

Phân tích gen biểu hiện của buồng trứng tôm sú (*Penaeus monodon*) liên quan đến tính trạng sinh sản dựa vào hệ gen tham chiếu

Lê Thị Hồng Thắm¹, Trần Thị Hải Yến¹, Võ Thị Minh Thu¹, Nguyễn Việt Tuấn², Nguyễn Minh Thành^{1*}

¹Khoa Công nghệ Sinh học, Trường Đại học Quốc tế, Đại học Quốc gia TP Hồ Chí Minh, khu phố 6, phường Linh Trung, TP Thủ Đức, TP Hồ Chí Minh, Việt Nam

²Bộ Nông nghiệp bang Victoria, Melbourne, 5 Ring Road, Bundoora Victoria, Australia

Ngày nhận bài 8/6/2023; ngày chuyển phản biện 10/6/2023; ngày nhận phản biện 21/6/2023; ngày chấp nhận đăng 25/6/2023

Tóm tắt:

Tôm sú (*Penaeus monodon*) là loài được nuôi phổ biến ở nước ta và trên thế giới. Hiện nay, nguồn tôm bố mẹ phục vụ cho sản xuất giống vẫn phụ thuộc vào tôm sú tự nhiên bởi vì chất lượng sinh sản của tôm tự nhiên cao hơn tôm gia hóa. Mục tiêu của nghiên cứu là phân tích sự biểu hiện gen của buồng trứng tôm sú tự nhiên ở giai đoạn chưa thành thực và thành thực sinh sản bằng các phương pháp phân tích dựa vào hệ gen tham chiếu, và so sánh kết quả từ các phương pháp phân tích này. cDNA của các mẫu buồng trứng được giải trình tự bằng thiết bị Illumina. Kết quả lắp ráp bằng phương pháp có hệ gen tham chiếu đạt 71.678 contig, trong đó độ dài N50 của contig là 1.587 bp. Kết quả phân tích sự biểu hiện gen dựa vào hệ phiên mã vừa xây dựng xác định được 288 gen có biểu hiện khác biệt khi so sánh mẫu buồng trứng của giai đoạn thành thực và chưa thành thực. Trường hợp phân tích gen biểu hiện bằng cách căn chỉnh trực tiếp lên hệ gen tham chiếu chỉ xác định được 49 gen có biểu hiện khác biệt. Kết quả phân tích sự biểu hiện gen cũng cho thấy những gen liên quan đến tính trạng sinh sản có biểu hiện tăng ở giai đoạn thành thực, bao gồm vitellogenin, NADH dehydrogenase subunit 2, anti-lipopolysaccharide factor-like. Kết quả biểu hiện gen không đồng nhất giữa 2 phương pháp phân tích được thảo luận.

Từ khóa: hệ gen tham chiếu, *Penaeus monodon*, sự biểu hiện gen, tính trạng sinh sản, tôm sú.

Chỉ số phân loại: 4.5, 4.6

Gene expression analyses of the ovaries in black tiger shrimp (*Penaeus monodon*) for reproductive trait using reference genome

Thi Hong Tham Le¹, Thi Hai Yen Tran¹, Thi Minh Thu Vo¹, Viet Tuan Nguyen², Minh Thanh Nguyen^{1*}

¹School of Biotechnology, International University, Vietnam National University - Ho Chi Minh City, Quarter 6, Linh Trung Ward, Thu Duc City, Ho Chi Minh City, Vietnam

²Agriculture Victoria, Melbourne, 5 Ring Road, Bundoora Victoria, Australia

Received 8 June 2023; revised 21 June 2023; accepted 25 June 2023

Abstract:

The black tiger shrimp (*Penaeus monodon*) is widely cultured in Vietnam and worldwide. Broodstock source for post-larval production still relies on the wild shrimps due to their higher reproductive quality in comparison with the domesticated broodstock. The aims of the current study were to analyse differential gene expression in the ovaries of black tiger shrimp between the stages of reproductive maturation using the reference genome-based approaches and to compare the results generated by these approaches. cDNAs of the ovarian samples were sequenced using the Illumina platform. Genome-guided assembly (GGA) was employed to generate 71,678 contigs with an N50 length of 1,587 bp. For gene expression analysis, the newly built-transcriptome-based approach identified a total number of 288 differentially expressed genes (DEGs) when the samples between the non-mature and mature stages were compared. The gene expression analysis using the reference mapping approach resulted in only 49 DEGs. The analyses indicated that many genes involved in the reproductive maturation were up-regulated, including vitellogenin, NADH dehydrogenase subunit 2, and anti-lipopolysaccharide factor-like. Different results of gene expression from both approaches are discussed.

Keywords: black tiger shrimp, gene expression, *Penaeus monodon*, reference genome, reproductive trait.

Classification numbers: 4.5, 4.6

*Tác giả liên hệ: Email: nmthanh@hcmiu.edu.vn

1. Đặt vấn đề

Hơn một thập kỷ qua, sự phát triển của kỹ thuật giải trình tự gen thế hệ mới (NGS) đã tạo điều kiện thuận lợi giải trình tự toàn bộ hệ gen hoặc hệ gen phiên mã cho nhiều loài không phải là loài nghiên cứu mẫu, trong đó có các loài thủy sản. NGS cũng tạo ra cuộc cách mạng trong lĩnh vực nghiên cứu hệ gen chức năng và sự biểu hiện của các gen. Tuy nhiên, hạn chế của đa số thiết bị NGS hiện nay là chỉ giải trình tự các đoạn ngắn nên gây khó khăn cho bước lắp ráp tiếp theo, đặc biệt là các hệ gen có nhiều vùng lặp lại như tôm sú [1]. Phần lớn các nghiên cứu sử dụng phương pháp lắp ráp không có hệ gen tham chiếu vì nhiều loài chưa có hệ gen hoàn chỉnh được công bố. Đối với những loài có hệ gen tham chiếu, phương pháp lắp ráp dựa vào hệ gen tham chiếu (GGA) để hướng dẫn quá trình định vị chính xác hơn và cho kết quả lắp ráp tốt hơn. Tất cả các bước chú giải và phân tích gen biểu hiện phụ thuộc hoàn toàn vào kết quả lắp ráp. Vì vậy chất lượng của kết quả lắp ráp tác động trực tiếp đến kết quả của các phân tích về sau. Phương pháp lắp ráp dựa vào hệ gen tham chiếu có thể là một lựa chọn để đạt chất lượng lắp ráp tốt hơn.

