

Phân tích hệ gen phiên mã của tôm sú cái (*Penaeus monodon*) ở Việt Nam liên quan đến tính trạng sinh sản bằng kỹ thuật giải trình tự gen thế hệ mới

Nguyễn Minh Thành¹, Trần Thị Hải Yến¹, Võ Thị Minh Thu^{1,2},
Lê Thị Hồng Thắm¹, Nguyễn Văn Sáng³, Nguyễn Thành Luân³, Nguyễn Việt Tuấn^{2*}

¹Khoa Công nghệ Sinh học, Trường Đại học Quốc tế, Đại học Quốc gia TP Hồ Chí Minh

²Trung tâm Nghiên cứu GeneCology, Đại học Sunshine Coast, Australia

³Viện Nghiên cứu Nuôi trồng Thủy sản 2 (RIA2)

Ngày nhận bài 20/7/2021; ngày chuyển phản biện 22/7/2021; ngày nhận phản biện 25/8/2021; ngày chấp nhận đăng 30/8/2021

Tóm tắt:

Tôm sú (*Penaeus monodon*) là loài tôm được nuôi phổ biến ở nước ta và trên thế giới. Hiện nay, sản xuất giống tôm sú vẫn phụ thuộc vào nguồn tôm bố mẹ tự nhiên bởi vì chất lượng sinh sản của tôm tự nhiên cao hơn tôm sú gia hóa. Trước thực tế đó, nhóm tác giả đã tiến hành nghiên cứu hệ gen phiên mã của tôm sú cái tự nhiên và gia hóa ở giai đoạn buồng trứng tiền thành thực bằng kỹ thuật giải trình tự gen thế hệ mới. Giải trình tự bằng thiết bị Illumina đạt số lượng đoạn trình tự (read) sau tinh sạch là 20.977.708 cho tôm tự nhiên và 31.185.197 cho tôm gia hóa. Kết quả lắp ráp theo phương pháp không có hệ gen tham chiếu đạt 35.870 contig, trong đó độ dài trung bình và N50 của contig lần lượt là 1.018 và 1.488 bp. Chú giải gen chức năng dựa vào 7 cơ sở dữ liệu đạt tỷ lệ tương đồng từ 19,74 đến 47,77%. Hệ gen phiên mã của nghiên cứu này có tỷ lệ tương đồng cao nhất với loài *Hyaella azteca*, *Cryptotermes secundus*, *Zootermopsis nevadensis* và các loài tôm thuộc họ *Penaeus*. Nghiên cứu cũng xác định được tổng cộng 5.788 gen có biểu hiện khác biệt ở tôm tự nhiên khi so sánh với tôm gia hóa. Các gen có biểu hiện khác biệt được tiếp tục phân loại dựa theo khóa thuật ngữ của Gene Ontology (GO). Kết quả phân loại chức năng theo GO ở tôm sú tự nhiên cho thấy, các gen hemolymph clottable, peritrophin, ecdysteroid có biểu hiện tăng và các gen serine protease, alpha-L-fucosidase-like, actin, catenin alpha có biểu hiện giảm. Nghiên cứu này là cơ sở quan trọng bổ sung thông tin hệ phiên mã của tôm sú liên quan đến tính trạng sinh sản và có thể ứng dụng để cải thiện chất lượng sinh sản của tôm sú gia hóa.

Từ khóa: hệ gen phiên mã, *Penaeus monodon*, RNA-seq, tính trạng sinh sản.

Chỉ số phân loại: 4.6

Đặt vấn đề

Tôm sú (*Penaeus monodon*) là loài thủy sản mang lại giá trị kinh tế lớn, đóng góp đáng kể vào phát triển kinh tế - xã hội của đất nước. Năm 2020, sản lượng tôm sú đạt 267.700 tấn và giá trị xuất khẩu đạt 575,44 triệu USD [1]. Tôm sú được xác định là một trong hai đối tượng tôm nuôi nước lợ chủ lực của nước ta (cùng với tôm thẻ chân trắng), do đó nhu cầu giống tôm sú ngày càng tăng về số lượng và chất lượng. Tôm sú bố mẹ hiện nay chủ yếu đánh bắt từ tự nhiên là nguồn lây truyền mầm bệnh, làm cho nghề nuôi tôm sú của Việt Nam không bền vững do dịch bệnh dễ bùng phát. Việc chủ động phát triển đàn tôm sú bố mẹ gia hóa chất lượng cao để phục vụ nghề nuôi tôm bền vững là cần thiết tại Việt Nam. Tuy nhiên, chất lượng sinh sản của tôm sú bố mẹ gia hóa thấp hơn tôm sú bố mẹ tự nhiên. Các cơ sở sản xuất giống tôm sú (năm 2020 có 1.750 cơ sở [1]) vẫn có khuynh hướng sử dụng tôm bố mẹ tự nhiên. Vì vậy, chúng tôi đã tiến hành nghiên cứu hệ phiên mã của tôm sú cái liên quan đến tính trạng sinh sản nhằm cải thiện đặc điểm sinh sản của tôm sú gia hóa. Nghiên cứu áp dụng RNA-sequencing là phương pháp nhờ vào ưu điểm vượt trội của kỹ thuật giải trình tự gen thế hệ mới để phát hiện toàn bộ bản phiên mã mRNA của một loài. Mục tiêu của nghiên cứu là xây dựng cơ sở dữ liệu hệ gen phiên mã

