

## XÁC ĐỊNH CẤU TRÚC VÀ ĐẶC ĐIỂM GEN HỌC HỆ GEN TY THỂ CỦA SÁN LÁ RUỘT NHỎ *HAPLORCHIS TAICHUI* (TREMATODA: HETEROPHYIDAE), MẪU VIỆT NAM

Lê Thanh Hòa<sup>1</sup>, Nguyễn Thị Khuê<sup>1</sup>, Nguyễn Thị Bích Nga<sup>1</sup>, Đỗ Thị Roan<sup>1</sup>, Đỗ Trung Dũng<sup>2</sup>, Lê Thị Kim Xuyên<sup>1</sup>, Đoàn Thị Thanh Hương<sup>1</sup>

<sup>1</sup>Viện Công nghệ sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

<sup>2</sup>Viện Sốt rét - Ký sinh trùng - Côn trùng Trung ương, Bộ Y tế

Ngày nhận bài: 03.4.2016

Ngày nhận đăng: 15.6.2016

### TÓM TẮT

Sán lá ruột nhỏ *Haplorchis taichui* (Nishigori, 1924) thuộc họ Heterophyidae, lớp Trematoda, ngành Platyhelminthes, kí sinh và gây bệnh ở người và động vật. Toàn bộ hệ gen ty thể (mitochondrial DNA genome, mtDNA) của *H. taichui*, chủng QT (HTAQT, Quảng Trị) đã được xác định cấu trúc và đặc điểm gen học, dữ liệu có giá trị phục vụ nghiên cứu dịch tễ học phân tử, thành phần loài, chẩn đoán, phân loại và phòng chống bệnh. Hệ gen mtDNA chủng HTAQT có độ dài 15.119 bp, chứa 36 gen, trong đó có 12 gen mã hóa protein (*cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4L*, *nad4*, *nad5*, *nad6*, *atp6* và *cob*); 2 gen RNA ribosome (rRNA) gồm *rrnL* (16S) và *rrnS* (12S); 22 gen vận chuyển RNA (tRNA hay *trn*); và một vùng không mã hóa (non-coding, NR), chia thành 2 tiểu vùng là *không mã hóa ngắn* (SNR) và *không mã hóa dài* (LNR). Vùng LNR, độ dài 1.692 bp, nằm giữa vị trí của *trnG* (vận chuyển Glycine) và *trnE* (Glutamic acid), chứa 6 cấu trúc lặp liên kế tiếp nhau (tandem repeat, TR), sắp xếp lần lượt là: TR1A, TR2A, TR1B, TR2B, TR3A, TR3B. Từng gen mã hóa protein (12 gen), rRNA ribosome (2 gen) và tRNA vận chuyển (22 gen) đã được phân tích, cụ thể gen mã hóa protein được xác định kích thước, cách sử dụng bộ mã khởi đầu và kết thúc; các gen rRNA và tRNA được xác định cấu trúc bậc hai.

**Từ khóa:** *H. taichui*, hệ gen ty thể, mtDNA, sán lá ruột nhỏ.

### MỞ ĐẦU

Bệnh sán lá ruột nhỏ (haplorchiasis) ở người chủ yếu do loài sán lá ruột nhỏ *Haplorchis taichui* và *H. pumilio* gây nên, được phát hiện phổ biến ở các nước châu Á, một phần châu Phi và Trung Nam Mỹ (Chai *et al.*, 2009; 2010; Dung *et al.*, 2007). Tại Việt Nam, do tập quán ăn gỏi cá nên tỷ lệ người mắc bệnh kí sinh trùng truyền qua thực phẩm rất cao, trong đó có bệnh do *Haplorchis* spp gây ra. Sán lá ruột nhỏ *H. taichui* và *H. pumilio* thuộc giống *Haplorchis* trong họ Heterophyidae, *H. taichui* được Nishigori phát hiện lần đầu tiên vào năm 1924, ký sinh trên người ở Philippin. Cho đến nay, *H. taichui*, *H. pumilio* được ghi nhận là lưu hành rộng rãi ở nhiều quốc gia trên thế giới trong đó có Việt Nam (Dung *et al.*, 2007; Chai *et al.*, 2012; De, Le, 2011; Nguyễn Thị Khuê *et al.*, 2015). Tại Việt Nam, haplorchiasis do *H. taichui* và *H. pumilio* được phát hiện ở hầu hết các tỉnh thành trong cả nước, đặc biệt ở các tỉnh phía Bắc và miền Trung (Dung *et al.*,

2007; Kim Văn Vạn *et al.*, 2007; Nguyễn Thị Khuê *et al.*, 2015).

Hệ gen ty thể (mitochondrial genome) của sán dẹt (platyhelminth) có cấu trúc là một vòng DNA hai sợi khép kín, kích thước khoảng 13.5- - 21 kb, chứa một hệ gen riêng gồm 12 gen mã hóa cho sản phẩm protein, 2 gen RNA ribosome (rRNA), 22 gen RNA vận chuyển (tRNA hoặc *trn*) và một phần không mã hóa (non-coding region, NR) dài ngắn khác nhau (Le *et al.*, 2002). Nghiên cứu giải mã hệ gen ty thể và phân tích đặc điểm của từng gen riêng biệt, đang được quan tâm và thực hiện trên toàn thế giới, nhằm mục đích để giải đáp các câu hỏi về chức năng, tiến hoá, nguồn gốc, phá hệ và di truyền quần thể (Bernt *et al.*, 2013). Chính vì vậy, đã có hàng chục hệ gen ty thể của sán lá đã được giải mã thành công và được ứng dụng hiệu quả trong nghiên cứu bệnh động vật lây sang người (Le *et al.*, 2002; Jia *et al.*, 2012).