Hiện tại, tôm sú (*Penaeus monodon*) vẫn là đối tượng nuôi chủ lực ở nước ta và đóng góp đáng kể vào giá trị xuất khẩu của ngành thủy sản. Năm 2022, sản lượng tôm sú đạt 271.000 tấn và xuất khẩu đạt 574 triệu USD, chiếm 13,3% cơ cấu các sản phẩm tôm xuất khẩu của Việt Nam [2]. Tôm sú bố mẹ chủ yếu đánh bắt từ tự nhiên vì sức sinh sản của tôm mẹ và chất lượng tôm giống vượt trội so với tôm bố mẹ gia hóa. Nhiều nghiên cứu so sánh hệ phiên mã của tôm sú tự nhiên và gia hóa để tìm kiếm các gen tiềm năng liên quan đến tính trạng sinh sản [3, 4]. Nghiên cứu hiện tại phân tích sự biểu hiện gen khác biệt ở buồng trứng giai đoạn tiền thành thực và thành thực của tôm mẹ tự nhiên. Đối với tôm sú, hệ gen hoàn chỉnh đã được công bố bởi T. Uengwetwanit và cs (2021) [5]. Vì vậy, phương pháp lắp ráp hệ gen phiên mã và phân tích sự biểu hiện gen của mô buồng trứng sử dụng phương pháp có hệ gen tham chiếu. Mục tiêu của nghiên cứu là xác định các gen tiềm năng đóng góp vào sự thành thực sinh sản của tôm sú tự nhiên và so sánh kết quả của các phương pháp phân tích sự biểu hiện gen.

2. Đối tượng và phương pháp nghiên cứu

2.1. Mẫu tôm thí nghiệm

Mẫu tôm sú cái tự nhiên ở giai đoạn tiền thành thực (giai đoạn 1, n=3) và giai đoạn thành thực sinh sản (giai đoạn 4, n=3) được thu thập tại vùng biển tỉnh Bà Rịa - Vũng Tàu

do ngư dân địa phương đánh bắt. Các giai đoạn phát triển buồng trứng của tôm cái được xác định bằng phương pháp soi đèn dưới bụng để kiểm tra kích thước của buồng trứng [6]. Buồng trứng là cơ quan đích của nghiên cứu hệ gen phiên mã được lưu trữ trong RNAlater (Invitrogen, Hoa Kỳ) và bảo quản ở -80°C cho đến khi thực hiện ly trích RNA.

2.2. Ly trích RNA tổng số, tổng hợp thư viện cDNA và giải trình tự bằng Illumina

Mẫu buồng trứng được ly trích RNA tổng số bằng phương pháp TRIzol/Chloroform (Invitrogen) [7]. RNA tổng số được định tính và định lượng bằng máy quang phổ Nanodrop™ (ThermoFisher) và Bioanalyser 2100 (Agilent). mRNA được ly tách khỏi hỗn hợp RNA tổng số bằng Dynabeads mRNA purification kit (Invitrogen) theo hướng dẫn của nhà sản xuất. Sau đó mRNA được phân cắt đoạn ngẫu nhiên và được sử dụng làm khuôn mẫu để tổng hợp cDNA bằng TruSeq® Stranded mRNA Library Prep kit của Illumina theo hướng dẫn của nhà sản xuất. Thư viện cDNA được giải trình tự bằng máy Illumina NovaSeq 6000 tại Novogene (25 Pandan Crescent #05-15 TIC Centre, Singapore 128477) vào năm 2022.

2.3. Lắp ráp các đoạn trình tự có hệ gen tham chiếu và chú giải các đoạn mRNA

Chất lượng các đoạn trình tự thô được đánh giá bằng phần mềm FastQC phiên bản 0.11.6 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Sau đó, các đoạn trình tự thô kém chất lượng và các đoạn adapter được loại bỏ bằng phần mềm Trimmomatic [8] với các tham số được đặt ra như sau: ILLUMINACLIP: TruSeq3-PE-2.fa: 2: 30: 10 HEADCROP:13 LEADING:5 TRAILING:5 SLIDINGWINDOW:5:15 MINLEN:50. Tất cả các đoạn trình tự ngắn đạt chất lượng sau đánh giá được tập hợp và lắp ráp thành bộ phiên mã theo trình tự của hệ gen tham chiếu. Đầu tiên, các tập tin danh mục gen được xây dựng từ hệ gen tôm sú *P. monodon* lưu trữ trên GenBank với mã số PRJNA679074 bằng công cụ HISAT2 phiên bản 2.1.0 [9]. Tiếp theo, các đoạn trình tự ngắn tinh sạch được sắp xếp và định vị vào các tập tin danh mục gen vừa tạo ra bằng công cụ HISAT2 phiên bản 2.1.0 với những thông số mặc định của công cụ. Sau đó, các đoạn trình tự ngắn đã được sắp xếp theo trình tự trên hệ gen được lắp ráp để xây dựng hệ gen phiên mã tôm sú cho nghiên cứu của chúng tôi bằng Trinity [10] dựa trên hệ gen tham chiếu. Cuối cùng, các contig trùng lặp được nhóm lại bằng CD-HIT phiên bản 4.8.1 [11]. Chất lượng của hệ gen phiên mã vừa lắp ráp được

đánh giá căn cứ vào cơ sở dữ liệu bộ gen của động vật chân khớp (arthropoda_odb10 lineage) sử dụng BUSCO [12]. Các contig được chú giải bằng BLASTx của NCBI [13] dựa vào cơ sở dữ liệu NR (Non-redundant) với giá trị $E < 1e^{-6}$.

