

ỨNG DỤNG KỸ THUẬT XÁC ĐỊNH KIỂU GEN BẰNG GIẢI TRÌNH TỰ (GBS) ĐỂ SÀNG LỌC CÁC ĐA HÌNH NUCLEOTIDE ĐƠN (SNPs) LIÊN QUAN ĐẾN TÍNH TRẠNG TĂNG TRƯỞNG Ở TÔM SÚ (*PENAEUS MONODON*)

Nguyễn Thị Minh Thanh¹, Nguyễn Quyết Tâm², Nguyễn Văn Hảo², Nguyễn Văn Sáng², Nguyễn Đăng Tôn³, Ma Thị Huyền Thương³, Kim Thị Phương Oanh³, Nông Văn Hải³, Nguyễn Thị Hoa¹, Hà Thị Thu¹, Vũ Thị Hiền¹, Nguyễn Đình Duy¹, Trần Xuân Thạch¹, Nguyễn Thị Tuyết Nhung¹, Nguyễn Hữu Ninh⁴, Đồng Văn Quyền¹, Đinh Duy Kháng¹✉

¹Viện Công nghệ sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

²Viện Nghiên cứu nuôi trồng thủy sản 2, Bộ Nông nghiệp và Phát triển nông thôn

³Viện Nghiên cứu hệ gen, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

⁴Viện Nghiên cứu nuôi trồng thủy sản 3, Bộ Nông nghiệp và Phát triển nông thôn

✉Người chịu trách nhiệm liên lạc. E-mail: khangvspt@ibt.ac.vn

Ngày nhận bài: 07.11.2017

Ngày nhận đăng: 20.3.2018

TÓM TẮT

Hiện nay, ngành nuôi tôm sú (*Penaeus monodon*) ở Việt Nam vẫn chưa thực sự phát triển bền vững do gặp nhiều khó khăn, trong đó có thách thức lớn về nguồn tôm sú giống. Để có được giống tôm sú chất lượng cao đòi hỏi phải có những nghiên cứu di truyền cải thiện chất lượng giống. Ứng dụng chỉ thị phân tử kết hợp với di truyền số lượng trong chọn giống là phương pháp hiệu quả nhờ thực hiện việc chọn lọc không cần trực tiếp trên tính trạng mong muốn mà thông qua chỉ thị phân tử liên kết với tính trạng đó. Công nghệ giải trình tự gen thế hệ mới (Next Generation Sequencing, NGS) kết hợp với các tính trạng số lượng để tìm ra các chỉ thị phân tử liên kết các tính trạng quan trọng là hướng đi đang được nhiều phòng thí nghiệm trên thế giới quan tâm. Trong nghiên cứu này, chúng tôi sử dụng kỹ thuật xác định kiểu gen bằng giải trình tự (Genotyping by Sequencing, GBS) với công nghệ NGS của Illumina để sàng lọc các đa hình nucleotide đơn (Single Nucleotide Polymorphisms, SNPs) liên quan đến tính trạng tăng trưởng nhằm tạo cơ sở dữ liệu cho các nghiên cứu chọn giống tôm sú hướng đến tính trạng tăng trưởng nhanh. Dựa trên hệ gen tham chiếu tạm thời của tôm sú được thiết lập *de novo*, tổng số 2.887 SNPs đã được sàng lọc, trong đó có 1.799 SNPs chỉ xuất hiện ở nhóm tôm sú lớn nhanh. Trong số 287 contig được chú giải ở nhóm tôm sú lớn nhanh có hai contig chứa SNPs có sự tương đồng với gen mã hóa protein myosin heavy chain (MHC) liên quan đến tính trạng tăng trưởng ở họ giáp xác, trong đó contig83953 tương đồng với MHCa và contig260347 tương đồng với MHC1. Hai SNPs đã được xác định là Contig83953:g.20T>C và Contig260347:g.19G>A. Đây là những kết quả bước đầu bổ sung vào Cơ sở dữ liệu hệ gen tôm sú, định hướng cho các nghiên cứu tiếp theo nhằm khai thác các chỉ thị phân tử ứng dụng trong chọn giống tôm sú.

Từ khóa: Đa hình nucleotide đơn, GBS, giải trình tự gen thế hệ mới, tính trạng tăng trưởng, tôm sú

MỞ ĐẦU

Tôm sú (*P. monodon*) là một trong những loài tôm có kích thước lớn và được coi là loài thủy sản nuôi kinh tế quan trọng ở nhiều nước trên thế giới. Ở Việt Nam, từ lâu tôm sú đã và sẽ tiếp tục là một trong những đối tượng nuôi chủ lực cho xuất khẩu thủy sản đem về nguồn ngoại tệ đáng kể cho đất nước. Ngành công nghiệp nuôi tôm sú hiện nay với sản lượng hơn 300.000 tấn/năm đòi hỏi khoảng 30 tỉ tôm giống chất lượng cao. Hiện nay, một số cơ sở ở Việt Nam đã gia

hóa thành công tôm sú, tuy nhiên, chất lượng tôm giống vẫn còn là vấn đề hết sức nan giải. Tôm sú giống không được kiểm soát chặt chẽ về mầm bệnh cũng như chất lượng di truyền dẫn đến những tổn thất to lớn cho người nuôi tôm. Đây là những thách thức lớn trong nỗ lực phát triển bền vững ngành công nghiệp nuôi tôm sú của Việt Nam.

Xác định được tầm quan trọng của chất lượng tôm giống trong chuỗi sản xuất, Việt Nam cũng như nhiều quốc gia nuôi tôm trên thế giới đang tập trung

đầu tư cho các chương trình nghiên cứu và phát triển tômgia hóa nhằm tăng sản lượng và khả năng kháng bệnh của tôm. So với tôm thẻ thì các nghiên cứu trên đối tượng tôm sú còn thua kém cả về số lượng và chất lượng, cũng như chưa tương xứng với vai trò của một đối tượng kinh tế quan trọng. Do những hạn chế về mặt sinh học nên quá trình nghiên cứu gia hóa, khép kín vòng đời tôm sú dù đã thành công bước đầu, nhưng việc chủ động tạo ra được nguồn tôm bố mẹ nhân tạo vẫn đang gặp rất nhiều khó khăn. Hiện nay chỉ có một vài nơi trên thế giới được ghi nhận thành công đối với loài này như Công ty Moana (Bi), CSIRO (Úc) và một số cơ sở của Việt Nam. Tuy nhiên, để có thể thương mại hóa thì đòi hỏi chúng ta phải có một chương trình chọn giống hiệu quả và liên tục qua nhiều thế hệ (<http://thuysanvietnam.com.vn>).