Nhằm cung cấp dữ liệu cấu trúc hệ gen ty thể và

đặc điểm gen học các loài sán lá ruột nhỏ đang kí sinh và gây bệnh trên người và động vật chúng tôi tiến hành giải trình tự và phân tích đặc điểm gen của một số kí sinh trùng thường gặp tại Việt Nam, thuộc các đề tài NAFOSTED (giai đoạn 2010- - 2016). Trong bài báo này, hệ gen ty thể của loài sán lá ruột nhỏ *H. taichui*, mẫu thu thập tại Quảng Trị, Việt Nam (kí hiệu HTAQT) đã được thu nhận và phân tích đầy đủ cung cấp dữ liệu cho các hướng nghiên cứu khác nhau sử dụng chuỗi gen/chỉ thị phân tử có từ hệ gen ty thể.

## NGUYÊN LIỆU VÀ PHƯƠNG PHÁP

### Mẫu *Haplorchis taichui* trong nghiên cứu và tách chiết DNA tổng số

Mẫu sán lá ruột nhỏ *H. taichui* trưởng thành thu từ một bệnh nhân tại Quảng Trị (Việt Nam), ký hiệu HTAQT, được sử dụng để giải trình tự hệ gen ty thể. Mẫu đã được xác định loài dựa vào đặc điểm hình thái học (De, Le, 2011) và khẳng định bằng sinh học phân tử sử dụng chỉ thị *cox1* và phân tích phá hệ xác định quan hệ về loài (Nguyễn Thị Khuê *et al.*, 2015).

DNA tổng số được tách chiết bằng bộ sinh phẩm bộ sinh phẩm DNeasy Blood and Tissue kit (hãng Qiagen) theo hướng dẫn của nhà sản xuất, như đã sử dụng trước đây (Nguyễn Thị Khuê *et al.*, 2015). DNA tổng số của mẫu được bảo quản ở -20°C và khuôn DNA tổng số được pha ở nồng độ 50 ng/μl, sử dụng cho phản ứng PCR (1-2 μl/phản ứng).

**Bảng 1.** Danh sách mồi sử dụng trong PCR và giải trình tự toàn bộ hệ gen ty thể của sán lá ruột nhỏ *Haplorchis taichui* (mẫu Việt Nam).

Tên mồi	Chuỗi mồi (5' -> 3')	Vị trí
PNAD1F	GATTATGGTGAGTCTGAAAGTGAG	<i>nad1</i>
JNAD1R	CTCACTTTCAGACTCACCATAATC	<i>nad1</i>
TRECOBF	CAGATGTCYTATTGGGCTGC	<i>cob</i>
TRECOBR	GAACHRVCCACARYCCCTTAAAC	<i>cob</i>
URNLF	AGCCAGGTTGGTCTTATCTAT	<i>rrnL</i>
URNSR	TACCWTGTTACGACTTAHCWCA	<i>rrnS</i>
JB3F	TTTTTTGGGCATCCTGAGGTTTAT	<i>cox1</i>
JB4.5R	TAAAGAAAGAACATAATGAAAATG	<i>cox1</i>
UNI16F	TGGCCGCAGTATHTTGACTGTGC	<i>rrnL</i>
UNI16R	TCTCGGGGTCTTCCGTCT	<i>rrnL</i>
UNI12F	GGTACAYACCGCCCGTC	<i>rrnS</i>
UNI12R	GACGGGCGGTRTGACC	<i>rrnS</i>
GLYF	AGTATKYGTCTTTCCAAGTC	<i>trnG</i>
GLYR	ACKAGACCHCYGACTTGGAAAGAC	<i>trnG</i>
HTA1F	TCTTATTTACTATCGGGGGGGTGAC	<i>cox1</i>
HTA2R	TCAAACAACACTCCCCAACAGAC	<i>cox1</i>
HTA3F	GACTTGGGGGATTATGCGTFACTG	<i>rrnS</i>
HTA4R	TGTAGACCGCTCACGAGAAACTTC	<i>rrnS</i>
HTA5F	GGTCGTTTTTGGGGTTTGCG	<i>nad1</i>
HTA6F	GTGAATAAGGGTTATGTCGAG	<i>nad3</i>
HTA6R	CTCGACATAACCCTTATTCAC	<i>nad3</i>
HTAF	CGTGTGTCATAAGAGATGGTCCG	<i>rrnS</i>
HTAR	CCCCAACTAAATCCCCCTTC	<i>rrnS</i>
Hsp9F	ATGATGTGGTGGCTACTCG	<i>nad4</i>
HTA12R	GCCCCATGTGAACCAAGAAC	<i>cob</i>
HTA13R	CTATACCCACCATAACCACTC	<i>cox3</i>
HTA14F	GGGACCTGATGGGCTTATCG	<i>nad5</i>
HTA15R	CGACCCACATAACGGGACAA	<i>nad5</i>
HTA16F	ACCTCGTGGTAACCTTCGTCG	<i>rrnS</i>
HTA17R	CACCCTCACCCCAATAACC	<i>nad5</i>
HTA18F	CCTTTTATTGGGGGAGGGGCT	<i>trnE</i>

