATGTGCTGAGG
2	<i>Carotenoid isomer oxygenase-like</i>	F: AAAGGCGATCAGGCATCCAG R: AAAGGCGATCAGGCATCCAG
3	<i>Cytochrome P450 2L1-like</i>	F: CGACACTTTGCCCTCATACC R: ATCCCGAACACCCGAACAC
4	<i>Crustacyanin-A2 subunit-like</i>	F: GTGACCACAGCAGGATTGAC R: GGCGAAAGCAGAAGGGAAAT
5	<i>Fatty acid binding protein 1-B.1-like</i>	F: ACGGTCTTGATGGTGACAAC R: ACTTACGAGTATGCTTCCCAAG
6	<i>NADH dehydrogenase subunit 1</i>	F: AGAGACTAATCGAACCCCTTTTG R: TCAGCTATAAAGATTAAGGCAAAACC
7	<i>Neuroparsin-A-like</i>	F: CTTCCACCAACAGCCTTCG R: AGACTTCGCAGCAGAGATGA
8	<i>Oplophorus-luciferin 2-monooxygenase non-catalytic subunit-like</i>	F: CTCACTGTCTTCCTTTCCCTTC R: GAGACGTTGCTGACTAGGTTG
9	<i>Sorbitol dehydrogenase-like</i>	F: TGGGTCTGCGGTTTCATAATG R: GATAGACAGGGAGACAAACGC
10	<i>Phospholipase ABHD3-like</i>	F: CCATCAGGAGCATGAGGAAG R: GCAGAGAGGCACAACAAAG
11	<i>Vitellogenin</i>	F: TCAGGAAACACTCTCGCAAC R: CAGATCAATAGAGCGAGGAAGG
12	<i>Elongation factor-1α</i>	F: TTCCGACTCCAAGAACGACC R: GAGCAGTGTGGCAATCAAGC

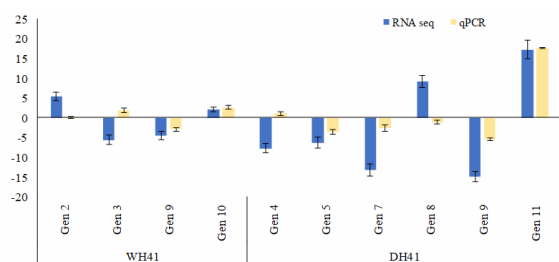
Ghi chú: Gen 1, 4, 5, 9, 11 đã kiểm chứng bằng kỹ thuật qPCR trong công bố của Nguyen và cs (2025) [8].

3. KẾT QUẢ NGHIÊN CỨU

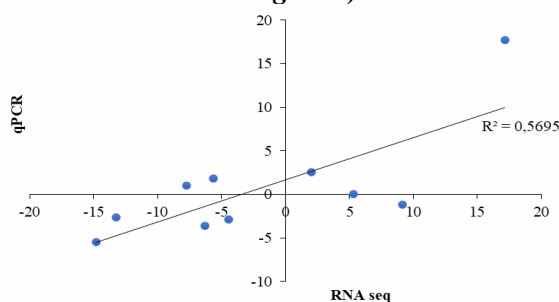
3.1. Sự biểu hiện gen của các gen tiềm năng liên quan đến tính trạng sinh sản ở mô gan tụy

Kết quả phân tích sự biểu hiện gen của các gen ở mô gan tụy được trình bày ở hình 1. Đối với nhóm tôm cái tự nhiên, gen *sorbitol dehydrogenase-like* (gen 9) và *phospholipase ABHD3-like* (gen 10) có biểu hiện gen tương đồng giữa phương pháp RNA-seq và qPCR. Trong đó, gen *sorbitol dehydrogenase-like* có biểu hiện giảm và gen *phospholipase ABHD3-like* có biểu hiện tăng khi so sánh mẫu ở giai đoạn thành thực sinh sản (giai đoạn 4) với mẫu ở giai đoạn tiền thành thực (giai đoạn 1). Tuy nhiên, gen *cytochrome P450 2L1-like* (gen 3) có biểu hiện giảm ($\log_2FC = -5.6$) khi phân tích bằng phương pháp RNA-seq và có biểu hiện tăng ($\log_2FC = 1,8$) khi phân tích bằng phương pháp qPCR.