của tôm sú tự nhiên và gia hóa, đồng thời tìm kiếm các gen tiềm năng liên quan đến tính trạng sinh sản ở giai đoạn buồng trứng tiền thành thực. Đây sẽ là nguồn thông tin gen chức năng giá trị có thể sử dụng cho những chương trình chọn giống nâng cao sức sinh sản của tôm sú sau này.

Đối tượng và phương pháp nghiên cứu

Mẫu tôm thí nghiệm

Nghiên cứu thu thập mẫu tôm sú cái tự nhiên (do cơ sở thu mua tôm sú bố mẹ cung cấp) và gia hóa (do Chương trình chọn giống của Viện Nghiên cứu Nuôi trồng Thủy sản 2 cung cấp). Buồng trứng ở giai đoạn tiền thành thực (giai đoạn 0) là cơ quan đích của nghiên cứu hệ gen phiên mã được lưu trữ trong RNAlater (Invitrogen, USA) và bảo quản ở -80°C cho đến khi thực hiện ly trích RNA tại Phòng thí nghiệm, Khoa Công nghệ Sinh học, Trường Đại học Quốc tế.

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) [2]. RNA tổng số được

*Tác giả liên hệ: Email: tnguyen@usc.edu.au

Transcriptome analyses of female black tiger shrimps (*Penaeus monodon*) for reproductive trait using RNA-sequencing approach

Minh Thanh Nguyen¹, Thi Hai Yen Tran¹,
Thi Minh Thu Vo^{1,2}, Thị Hồng Tham Lê¹,

Van Sang Nguyen³, Thanh Luan Nguyen³, Viet Tuan Nguyen^{2*}

¹School of Biotechnology, International University, VNU, Ho Chi Minh city

²GeneCology Research Centre, University of the Sunshine Coast, Australia

³Research Institute for Aquaculture No.2 (RIA2)

Received 20 July 2021; accepted 30 August 2021

Abstract:

The black tiger shrimp (*Penaeus monodon*) is the widely cultured aquaculture species in Vietnam and worldwide. Production of *P. monodon* postlarvae still relies on the wild broodstock due to their higher fecundity and larval quality in comparison with the domesticated broodstock. Therefore, the current study applied an RNA-sequencing approach by the Illumina platform to generate the transcriptomic resources for the wild and domesticated females at the previtellogenic stage of ovaries (stage 0). Total clean reads were 20,977,708 for the wild female and 31,185,197 for the domesticated female. *De novo* assembly was employed to generate 35,870 contigs with an average length of 1,018 bp and N50 length of 1,488 bp. The ratios of contigs possessing significant similarity through annotation across seven databases ranged from 19.74 to 47.77%. Top hit species from BLASTx searches included *Hyalella azteca*, *Cryptotermes secundus*, *Zootermopsis nevadensis*, followed by *Penaeus* sp.. We identified a total of 5,788 differentially expressed transcripts between the ovaries of wild and domesticated shrimps. The differentially expressed transcripts were further enriched according to the classification terms of Gene Ontology (GO). Results of GO enrichment analysis in the wild female indicated that many genes such as hemolymph clottable, peritrophin, ecdysteroid were up-regulated while the following genes, including serine protease, alpha-L-fucosidase-like, actin, catenin alpha were down-regulated. The current study provides more transcriptomic resources for the reproductive trait in *P. monodon*. These resources are potentially applied for the improvement of reproduction in the domesticated shrimp.

Keywords: *Penaeus monodon*, reproductive trait, RNA-sequencing, transcriptome.