Trong những năm gần đây, một số cơ sở nghiên cứu ở nước ta đã bắt đầu chú ý đến các nghiên cứu nhằm nâng cao chất lượng di truyền và kiểm soát dịch bệnh ở tôm sú giống. Để có được giống tôm sú có chất lượng cao đòi hỏi phải có những nghiên cứu di truyền cải thiện chất lượng giống, trước hết tập trung vào các tính trạng kinh tế quan trọng như sinh trưởng, kháng bệnh và chất lượng tôm. Các chỉ thị phân tử (CTPT) ngày nay đang được sử dụng rộng rãi như một công cụ chọn lọc đặc lực đối với các tính trạng quan tâm trong các chương trình chọn giống ở nhiều loài vật nuôi, cây trồng và các loài thủy sản. Việc ứng dụng các CTPT giúp cho các nhà chọn giống nhận diện chính xác các gen quan tâm nhằm rút ngắn thời gian chọn giống. Hiệu quả cải tiến giống sẽ gia tăng gấp nhiều lần so với các phương pháp chọn giống cổ điển nhờ thực hiện việc chọn lọc không cần trực tiếp trên tính trạng mong muốn mà thông qua CTPT liên kết với tính trạng đó (Bùi Chí Bửu, Nguyễn Thị Lang, 2004).

SNPs là những CTPT đã và đang được ứng dụng hiệu quả trong công tác chọn giống. Các vị trí thể hiện SNPs trong hệ gen là nơi mà ở đó chuỗi DNA được phân biệt bởi một base duy nhất khi hai hoặc nhiều cá thể được so sánh. Sự khác nhau về nucleotide này có thể dẫn đến thay đổi tính trạng và được sử dụng để đánh giá sự đa dạng trong tiến hóa. SNPs là dạng thay đổi trình tự trong hệ gen phổ biến nhất cho tới nay (Nguyễn Đức Thành, 2014).

SNPs có số lượng lớn, ổn định thích hợp cho việc tự động hóa trong phân tích và chỉ ra được các đa hình có thể không phát hiện được bởi các phương pháp khác (Liu and Cordes, 2004; Vaseeharan *et al.*, 2013). Điều quan trọng là SNPs có thể xuất hiện ở cả

các vùng mã hóa và không mã hóa trong hệ gen của sinh vật (Nguyễn Đức Thành, 2014), tác động trực tiếp đến tính trạng quan tâm, rất hiệu quả trong việc xác định mối tương quan giữa SNPs và tính trạng nào đó (Beuzen *et al.*, 2000). SNPs đã được sử dụng để sàng lọc các gen tiềm năng liên quan đến tính trạng tăng trưởng của một số đối tượng thủy sản (Nguyễn Minh Thành *et al.*, 2011), bao gồm cả hồi *Salvelinus alpinus* (Tao, Boulding, 2003), cá chêm (Xu *et al.*, 2006), tôm thẻ chân trắng *Litopenaeus vannamei* và tôm sú *P. monodon* (Glenn *et al.*, 2005).

Có nhiều phương pháp để phát hiện SNPs (Liu and Cordes, 2004; Jehan & Lakhampaul, 2006; Nguyễn Đức Thành, 2014), trong đó giải trình tự DNA là phương pháp chính xác và được sử dụng nhiều nhất cho đến nay (Liu and Cordes, 2004). Hiện nay, GBS sử dụng công nghệ NGS đã và đang được ứng dụng rộng rãi nhiều đối tượng sinh vật khác nhau. GBS cho phép giải quyết được đồng thời hai mục đích là phát hiện CTPT và xác định kiểu gen. GBS đã và đang trở thành phương pháp khả thi được áp dụng đối với các loài có hệ gen lớn và có tính đa hình cao (Elshire *et al.*, 2011). Những tiến bộ của công nghệ GBS cho phép đưa ra một lượng lớn các SNPs để sử dụng trong các phân tích di truyền (Beissinger *et al.*, 2013).

Trong nghiên cứu này, chúng tôi sử dụng kỹ thuật GBS để sàng lọc các SNPs liên quan tới tính trạng tăng trưởng ở tôm sú nhằm tạo cơ sở dữ liệu cho các nghiên cứu chọn giống tôm sú hướng đến tính trạng tăng trưởng nhanh.

VẬT LIỆU VÀ PHƯƠNG PHÁP

Mẫu tôm sú

Vật liệu gốc bao gồm bốn dòng tôm sú bố mẹ có nguồn gốc địa lý khác nhau, trong đó ba dòng có nguồn gốc tự nhiên là tôm sú Ấn Độ Dương (kí hiệu: A) nhập từ Thái Lan (vùng biển Amanda) và từ Myanmar, tôm sú Thái Bình Dương (T) nhập từ Singapore, tôm tự nhiên của Việt Nam (thu thập từ Cà Mau, Đà Nẵng, Phú Yên) gọi chung là tôm nội địa (N) và nhóm thứ tư là dòng tôm Gia hóa (G) được cung cấp bởi Công ty Moana Ninh Thuận - Đây là dòng tôm đã được thương mại hóa dùng làm tôm bố mẹ sản xuất tôm giống phục vụ cho nuôi thương phẩm. Các gia đình tôm sú (thế hệ G_0 và G_1) được sinh ra từ các đàn tôm bố mẹ này bằng cách phối ghép hỗn hợp giữa bốn dòng tôm bố mẹ. Từ các cá thể trong đàn con của các gia đình tôm sú thế hệ

G₀ và G₁ chọn ra các cá thể (bao gồm cả tôm cái - ♀ và tôm đực - ♂) thuộc hai nhóm khác nhau theo tiêu chí mức độ tăng trưởng: Lớn nhanh (Fast - F) và Lớn chậm (Slow - S) để sử dụng cho nghiên cứu sàng lọc

SNP (Bảng 1). Từ dữ liệu về nguồn gốc tôm giống, chỉ các cá thể tôm sú có tôm bố hoặc tôm mẹ hoặc cả tôm bố và tôm mẹ nguồn gốc nội địa (N) được chọn để sử dụng cho nghiên cứu.

Bảng 1. Các nhóm tôm sú dùng cho nghiên cứu sàng lọc SNP.

Kí hiệu nhóm tôm	Thế hệ	Tình trạng	Số lượng cá thể Tổng số (♀, ♂)
I	G ₀	Lớn nhanh (F)	60 (30, 30)
II		Lớn chậm (S)	60 (30, 30)
III	G ₁	Lớn nhanh (F)	60 (30, 30)
IV		Lớn chậm (S)	60 (30, 30)

Quá trình ghép phôi, sinh sản nhân tạo, ương nuôi, phân loại (lớn nhanh, lớn chậm) và truy xuất nguồn gốc tôm sú được thực hiện theo quy trình sản xuất giống đã được nghiên cứu và thực hiện thành công nhiều năm tại Trung tâm Quốc gia giống hải sản Nam Bộ, thuộc Viện Nghiên cứu nuôi trồng thủy sản 2.