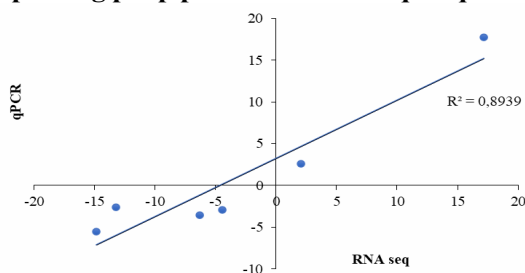
Ngoài ra, gen *carotenoid isomer oxygenase-like* (gen 2) có biểu hiện tương đồng giữa 2 phương pháp nhưng không có sai khác về ý nghĩa thống kê ở phương pháp qPCR ($p > 0,05$). Đối với nhóm tôm cái gia hóa, 4 gen (*fatty acid binding protein 1-B.1-like* (gen 5), *neuroparsin-A-like* (gen 7), *sorbitol dehydrogenase-like* (gen 9), *vitellogenin* (gen 11)) trong số 6 gen khảo sát có biểu hiện tương đồng khi so sánh giữa 2 phương pháp phân tích. Gen *crustacyanin-A2 subunit-like* (gen 4) và *oplophorus-luciferin 2-monooxygenase non-catalytic subunit-like* (gen 8) có biểu hiện nhưng không có ý nghĩa thống kê khi phân tích bằng qPCR ($p > 0,05$). Phân tích mối tương quan giữa kết quả của phương pháp RNA-seq và qPCR cho thấy, hệ số tương quan đạt giá trị trung bình ($r^2 = 0,5695$; hình 2) và tăng lên rõ rệt ($r^2 = 0,8939$; hình 3) khi loại trừ kết quả của các gen có biểu hiện ngược nhau hoặc biểu hiện không có ý nghĩa thống kê (4 gen).



Hình 1. So sánh sự biểu hiện gen ở mô gan tụy của tôm sú cái tự nhiên và gia hóa bằng phương pháp RNA-seq và qPCR. (Kết quả trình bày: Trung bình $\pm$ sai số chuẩn (trung bình của 3 mẫu lặp lại của giai đoạn 1 hoặc 4 của nhóm tôm tự nhiên hoặc gia hóa). WH41: Mẫu gan tụy của tôm tự nhiên so sánh giữa giai đoạn 4 và 1; DH41: Mẫu gan tụy của tôm gia hóa so sánh giữa giai đoạn 4 và 1. Gen 2: Carotenoid isomer oxygenase-like; gen 3: Cytochrome P450 2L1-like; gen 4: Crustacyanin-A2 subunit-like; gen 5: Fatty acid binding protein 1-B.1-like; gen 7: Neuroparsin-A-like; gen 8: Oplophorus-luciferin 2-monooxygenase non-catalytic subunit-like; gen 9: Sorbitol dehydrogenase-like; gen 11: Vitellogenin.)



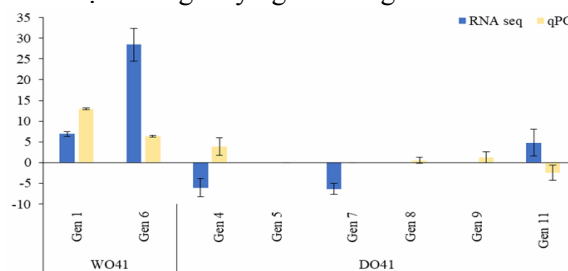
Hình 2. Mối tương quan của gen biểu hiện giữa phương pháp phân tích RNA-seq và qPCR



Hình 3. Mối tương quan của gen biểu hiện giữa phương pháp phân tích RNA-seq và qPCR (loại trừ gen biểu hiện không tương đồng giữa 2 phương pháp và gen biểu hiện không sai khác có ý nghĩa thống kê)