2.4. Phân tích sự biểu hiện gen của mẫu buồng trứng ở giai đoạn tiền thành thực và thành thực

Trong nghiên cứu này, chúng tôi phân tích so sánh sự biểu hiện gen của buồng trứng giữa giai đoạn tiền thành thực và thành thực bằng 2 phương pháp: (i) Sắp xếp và định vị trực tiếp vào hệ gen tham chiếu (mã số PRJNA679074) (reference mapping - phương pháp RM); (ii) Sắp xếp và định vị dựa vào hệ gen phiên mã vừa được xây dựng từ phương pháp lắp ráp có hệ gen tham chiếu ở bên trên (phương pháp GGA).

Đối với phương pháp RM, các đoạn trình tự ngắn của từng mẫu buồng trứng được sắp xếp và định vị vào các tập tin danh mục gen bằng công cụ HISAT2 phiên bản 2.1.0. Tiếp đó, các đoạn trình tự đã được định vị được sắp xếp thứ tự bằng công cụ Samtools phiên bản 1.15 [14]. Công cụ HTSeq phiên bản 0.9.1 [15] được sử dụng để xác định số lượng đoạn trình tự được sắp xếp lên mỗi gen ở các mẫu buồng trứng có giai đoạn thành thực khác nhau. Kết thúc bước này tạo ra các tập số liệu định lượng các đoạn trình tự ngắn cho mỗi gen biểu hiện (transcript).

Đối với phương pháp GGA, các đoạn trình tự ngắn của từng mẫu buồng trứng được sắp xếp trình tự dựa vào thư viện contig thay cho các tập danh mục gen bằng công cụ Bowtie 2 [16]. Các đoạn trình tự ngắn đã được định vị vào thư viện contig cũng tiếp tục được sắp xếp theo thứ tự bằng Samtools. Công cụ RSEM [17] được sử dụng để xác định số lượng đoạn trình tự ngắn của các mẫu buồng trứng được sắp xếp lên mỗi contig. Tương tự bên trên, kết thúc bước này tạo ra các tập số liệu định lượng các đoạn trình tự ngắn cho mỗi contig.

Các tập số liệu định lượng các đoạn trình tự ngắn từ mỗi phương pháp phân tích được sử dụng để phân tích sự biểu hiện gen khác biệt giữa mẫu buồng trứng ở giai đoạn 1 và giai đoạn 4 bằng công cụ DESeq2 phiên bản 3.5.3 [18]

trên nền tảng R. Những transcript hay contig có tổng số các đoạn trình tự ngắn từ tất cả mẫu nhỏ hơn 10 bị loại bỏ. Một transcript hay contig được xác định là có biểu hiện gen khác biệt giữa giai đoạn 1 và 4 dựa vào các giá trị: $|\log_2(\text{fold change})| \geq 2$, giá trị p điều chỉnh (p-adjusted value) $< 0,05$.

3. Kết quả và bàn luận

3.1. Kết quả lắp ráp bằng phương pháp có hệ gen tham chiếu

Kết quả giải trình tự thư viện cDNA của các mẫu mô buồng trứng và chất lượng của các đoạn trình tự được trình bày ở bảng 1. Số lượng đoạn trình tự thô lần lượt là 43.284.300-45.841.714 cho mẫu buồng trứng ở giai đoạn 1 và 44.647.570-55.300.914 cho mẫu buồng trứng ở giai đoạn 4. Bước sàng lọc bằng công cụ Trimmomatic đã loại bỏ đoạn adapter và các đoạn trình tự chất lượng kém. Số lượng đoạn trình tự sau đánh giá đạt 41.194.378-44.154.658 cho mẫu buồng trứng ở giai đoạn 1 và 42.517.950-52.755.746 cho mẫu buồng trứng ở giai đoạn 4. Kích thước của các đoạn trình tự sau đánh giá là 50-137 bp.

Nghiên cứu của chúng tôi sử dụng phương pháp lắp ráp dựa vào hệ gen tham chiếu. Vì vậy chỉ các đoạn trình tự được sắp xếp và định vị vào hệ gen tham chiếu mới được giữ lại cho bước lắp ráp. Các chỉ số chính của hệ gen phiên mã được trình bày ở bảng 2. Kết quả lắp ráp đạt 71.678 contig (đã loại bỏ các contig lặp lại) từ các đoạn trình tự tinh sạch cộng gộp của tất cả mẫu buồng trứng ở 2 giai đoạn. Chiều dài trung bình của các contig là 825 bp (dao động từ 191 đến 25.504 bp). Chiều dài contig N50 là chỉ tiêu để đánh giá chất lượng của phần mềm lắp ráp đạt 1.587 bp. Nghiên cứu hiện tại có chiều dài N50 lớn hơn các nghiên cứu trước đây công bố cho chỉ tiêu này trên cùng đối tượng nghiên cứu [19, 20]. Đối với các nghiên cứu sử dụng kỹ thuật giải trình tự đoạn dài bằng thiết bị PacBio hoặc Nanopore, chiều dài N50 lần lượt là 1.919 bp [21] và 1.728 bp [22], không lớn hơn nhiều so với N50 của nghiên cứu hiện tại. Đối với hệ gen có nhiều vùng lặp lại, phần mềm thường bị lỗi trong xác định vị trí khi sắp xếp các đoạn trình tự ngắn và dẫn đến chất lượng hệ gen lắp ráp thấp [23]. Vì vậy, nghiên cứu hiện

Bảng 1. Tóm tắt kết quả giải trình tự và đánh giá chất lượng các đoạn trình tự ngắn.