Tách DNA tổng số

DNA tổng số của các mẫu tôm sú sử dụng cho nghiên cứu sàng lọc SNPs được tách chiết theo quy trình của Eurobiobank (2004), tinh sạch bằng kit PureLink Genomic DNA của Life Technologies và xác định nồng độ bằng máy quang phổ *NanoDrop® Spectrophotometer* (Desjardins, Conklin, 2010). Các bước chi tiết của quy trình tách DNA tổng số từ mô cơ tôm sú đã được mô tả trong công bố của Nguyễn Thị Minh Thanh *et al.*, (2015).

Xây dựng thư viện GBS và giải trình tự DNA

Kỹ thuật GBS được thực hiện theo các bước như mô tả trong công bố của Elshire và đồng tác giả (2011). Enzyme cắt hạn chế (RE) được sử dụng là *ApeKI* (New England Biolabs), cắt tại trình tự nhận biết GCWGC. DNA tổng số sau khi được cắt và gắn với adapter được tinh sạch bằng QIAquick PCR Purification Kit (Qiagen). Các đoạn DNA tinh sạch từ các mẫu được trộn lại với nhau với tỷ lệ tương đương. Khoảng 100 ng sản phẩm DNA trộn từ các mẫu được sử dụng làm khuôn để tạo thư viện cho NGS với cặp mỗi primer1 và primer2 bắt cặp bổ sung với barcode adapter và common adapter:

Primer1: (5'-3')
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCT
ACACGACGCTCTTCCGATCT

Primer2: (5'-3')
CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATT
CCTGCTGAACCGCTCTTCCGATCT

Sản phẩm PCR có kích thước từ 300-500 bp được thu nhận bằng phương pháp cắt gel và tinh sạch bằng kit QIAquick PCR Purification Kit. Kích thước và độ sạch của thư viện DNA sau đó được đánh giá trên Bioanalyzer 2100 (Agilent Technology). Thư viện DNA được làm giàu và xác định trình tự cả hai chiều (paired) trên hệ thống Illumina NextSeq® 500, theo hướng dẫn của nhà sản xuất sử dụng kit NextSeq 500/550 High Output v2 kit (300 cycles).

Xử lý dữ liệu GBS, lắp ráp hệ gen tham chiếu tạm thời và xác định SNPs

Dữ liệu GBS được xử lý và phân tích để phát hiện các SNPs dựa theo phương pháp GBS-SNP-CROP (GBS SNP-Calling Reference Optional Pipeline) được mô tả bởi Melo *et al.*, (2016). Dữ liệu thô được chuyển thành định dạng fastq bằng công cụ BCL2FASTQ 2.1. Chất lượng trình tự được đánh giá bằng FastQC. Trình tự adapter được loại bỏ bằng phần mềm Trimmomatic software v.0.3.6 (Bolger *et al.*, 2014). Trình tự của từng mẫu được tách ra trên cơ sở nhận biết trình tự barcode sử dụng công cụ in-house script do nhóm tác giả tự phát triển. Do hệ gen tôm sú chưa được công bố trình tự tham chiếu nên chúng tôi đã sử dụng các mẫu lắp ráp *de novo* để tạo trình tự tham chiếu tạm thời như mô tả bởi Melo *et al.*, (2016), sử dụng các công cụ PEAR v.0.96 và USEARCH v.8.0.162 (Edgar 2010; Zhang *et al.*, 2014), với độ sâu bao phủ (read depth) ≥ 10 (Melo *et al.*, 2016). Sau khi có trình tự tham chiếu, trình tự của các mẫu tôm sú được so sánh với trình tự tham chiếu bằng BWA-MEM v.0.7 (Li *et al.*, 2009a). Dữ liệu SNP được truy xuất bằng công cụ SAMtools v.1.2 (Li *et al.*, 2009b). SNP được xác định khi giống hệt các contig và sự sai khác nucleotide được phát hiện trên ít nhất bốn trình tự (read). Chỉ các SNPs hai allele đáp ứng mức độ lặp lại (read depth) như trên mới được xác định là SNP giả định (Kumar, 2014;

Nguyễn Minh Thành *et al.*, 2015; Melo *et al.*, 2016). Phương pháp sàng lọc để xác định được bộ dữ liệu SNPs có độ tin cậy cao đã được mô tả chi tiết trong công bố của Melo *et al.*, (2016).

Chú giải các đoạn trình tự và xác định gen liên quan

Công cụ BlastX E-value < 1e⁻⁶) được sử dụng để so sánh các contig với cơ sở dữ liệu NCBI non-redundant (Nr-NCBI) nhằm tìm kiếm sự tương đồng

giữa các contig cần phân tích với các trình tự protein trên Ngân hàng cơ sở dữ liệu NCBI (<http://www.ncbi.nlm.nih.gov/>) (Jung *et al.*, 2011). Kết quả tìm kiếm được truy xuất dưới dạng danh sách các gen chức năng với các thông tin về genID, protein được mã hóa và tên loài tương ứng. Danh sách này được sử dụng để rà soát tìm kiếm gen mã hóa protein liên quan đến tăng trưởng ở tôm sú bằng cách đối chiếu với 22 loại protein đã biết liên quan đến tăng trưởng trong họ giáp xác (Bảng 2) (Jung *et al.*, 2011).

Bảng 2. Danh sách 22 loại protein liên quan đến tính trạng tăng trưởng ở họ giáp xác (Jung *et al.*, (2013).