3.2. Sự biểu hiện gen của các gen tiềm năng liên quan đến tính trạng sinh sản ở mô buồng trứng

Kết quả phân sự biểu hiện gen của các gen ở mô buồng trứng được trình bày ở hình 4. Gen *anti-lipopolysaccharide factor-like* và gen *NADH dehydrogenase subunit 1* đều có biểu hiện tăng ở mẫu buồng trứng của tôm tự nhiên khi phân tích bằng cả 2 phương pháp. Đối với mẫu buồng trứng của tôm gia hóa, các gen *fatty acid binding protein 1-B.1-like* (gen 5), *oplophorus-luciferin 2-monooxygenase non-catalytic subunit-like* (gen 8) và *sorbitol dehydrogenase-like* (gen 9) không có sự biểu hiện khi phân tích bằng phương pháp RNA-seq ($\log_2FC = 0$). Các gen này cũng có kết quả tương tự khi phân tích bằng qPCR: Gen 5 không biểu hiện, gen 8 và 9 có biểu hiện rất thấp ($\log_2FC = 0,6$ và $1,3$) và không có ý nghĩa thống kê ($p > 0,05$). Gen *neuroparsin-A-like* (gen 7) có biểu hiện giảm khi phân tích bằng RNA-seq nhưng không có biểu hiện khi phân tích bằng qPCR. Gen *crustacyanin-A2 subunit-like* (gen 4) và gen *vitellogenin* (gen 11) có biểu hiện khi sử dụng 2 phương pháp phân tích khác nhau nhưng kết quả biểu hiện không có ý nghĩa thống kê ($p > 0,05$). Mối tương quan của sự biểu hiện gen giữa phương pháp RNA-seq và qPCR không được trình bày vì nhiều gen không có biểu hiện hoặc biểu hiện không có ý nghĩa thống kê.



Hình 4. So sánh sự biểu hiện gen ở mô buồng trứng của tôm sú cái tự nhiên và gia hóa bằng phương pháp RNA-seq và qPCR. (Kết quả trình bày: Trung bình $\pm$ sai số chuẩn (trung bình của 3 mẫu lặp lại của giai đoạn 1 hoặc 4 của nhóm tôm tự nhiên hoặc gia hóa). WO41: mẫu buồng trứng của tôm tự nhiên so sánh giữa giai đoạn 4 và 1; DO41: Mẫu buồng trứng của tôm gia hóa so sánh giữa giai đoạn 4 và 1. Gen 1: Anti-lipopolysaccharide factor-like; gen 6: NADH dehydrogenase subunit 1; gen 4: Crustacyanin-A2 subunit-like; gen 5: Fatty acid

binding protein 1-B.1-like; gen 7: Neuroparsin-A-like; gen 8: Oplophorus-luciferin 2-monooxygenase non-catalytic subunit-like; gen 9: sorbitol dehydrogenase-like; gen 11: Vitellogenin).

4. THẢO LUẬN

Sự biểu hiện gen được phân tích bằng kỹ thuật RNA-seq phụ thuộc rất lớn vào bước định vị các đoạn trình tự lên thư viện gen và kết quả định vị là nguồn dữ liệu cho định lượng mức độ biểu hiện gen. Nghiên cứu của Corchete và cs (2020) [12] khẳng định bước định lượng đoạn trình tự được định vị và chuẩn hóa kết quả ước tính là bước chính ảnh hưởng đáng kể đến kết quả phân tích sự biểu hiện gen. Nghiên cứu sử dụng nhiều công cụ định vị và thuật toán khác nhau để chuẩn hóa kết quả. Công cụ Cufflinks và RSEM có mức độ chính xác cao nhất trong bước định lượng và theo sau là HTSeq và Stringtie. Thuật toán chuẩn hóa kết quả ước tính thành chỉ số TMM (Trimmed Mean of M value), tiếp theo là các chỉ số RLE (Relative Log Expression), TPM (Transcripts Per Million) có độ chính xác cao trong định lượng gen biểu hiện. Nghiên cứu hiện tại sử dụng RSEM là công cụ định lượng dựa trên nền tảng kết quả định vị từ công cụ Bowtie 2. Trong quá trình định vị, 1 đoạn trình tự ngắn có thể được định vị ở nhiều vị trí khác nhau của hệ gen hoặc hệ phiên mã tham chiếu gọi là trình tự đa định vị. Một số công cụ mặc định loại bỏ các đoạn trình tự đa định vị dẫn đến ước tính thấp hơn số lượng trình tự biểu hiện. Trong khi đó, RSEM khắc phục được hạn chế này nên có thể nâng cao mức độ chính xác của bước ước tính định lượng. Công cụ RSEM sử dụng phương pháp thống kê không dựa vào trình tự của toàn bộ hệ gen tham chiếu mà chỉ dựa vào thư viện trình tự phiên mã. Điều này loại trừ được lỗi định lượng thấp hơn giá trị thực của gen biểu hiện khi các đoạn trình tự sai khác nhiều với hệ gen tham chiếu trong quá trình định vị [13]. Ngoài ra, kết quả định vị cũng phụ thuộc đáng kể vào chiều dài của các trình tự phiên mã (transcript). Theo Musich và cs (2021) [14], tùy thuộc vào công cụ định vị khác nhau sẽ cho kết quả định vị cao hơn khi chiều dài trình tự > 1.000 base. Nghiên cứu này phù hợp với kết quả nghiên cứu của Musich và cs (2021) [14] khi các gen biểu hiện có sự tương đồng giữa 2 kỹ thuật phân tích đều có