## Thiết kế mồi và thực hiện PCR

Trước tiên, một số cặp mồi cơ bản được thiết kế dựa vào so sánh trình tự của các gen chọn lọc lấy từ các hệ gen ty thể của các loài sán dẹt có trong Ngân hàng gen (Le *et al.*, 2002); để thực hiện PCR dài (long, LPCR) thu nhận đoạn DNA hệ gen ty thể kích thước 3-8 kb và PCR đơn thuần thu PCR sản phẩm dưới 3 kb. Bảng 1 liệt kê tất cả mồi chung (14 mồi đầu tiên trong danh sách) và mồi đặc hiệu (17 mồi còn lại) được thiết kế và sử dụng. Sau đó trên cơ sở trình tự nucleotide của các chuỗi đã thu nhận tiếp tục thiết kế mồi và thực hiện giải trình tự bằng phương pháp “lao mồi” (primer-walking); hoặc/và sau khi tách dòng. PCR được thực hiện với chu trình nhiệt gồm các bước 94°C/5 phút trong 1 chu kỳ; tiếp theo là 35 chu kỳ ở 94°C/1 phút, 50°C/1 phút và 72°C/1-4 phút; chu kỳ cuối kéo dài 10 phút ở 72°C. Đối với LPCR, thực hiện bước kéo dài ở 72°C/6--10 phút. Sản phẩm PCR được tinh sạch bằng QIAquick PCR Purification Kit (QIAGEN) và được dòng hoá vào vector pCR2.1 hoặc pCR2.1-TOPO TA-cloning Kit (Invitrogen), chuyển nạp vào dòng tế bào IVN $\alpha$ F’ để chọn lọc khuẩn lạc. DNA của plasmid tái tổ hợp được tách chiết với bộ hoá chất QIAprep Spin Plasmid Extraction Kit (QIAGEN).

## Giải trình tự và phân tích chuỗi gen

Trình tự nucleotide từ giải trình tự trực tiếp sản phẩm PCR hoặc từ DNA plasmid tái tổ hợp trên máy tự động ABI Avant 3100 Genetic Analyzer (Mỹ), được xử lý bằng chương trình SeqEd1.3; sau đó so sánh sử dụng chương trình AssemblyLIGN1.9 và MacVector8.2 (Accelrys Inc.). Xác định cấu trúc gen bằng phần mềm MacVector8.2 và GENEDOC2.7 (<http://iubio.bio.indiana.edu/soft/molbio/ibmpc/genedoc-readme.html>), xác định cấu trúc RNA vận chuyển bằng chương trình tRNAscan-SE1.21 (<http://lowelab.ucsc.edu/tRNAscan-SE/>) hoặc và chương trình ARWEN (<http://mbio-serv2.mbioekol.lu.se/ARWEN/>), cấu trúc của RNA ribosome sử dụng chương trình RNAviz (<http://rnviz.sourceforge.net/>). Trình tự amino acid từ trình tự nucleotide các gen mã hóa protein của hệ gen ty thể được suy diễn sử dụng bộ mã của sán dẹt (flatworm mitochondrial code) có trong chương trình GENEDOC2.7.

## KẾT QUẢ VÀ THẢO LUẬN

### Cấu trúc và sắp xếp hệ gen ty thể của sán lá ruột nhỏ *H. taichui*

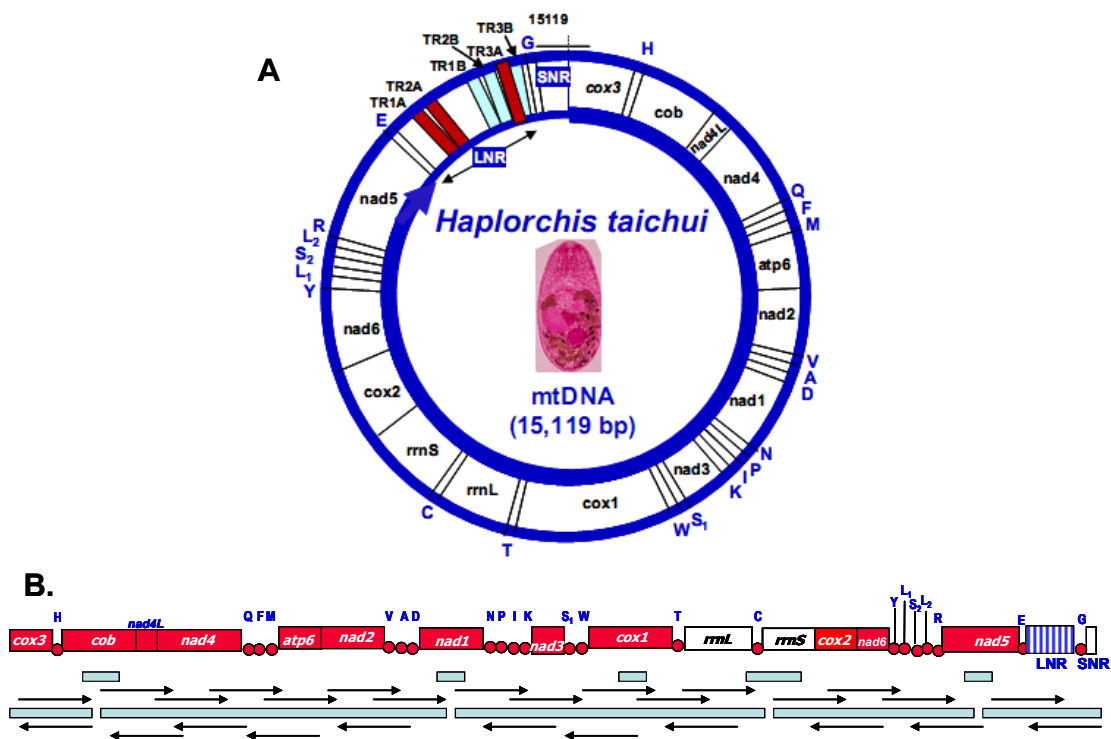
Toàn bộ chuỗi nucleotide hệ gen ty thể hoàn chỉnh (mtDNA) của sán lá ruột nhỏ *H. taichui* mẫu Việt Nam (HTAQT) có độ dài 15.119 bp, đã được thu nhận. Phân tích đặc điểm toàn bộ hệ gen và của từng gen cho thấy, mtDNA *H. taichui* có cấu trúc là vòng DNA khép kín chứa 36 gen, gồm 12 gen mã hóa protein, 2 gen RNA ribosome, 22 gen RNA vận chuyển amino acid và một vùng DNA không mã hóa. Sắp xếp thứ tự các gen được liệt kê ở Bảng 2, sắp xếp dạng vòng tròn (circular) trình bày ở Hình 1A và dạng thẳng (linear gene arrangement) ở Hình 1B. Tất cả các gen ty thể đều sắp xếp cùng chiều xuôi và gần như kế tiếp nhau, chỉ cách nhau rất gần với khoảng cách giao gen (là chuỗi nucleotide nối giữa gen trước với gen liền kề sau đó) là 0-25 nucleotide (Bảng 2). Một số gen sử dụng 1-3 nucleotide của nhau, lồng vào nhau, để kết thúc hoặc khởi đầu gen (Bảng 2). Gen *atp8* không có trong cấu trúc mtDNA của *H. taichui* và đó cũng là đặc trưng về cấu trúc của mtDNA ở các loài sán lá (trematode) và của tất cả các loài sán dẹt (platyhelminth) (Le *et al.*, 2002).