STT	Tên protein	STT	Tên protein
1	5-Hydroxytryptamine receptor (5-HT)	12	Translin-associated factor X (TRAX)
2	Alpha-amylase	13	Candidate growth genes effect on moulting
3	Cathepsin L	14	Actin
4	Cyclophilin (CyPs)	15	Crustacean hyperglycaemic hormone (CHH)
5	Fatty acid-binding protein (FANTs)	16	Eyestalk factors
6	Fibrillarin	17	Moult inhibiting hormone (MIH)
7	Glyceradehyde-3-phosphate dehydrogenase (GAPDH)	18	Candidate muscle build-up or degradation genes involved in moulting
8	Growth hormone and insulin-like growth factor	19	Methyl farnesoate (MF) and farneosoic acid O-methyltransferase (FAMeT)
9	Myostatin and growth differentiation factor 8/11	20	Heat shock proteins (HSPs)
10	Signal transducer and activator of transcription (STAT)	21	Myosin heavy chain
11	Secreted protein acidic and rich in cysteine (SPARC)	22	Ubiquitin

Xác định trình tự aminoacid của contig và đánh giá mức độ tương đồng với gen mã hóa protein quan tâm

Công cụ Expasy (<http://web.expasy.org>) được sử dụng để xác định trình tự aminoacid tương ứng với trình tự nucleotide của contig quan tâm bằng cách kiểm tra toàn bộ 6 khung đọc. Sáu khung đọc này tạo ra 6 kiểu trình tự aminoacid khác nhau đối với mỗi contig, tiếp tục sử dụng công cụ Uniprot (<http://www.uniprot.org>) để chọn ra kiểu trình tự amino acid có tỷ lệ tương đồng cao nhất so với trình tự amino acid của protein quan tâm.

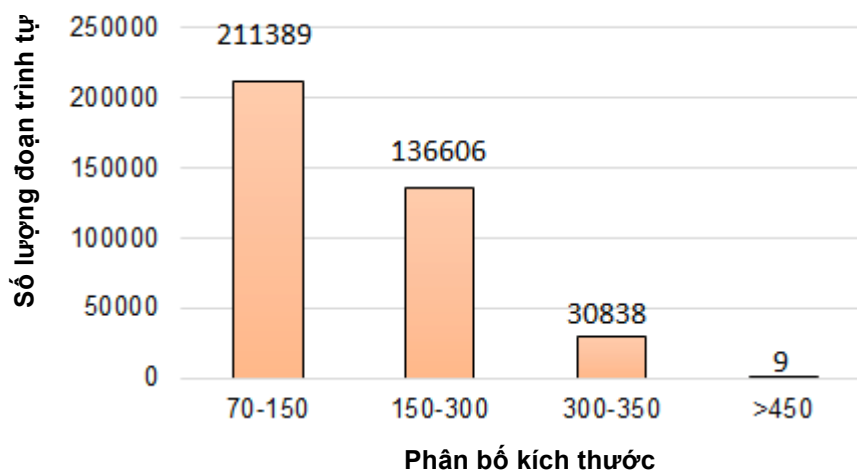
KẾT QUẢ VÀ THẢO LUẬN

Giải trình tự, kết nối các đoạn trình tự và thiết lập hệ gen tham chiếu tạm thời

Giải trình tự hệ gen tôm sú thuộc hai nhóm tôm tăng trưởng nhanh và tôm tăng trưởng chậm trên hệ thống Illumina NextSeq® 500, chúng tôi thu được 145.836.644 đoạn trình tự thô (raw read); trong số đó có 120.438.739 (chiếm 83%) đoạn trình tự mang barcode và điếm cắt của *ApeKI*. Sau khi tinh sạch loại bỏ các đoạn adapter, các đoạn trình tự chất lượng thấp, chúng tôi thu được 102.505.713 (70,2%) đoạn trình tự có chất lượng tốt để sử dụng cho việc kết nối contig và thiết lập hệ gen tham chiếu tạm thời, với dung lượng dữ liệu 100GB, Từ 102.505.713 đoạn trình tự sau khi kết nối thu được 510.076 constig với sự phân bố kích thước các contig được thể hiện ở hình 1. Các contig có chiều dài 70 - 150 bp chiếm số lượng lớn nhất (211.389), tiếp đến là các contig có chiều dài 150 - 300 bp và ít nhất là các contig có chiều dài 450 - 547 bp.

Bảng 3. Tóm tắt kết quả xử lý dữ liệu giải trình tự.

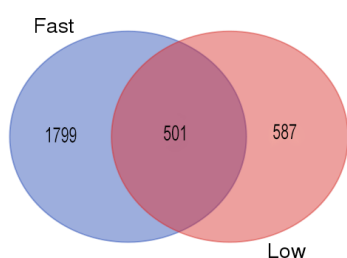
Chỉ số phân tích	Giá trị
Dung lượng dữ liệu (GB)	100
Số lượng đoạn trình tự (read) sử dụng kết nối contig	102.505.713
Tổng số contig	510.076
Kích thước hệ gen tham chiếu tạm thời (bp)	76.229.742



Hình 1. Sự phân bố theo kích thước của các contig sau kết nối.

Sàng lọc các SNP giả định

Chúng tôi đã phát hiện và sàng lọc được tổng cộng 2.887 SNPs, trong đó có 1.799 SNPs chỉ xuất hiện ở nhóm tôm tăng trưởng nhanh, 587 SNPs chỉ xuất hiện ở nhóm tôm tăng trưởng chậm và 501 SNPs xuất hiện ở cả hai nhóm (Hình 2).



Hình 2. Số lượng SNPs ở hai nhóm tôm lớn nhanh và lớn chậm.

Tổng số SNPs xác định được trong nghiên cứu này nhiều hơn so với một số nghiên cứu của các tác giả khác trên các đối tượng tôm như nghiên cứu của Kumar (2014) trên tôm sú *P. monodon* (422 SNPs), nghiên cứu của Du *et al.*, (2010)

trên tôm thẻ chân trắng *L. vannamei* (1.344 SNPs); tuy nhiên ít hơn so với nghiên cứu của Nguyễn Minh Thành *et al.*, (2015) khi nghiên cứu trên đối tượng cá tra (21.302 SNPs).

Kết quả sàng lọc SNPs trên hệ gen của cả hai nhóm tôm sú tăng trưởng nhanh và tăng trưởng chậm cho thấy số lượng SNPs chỉ xuất hiện ở nhóm tôm tăng trưởng nhanh (mà không xuất hiện ở nhóm tôm tăng trưởng chậm) là tương đối nhiều (1.799 SNPs). Điều này mở ra một hướng phân tích và khai thác dữ liệu rất có ý nghĩa thực tiễn, đó là tìm kiếm các chỉ thị SNPs có khả năng liên quan đến tính trạng tăng trưởng ở tôm sú. Vì hệ gen của đa số loài thủy sản nói chung và tôm sú nói riêng hiện nay chưa được giải mã hoàn toàn nên việc xác định được các SNPs liên kết với các gen chức năng sẽ mở ra nhiều tiềm năng ứng dụng đối với ngành nuôi trồng thủy sản. Theo Salem *et al.*, (2012), SNPs giải thích 90% sự khác biệt di truyền giữa các cá thể và quá trình trao đổi chéo trong phân bào giảm nhiễm rất hiếm khi tách rời chỉ thị SNPs khỏi gen chức năng khi SNPs được xác định nằm trên hoặc gần gen chức năng (Nguyễn Minh Thành *et al.*, 2015).