chiều dài > 1.000 base (1.024 - 7.997 base (kết quả trích xuất từ tệp RSEM)). Công cụ RSEM cũng có thể ước tính sự biểu hiện gen ở mức độ gen và mức độ biến thể gen (isoform) [10]. Các biến thể gen là các gen có nguồn gốc từ một gen và được tạo ra thông qua quá trình cắt nối các vùng mã hóa có lựa chọn (alternative splicing). Sự biểu hiện gen có thể xác định ở mức độ biến thể gen giúp hiểu được cơ chế điều hòa biểu hiện gen ở từng mô nhất định hoặc trong các điều kiện thí nghiệm khác nhau. Đây là một ưu thế khi nghiên cứu sự biểu hiện gen sử dụng kỹ thuật RNA-seq.

Tuy nhiên, nghiên cứu hiện tại có những gen biểu hiện không tương đồng khi so sánh giữa 2 kỹ thuật phân tích, cụ thể là gen *cytochrome P450 2L1-like* ở mẫu gan tụy của tôm tự nhiên và gen *neuroparsin-A-like* ở mẫu buồng trứng của tôm gia hóa. Nghiên cứu của Su và cs (2014) [15] thực hiện đánh giá so sánh toàn diện các kỹ thuật nghiên cứu sự biểu hiện gen bao gồm RNA-seq, kỹ thuật vi mảng và qPCR. Nghiên cứu khẳng định, không có kỹ thuật nào trong số 3 kỹ thuật được so sánh đạt độ tin cậy tuyệt đối trong định lượng, kể cả kỹ thuật qPCR vẫn thường được xem là “tiêu chuẩn vàng” trong nghiên cứu định lượng gen. Nghiên cứu của Aguiar và cs (2023) [13] cũng có kết quả tương tự khi so sánh kỹ thuật qPCR và RNA-seq trong nghiên cứu định lượng sự biểu hiện gen HLA. Nghiên cứu này giải thích một số nguyên nhân có thể làm cho kết quả khác nhau trong ước tính định lượng giữa 2 kỹ thuật. Đầu tiên, sự đa dạng di truyền là đặc tính hiển nhiên hiện diện ở sinh vật nên các mẫu lặp lại của tôm tự nhiên hoặc gia hóa không hoàn toàn giống nhau. Tùy theo kỹ thuật RNA-seq hoặc qPCR có khuynh hướng thiên về 1 biến thể gen nhất định hoặc alen của 1 gen dẫn đến kết quả khác nhau giữa 2 kỹ thuật. Nghiên cứu của Bustin (2010) [16] xác nhận kết quả định lượng gen biểu hiện bằng qPCR thường phụ thuộc vào đối tượng nghiên cứu, loại mô và thời gian thu mẫu. Do đó, bất kỳ kết luận nào phải được diễn giải trong điều kiện thí nghiệm cụ thể. Tiếp theo, bước chuẩn hóa số liệu định lượng cũng góp phần vào sai khác kết quả giữa các phương pháp. Đối với kỹ thuật RNA-seq, thuật toán chuẩn hóa tính toán chỉ số TPM sẽ phụ thuộc vào chiều dài của gen tham chiếu. Nếu chiều dài của gen tham chiếu dài hơn chiều dài thật của gen phiên mã, kết