Mẫu	Buồng trứng giai đoạn 1			Buồng trứng giai đoạn 4		
	1	2	3	1	2	3
Số lượng trình tự thô	45.841.714	43.284.300	44.784.006	44.647.570	47.042.688	55.300.914
Số lượng trình tự sau đánh giá	44.154.658	41.194.378	43.299.180	42.517.950	44.134.854	52.755.746
Kích thước (bp)	50-137	50-137	50-137	50-137	50-137	50-137
Q30 (%)	91,09	92,50	91,88	91,84	91,54	91,50
Tỷ lệ sắp xếp và định vị vào hệ gen tham chiếu (%)	84,36	87,51	83,50	87,77	87,17	86,69

Q30: chỉ số Phred đánh giá chất lượng các đoạn trình tự.

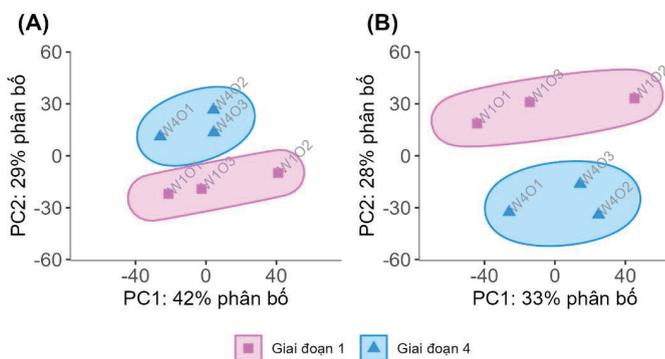
tại lựa chọn phương pháp lắp ráp có hệ gen tham chiếu để cải thiện chất lượng hệ gen phiên mã được lắp ráp. Kết quả phân tích BUSCO cho thấy mức độ hoàn chỉnh của hệ gen phiên mã mới lắp ráp đạt 79,3%.

Bảng 2. Kết quả lắp ráp và sàng lọc các contig.

Chỉ số phân tích	Giá trị
Tổng số contig	71.678
Tổng số base (bp)	59.141.777
Chiều dài contig trung bình (bp)	825
Chiều dài contig ngắn nhất (bp)	191
Chiều dài contig lớn nhất (bp)	22.504
Chiều dài contig N50 (bp)	1.587
Số lượng contig ≥ 1000 bp	16.491

3.2. Phân tích sự biểu hiện gen của mẫu buồng trứng ở giai đoạn tiền thành thực (giai đoạn 1) và giai đoạn thành thực (giai đoạn 4)

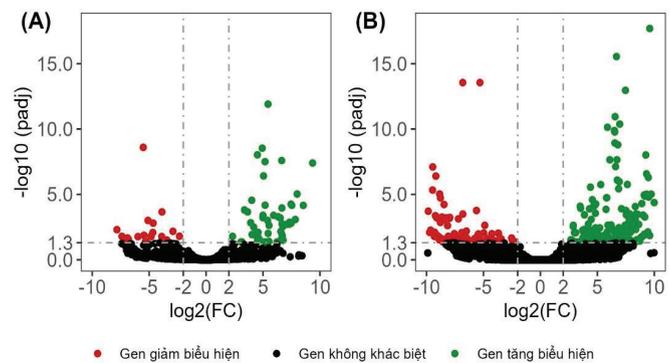
Sự phân bố số liệu của thư viện hệ gen phiên mã của các mẫu buồng trứng ở giai đoạn 1 và 4 được trình bày bằng đồ thị PCA (hình 1). Đồ thị PCA minh họa sự phân bố số liệu của các mẫu dựa vào mức độ tương đồng số liệu giữa các mẫu phân tích. Hình 1 cho thấy sự biểu hiện gen của hệ phiên mã có thể phân thành nhóm riêng biệt giữa giai đoạn 1 và 4. Trục PC1 của hình 1A giải thích 42% sai khác giữa các nhóm mẫu khi phân tích bằng phương pháp RM. Trong khi đó, trục này của hình 1B giải thích 33% sai khác giữa các nhóm mẫu khi phân tích bằng phương pháp GGA. Tuy nhiên, phương pháp GGA có phân bố số liệu khác biệt rõ rệt giữa giai đoạn 1 và 4. Đồ thị PCA cũng thể hiện số liệu của 3 mẫu lặp trong cùng 1 giai đoạn có thể gom thành nhóm tách biệt. Điều này cho thấy sự tương đồng của biểu hiện gen giữa các mẫu lặp sinh học ở cùng 1 giai đoạn nên đủ tin cậy cho các phân tích tiếp theo.



Hình 1. Đồ thị PCA của thư viện hệ gen phiên mã của các mẫu buồng trứng. (A) Phương pháp RM; (B) Phương pháp GGA.