Classification number: 4.6

định tính và định lượng bằng máy quang phổ Nanodrop™ (ThermoFisher) và Bioanalyser 2100 (Agilent). RNA tổng số cũng được kiểm tra độ tinh sạch (nhiễm DNA hoặc protein) và sự phân hủy bằng phương pháp điện di agarose. 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. cDNA tiếp tục được định tính và định lượng bằng Qubit 2.0 (Invitrogen) và Bioanalyser, cuối cùng thư viện cDNA được giải trình tự bằng máy Illumina NovaSeq 6000 tại Novogene (Singapore).

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

Các đoạn trình tự thô sau khi giải trình tự được xử lý loại bỏ các adapter gắn vào khi chuẩn bị các thư viện cDNA để giải trình tự. Phần mềm FastQC được sử dụng để đánh giá chất lượng các đoạn trình tự thô. Các đoạn trình tự có tỷ lệ nucleotide không thể xác định >10% và chất lượng thấp (QC<20) được loại bỏ bằng phần mềm Trimmomatic [3]. Các đoạn trình tự ngắn đạt chất lượng được lắp ráp thành bộ phiên mã theo định dạng không có hệ gen tham chiếu bằng phần mềm Trinity [4]. Các contig lắp lại sau khi lắp ráp từ Trinity được loại bỏ bằng phần mềm CORSET và các transcript dài nhất của mỗi gen được báo cáo là các unigene.

Các công cụ tin sinh học khác nhau được sử dụng để chú giải các transcript dựa vào 7 cơ sở dữ liệu. Đầu tiên công cụ BLASTX của NCBI [5] được sử dụng cho cơ sở dữ liệu NR (Non-redundant protein sequences) (giá trị $E < 1e^{-5}$). Sau đó chương trình DIAMOND [6] sử dụng cho các cơ sở dữ liệu NT (NCBI nucleotide sequences), KOG (euKaryotic orthologous groups) và Swiss-Prot (giá trị $E < 1e^{-5}$). Chương trình HMMER sử dụng cho cơ sở dữ liệu PFAM (Protein family) (giá trị $E < 0,01$). Phần mềm Blast2GO [7] áp dụng cho cơ sở dữ liệu GO (giá trị $E < 1e^{-6}$) và KAAS - KEGG Automatic annotation server sử dụng cho cơ sở dữ liệu KEGG (Kyoto encyclopedia of genes and genome) (giá trị $E < 1e^{-5}$).

Sơ sánh sự biểu hiện gen giữa tôm tự nhiên và gia hóa

Phần mềm Bowtie [8] được sử dụng để giống hàng các transcript và phần mềm RSEM [9] được sử dụng cho định lượng gen biểu hiện của mẫu tôm tự nhiên và gia hóa. Phân tích biểu hiện gen khác biệt giữa mẫu tôm tự nhiên và gia hóa bằng phần mềm EgdeR [10] dựa vào các giá trị: $|\log_2\text{FoldChange}| > 2$ và giá trị p điều chỉnh (p-adjusted value) $< 0,005$ [11]. Sau khi xác định được các gen có biểu hiện khác biệt giữa 2 nhóm tôm, các gen này được tiếp tục phân loại thành các nhóm gen chức năng theo các khóa thuật ngữ của GO bằng phần mềm

GOSeq (V1.32.0) [12] và topGO (2.32.0) [13]. Bước phân tích này nhằm xác định các nhóm chức năng mà các gen có biểu hiện sai khác tham gia liên quan đến tính trạng nghiên cứu.

Kết quả và bàn luận

Kết quả giải trình tự và lắp ráp các đoạn trình tự

Thư viện cDNA chuẩn bị từ mô buồng trứng của tôm sú tự nhiên và gia hóa được giải trình tự bằng thiết bị Illumina. Kết quả giải trình tự đạt được dữ liệu 6,5 Gbp cho mẫu tôm tự nhiên và 9,5 Gbp cho mẫu tôm gia hóa. Số lượng đoạn trình tự ngắn thô lần lượt là 21.642.150 và 31.742.591 cho mẫu tôm tự nhiên và gia hóa (bảng 1). Kích thước trung bình của các đoạn trình tự thô là 150 bp theo tiêu chuẩn kỹ thuật của thiết bị Illumina HiSeq4000. Sau khi sàng lọc các đoạn trình tự không đạt chất lượng, tỷ lệ đoạn trình tự ngắn được sử dụng cho lắp ráp là 96,93 và 98,24% cho từng nhóm tôm.

Bảng 1. Tóm tắt kết quả giải trình tự bằng thiết bị Illumina.