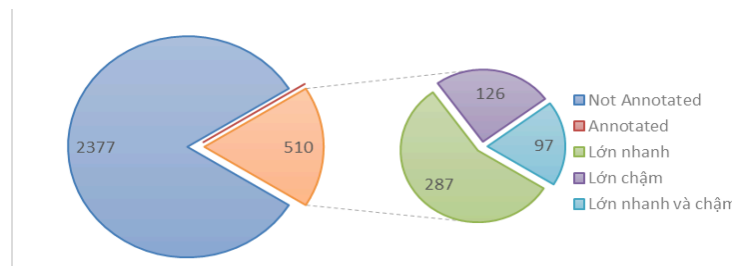
Chú giải gen chức năng từ dữ liệu giải trình tự gen tôm sú

Từ dữ liệu giải trình tự sau tinh sạch và dữ liệu sàng lọc các SNPs, chỉ những đoạn trình tự chứa SNPs được sử dụng để chú giải nhằm tìm kiếm sự tương đồng với các trình tự gen chức năng được lưu trữ trong GenBank (NCBI). Toàn bộ 2.887 đoạn trình tự chứa SNPs ở cả hai nhóm tôm sú lớn nhanh và lớn chậm được chú giải chức năng bằng công cụ BlastX (nr-NCBI).

Kết quả có 510 (17,67%) contig chứa SNPs có trình tự nucleotide tương đồng với các trình tự trên cơ sở dữ liệu nr-NCBI (với tham số E-value $1e^{-6}$); trong đó có 287 (56,27 %) trình tự contig chứa SNPs thuộc nhóm tôm sú lớn nhanh, 126 (24,71 %) trình tự contig chứa SNPs thuộc nhóm tôm sú lớn chậm và 97 (19,02 %) trình tự contig chứa SNPs thuộc cả hai nhóm (Hình 3).

Từ kết quả BlastX, chúng tôi thu được danh sách chú giải gồm các protein đã biết chức năng hoặc các

protein dự đoán (predicted proteins) cùng với tên loài tương ứng. Với mục tiêu nghiên cứu là tìm kiếm và xác định sự liên kết giữa các gen mã hóa các protein có liên quan đến tính trạng tăng trưởng ở tôm sú với các SNPs đã sàng lọc được, chúng tôi đã đối chiếu lần lượt 22 protein liên quan đến tính trạng tăng trưởng trong họ giáp xác (Jung *et al.*, 2013) với các protein đã chú giải được để rà soát và tìm ra loại protein tương ứng với trình tự contig chứa SNPs của hai nhóm tôm sú nghiên cứu. Chúng tôi đã xác định được hai contig (contig 83953 dài 98 bp và contig 260347 dài 80 bp) của nhóm tôm sú lớn nhanh có trình tự amino acid tương ứng tương đồng với trình tự amino acid của protein Myosin Heavy Chain (MHC). Trong đó, contig 83953 tương đồng với MHC type a (MHCa) và contig 260347 tương đồng với MHC type 1 (MHC1). Trong nghiên này, chúng tôi không tìm thấy contig nào của nhóm tôm sú lớn chậm có sự tương đồng với 22 protein liên quan đến tính trạng tăng trưởng trong họ giáp xác đã được công bố bởi Jung *et al.*, (2013).



Hình 3. Số lượng contig được chú giải chức năng

Bảng 5. Kết quả chú giải các gen liên quan tính trạng tăng trưởng ở tôm sú.

Contig được chú giải	Gen bank ID	Proteintương ứng	Loài tương ứng	Trình tự nucleotide (đối với contig83953 là trình tự bổ sung)
Contig83953 (98 bp)	gi 343183153 dbj BA K61429.1	Myosin heavy chain type a	<i>Marsupenaeus japonicus</i>	CAAGGTCGTCGATCT TCAGCTTCCATTTCAG AGATGATCTTATCGA AGTTCTTCTGTTTCT TCTCAGCGGAGTTG GCCAGAGTCTGTGC ACGTTTCAGCA
Contig260347 (80 bp)	gi 410509306 dbj BA M65719.1	Myosin heavy chain type 1	<i>Penaeus monodon</i>	CTCGTCTCGAGGAA GCCGAAATGCAGAT TGAGTCTCTCAATGT TAAGAAGTTCATTT GGAGAAGACCAAGA TGCGTGCG

Theo kết quả chú giải, gen mã hóa protein MHCa (gi|343183153|dbj|BAK61429.1) xuất hiện ở loài tôm *Marsupenaeus japonicus* và gen mã hóa protein MHC1 (gi|410509306|dbj|BAM65719.1) xuất hiện ở tôm sú *P. monodon*. Hai gen này đóng vai trò quan trọng trong quá trình hình thành phức hệ cơ. Chúng liên kết với actin tạo nên bó cơ dày và giúp thủy phân ATP. Ngoài ra đối với sự tích lũy các khối cơ ở giáp xác, mức độ biểu hiện của các gen myosin có thể coi như chỉ thị phân tử cho tiềm năng tăng trưởng của cá thể (Jung *et al.*, 2011; Jung *et al.*, 2013). Trình tự nucleotide của contig 83953 và contig 260347 được trình bày ở bảng 5.

Xác định chỉ thị SNPs và tương quan giữa các contig chứa SNPs với gen MHC

Phối hợp các kết quả sàng lọc SNPs và kết quả chú giải protein liên quan đến tính trạng tăng trưởng ở tôm sú (chỉ các contig chứa SNPs được chọn để

chú giải gen chức năng, sau đó so sánh trình tự của contig chứa SNPs với trình tự của gen chức năng tương ứng để xác định loại nucleotide thay thế), chúng tôi đã xác định được vị trí SNP trên contig83953 và contig260347 như sau: SNP T → C (Bảng 5) tại vị trí nucleotide thứ 20 trên mạch bổ sung với contig83953, tức là SNP A → G trên mạch gốc contig83953 và SNP G → A (Bảng 5) tại vị trí nucleotide thứ 19 trên contig260347. Ký pháp HGVS của hai SNP trên đây được thể hiện ở bảng 6.

Để xác định tương quan giữa các contig 83953 và contig 260347 chứa SNP với gen mã hóa protein MHC liên quan đến tăng trưởng ở tôm sú, chúng tôi đã tiến hành dịch mã các trình tự nucleotide của contig 83953 và contig 260347 thành các chuỗi amino acid tương ứng, sau đó so sánh với trình tự amino acid của gen mã hóa protein MHC. Kết quả dịch mã được trình bày ở bảng 7.

Bảng 6. Kết quả xác định SNP trên contig83953 và contig260347.