Số lượng gen biểu hiện khác biệt khi so sánh mẫu buồng trứng ở giai đoạn 4 với 1 được trình bày bằng đồ thị volcano. Đối với phương pháp RM, tổng cộng 49 gen có biểu hiện

khác biệt được xác định, bao gồm 37 gen tăng cường biểu hiện và 12 gen giảm biểu hiện (hình 2A). Trường hợp phân tích theo phương pháp GGA, tổng cộng 288 gen có biểu hiện khác biệt khi so sánh mẫu của giai đoạn 4 với 1. Trong đó, 182 gen có biểu hiện tăng cường và 106 gen có biểu hiện giảm (hình 2B). Số lượng gen biểu hiện khác biệt giữa 2 giai đoạn thành thực phân tích bằng phương pháp GGA cao gấp 6 lần so với phương pháp RM. Phương pháp GGA cũng xác định được 8 gen có biểu hiện tăng gấp 10 lần, trong đó có gen NADH dehydrogenase subunit 2 (mitochondrion). A. Conesa và cs (2016) [24] cho biết phương pháp định lượng bằng cách sắp xếp và định vị các đoạn trình tự trực tiếp lên hệ gen tham chiếu có thể sử dụng nếu tỷ lệ định vị vào hệ gen tham chiếu đạt 70-90%. Tỷ lệ định vị của bộ dữ liệu trong nghiên cứu hiện tại đạt 83,50-87,77% (bảng 1). Tỷ lệ này cho thấy phương pháp RM có thể áp dụng để phân tích sự biểu hiện gen. Tuy nhiên các mẫu buồng trứng nghiên cứu có thể có khác biệt về mặt di truyền so với mẫu tôm của hệ gen tham chiếu (mẫu tôm Thái Lan) được sử dụng để sắp xếp và định vị các đoạn trình tự. Các biến đổi hay đột biến về mặt di truyền có thể là sai khác nucleotide đơn hoặc đa nucleotide và sai khác các cấu trúc gen lớn đều ảnh hưởng đến kết quả sắp xếp và định vị trình tự. Ngoài ra, các đoạn trình tự của bộ dữ liệu hệ gen phiên mã đang nghiên cứu có thể không trùng khớp với các đoạn chú giải của hệ gen tham chiếu cũng bị loại trong quá trình định vị và dẫn tới kết quả định lượng các gen bị giảm [25]. Số lượng đoạn trình tự bị loại bỏ trong quá trình định vị đã làm giảm mức độ chính xác của phân tích sự biểu hiện gen ở mẫu buồng trứng khi so sánh các mẫu có giai đoạn phát triển khác nhau. Nghiên cứu hiện tại cho thấy phương pháp RM xác định được số lượng gen có biểu hiện khác biệt thấp hơn nhiều lần so với phương pháp GGA. Kết quả này phù hợp với nhận định của A. Srivastava và cs (2020) [25].

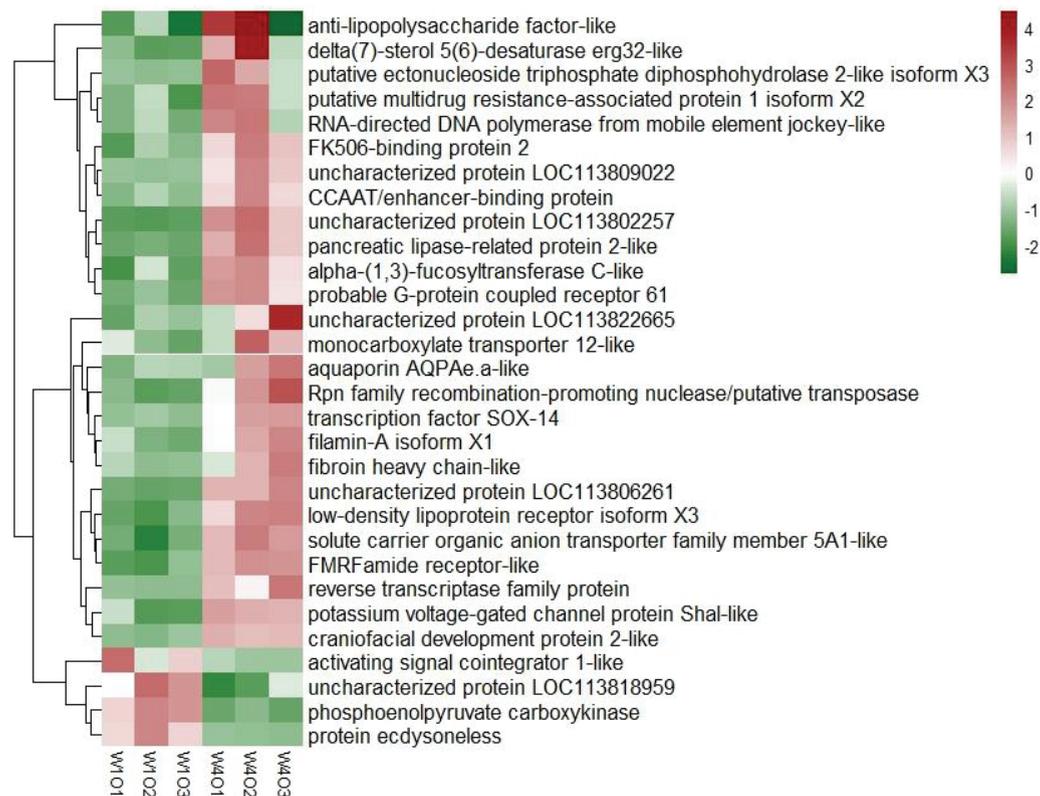


Hình 2. Đồ thị volcano. (A) Phương pháp RM; (B) Phương pháp GGA. Chấm xanh (bên phải đồ thị) cho thấy gen tăng cường biểu hiện, chấm đỏ (bên trái đồ thị) cho thấy gen suy giảm biểu hiện, chấm đen (sát trục hoành) cho thấy các gen biểu hiện không khác biệt giữa 2 giai đoạn.