Thứ tự sắp xếp dạng thẳng (linear gene arrangement) của 36 gen và vùng không mã hóa (LNR và SNR), được mở vòng tại *cox3*, như sau: *cox3-H-cob-nad4L-nad4-Q-F-M-atp6-nad2-V-A-D-nad1-N-P-I-K-nad3-S<sub>1</sub>-W-cox1-T-rrnL-C-rrnS-cox2-nad6-Y-L<sub>1</sub>-S<sub>2</sub>-L<sub>2</sub>-R-nad5-E-LNR[TR1A-3A;TR1B-3B]-G-SNR*. Vùng DNA không mã hóa (NR) được chia thành 2 tiểu vùng: một tiểu vùng *không mã hóa dài*, LNR (long non-coding region) và một tiểu vùng *không mã hóa ngắn*, SNR (short). Vị trí, kích thước, đặc điểm của từng gen và vùng không mã hóa trong hệ gen ty thể của loài sán lá ruột nhỏ *H. taichui* HTAQT được liệt kê trong Bảng 1 và minh họa ở Hình 1. Hai gen *nad4L* (264 bp) và *nad4* (1.278 bp) sắp xếp kế tiếp lồng vào nhau trên hai khung gen khác nhau, trong đó gen *nad4* gối vào đầu 3’ của gen *nad4L* và chung nhau 40 nucleotide. Tương tự, gen *cox2* (624 bp) và gen *nad6* (459 bp) có 14 nucleotide ở đầu 3’ kết thúc của *cox2* gối vào đầu 5’ khởi đầu của *nad6* (Bảng 2), một trường hợp hiếm gặp ở sán dẹt được biết cho đến nay.

*H. taichui* sử dụng A = 19,56%, T = 39,71%, G =

28,34%, C = 12,39%, để kiến tạo hệ gen ty thể (15.119 bp) với tổng thể thành phần A+T là 59,27% và G+C là 40,73%. Tổng thể thành phần A+T ở mtDNA của *H. taichui* cũng tương tự như ở một số loài sán lá gan lớn

(fasciolid) và sán lá gan nhỏ (opisthorchiid), vào khoảng 60%, nhưng khác xa và ít hơn nhiều so với các loài sán máng họ Schistosomatidae (thường là khoảng 70% A+T) (Le *et al.*, 2002).



**Hình 1.** Cấu trúc dạng vòng tròn (A), sắp xếp gen của hệ gen ty thể dạng thẳng (B, dãy trên) và sơ đồ giải trình tự toàn bộ hệ gen ty thể loài sán lá ruột nhỏ *Haplorchis taichui*, mẫu Việt Nam. Các mũi tên chỉ chiều giải trình tự xuôi và ngược; các vạch là sản phẩm PCR ngắn tạo cơ sở cung cấp chuỗi nucleotide thiết kế mồi và PCR dài cung cấp DNA để giải trình tự. Ảnh hình thái *H. taichui* trưởng thành được trình bày ở chính giữa. Gen được sắp xếp theo chiều kim đồng hồ bắt đầu là *cox3* và phiên mã một chiều (thể hiện bởi mũi tên tròn ở bên trong hình). Gen mã hóa protein và tiểu đơn vị ribosome lớn và nhỏ (*rrnL* và *rrnS*) được viết tắt bằng danh pháp quốc tế về hệ gen ty thể (Boore, 1999; Le *et al.*, 2002). Vị trí các gen vận chuyển amnio acid (tRNA hoặc *trn*) được viết tắt bằng một chữ cái mã hóa cho từng loại amino acid tương ứng và cho hai amino acid Serine, S<sub>1</sub> và S<sub>2</sub>; và Leucine, L<sub>1</sub> và L<sub>2</sub> (xem Bảng 2). LNR là vùng không mã hóa dài (long non-coding region), nằm giữa *trnE* và *trnG*, bên trong có các cấu trúc lặp TR1A-3A và TR1B-3B; và SNR, vùng không mã hóa ngắn (short non-coding region), định vị ở giữa hai gen *trnC* và *cox3*.