Chỉ số phân tích	Tôm tự nhiên	Tôm gia hóa
Tổng số base (Gbp)	6,5	9,5
Tổng số base đạt chuẩn >Q20 (Gbp)	6,3	9,4
Số lượng đoạn trình tự ngắn	21.642.150	31.742.591
Số lượng đoạn trình tự sau tinh sạch	20.977.708 (96,93%)	31.185.197 (98,24%)

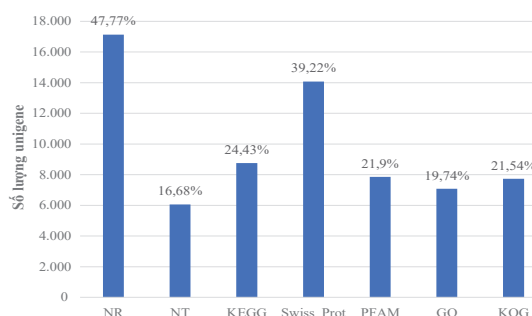
Kết quả lắp ráp đạt 35.870 contig (đã sàng lọc các contig lặp lại) từ các đoạn trình tự tinh sạch cộng gộp của tôm tự nhiên và gia hóa. Chiều dài trung bình của các contig là 1.018 bp, dao động từ 301 bp đến 9.245 bp (bảng 2). Ngoài ra, chiều dài contig N50 là 1.488 bp. Chiều dài trung bình của các contig và chiều dài N50 là các chỉ tiêu để đánh giá chất lượng của phần mềm lắp ráp [14]. Nghiên cứu của Sadamoto và cs (2012) [15] cho rằng, chiều dài contig N50 càng dài cho thấy chất lượng càng cao của kết quả lắp ráp. Chiều dài N50 của nghiên cứu hiện tại lớn hơn N50 của nghiên cứu hệ gen tôm sú cũng sử dụng thiết bị Illumina giải trình tự [16]. Tỷ lệ các unigene ≥ 1.000 bp đạt tương đối cao (33%), rất có ý nghĩa cho các bước phân tích tiếp theo.

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

Chỉ số phân tích	Giá trị
Tổng số contig	35.870
Tổng số base (bp)	36.498.375
Chiều dài contig trung bình (bp)	1.018
Chiều dài contig ngắn nhất (bp)	301
Chiều dài contig lớn nhất (bp)	9.245
Chiều dài contig N50 (bp)	1.488
Số lượng contig ≥ 1.000 bp	11.856 (33%)

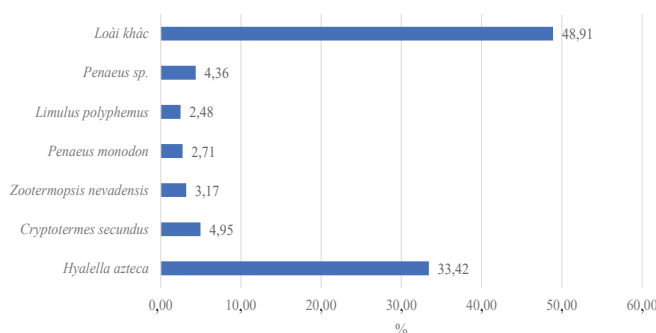
Kết quả chú giải và phân loại các gen chức năng

Kết quả chú giải các gen chức năng sau khi lắp ráp dựa vào 7 cơ sở dữ liệu khác nhau được trình bày ở hình 1.



Hình 1. Tỷ lệ chú giải các unigene dựa vào các cơ sở dữ liệu khác nhau.

Kết quả hình 1 cho thấy, tỷ lệ chú giải đạt cao nhất là 47,77% khi dựa vào cơ sở dữ liệu NR của NCBI. Đây là cơ sở dữ liệu tập hợp trình tự protein từ nhiều nguồn dữ liệu khác nhau trên thế giới. Vì vậy cơ sở dữ liệu này luôn được lựa chọn để chú giải gen chức năng nhằm tìm kiếm mức độ tương đồng cao nhất giữa các trình tự đang được nghiên cứu và các gen đã được công bố lưu trữ trên các cơ sở dữ liệu. Tỷ lệ chú giải cao nhất khi dựa vào NR là điều hiển nhiên. Kế tiếp, tỷ lệ chú giải đạt 39,22% khi dựa vào cơ sở dữ liệu Swiss_Prot. Cơ sở dữ liệu này chuyên biệt cho các trình tự protein và đã được kiểm chứng bởi các thực nghiệm nên là cơ sở dữ liệu tin cậy cho chú giải gen chức năng. Tỷ lệ chú giải dựa vào Swiss_Prot trong nghiên cứu của chúng tôi cao hơn công bố của Huerlimann và cs (2018) [17] cũng nghiên cứu hệ gen phiên mã của tôm sú đạt tỷ lệ chú giải khoảng 30%. Nhóm nghiên cứu khẳng định, tỷ lệ chú giải thấp là kết quả thông thường khi nghiên cứu hệ gen phiên mã của các loài không phải là các loài nghiên cứu tiêu biểu (non-model species). Các cơ sở dữ liệu còn lại cũng được chúng tôi sử dụng để chú giải nhằm phân loại các nhóm protein (PFAM), nhóm gen chức năng (GO và KOG) và các chuỗi chuyển hóa sinh học (KEGG). Kết quả so sánh mức độ tương đồng của các unigene của nghiên cứu hiện tại với các loài đã được nghiên cứu được trình bày ở hình 2.