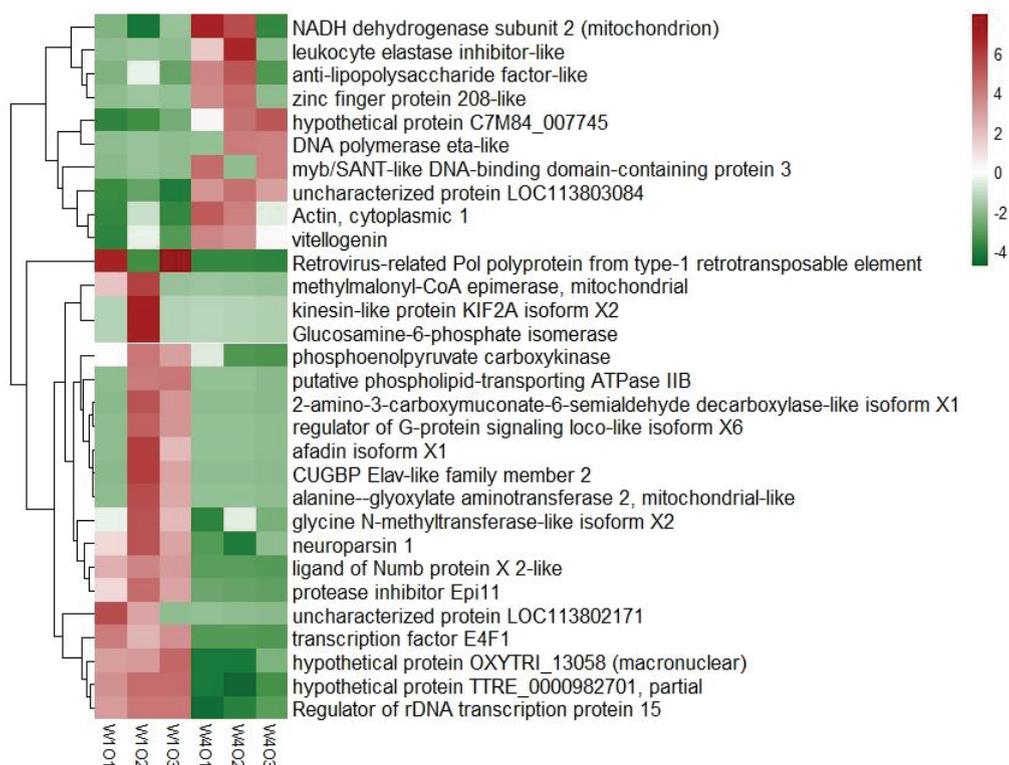
Kết quả top 30 của các gen có biểu hiện tăng hoặc giảm nhiều nhất ở giai đoạn 4 khi so sánh với giai đoạn 1 bằng phương pháp phân tích RM được trình bày ở đồ thị heatmap (hình 3). Các gen có biểu hiện tăng nhiều nhất bao gồm anti-lipoplysaccharide factor-like, low-density lipoprotein receptor isoform X3, pancreatic lipase-related protein 2-like, probable G-protein coupled receptor 61. Các gen có biểu hiện giảm ở giai đoạn 4 bao gồm phosphoenolpyruvate carboxykinase, activating signal cointegrator 1-like, protein ecdysoneless. Tương tự, đồ thị heatmap (hình 4) của phương pháp phân tích GGA cho thấy gen có biểu hiện tăng ở giai đoạn 4 là anti-lipoplysaccharide factor-like, vitellogenin, NADH dehydrogenase subunit 2 (mitochondrion), actin, cytoplasmic 1, leukocyte elastase inhibitor-like, myb/SANT-like DNA-binding domain-containing protein 3, zinc finger protein 208-like và DNA polymerase eta-like. Các gen có biểu hiện giảm ở giai đoạn 4 là neuroparsin 1, regulator of G-protein signaling loco-like isoform X6, phosphoenolpyruvate carboxykinase.

Phương pháp RM và GGA đều xác định gen anti-lipoplysaccharide factor-like có biểu hiện tăng ở giai đoạn buồng trứng thành thực. S. Li và cs (2015) [26] báo cáo gen

anti-lipoplysaccharide factor tổng hợp 1 protein có kích thước nhỏ có hoạt tính kháng khuẩn và được xác định tồn tại ở nhiều loại giáp xác, kể cả tôm sú. H. Ye và cs (2016) [27] công bố rằng, nhiều gen tham gia vào quá trình miễn dịch cũng đồng thời tham gia vào quá trình thành thực sinh sản của buồng trứng. Tuy nhiên, cho tới thời điểm hiện tại chưa có nghiên cứu nào khẳng định gen anti-lipoplysaccharide factor-like có ảnh hưởng tới quá trình thành thực sinh sản ở giáp xác. Do đó kết quả của chúng tôi có thể bổ sung thêm chức năng của gen này liên quan đến sự thành thực sinh sản ở tôm sú. Đặc biệt phương pháp GGA xác định được gen vitellogenin có biểu hiện tăng khác biệt ở giai đoạn 4. Theo S.J. Gutierrez và cs (2019) [28] buồng trứng thành thực bao gồm quá trình tổng hợp protein noãn hoàng vitellin là nguồn năng lượng và dinh dưỡng quan trọng nhất cho quá trình phát triển phôi. Vitellin được sản xuất từ sự phân giải tiền chất vitellogenin. Quá trình tổng hợp và tích lũy vitellogenin ở buồng trứng quyết định sự thành thực của buồng trứng. Quá trình tổng hợp vitellogenin xảy ra ở buồng trứng và/hoặc cơ quan gan tùy thuộc vào loài giáp xác và môi trường sống của nó. Gen vitellogenin biểu hiện tăng khác biệt ở buồng trứng giai đoạn 4 của nghiên cứu hiện tại là tương đồng với công bố trước đây [29]. Phương pháp GGA cũng phát hiện



Hình 3. Đồ thị heatmap của phương pháp RM. Màu xanh cho thấy sự biểu hiện giảm, màu đỏ cho thấy sự biểu hiện tăng; W1O1-W1O3: mẫu buồng trứng ở giai đoạn 1; W4O1-W4O3: mẫu buồng trứng ở giai đoạn 4.