**Phân tích đặc điểm gen RNA ribosome và RNA vận chuyển**

RNA ribosome (gồm hai tiểu phần, *rrnL* và *rrnS*) và RNA vận chuyển (gồm 22 tRNA) lần lượt, có độ dài, cấu trúc, thứ tự sắp xếp cơ bản là giống nhau giữa các loài sán

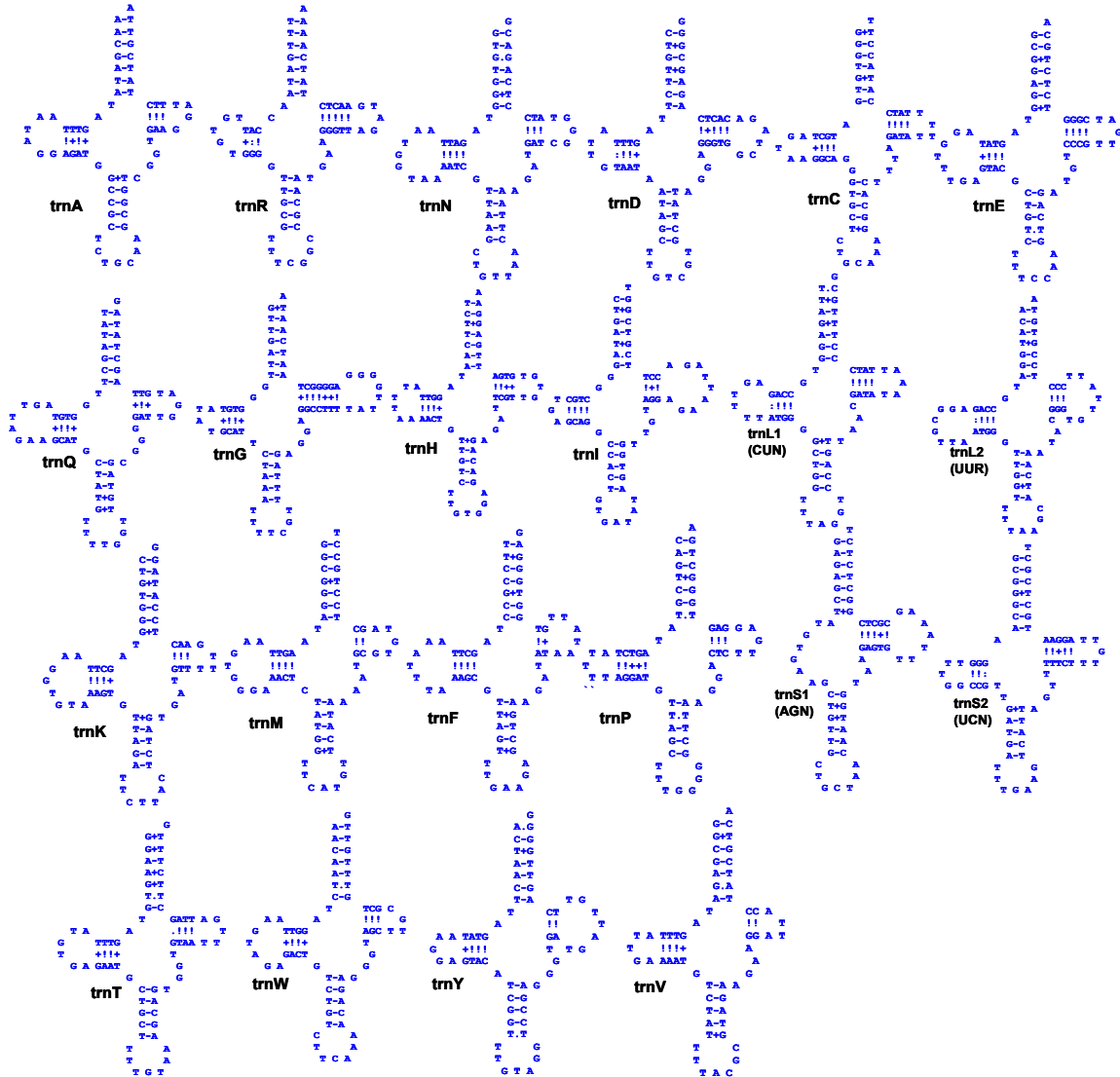
dẹt, nằm xen giữa các gen mã hóa protein (Hình 1B). Hai gen RNA ribosome, *rrnL* và *rrnS*, cách nhau bằng một chuỗi gen của tRNA-Cysteine (*trnC*) xen vào giữa (Bảng 1). Cấu trúc bậc hai suy diễn hình “lá sồi” của tRNA ở *H. taichui* tương tự của các loài sán dẹt và giống như ở sán lá gan lớn *F. hepatica* (Le *et al.*, 2002).

**Bảng 2.** Xác định vị trí và đặc điểm của từng gen (mã hoá protein, RNA vận chuyển và RNA ribosome) và vùng không mã hoá trong hệ gen ty thể của sán lá ruột nhỏ *H. taichui* (HTAQT, mẫu thu thập tại Việt Nam).

<i>Haplorchis taichui</i> mtDNA (HTQT, 15.119 bp)							
<i>Gen</i>	Vị trí (5'>3')	Đặc điểm gen		Bộ mã		ĐỐI mã tRNA (anticodon)	Độ dài vùng giao gen (bp)
		bp	aa	KĐ	KT		
<i>cox3</i>	1-645	645	212	ATG	TAG		+2
<i>trnH</i>	648-713	66				GTG	+6
<i>cob</i>	720-1829	1110	369	ATG	TAG		+1
<i>nad4L</i>	1831-2094	264	88	ATG	TAG		-40
<i>nad4</i>	2055-3335	1281	426	GTG	TAA		+9
<i>trnQ</i>	3345-3410	66				TTG	+3
<i>trnF</i>	3414-3477	64				GAA	-1
<i>trnM</i>	3477-3540	64				CAT	0
<i>atp6</i>	3541-4056	516	171	ATG	TAG		+25
<i>nad2</i>	4083-4951	870	289	ATG	TAA		+4
<i>trnV</i>	4955-5015	60				TAC	+1
<i>trnA</i>	5017-5079	63				TGC	+2
<i>trnD</i>	5082-5148	67				GTC	0
<i>nad1</i>	5149-6054	906	301	GTG	TAG		+2
<i>trnN</i>	6057-6122	66				GTT	+3
<i>trnP</i>	6126-6190	65				TGG	-4
<i>trnI</i>	6187-6251	65				GAT	+3
<i>trnK</i>	6255-6319	65				CTT	0
<i>nad3</i>	6320-6678	366	121	GTG	TAG		+13
<i>trnS1</i> (AGN)	6692-6753	63				GCT	+7
<i>trnW</i>	6761-6824	64				TCA	+4
<i>cox1</i>	6829-8370	1542	513	ATG	TAG		-1
<i>trnT</i>	8370-8433	64				TGT	0
<i>rrnL</i> (16S)	8434-9412	979					0
<i>trnC</i>	9413-9476	64					0
<i>rrnS</i> (12S)	9477-10225	749					0
<i>cox2</i>	10226-10839	624	307	ATG	TAG		-14
<i>nad6</i>	10836-11294	459	152	ATG	TAG		+1
<i>trnY</i>	11294-11358	63				GTA	0
<i>trnL1</i> (CUN)	11359-11424	66				TAG	-2
<i>trnS2</i> (UCN)	11423-11486	64				TGA	+8
<i>trnL2</i> (UUN)	11495-11562	68				TAA	+11
<i>trnR</i>	11574-11638	64				TCG	+1
<i>nad5</i>	11640-13226	1587	528	GTG	TAA		+11
<i>trnE</i>	13238-13310	73				TTC	0
LNR	13318-15009	1692					0
1. TR1A	13557-13738	182					0
2. TR2A	13739-13920	182					0
interTR-NR	13921-14407	487					0
3. TR1B	14408-14594	187					0
4. TR2B	14595-14777	183					0
5. TR3A	14731-14913	182					0
6. TR3B	14778-14960	183					+42
<i>trnG</i>	15003-15070	68				TCC	0
SNR	15071-15119	49					0