Tên Contig	Ký pháp HGVS của SNP tương ứng	Nucleotide thay thế	Nucleotide tham chiếu	Ghi chú
Contig83953	Contig83953:g.20T>C	C	T	Trình tự bổ sung
		G	A	Trình tự mạch gốc
Contig260347	Contig260347:g.19G>A	A	G	Mạch gốc

Bảng 7. Trình tự acid amin tương ứng của contig83953 và contig260347.

Tên contig	Trình tự amino acid tương ứng (http://web.expasy.org)	Tổng số amino acid của chuỗi
Contig83953	AERAQTLANSAEKKQKNFDKIISEWKLKIDDL	32
Contig260347	RLEEAEMETQIESLNVKNLHLEKTKMETRA	30

Để kiểm tra sự chính xác của các trình tự amino acid, chúng tôi đã tiến hành Blast (Basic Local Alignment Search Tool) bằng công cụ Uniprot (<http://www.uniprot.org>) với cả hai loại dữ liệu đầu vào là trình tự amino acid và trình tự nucleotide của

contig83953 và contig260347. Kết quả thu được từ hai loại dữ liệu đầu vào hoàn toàn trùng khớp nhau về tỷ lệ tương đồng giữa chuỗi amino acid của contig83953 và contig260347 so với protein MHC ở các loài khác nhau (Bảng 8 và Bảng 9).

Bảng 8. Các tỷ lệ tương đồng cao nhất (top 5) giữa contig83953 với protein MHC (<http://www.uniprot.org>).

STT	GenBank ID	Protein MHC (loài tương ứng)	Tỷ lệ tương đồng (%)	Vị trí amino acid trên protein MHC
1	F8WR03	Myosin heavy chain type a (<i>Penaeus japonicus</i>)	90,6	1431-1462
2	K4Q111	Myosin heavy chain type 1 (<i>Litopenaeus vannamei</i>)	87,5	1431-1462
3	K4Q4N8	Myosin heavy chain type 1 (<i>Penaeus monodon</i>)	87,5	1432-1463
4	Q6XGZ8	Slow muscle myosin S1 heavy chain (<i>Homarus americanus</i>)	83,9	30-60
5	B0W188	Myosin heavy chain (<i>Culex quinquefasciatus</i>)	80,6	1399-1429

Bảng 9. Các tỷ lệ tương đồng cao nhất (top 5) giữa contig260347 với protein MHC (<http://www.uniprot.org>).

STT	GeneBank ID	Protein MHC(loài tương ứng)	Tỷ lệ tương đồng (%)	Vị triaminoacid trên protein MHC
1	K4Q4N8	Myosin heavy chain type 1 (<i>Penaeus monodon</i>)	96,2	1396-1425
2	F8WR03	Myosin heavy chain type a (<i>Penaeus japonicus</i>)	96,2	1395-1420
3	K4Q111	Myosin heavy chain type 1 (<i>Litopenaeus vannamei</i>)	96,2	1395-1420
4	K4Q2Y1	Myosin heavy chain type 2 (<i>Penaeus monodon</i>)	76,0	1396-1418
5	K4Q2S1	Myosin heavy chain type 2 (<i>Litopenaeus vannamei</i>)	76,0	1396-1418

Kết quả ở bảng 8 cho thấy: trình tự aminoacid tương ứng của contig83953 đạt tỷ lệ tương đồng cao nhất (90,6%) với trình tự amino acid của protein MHCa ở loài tôm *P. japonicus*, tỷ lệ tương đồng với MHC1 ở hai loài tôm khác là *P. monodon* và *L. vannamei* cũng tương đối cao (đều đạt 87,5%). Trong khi đó, kết quả ở bảng 9 cho thấy: trình tự amino acid tương ứng của contig260347 đạt tỷ lệ tương đồng cao nhất (96,2%) với trình tự amino acid của protein MHC ở cả ba loài tôm *P. monodon*, *P. japonicus* và *L. vannamei*.

Như vậy, với việc sử dụng hệ gen tham chiếu tạm thời của tôm sú *P. monodon* được thiết lập *de novo* để sàng lọc SNPs, nghiên cứu của chúng tôi đã xác định được hai SNPs giả định ở nhóm tôm sú lớn nhanh.

Hiện nay chỉ có vài công bố về mối tương quan có ý nghĩa giữa SNPs với các tính trạng có giá trị kinh tế ở đối tượng thủy sản. Đó là SNPs xuất hiện ở gen amylase có liên quan đến tốc độ tăng trưởng của hàu *Crassostrea gigas* (Prudence *et al.*, 2006), SNPs xuất hiện ở gen parvalbumin ảnh hưởng đến tăng trưởng của cá chêm *Lates calcarifer* (Xu *et al.*, 2006). Nhóm nghiên cứu Zeng *et al.*, (2008) đã công bố SNPs ở gen Hsp70 có liên quan đến khả năng kháng bệnh của tôm thẻ chân trắng *L. vannamei*. Nghiên cứu của Nguyễn Minh Thành *et al.*, (2011) đã sàng lọc được SNPs xuất hiện ở các đoạn không mã hóa của gen CHH. Việc ứng dụng SNPs trong các chương trình chọn giống còn khá mới mẻ. Đa số các nghiên cứu sàng lọc chỉ thị SNPs được thực hiện đối với hệ phiên mã (transcriptome) (Jung *et al.*, 2011; Gao *et al.*, 2012; Nguyễn Minh Thành *et al.*, 2015)... Một số nghiên cứu sử dụng SNP để sàng lọc các gen tiềm năng liên quan đến tính trạng tăng trưởng ở một số đối tượng thủy sản như ở cá hồi *Salvelinus alpinus* (Tao and Boulding, 2003), tôm thẻ chân trắng *L. vannamei* và tôm sú *P. monodon* (Glenn *et al.*, 2005)...

Kết quả nghiên cứu của chúng tôi là phát hiện đầu tiên về chỉ thị SNPs liên quan đến gen MHC ở

tôm sú Việt Nam. Tuy nhiên, để có thể khẳng định được hai SNPs giả định này có phải là SNPs thực sự hay không thì cần phải tiến hành thêm các nghiên cứu sâu hơn nữa về lai giống và di truyền số lượng để kiểm tra sự di truyền của các SNPs, đồng thời cần sàng lọc SNPs trên quần đàn lớn hơn. Việc khẳng định sự tồn tại của các SNPs này sẽ chính xác hơn nếu có hệ gen tham chiếu đã được giải mã của tôm sú *P. monodon* trên GenBank. Việc phát hiện được SNPs liên kết với gen liên quan đến tính trạng tăng trưởng có ý nghĩa quan trọng và thiết thực đối với nghiên cứu và thực tiễn chọn giống dựa trên chỉ thị phân tử (MAS) nhằm làm tăng hiệu quả chọn giống.