Hình 4. Đồ thị heatmap của phương pháp GGA. Màu xanh cho thấy sự biểu hiện giảm, màu đỏ cho thấy sự biểu hiện tăng; W1O1-W1O3: mẫu buồng trứng ở giai đoạn 1; W4O1-W4O3: mẫu buồng trứng ở giai đoạn 4.

gen NADH dehydrogenase subunit 2 của ty thể có biểu hiện tăng khác biệt ở giai đoạn 4. Gen NADH dehydrogenase subunit 1-4 là gen liên quan đến sản sinh năng lượng và được xác nhận có biểu hiện tăng ở tôm sú bị cắt mắt để kích thích thành thực sinh sản [22]. Hai phương pháp phân tích đều xác định gen phosphoenolpyruvate carboxykinase có biểu hiện giảm ở giai đoạn 4. Đây là gen tổng hợp enzyme tham gia vào quá trình tạo glucose và được báo cáo tăng biểu hiện ở nghiên cứu của K. Sittikankaew và cs (2020) [22]. Phương pháp GGA cũng xác định gen neuroparsin 1 có biểu hiện khác biệt khi so sánh mẫu buồng trứng của giai đoạn 4 và giai đoạn 1. Gen neuroparsin tổng hợp các peptide thần kinh tham gia vào quá trình lột xác, sinh sản và trao đổi chất ở các loài giáp xác. Nghiên cứu hiện tại cho thấy gen neuroparsin 1 có biểu hiện giảm ở giai đoạn 4, trong khi gen này được báo cáo không có sự biểu hiện khác biệt có ý nghĩa giữa giai đoạn buồng trứng non và giai đoạn tích lũy noãn hoàng [29]. Vì vậy kết quả phân tích gen biểu hiện khác biệt phụ thuộc hoàn toàn vào phương pháp phân tích, giai đoạn thành thực của buồng trứng, nguồn gốc của mẫu phân tích (tôm tự nhiên hay tôm gia hóa) và mức độ đa dạng di truyền của mẫu phân tích. Tuy nhiên, các gen biểu hiện khác biệt phải được kiểm định bằng phương pháp PCR định lượng (qRT-PCR) để khẳng định mức độ chính xác của phương pháp phân tích bằng kỹ thuật giải trình tự RNA.

4. Kết luận

Nghiên cứu hiện tại giải trình tự hệ gen phiên mã đạt 41.194.378-52.755.746 đoạn trình tự tinh sạch tùy thuộc vào mẫu buồng trứng ở các giai đoạn khác nhau. Kết quả lắp ráp dựa vào hệ gen tham chiếu đạt 71.678 contig với chiều dài N50 là 1.587 bp. Phân tích sự biểu hiện gen khác biệt giữa giai đoạn buồng trứng tiền thành thực và thành thực của tôm sú tự nhiên được thực hiện bằng 2 phương pháp. Phương pháp phân tích dựa vào hệ phiên mã mới xây dựng xác định được số lượng gen biểu hiện khác biệt cao gấp nhiều lần so với phương pháp căn chỉnh các đoạn trình tự trực tiếp lên hệ gen tham chiếu. Đồng thời, nghiên cứu cũng xác định được một số gen liên quan đến quá trình phát triển và thành thực của buồng trứng có biểu hiện tăng ở giai đoạn thành thực sinh sản. Nghiên cứu hiện tại đã đề xuất các phương pháp khả thi trong phân tích sự biểu hiện gen cho các nghiên cứu sắp tới.

LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Quỹ Phát triển Khoa học và Công nghệ Quốc gia (NAFOSTED) thông qua đề tài mã số 106.05-2019.36. Nhóm tác giả xin trân trọng cảm ơn.