Hình 2. Tỷ lệ unigene tương đồng với các loài bằng phân tích BLASTX.

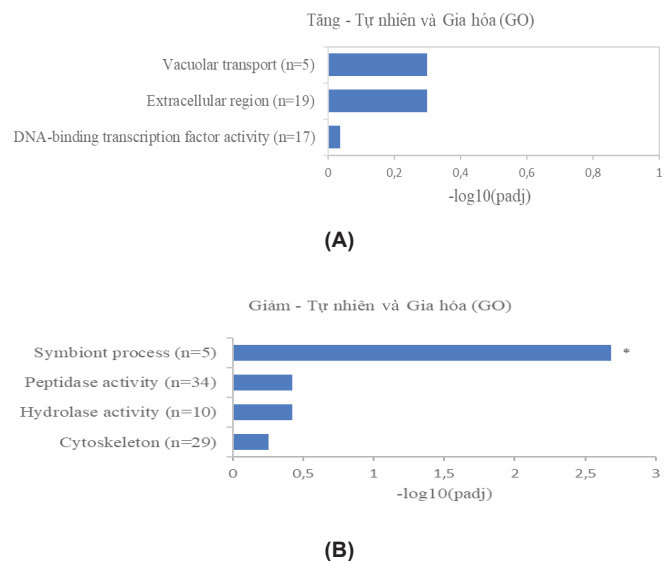
Kết quả ở hình 2 cho thấy, loài có mức độ tương đồng cao nhất là phiêu sinh động vật *Hyalella azteca*. Đây là một trong những loài giáp xác của dự án i5k với mục tiêu giải mã toàn bộ hệ gen của 5.000 loài thuộc ngành động vật chân khớp [18]. Hệ gen của *H. azteca* đã được giải mã hoàn toàn và công bố nên trình tự gen của loài này phong phú trên các cơ sở dữ liệu. Các loài kế tiếp có mức độ tương đồng cao với tôm sú là *Cryptotermes*

secundus và *Zootermopsis nevadensis*. Hệ gen nhân và hệ gen phiên mã của các loài này cũng đã được công bố gần đây nên có sẵn trên các cơ sở dữ liệu [19-21]. Đối với tôm sú, nghiên cứu của chúng tôi đạt mức độ tương đồng chỉ 2,71% mặc dù hệ gen phiên mã của tôm sú đã được công bố [11, 17]. Điều này cho thấy, nghiên cứu của chúng tôi chỉ giải trình tự một loại mô nên không thể thu nhận tỷ lệ lớn hệ gen phiên mã của tôm sú. Điểm đáng lưu ý là hệ gen phiên mã của tôm sú đã công bố có tỷ lệ lặp lại cao (51,3%) [17] nên mức độ đa dạng nguồn gen không cao và có thể ảnh hưởng đến kết quả so sánh tương đồng nghiên cứu của chúng tôi và cơ sở dữ liệu hiện có. Ngoài ra, kết quả lắp ráp các đoạn trình tự dựa vào phương pháp lắp ráp không có hệ gen tham chiếu và phần mềm lắp ráp khác nhau cũng có thể ảnh hưởng đến mức độ sai khác của kết quả lắp ráp khi so sánh các nghiên cứu khác nhau. Tương tự, tỷ lệ tương đồng của các unigene tôm sú của chúng tôi cũng chỉ đạt 4,36% khi so sánh với các loài thuộc giống *Penaeus*, bao gồm *P. japonicus*, *P. merguensis*, *P. chinensis*, *P. indicus*, *P. semisulcatus*, *P. stylirostris*, *P. setiferus* và *Litopenaeus vannamei*. Kết quả này thể hiện cơ sở dữ liệu của các loài tôm thuộc họ *Penaeus* còn hạn chế và chưa được nghiên cứu đáng kể mặc dù chúng là các loài thủy sản có giá trị thương mại cao. Mức độ tương đồng cao của hệ gen phiên mã tôm sú với các loài *H. azteca*, *Z. nevadensis*, *L. polyphemus*, *L. vannamei* cũng tương tự công bố của Nguyen và cs (2020) [11] khi nghiên cứu hệ gen phiên mã của tôm sú gia hóa ở các giai đoạn thành thực sinh sản.