KẾT LUẬN

Kỹ thuật GBS với công nghệ giải trình tự của Illumina sử dụng thiết bị NextSeq500 và các công cụ tin sinh học cho phép thiết lập *de novo* hệ gen tham chiếu tạm thời của tôm sú *P. monodon* có kích thước 76.229.742 bp. Bộ dữ liệu SNPs ở tôm sú đã được sàng lọc với tổng cộng 2.887 SNPs, trong đó có 1.799 SNPs chỉ xuất hiện ở nhóm tôm sú lớn nhanh, 587 SNPs chỉ xuất hiện ở nhóm tôm sú lớn chậm và 501 SNPs xuất hiện ở cả hai nhóm. Có 510 trình tự contig được chú giải thành công bao gồm 287 contig ở nhóm tôm sú lớn nhanh và 126 contig ở nhóm tôm sú lớn chậm. Đặc biệt chúng tôi đã xác định được hai contig chứa SNP có sự tương đồng cao với gen mã hóa protein myosin heavy chain (MHC) liên quan đến tính trạng tăng trưởng ở họ giáp xác, trong đó contig83953 tương đồng với MHCa và contig260347 tương đồng với MHC1. Hai SNPs đã được xác định là contig 83953:g.20T>C và contig260347:g.19G>A Đây là những kết quả bước đầu có ý nghĩa thiết thực và có tiềm năng ứng dụng thực tiễn đối với công tác chọn giống tôm sú Việt Nam nói riêng và các loài thủy sản nói chung dựa trên chỉ thị phân tử (MAS) nhằm làm tăng hiệu quả chọn giống.

Lời cảm ơn: Công trình này được thực hiện với sự tài trợ kinh phí của Bộ Khoa học và Công nghệ thông qua nhiệm vụ “Lập bản đồ gen tôm sú (*Penaeus monodon*)”.

TÀI LIỆU THAM KHẢO

- Beuzen ND, Stear MJ, Chang KC (2000) Molecular markers and their use in animal breeding. *Vet J* 160: 42-52.
- Beissinger TM, Hirsch CN, Sekhon RS, Foerster JM, Johnson JM, Muttoni G, Vaillancourt B, Buell CR, Kaeppler SM, de Leon N (2013) Marker density and read depth for genotyping populations using genotyping by sequencing. *Genetics* 193:1073-1081. Doi:10.1534/genetics.112.147710.
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15): 2114-2120.
- Bùi Chí Bửu, Nguyễn Thị Lang (2004) *Di truyền phân tử*. Nhà xuất bản Nông nghiệp, TP. HCM: 288-305.
- De-Santis C, Jerry DR (2007) Candidate growth genes in finfish - Where should we be looking? *Aquaculture* 272: 22-38.
- Desjardins P, Conklin D (2010) Nanodrop microvolume quantitation of nucleic acids. *J Vis Exp* 2010 Nov 22;(45). pii: 2565. doi: 10.3791/2565.
- Du ZQ, Ciobanu DC, Onteru SK, Gorbach D, Mileham AJ, Jaramillo G, Rothschild MF (2010) A gene-based SNP linkage map for pacific white shrimp, *Litopenaeus vannamei*. *Anim Genet* 41(3):286-94. doi: 10.1111/j.1365-2052.2009.02002.x.
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19): 2460-2461.
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:e19379. Doi:10.1371/journal.pone.0019379.
- Gao Z, Luo W, Liu H, Zeng C, Liu X, Yi S, Wang W (2012) Transcriptome analysis and SSR/SNP markers information of the blunt snout bream (*Megalobrama amblycephala*). *PLoS ONE* 7: e42637.
- Glenn KL, Grapes L, Suwanasopee T, Harris DL, Li Y, Wilson K, Rothschild MF (2005) SNP analysis of AMY2 and CTSL genes in *Litopenaeus vannamei* and *Penaeus monodon* shrimp. *Animal Genetics* 36: 235-236.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol* 29: 644-652.
- Hou R, Bao Z, Wang S, Su H, Li Y, Du H, Hu J, Wang S, Hu X (2011) Transcriptome sequencing and de novo analysis for Yesso Scallop (*Patinopecten yessoensis*) using 454 GS FLX. *PLoS ONE* 6: e21560.
- Huang XD, Zhao M, Liu WG, Guan YY, Shi Y, Wang Q, Wu SZ, He MX (2013) Gigabase-scale transcriptome analysis on four species of Pearl oysters. *Marine Biotechnology* 15:253-264.
- Jehan T, Lakhnpaul S (2006) Single nucleotide polymorphism (SNP) - methods and applications in plant genetics: A review. *Indian J Biotechnol*.5: 435-459.
- Jung H, Lyons RE, Dinh H, Hurwood DA, McWilliam S *et al.* (2011) Transcriptomics of a Giant Freshwater Prawn (*Macrobrachium rosenbergii*): De Novo Assembly, Annotation and Marker Discovery. *PLoS ONE* 6(12): e27938. doi:10.1371/journal.pone.0027938.
- Koyama H, Akolkar DB, Shiokai T, Nakaya M, Piyapattanakorn S, Watabe S (2012) The occurrence of two types of fast skeletal myosin heavy chains from abdominal muscle of kuruma shrimp *Marsupenaeus japonicus* and their different tissue distribution. *J Exp Biol* 215: 14-21.
- Kumar KV (2014) SNP markers in growth-related candidate genes of Black tiger shrimp and their significance. *Abstract, 2nd International Conference on Animal & Dairy Sciences*. September 15-17, 2014, HICC, Hyderabad, India.
- Li H, Durbin R (2009a) Fast and accurate short read alignment with BurrowsWheeler Transform. *Bioinformatics* 25: 1754-1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009b) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16): 2078-2079.
- Liu ZJ, Cordes JF (2004) DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* 238: 1-37.
- Liu Z (2007) Single nucleotide polymorphism (SNP). In: Liu, Z. (Ed.), *Aquaculture Genome Technologies*. Blackwell, USA, pp. 59-72.
- Liu S, Zhang Y, Zhou Z, Waldbieser G, Sun F, Lu J, Zhang J, Jiang Y, Zhang H, Wang X, Rajendran KV, Khoo L, Kucuktas H, Peatman E, Liu Z (2013) Efficient assembly and annotation of the transcriptome of catfish by RNA-Seq analysis of a doubled haploid homozygote. *BMC Genomics* 13:595.
- Melo TO, Radhika Bartaula, Iago Hale (2016) GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable

- length, paired-end genotyping-bysequencing data. *BMC Bioinformatics* 17:29.doi: 10.1186/s12859-016-0879-y.
- Prudence M, Moal J, Boudry P, Daniel JY, Quere C, Jeffroy F, Mingant C, Ropert M, Bedier E, Wormhoudt AV, Samain JF, HuvetA (2006) An amylase gene polymorphism is associated with growth differences in the Pacific cupped oyster *Crassostrea gigas*. *Anim Genet* 37: 348-351.
- Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, Mungall K, Lee S, Okada HM, Qian JQ, Griffith M, Raymond A, Thiessen N, Cezard T, Butterfield YS, Newsome R, Chan SK, She R, Varhol R, Kamoh B, Prabhu AL, Tam A, Zhao Y, Moore RA, Hirst M, Marra MA, Jones SJ, Hoodless PA, Birol I (2010) De novo assembly and analysis of RNA-seq data. *Nat Method* 7: 909-12.
- Tao WJ, Boulding EG (2003) Associations between single nucleotide polymorphisms in candidate genes and growth rate in Arctic charr (*Salvelinus alpinus* L.). *Heredity* 91: 60-69.
- Nguyễn Thị Minh Thanh, Nguyễn Hải Triều, Phạm Thị Hoa, Nguyễn Thị Hoa, Hà Thị Thu, Đậu Huy Tùng, Vũ Thị Hiền, Nguyễn Hữu Ninh, Nguyễn Thị Tuyết Nhung, Đồng Văn Quyền, Đinh Duy Khang (2015) Đánh giá tính đa hình AFLP của các quần đàn tôm sú (*Penaeus monodon*) thu nhận từ các vùng biển của Việt Nam. *Tạp chí Công nghệ Sinh học* 13(1): 31-38.
- Nguyễn Đức Thành (2014) Các kỹ thuật chỉ thị DNA trong nghiên cứu và chọn lọc thực vật. *Tạp chí Sinh học* 36(3): 265-294.
- Nguyễn Minh Thành, Andrew CB, Peter BM, Yutao L, Russell EL (2011) Mối tương quan giữa SNP (single nucleotide polymorphisms) của gen CHH (crustacean hyperglycemic hormone) và tính trạng tăng trưởng của tôm càng xanh, *Macrobrachium rosenbergii*. *Kỷ yếu Hội nghị Khoa học thủy sản toàn quốc*, 16/12/2011, Đại học Nông lâm TP. Hồ Chí Minh, 49-58.
- Nguyễn Minh Thành, Võ Thị Minh Thu, Jung H, Mather P (2015) Phân tích hệ gen chức năng từ mô thận cá tra (*Pangasianodon hypophthalmus*) nuôi ở điều kiện mặn: lắp ráp, chú giải, phân tích chỉ thị SNP. *Tạp chí Sinh học* 37(2): 220-227.
- Vaseeharan B, Rajakamaran P, Jayaseelan D, Vincent AY (2013) Molecular markers and their application in genetic diversity of penaeid shrimp. *Aquacult Int* 21: 219-241.
- Xu YX, Zhu ZY, Lo LC, Wang CM, Lin G, Feng F, Yue GH (2006) Characterization of two parvalbumin genes and their association with growth traits in Asian seabass (*Lates calcarifer*). *Anim Genet* 37: 266-268.
- Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30(5): 614-620.
- Zhou Y, Gao F, Liu R, Feng J, Li H (2012) De novo sequencing and analysis of root transcriptome using 454 pyrosequencing to discover putative genes associated with drought tolerance in *Ammopiptanthus mongolicus*. *BMC Genomics* 13: 266.
- Zeng D, Chen X, Li Y, Peng M, Ma N, Jiang W, Yang C, Li M (2008) Analysis of Hsp70 in *Litopenaeus vannamei* and detection of SNPs. *J Crustacean Biol* 28: 727-730.
- <http://thuysanvietnam.com.vn>
- <http://www.ncbi.nlm.nih.gov/>
- <http://web.expasy.org>
- <http://www.uniprot.org>

APPLICATION OF GENOTYPING BY SEQUENCING (GBS) FOR SCREENING SINGLE NUCLEOTIDE POLYMORPHISM (SNPs) ASSOCIATED WITH BLACK TIGER SHRIMP (*PENAEUS MONODON*) GROWTH TRAIT

Nguyen Thị Minh Thanh¹, Nguyen Quyet Tam², Nguyen Van Hao², Nguyen Van Sang², Nguyen Dang Ton³, Ma Thi Huyen Thuong³, Kim Thi Phuong Oanh³, Nong Van Hai³, Nguyen Thi Hoa¹, Ha Thi Thu¹, Vu Thi Hien¹, Nguyen Dinh Duy¹, Tran Xuan Thach¹, Nguyen Thi Tuyet Nhung¹, Nguyen Huu Ninh², Dong Van Quyen¹, Dinh Duy Khang¹

¹Institute of Biotechnology, Vietnam Academy of Science and Technology

²Research Aquaculture No.2

³Institute of Genome Research, Vietnam Academy of Science and Technology

⁴ Research Aquaculture No.3

SUMMARY

Nowadays, the black tiger shrimp (*Penaeus monodon*) farming in Vietnam still have to face a lot of problems, especially in the Country's Shrimp Broodstock Production. To overcome these problems, we need a strategy for development of domesticated stocks of *P. monodon* and rational breeding programs for improved

growth and disease resistance. Marker-assisted selection is a procedure that has been developed to avoid these problems associated with phenotypic selection, replacing the selection of the phenotype by selection using markers linked with quantitative trait loci (QTL). Application of next generation sequencing (NGS) associated with QTL to discover important molecular markers has been applied in many laboratories. In this study, we used genome typing by sequencing (GBS) with NGS for screening the SNPs related with growth trait to create a database for father study on the black tiger shrimp fast growth selection. Based on the putative genome sequences created by *de novo* assembly, 2887 SNPs have been screened, among them, 1799 SNPs appeared only in the fast growth population. Among the 287 annotated contigs, the two contigs were homologous with myosin heavy chain (MHC) associated with growth trait in crustacean species, contig83953 and contig260347 were homologous with MHCa and MHC1, respectively. The two SNPs Contig83953:g.20T>C and Contig260347:g.19G>A were indicated. We hope that, these results will contribute to the black tiger shrimp genome database for further study to exploit the molecular markers for black tiger shrimp breeding and selections.

Keywords: *GBS, Penaeus monodon, single nucleotide polymorphisms (SNPs), growth trait, Next generation sequencing (NGS)*