quả ước tính định lượng sẽ thấp hơn sự biểu hiện thực tế. Đối với kỹ thuật qPCR, bước chuẩn hóa số liệu định lượng sẽ phụ thuộc vào gen đối chứng được lựa chọn. Do đó gen đối chứng phải được kiểm tra sự biểu hiện gen ổn định ở nhiều mẫu khác nhau trước khi lựa chọn là gen đối chứng. Ngoài ra, kết quả của chú giải gen và bước loại bỏ các trình tự phiên mã trùng lặp cũng có thể giải thích cho sự sai khác giữa các kỹ thuật phân tích định lượng RNA. Kỹ thuật qPCR có thể khuếch đại các trình tự phiên mã không được chú giải hoặc bị loại bỏ ở bước sàng lọc sự trùng lặp của các trình tự nên không được xác định tồn tại bằng kỹ thuật RNA-seq [2]. Nghiên cứu của Aguiar và cs (2023) [13] kết luận mức độ tương đồng của nghiên cứu sự biểu hiện gen bằng kỹ thuật RNA-seq và qPCR ở thời điểm hiện tại vẫn chưa được rõ ràng vì mỗi kỹ thuật phụ thuộc vào điều kiện thực hiện thí nghiệm, phương pháp phân tích và chuẩn hóa số liệu, các bước xử lý số liệu của công cụ tin sinh học và lỗi thiên lệch kết quả của kỹ thuật RNA-seq và qPCR. Một nguyên nhân khác là số lượng nhỏ mẫu lặp lại của một điều kiện nghiên cứu (thông thường 3 mẫu lặp lại) dẫn đến độ tin cậy của phân tích thống kê bị giảm nên độ chính xác không cao. Vì vậy, kết quả nghiên cứu này không có sự tương đồng khi so sánh giữa 2 kỹ thuật được sử dụng là tương tự với các công bố trước đây.

Mặc dù mỗi kỹ thuật phân tích định lượng RNA có những ưu điểm và nhược điểm riêng, nhưng kỹ thuật RNA-seq vẫn được lựa chọn phổ biến để nghiên cứu sự biểu hiện gen ở thời điểm hiện tại. Vì kỹ thuật RNA-seq có thể định lượng các gen có sự biểu hiện thấp và những biến thể gen mới chưa được công bố [16]. Nghiên cứu hiện tại còn một số hạn chế chưa được giải quyết. Do đó, cần khai thác nhiều công cụ tin sinh học khác nhau để xử lý và phân tích số liệu để tìm ra công cụ phù hợp có thể cho kết quả phân tích có độ chính xác cao. Số lượng mẫu lặp lại cũng phải đủ lớn cho phân tích thống kê có ý nghĩa và tin cậy trong các kết luận về biểu hiện gen nghiên cứu. Khi nghiên cứu so sánh sự biểu hiện gen giữa kỹ thuật RNA-seq và qPCR, số lượng gen nghiên cứu ở nhiều loại mô khác nhau cũng phải lớn hơn để cho thấy rõ ràng mối tương quan của kết quả đạt được từ 2 phương pháp. Nếu kết quả phân tích biểu hiện gen có sự sai khác giữa 2

phương pháp, nghiên cứu cần thiết kế kiểm chứng ở các loại mẫu có nguồn gốc khác nhau (tự nhiên và gia hóa) để tìm kiếm bản chất của sự sai khác là do kỹ thuật phân tích hay loại mẫu phân tích.

5. KẾT LUẬN VÀ ĐỀ NGHỊ

Nghiên cứu này xác định đa số các gen liên quan đến tính trạng sinh sản có sự biểu hiện gen tương đồng ở mô gan tụy và buồng trứng của nhóm tôm cái tự nhiên, hoặc gia hóa khi so sánh kết quả giữa kỹ thuật RNA-seq và qPCR. Tuy nhiên, số lượng nhỏ gen được phân tích biểu hiện nên nghiên cứu chưa khẳng định được phương pháp RNA-seq hoặc qPCR có độ chính xác cao hơn trong nghiên cứu sự biểu hiện gen. Nghiên cứu tiếp theo cần bổ trí nhiều hơn số lượng mẫu lặp lại, tăng số lượng gen biểu hiện và kiểm chứng lại các gen biểu hiện có sự sai khác giữa 2 phương pháp phân tích. Ngoài ra, cũng cần khai thác nhiều công cụ tin sinh học khác nhau trong phân tích sự biểu hiện gen.