So sánh sự biểu hiện gen giữa tôm tự nhiên và gia hóa

Dựa vào giá trị $|\log_2\text{FoldChange}|$ và padj đã được thiết lập để so sánh sự biểu hiện gen khác biệt giữa tôm tự nhiên và tôm gia hóa, nghiên cứu của chúng tôi xác định được tổng cộng 5.788 gen có biểu hiện khác biệt. Trong đó tôm tự nhiên có 3.290 gen biểu hiện tăng và 2.498 gen biểu hiện giảm khi so sánh với gia hóa. Các gen có biểu hiện tăng ở tôm sú tự nhiên được phân loại thành 3 nhóm gen chức năng chính của GO, bao gồm quá trình sinh học: vận chuyển bằng không bào (vacuolar transport), thành phần tế bào: vùng ngoại bào (extracellular region) và chức năng phân tử: nhân tố tương tác DNA điều hòa phiên mã (DNA-binding transcription factor activity) (hình 3A). Trong nhóm vùng ngoại bào, gen hemolymph clottable có biểu hiện tăng, tương tự như công bố của [11] khi nghiên cứu tôm sú gia hóa ở giai đoạn tiền thành thực. Protein tổng hợp từ gen hemolymph clottable thuộc họ glycoprotein và có cấu trúc tương tự với vitellogenin. Những protein này giữ vai trò quan trọng trong quá trình thành thực buồng trứng của tôm sú [22]. Ngoài ra, nghiên cứu cũng xác định được gen tổng hợp peritrophin buồng trứng thuộc nhóm extracellular region là protein bảo vệ các trứng sau khi sinh sản [23]. Đối với nhóm chức năng phân tử, gen tổng hợp ecdysteroid được xác định có vai trò điều hòa các quá trình phát triển, thành thực sinh dục và sinh sản ở các loài chân khớp [24]. Tương tự, số lượng gen có biểu hiện giảm ở tôm sú tự nhiên cũng được phân loại thành nhóm gen chức năng, bao gồm quá trình sinh học: cộng sinh (symbiont process), thành phần tế bào: thành phần nâng đỡ

tế bào (cytoskeleton) và chức năng phân tử: hoạt tính phân cắt peptide (peptidase activity) và hoạt tính thủy phân (hydrolase activity) (hình 3B). Trong nhóm chức năng phân tử, nghiên cứu của chúng tôi tìm kiếm được gen tổng hợp serine protease có hoạt tính phân cắt peptide tham gia vào quá trình hoạt hóa hệ miễn dịch ở tôm [25]. Tác giả Ye và cs (2016) [26] công bố rằng, nhiều gen tham gia vào quá trình miễn dịch cũng tham gia vào quá trình thành thực buồng trứng. Tuy nhiên, gen này có biểu hiện giảm ở tôm sú tự nhiên trong nghiên cứu hiện tại. Ngoài ra, chúng tôi xác định được gen tổng hợp alpha-L-fucosidase-like protein có hoạt tính thủy phân cũng có biểu hiện giảm ở tôm sú tự nhiên. Fucosidase có tác dụng tiêu hóa các polysaccharide có trong thực vật [27]. Nghiên cứu của Sittikankaew và cs (2020) [28] báo cáo sử dụng thức ăn công nghiệp cho tôm sú mẹ hoạt hóa các gen tham gia vào các quá trình chuyển hóa khác với sử dụng thức ăn tươi sống nuôi vỗ tôm mẹ thành thực. Thức ăn tươi sống bao gồm giun nhiều tơ, mực, nhuyễn thể cung cấp nguồn dinh dưỡng giàu protein và chất béo không bão hòa (PUFA). Đây là nguồn dinh dưỡng cần thiết đã hoạt hóa các gen vitellogenin, collagen alpha-1V chain-like và heme-binding protein 2-like tham gia vào quá trình phát triển buồng trứng và thành thực sinh sản của tôm mẹ. Trong khi thức ăn công nghiệp chỉ hoạt hóa các gen tham gia vào quá trình chuyển hóa để duy trì hoạt động thông thường của tế bào. Điều này có thể giải thích gen tổng hợp enzyme tiêu hóa thức ăn có nguồn gốc từ thực vật giảm biểu hiện ở tôm sú tự nhiên trong quá trình thành thực sinh sản ở nghiên cứu của chúng tôi. Các gen thuộc nhóm cytoskeleton bao gồm actin và catenin alpha là các gen tham gia vào quá trình giảm stress sau khi cắt mắt ở tôm sú mẹ nhằm kích thích buồng trứng thành thực [28]. Các gen này cũng có biểu hiện giảm ở tôm sú tự nhiên khi so sánh với tôm sú gia hóa.