**Ghi chú:** bp: base pair; aa: amino acid; KĐ: khởi đầu; KT: Kết thúc. Độ dài chuỗi giao gen (bp) (đánh dấu +, với gen tiếp theo; -, lồng vào gen trước; RNA vận chuyển tRNA (transfer RNA). LNR: Long non-coding region; SNR: Short non-coding region. TRA: Chuỗi lặp liền kề (Short Tandem Repeat) thuộc type A; TRB: thuộc type B; inter TR-NR: vùng không mã hóa (non-coding region) nằm giữa các nhóm cấu trúc lặp.

**Haplorchis taichui mtDNA (tRNAs)**



**Hình 2.** Cấu trúc bậc hai suy diễn của 22 RNA vận chuyển trong hệ gen ty thể của loài *Haplorchis taichui* sắp xếp theo thứ tự kí hiệu chữ cái (ABC). Trật tự sắp xếp RNA vận chuyển trình bày ở Hình 1 và Bảng 2. Ngoại trừ cấu trúc tRNA của Serine 1 (*trnS<sub>1</sub>*) bị khuyết cánh tay dihydrouridine (DHU-arm) thường gặp ở hệ gen ty thể của một số loài sán dẹt (Le *et al.*, 2002; Lê Thanh Hòa *et al.*, 2013), tất cả 21 tRNA còn lại đều có cấu trúc đầy đủ.

Độ dài của tRNA ở *H. taichui* biến động từ 60 bp (ở tRNA-Valine) đến 73 bp (ở tRNA-Glutamic acid), phần lớn tập trung vào khoảng 62- - 64 bp (Bảng 1). Vùng tiếp nhận amino acid (amino-acyl acceptor stem) chứa 7 bp (ít khi 6 bp), thường thấy ở các loài sinh vật

nhân thật; vùng “*biến đổi*” (variable loop) nổi vùng “*đôi mã*” (anticodon) và vùng tay TΨC (TΨC-stems), chỉ chứa 4 bp (ít khi 5 hoặc 6 bp) (Hình 2). Về cấu trúc, ngoại trừ tRNA của Serine 1 (*trnS<sub>1</sub>*), hầu như tất cả tRNA vận chuyển (21 gen) của mtDNA *H. taichui*

đều có cấu trúc bậc hai suy diễn hình “lá sồi” (clover-leaf) hoàn toàn đặc trưng của sản dẹt, kể cả tRNA của Serine 2 (*trnS*<sub>2</sub>), của Leucine (*trnL*<sub>1</sub>; *trnL*<sub>2</sub>) và của Cystein (*trnC*). Chỉ duy nhất cấu trúc của *trnS*<sub>1</sub> của *H. taichui* bị khuyết cánh tay dihydrouridine (DHU-arm), tạo nên một vòm nucleotide, thường gặp ở hệ gen ty thể của một số loài sản dẹt (Le *et al.*, 2002; Lê Thanh Hòa *et al.*, 2013) (Hình 2).

### Phân tích đặc điểm gen mã hóa protein

Bảng mã di truyền được sử dụng tương tự như ở các loài sản dẹt (trematode) được nghiên cứu cho đến nay (Le *et al.*, 2002; 2004; Lee *et al.*, 2013). Tất cả 64 bộ mã đều được sử dụng cho mã hóa gen protein của *H. taichui*, trong đó bộ mã ATG (sử dụng cho 8 gen) và GTG (cho 4 gen) làm bộ mã khởi đầu; và TAA và TAG làm các bộ mã kết thúc gen mã hóa protein. Bộ mã TGA, thông thường là bộ mã kết thúc ở các loài sinh vật nhân thật (metazoan), nhưng ở sản dẹt nói chung và sản lá ruột nhỏ *H. taichui* trong nghiên cứu này, mã hóa cho amino acid tryptophan (W) (Le *et al.*, 2002).

Không có bất kỳ trường hợp nào có bộ mã kết thúc khiếm khuyết (T hay TA) để kết thúc một gen như thường thấy ở hệ gen ty thể ở nhiều loài giun tròn (nematode) hoặc sinh vật nhân thật (metazoan) (Le *et al.*, 2004). Một số bộ mã bất thường khác, như ATA hoặc ATT sử dụng làm bộ mã khởi đầu, đã có ghi nhận ở một số loài sinh vật nhân thật; và TTG hoặc GTT ở một số loài giun tròn, cũng không được tìm thấy ở loài sản lá *H. taichui* đang nghiên cứu ở đây (Bảng 2).