TÀI LIỆU THAM KHẢO

- [1]. S. Liu *et al.* (2012). Functional genomics research in aquaculture: Principles and general approaches, in *Functional Genomics in Aquaculture*, M. Saroglia and Z. (John) Liu, Eds., Wiley-Blackwell, 41-77.
- [2]. C. Everaert *et al.* (2017). Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. *Sci Rep*, 7(1), DOI: 10.1038/s41598-017-01617-3.
- [3]. C. Trapnell, L. Pachter, and S. L. Salzberg (2009). TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*, 25(9), 1105-1111, DOI: 10.1093/bioinformatics/btp120.
- [4]. A. Dobin *et al.* (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15 - 21, doi: 10.1093/bioinformatics/bts635.
- [5]. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, and C. Kingsford (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*, 14(4), 417 - 419, DOI: 10.1038/nmeth.4197.
- [6]. N. L. Bray, H. Pimentel, P. Melsted, and L. Pachter (2016). Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*, 34(5), 525-527, DOI: 10.1038/nbt.3519.

- [7]. VASEP (2025). *Report on Vietnam seafood exports in 2025*.
- [8]. T. M. Nguyen *et al.* (2025). Transcriptomic comparison between wild-caught and domesticated black tiger shrimp (*Penaeus monodon*) in early and late-vitellogenic broodstock females. *Aquac Rep*, 41, DOI: 10.1016/j.aqrep.2025.102675.
- [9]. B. Langmead and S. L. Salzberg (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9(4), 357 - 359, DOI: 10.1038/nmeth.1923.
- [10]. B. Li and C. N. Dewey (2014). RSEM: Accurate transcript quantification from RNA-seq data with or without a reference genome. *Bioinformatics: The Impact of Accurate Quantification on Proteomic and Genetic Analysis and Research*, 41 - 74, DOI: 10.1201/b16589.
- [11]. M. I. Love, W. Huber, and S. Anders (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 15(12), DOI: 10.1186/s13059-014-0550-8.
- [12]. L. A. Corchete, E. A. Rojas, D. Alonso-López, J. De Las Rivas, N. C. Gutiérrez, and F. J. Burguillo (2020). Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis. *Sci Rep*, 10(1), DOI: 10.1038/s41598-020-76881-x.
- [13]. V. R. C. Aguiar *et al.* (2023). Comparison between qPCR and RNA-seq reveals challenges of quantifying HLA expression. *Immunogenetics*, 75(3), 249 - 262, DOI: 10.1007/s00251-023-01296-7.
- [14]. R. Musich, L. Cadle-Davidson, and M. V. Osier (2021). Comparison of Short-Read Sequence Aligners Indicates Strengths and Weaknesses for Biologists to Consider. *Front Plant Sci*, 12, DOI: 10.3389/fpls.2021.657240.
- [15]. Z. Su *et al.*, (2014). A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the sequencing quality control consortium. *Nat Biotechnol*, 32, (9), 903 - 914, DOI: 10.1038/nbt.2957.
- [16]. S. A. Bustin (2010). Why the need for qPCR publication guidelines?-The case for MIQE. *Methods*, 50(4), 217 -226, DOI: 10.1016/j.ymeth.2009.12.006.

EXPRESSION COMPARISON OF CANDIDATE GENES RELATED TO A REPRODUCTIVE TRAIT IN THE BLACK TIGER SHRIMP (*Penaeus monodon*) USING RNA-seq AND qPCR

Tran Thi Hai Yen¹, Nguyen Minh Thanh¹

¹International University, Vietnam National University of Ho Chi Minh city

Abstract

Expression analysis of candidate genes related to a reproductive trait was performed on various tissues of wild and domesticated female black tiger shrimps using RNA-seq and qPCR techniques. The aim of the current study was to compare the gene expression patterns between the different RNA-based techniques used. For hepatopancreas tissues, 6 genes out of 10 genes expressed similarly while gene *cytochrome P450 2L1-like* was down-regulated using the RNA-seq technique and up-regulated using qPCR. For ovarian tissues, genes *anti-lipopolysaccharide factor-like* and *NADH dehydrogenase subunit 1* had the same trend when two techniques were employed. The expressed genes in the hepatopancreas tissues were also studied in the ovaries. These genes were not expressed or expressed insignificantly using either the RNA-seq technique or qPCR. Furthermore, we discussed the influences of various bioinformatics tools on accuracy of quantification and the potential sources contributing to the expression differences between the RNA-seq and pPCR techniques.

Keywords: *Black tiger shrimp, Penaeus monodon, gene expression, RNA-seq, qPCR.*

Ngày nhận bài: 3/11/2025

Ngày chuyển phản biện: 4/12/2025

Ngày thông qua phản biện: 25/12/2025

Ngày duyệt đăng: 7/5/2026