Hình 3. Phân loại các gen có biểu hiện khác biệt ở tôm sú tự nhiên theo khóa thuật ngữ của GO. (A) Các gen có biểu hiện tăng; (B) Các gen có biểu hiện giảm. Hàm $-\log_{10}(\text{padj})$ được sử dụng để điều chỉnh thang đo của ý nghĩa xác suất thống kê.

Kết luận

Nghiên cứu đã xây dựng được cơ sở dữ liệu hệ gen phiên mã của tôm sú tự nhiên và gia hóa ở giai đoạn buồng trứng tiền thành thực, đồng thời xác định được một số gen liên quan đến quá trình phát triển và thành thực của buồng trứng khi so sánh gen biểu hiện giữa hai nhóm tôm sú. Đây là cơ sở dữ liệu hoàn chỉnh có ý nghĩa to lớn trong nghiên cứu cải thiện chất lượng sinh sản của tôm sú gia hóa, góp phần phát triển ngành công nghiệp nuôi tôm sú bền vững ở Việt Nam.

LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Quỹ Phát triển KH&CN Quốc gia (NAFOSTED) trong đề tài mã số 106.05-2019.36. Nhóm nghiên cứu xin trân trọng cảm ơn NAFOSTED và RIA2 đã cung cấp mẫu tôm sú gia hóa của Chương trình chọn giống tôm sú cho nghiên cứu này.

TÀI LIỆU THAM KHẢO

- [1] Hiệp hội Chế biến và Xuất khẩu Thủy sản Việt Nam (2021), *Báo cáo xuất khẩu thủy sản Việt Nam năm 2020*, 103tr.
- [2] P. Chromczynski, K. Mackey (1995), "Modification of TRIZOL reagent procedure for isolation of RNA from Polysaccharide-and proteoglycan-rich sources", *Biotechniques*, **19**, pp.942-945.
- [3] A.M. Bolger, M. Lohse, B. Usadel (2014), "Trimomatic: a flexible trimmer for Illumina sequence data", *Bioinformatics*, **30**(15), pp.2114-2120.
- [4] M.G. Grabherr, et al. (2011), "Full-length transcriptome assembly from RNA-seq data without a reference genome", *Nat. Biotechnol.*, **29**(7), pp.644-652.
- [5] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.*, **25**, pp.3389-3402.
- [6] B. Buchfink, C. Xie, D.H. Huson (2014), "Fast and sensitive protein alignment using DIAMOND", *Nature Methods*, **2**, pp.59-60.
- [7] S. Götz, J.M. Garcia-Gomez, J. Terol, T.D. William, S.H. Gagaraj (2008), "High-throughput functional annotation and data mining with Blast2GO suite", *Nucleic Acids Res.*, **36**, pp.3420-3435.
- [8] B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg (2009), "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome", *Genome Biology*, **10**(3), DOI: 10.1186/gb-2009-10-3-r25.
- [9] B. Li, C. Dewey (2011), "RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome", *BMC Bioinformatics*, **12**, DOI: 10.1186/1471-2105-12-333.
- [10] M.D. Robinson, A. Oshlack (2010), "A scaling normalization method for differential expression analysis of RNA-seq data", *Genome Biol.*, **11**, p.25.
- [11] T.V. Nguyen, L.W. Ryan, J. Nocillado, M. Le Groumellec, A. Elizur, T. Ventura (2020), "Transcriptomic changes across vitellogenesis in the black tiger prawn (*Penaeus monodon*), neuropeptides and G protein-coupled receptors repertoire curation", *General and Comparative Endocrinology*, **298**, DOI: 10.1186/gb-2010-11-2-114.
- [12] M.D. Young, M.J. Wakefield, G.K. Smyth, A. Oshlack (2010), "Gene ontology analysis for RNA-seq: accounting for selection bias", *Genome Biology*, **11**, pp.4.
- [13] A. Alexa, J. Rahnenfuhrer (2021), *topGO: Enrichment Analysis*

for Gene Ontology, R package version 2.44.0.