*H. taichui* sử dụng 10.164 bp cho mã hóa cho 3.376 amino acid kiến tạo 12 gen mã hóa protein (36 nucleotide cho 12 bộ mã kết thúc). *H. taichui* mtDNA sử dụng nhiều nhất, đến 301 bộ mã cho Phenylalanine (TTT-Phe), 224 cho Leucine (TTG-Leu) và 171 cho Valine (GTT-Val); và sử dụng ít nhất, chỉ có 5 bộ mã lần lượt cho Glutamine (CAA-Gln) và Arginine (CGC-Arg) và 6 cho Asparagine (AAC-Asn). Những bộ mã sử dụng nhiều nhất thuộc về Phenylalanine và Leucine, những bộ mã ít sử dụng nhất thường là Arginine và Glutamine (Chen *et al.*, 2013).

### Đặc điểm vùng không mã hóa protein

Vùng không mã hóa protein (non-coding region) ở mtDNA của sản lá ruột nhỏ *H. taichui*, được chia làm hai tiểu vùng: một tiểu vùng không mã hóa dài, LNR

(long non-coding region), có độ dài 1692 bp và một tiểu vùng không mã hóa ngắn, SNR, chỉ có 49 bp, có thể coi như là một vùng giao gen (Bảng 2). Vùng LNR nằm giữa gen tRNA-Glutamic acid (*trnE*) và tRNA-Glycine (*trnG*), chứa các cấu trúc “vòm” (loop) và “keo tóc” (hairpin) và đặc biệt, có hai nhóm cấu trúc lặp liên kề (tandem repeat, TR), chứa 3 cấu trúc giống nhau ở mỗi nhóm (Hình 1). Nhóm cấu trúc liên kề thứ nhất kí hiệu là nhóm A (type A, TR1A-TR2A-TR3A) có trình tự 182 nucleotide, đồng nhất rất cao giữa các chuỗi. Nhóm cấu trúc liên kề thứ hai kí hiệu là nhóm B (type B, TR1B-TR2B-TR3B) có trình tự 183- - 187 nucleotide, mức độ đồng nhất thấp hơn, trong đó TR3B và TR3A cùng chung nhau đến 136 bp ở mỗi chuỗi. Vùng không mã hóa có cấu trúc lặp liên kề là nét đặc trưng ở mtDNA của một số loài sản lá, như bắt gặp trong mtDNA của sản lá gan lớn *F. hepatica* và sản máng *Schistosoma* spp, tạo nên cấu trúc đa hình mang tính di truyền quần thể (Le *et al.*, 2002).

### So sánh đặc điểm hệ gen ty thể giữa các chủng *H. taichui*

Hệ gen ty thể của *H. taichui* thu thập tại Việt Nam trong nghiên cứu này được so sánh với trình tự và cấu trúc của toàn bộ hệ gen ty thể của một mẫu *H. taichui* mới công bố gần đây của các tác giả Hàn Quốc, được đăng kí Ngân hàng gen số KF214770 (mẫu thu từ Lào) (Lee *et al.*, 2013). Độ dài của mtDNA *H. taichui* (mẫu Lào) là 15.130 bp, có trật tự sắp xếp gen và số lượng gen giống với hệ gen ty thể của *H. taichui* mẫu Việt Nam. Các gen mã hóa protein, gen rRNA ribosome và tRNA có độ dài và đặc điểm giống với các gen tương tự của mẫu Việt Nam. Tuy nhiên, về cấu trúc lặp liên kề, ở mẫu Việt Nam có đến 6 TR (TR1A-TR2A-TR3A; và TR1B-TR2B-TR3B) đại diện đa hình (polymorphism) của hệ gen ty thể sản lá, đã được phân tích đầy đủ, nhưng ở mẫu của Lào (Lee *et al.*, 2013) không thấy được nhắc đến. Về cấu trúc tRNA, ở mẫu *H. taichui* của Việt Nam chỉ có cấu trúc của *trnS*<sub>1</sub>, trong khi đó cả hai tRNA (*trnS*<sub>1</sub> và *trnS*<sub>2</sub>) không có hình “lá sồi” đặc trưng mà hoàn toàn bị khuyết cánh tay dihydrouridine (DHU-arm) ở mẫu Lào trong nghiên cứu của Lee *et al* (2013).

Do tầm quan trọng của hệ gen ty thể trong giám định, chẩn đoán, phân loại, dịch tễ học phân tử và nghiên cứu di truyền quần thể, nên càng ngày càng có nhiều hệ gen ty thể của các loài từ bậc thấp đến bậc