TÀI LIỆU THAM KHẢO

- [1] R. Huerlimann, N.M. Wade, L. Gordon, et al. (2018), “De novo assembly, characterization, functional annotation and expression patterns of the black tiger shrimp (*Penaeus monodon*) transcriptome”, *Sci. Rep.*, **8(1)**, DOI: 10.1038/s41598-018-31148-4.
- [2] Vietnam Association of Seafood Exporters and Producers (2023), *Report on Vietnam Seafood Exports in 2022*, 102pp (in Vietnamese).
- [3] T. Uengwetwanit, P. Ponza, D. Sangsrakru, et al. (2018), “Transcriptome-based discovery of pathways and genes related to reproduction of the black tiger shrimp (*Penaeus monodon*)”, *Mar. Genomics*, **37**, pp.69-73, DOI: 10.1016/j.margen.2017.08.007.
- [4] G. Rotllant, N.M. Wade, S.J. Arnold, et al. (2015), “Identification of genes involved in reproduction and lipid pathway metabolism in wild and domesticated shrimps”, *Mar. Genomics*, **22**, pp.55-61, DOI: 10.1016/j.margen.2015.04.001.
- [5] T. Uengwetwanit, W. Pootakham, I. Nookaew, et al. (2021), “A chromosome-level assembly of the black tiger shrimp (*Penaeus monodon*) genome facilitates the identification of growth-associated genes”, *Mol. Ecol. Resour.*, **21(5)**, pp.1620-1640, DOI: 10.1111/1755-0998.13357.
- [6] S.A. Uddin, N.G. Das, A.H.M. Kamal (2005), “Maturation and spawning of penaeid shrimp *Penaeus monodon* fabricius collected from off shore water of the Bay of Bengal”, *Int. J. Agri. Biol.*, **7(6)**, pp.925-927.
- [7] P. Chomczynski, K. Mackey (1995), “Short technical reports: Modification of the TRI reagent procedure for isolation of RNA from polysaccharide-and proteoglycan-rich sources”, *Biotechniques*, **19(6)**, pp.942-945.
- [8] A.M. Bolger, M. Lohse, B. Usadel (2014), “Trimmomatic: A flexible trimmer for Illumina sequence data”, *Bioinformatics*, **30(15)**, pp.2114-2120, DOI: 10.1093/bioinformatics/btu170.
- [9] D. Kim, J.M. Paggi, C. Park, et al. (2019), “Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype”, *Nat. Biotechnol.*, **37(8)**, pp.907-915, DOI: 10.1038/s41587-019-0201-4.
- [10] M.G. Grabherr, B.J. Haas, M. Yassour, et al. (2011), “Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data”, *Nat. Biotechnol.*, **29(7)**, pp.644-652, DOI: 10.1038/nbt.1883.
- [11] W. Li, A. Godzik (2006), “Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences”, *Bioinformatics*, **22(13)**, pp.1658-1659, DOI: 10.1093/bioinformatics/btl158.
- [12] F.A. Simão, R.M. Waterhouse, P. Ioannidis, et al. (2015), “BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs”, *Bioinformatics*, **31(19)**, pp.3210-3212, DOI: 10.1093/bioinformatics/btv351.
- [13] S.F. Altschul, T.L. Madden, A.A. Schäffer, et al. (1997), “Gapped BLAST and PSI-BLAST: A new generation of protein database search programs”, *Nucleic Acids Research*, **25(17)**, pp.3389-3402, DOI: 10.1093/nar/25.17.3389.
- [14] H. Li, B. Handsaker, A. Wysoker, et al. (2009), “The Sequence Alignment/Map format and SAMtools”, *Bioinformatics*, **25(16)**, pp.2078-2079, DOI: 10.1093/bioinformatics/btp352.
- [15] S. Anders, P.T. Pyl, W. Huber (2015), “HTSeq-A python framework to work with high-throughput sequencing data”, *Bioinformatics*, **31(2)**, pp.166-169, DOI: 10.1093/bioinformatics/btu638.
- [16] B. Langmead, S.L. Salzberg (2012), “Fast gapped-read alignment with Bowtie 2”, *Nat. Methods*, **9**, pp.357-359, DOI: 10.1038/nmeth.1923.
- [17] B. Li, C.N. Dewey (2011), “RSEM: Accurate transcript quantification from RNA-seq data with or without a reference genome”, *BMC Bioinformatics*, **12**, DOI: 10.1186/1471-2105-12-323.
- [18] M.I. Love, W. Huber, 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.
- [19] N.M. Thanh, T.T.H. Yen, V.T.M. Thu, et al. (2021), Transcriptome analyses of female black tiger shrimps (*Penaeus monodon*) for reproductive trait using RNA-sequencing approach”, *Vietnam Journal of Science and Technology*, **63(12)**, pp.47-51, DOI: 10.31276/VJST.63(12).47-51 (in Vietnamese).
- [20] J. Yuan, X. Zhang, C. Liu, et al. (2018), “Genomic resources and comparative analyses of two economical penaeid shrimp species, *Marsupenaeus japonicus* and *Penaeus monodon*”, *Mar. Genomics*, **39**, pp.22-25, DOI: 10.1016/j.margen.2017.12.006.
- [21] D.V. Quyen, H.M. Gan, Y.P. Lee, et al. (2020), “Improved genomic resources for the black tiger prawn (*Penaeus monodon*)”, *Mar. Genomics*, **52**, DOI: 10.1016/j.margen.2020.100751.
- [22] K. Sittikankaew, W. Pootakham, C. Sonthirod, et al. (2020), “Transcriptome analyses reveal the synergistic effects of feeding and eyestalk ablation on ovarian maturation in black tiger shrimp”, *Sci. Rep.*, **10(1)**, DOI: 10.1038/s41598-020-60192-2.
- [23] H.E.L. Lischer, K.K. Shimizu (2017), “Reference-guided de novo assembly approach improves genome reconstruction for related species”, *BMC Bioinformatics*, **18(1)**, DOI: 10.1186/s12859-017-1911-6.
- [24] A. Conesa, P. Madrigal, S. Tarazona, et al. (2016), “A survey of best practices for RNA-seq data analysis”, *Genome Biol.*, **17**, DOI: 10.1186/s13059-016-0881-8.
- [25] A. Srivastava, L. Malik, H. Sarkar, et al. (2020), “Alignment and mapping methodology influence transcript abundance estimation”, *Genome Biol.*, **21(1)**, DOI: 10.1186/s13059-020-02151-8.
- [26] S. Li, S. Guo, F. Li, et al. (2015), “Functional diversity of anti-lipopolysaccharide factor isoforms in shrimp and their characters related to antiviral activity”, *Mar. Drugs*, **13(5)**, pp.2602-2616, DOI: 10.3390/md13052602.
- [27] H. Ye, X. Li, T. Zheng, et al. (2016), “The effect of the immune system on ovarian function and features of ovarian germline stem cells”, *SpringerPlus*, **5(1)**, DOI: 10.1186/s40064-016-2390-3.
- [28] S.J. Gutierrez, C.E.C. Caballero, C.B. Hernandez, et al. (2019), “Crustacean vitellogenin: A systematic and experimental analysis of their genes, genomes, mRNAs and proteins; and perspective to Next Generation Sequencing”, *Crustaceana*, **92(10)**, pp.1169-1205, DOI: 10.1163/15685403-00003930.
- [29] T.V. Nguyen, L.W. Ryan, J. Nocillado, et al. (2020), “Transcriptomic changes across vitellogenesis in the black tiger prawn (*Penaeus monodon*), neuropeptides and G protein-coupled receptors repertoire curation”, *Gen. Comp. Endocrinol.*, **298**, DOI: 10.1016/j.ygcen.2020.113585.