- [14] S. Liu, (2013), "Efficient assembly and annotation of the transcriptome of catfish by RNA-seq analysis of a doubled haploid homozygote", *BMC Genomics*, **13**, DOI: 10.1186/1471-2164-13-595.
- [15] H. Sadamoto, H. Takahashi, T. Okada, H. Kenmoku, M. Toyota, Y. Asakawa (2012), "De novo sequencing and transcriptome analysis of the central nervous system of mollusc *Lymnaea stagnalis* by deep RNA sequencing", *PLOS ONE*, **7**, DOI: 10.1371/Journal.pone.0042546.
- [16] D.V. Quyen, H.M. Gan, Y.P. Lee, D.D. Nguyen, T.H. Nguyen, X.T. Tran, V.S. Nguyen, D.D. Khang, C.M. Austin (2020), "Improved genomic resources for the black tiger prawn (*Penaeus monodon*)", *Marine Genomics*, **52**, DOI: 10.1016/j.margen.2020.100751.
- [17] R. Huerlimann, et al. (2018), "De novo assembly, characterization, functional annotation and expression patterns of the black tiger shrimp (*Penaeus monodon*) transcriptome", *Sci. Rep.*, **8**, pp.13553, DOI: 10.1038/s41598-018-31148-4.
- [18] M. Poelchau, C. Childers, G. Moore, V. Tsavatapalli, J. Evans, C.Y. Lee, H. Lin, J.W. Lin, K. Hackett (2015), "The i5k workspace@NAL-enabling genomic data access, visualization and curation of arthropod genomes", *Nucleic Acids Research*, **43**, pp.D714-D719.
- [19] C.L. Schoch, et. al. (2020), "NCBI Taxonomy: a comprehensive update on curation, resources and tools", *Database*, **2020**, DOI: 10.1093/database/baaa062.
- [20] Z.Q. Qian (2016), "The complete mitogenome of the dampwood termite *Zootermopsis nevadensis* (Insecta: Isoptera: Termitidae)", *Mitochondrial DNA Part A*, **27**(2), pp.1163-1164.
- [21] S.D. Simpson, J.S. Ramsdell, W.H. Watson, C.C. Chabot (2017), "The draft genome and transcriptome of the Atlantic horseshoe crab, *Limulus polyphemus*", *International Journal of Genomics*, **1017** DOI: 10.1155/2017/7636513.
- [22] W. Cheng, P.C. Chiang, C.Y. Lai, M.S. Yeh (2008), "Expression of clottable protein of tiger shrimp (*Penaeus monodon*) in gonads and its possible role as nutrient source for the embryo", *Dev. Comp. Immunol.*, **32**, pp.1422-1429.
- [23] W. Loongyai, J.C. Avarre, M. Cerutti, E. Lubzens, W. Chotigeat (2007), "Isolation and functional characterization of a new shrimp ovarian peritrophin with antimicrobial activity from *Fenneropenaeus merguensis*", *Marine Biotechnology*, **9**, pp.624-637.
- [24] Z. Qian, et al., (2014), "Identification of ecdysteroid signaling late-response genes from different tissues of the Pacific white shrimp, *Litopenaeus vannamei*", *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.*, **172**, pp.10-30.
- [25] Y. Liu, F. Hou, X. Liu (2017), "Characterization and expression analysis of serpin B3, the first clade B serine protease inhibitor in Pacific white shrimp, *Litopenaeus vannamei*", *Developmental and Comparative Immunology*, **72**, pp.103-111.
- [26] H. Ye, X. Li, T. Zheng, X. Liang, J. Li, J. Huang, Z. Pan, Y. Zheng (2016), "The effect of the immune system on ovarian function and features of ovarian germline stem cells", *SpringerPlus*, **5**, DOI: 10.1186/s40064-016-2390-3.
- [27] O. Berteau, I. McCort, N. Goasdoué, B. Tissot, R. Daniel (2002), "Characterization of a new alpha-L-fucosidase isolated from the marine mollusk *Pecten maximus* that catalyzes the hydrolysis of alpha-L-fucose from algal fucoidan (*Ascophyllum nodosum*)", *Glycobiology*, **12**(4), pp.273-282.
- [28] K. Sittikankaew, et al. (2020), "Transcriptome analyses reveal the synergistic effects of feeding and eyestalk ablation on ovarian maturation in black tiger shrimp", *Sci. Rep.*, **10**, DOI: 10.1038/s41598-020-62221-6.