cao được giải mã và phân tích (O'Brien *et al.*, 2009). Sán dẹt trong đó có loài sán lá ruột nhỏ *H. taichui*, được quan tâm hàng đầu do chúng kí sinh và gây bệnh ở động vật và người, kể cả *Clonorchis sinensis* và *Opisthorchis viverrini* gây ung thư túi mật tiên phát ở người (Sripa *et al.*, 2012). Hệ gen ty thể trình bày trong bài báo này của *H. taichui* có trật tự gen (gene order), cấu trúc gen (gene structure) gồm 12 gen mã hóa protein, 2 gen rRNA và 22 gen tRNA và những vùng gen không mã hóa. Ngoài những đặc điểm chung giống như một số loài sán dẹt trước đây đã được giải mã và phân tích, loài *H. taichui* (mẫu Việt Nam) có một số nét khác biệt đặc trưng. Ngoài bộ mã ATG, GTG cũng được sử dụng để khởi đầu và ngoài TAG, TAA dùng để kết thúc gen mã hóa protein, là đặc điểm chung về sử dụng bộ mã ở mtDNA của các loài sán dẹt được biết cho đến nay. Vùng không mã hóa phân chia thành 2 tiểu vùng, bao gồm vùng ngắn (SNR) và vùng dài (LNR), chứa 6 chuỗi lặp liên kề (TR1A-3A; và TR1B-3B). Ngoài trừ một vài loài (sán lá gan nhỏ, *O. viverrini*, *O. felineus*, *C. sinensis*) không có cấu trúc lặp liên kề (Shekhovtsov *et al.*, 2010; Cai *et al.*, 2012), còn trong hệ gen ty thể của tất cả hơn 20 loài sán dẹt (trematode) khác đều tìm thấy chuỗi lặp TR đặc biệt này, có độ dài và số lượng thay đổi tùy loài. Tại sao có cấu trúc lặp liên kề ở vùng không mã hóa, vai trò của chúng trong tiến hóa, trong di truyền, trong chức năng của loài, vẫn là câu hỏi cần giải đáp (Bernt *et al.*, 2013; Hammani *et al.*, 2014). RNA vận chuyển (tRNA) trong mtDNA của loài sán lá ruột nhỏ *H. taichui* mới thu nhận này cũng như của các loài sán dẹt trước đây, có kích thước ngắn hơn tRNA của hệ gen nhân, có cấu trúc hình “lá sồi” đặc trưng. Bên cạnh đó, chỉ có tRNA-Serine (*trnS*<sub>1</sub>) có cấu trúc khuyết cánh tay DHU (DHU-arm), số lượng ít hơn so với số lượng tRNA có cấu trúc không bình thường này (2 -3 tRNA) ở các loài sán dẹt khác.

## KẾT LUẬN

Hệ gen ty thể của sán lá ruột nhỏ *H. taichui* mẫu của Việt Nam (chủng HTAQT) đã được giải mã và phân tích đầy đủ đặc điểm gen học. Toàn bộ mtDNA của HTAQT có kích thước 15.119 bp, chứa 36 gen trong đó có 12 gen mã hóa protein (*cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4L*, *nad4*, *nad5*, *nad6*, *atp6* và *cob*); 2 gen RNA ribosome, *rrnL* (16S) và *rrnS* (12S); 22 gen vận chuyển RNA (tRNA hay *trn*); và một vùng

không mã hóa có các cấu trúc lặp liên kề (6 cấu trúc lặp). Đặc điểm của từng gen về độ dài, sử dụng bộ mã và phân bố, cũng như cấu trúc bậc hai của rRNA và tRNA cũng đã được xác định và phân tích đầy đủ và so sánh với mẫu của Lào do nhóm nghiên cứu Hàn Quốc phân tích.

**Lời cảm ơn:** Cảm ơn Quỹ Phát triển Khoa học và Công nghệ quốc gia (NAFOSTED) đã hỗ trợ kinh phí cho đề tài mã số 106.06-2012.05 (LTH) và Phòng Thí nghiệm trọng điểm Công nghệ gen (Viện Công nghệ sinh học) về trang thiết bị, để thực hiện nghiên cứu này.

## TÀI LIỆU THAM KHẢO

- Bernt M, Brabant A, Schierwater B, Stadler PF (2013) Genetic aspects of mitochondrial genome evolution. *Mol Phylogenet Evol* 69(2): 328-338.
- Boore JL (1999) Animal mitochondrial genomes. *Nucleic Acids Res* 27: 1767-1780.
- Cai XQ, Liu GH, Song HQ, Wu CY, Zou FC, Yan HK, Yuan ZG, Lin RQ, Zhu XQ (2012) Sequences and gene organization of the mitochondrial genomes of the liver flukes *Opisthorchis viverrini* and *Clonorchis sinensis* (Trematoda). *Parasitol Res* 110(1): 235-243.
- Chai JY, De NV, Sohn WM (2012) Foodborne Trematode Metacercariae in Fish from Northern Vietnam and Their Adults Recovered from Experimental Hamsters. *Korean J Parasitol* 50(4): 317-325.
- Chai JY, Yong TS, Eom KS, Min DY, Shin EH, Banoung V, Insisiengmay B, Insisiengmay S, Phommasack B, Rim HJ (2010) Prevalence of the intestinal flukes *Haplorchis taichui* and *H. yokogawai* in a mountainous area of Phongsaly Province, Lao PDR. *Korean J Parasitol* 48(4): 339-342.
- Chai JY, Shin EH, Lee SH, Rim HJ (2009) Foodborne intestinal flukes in Southeast Asia. *Korean J Parasitol* 47: Supplement: S69-S102.
- Chen L, Yang DY, Liu TF, Nong X, Huang X, Xie Y, Fu Y, Zheng WP, Zhang RH, Wu XH, Gu XB, Wang SX, Peng XR, Yang GY (2013) Synonymous codon usage patterns in different parasitic platyhelminth mitochondrial genomes. *Genet Mol Res* 12(1): 587-596.
- De NV, Le TH (2011) Human infections of fish-borne trematodes in Vietnam: prevalence and molecular specific identification at an endemic commune in Nam Dinh province. *Exp Parasitol* 129(4): 355-361.



- Dung DT, De NV, Waikagul J, Dalsgaard A, Chai JY, Sohn WM, Murrell KD (2007) Fishborne zoonotic intestinal trematodes, Vietnam. *Emerg Infect Dis* 13(12): 1828-1833.
- Hammani K, Bonnard G, Bouchoucha A, Gobert A, Pinker F, Salinas T, Giegé P (2014) Helical repeats modular proteins are major players for organelle gene expression. *Biochimie* 100: 141-150.
- Jia W, Yan H, Ni X, Lou Z, Li H, Cao P, Cai X (2012) Advances in the study of helminth mitochondrial genomes and their associated applications. *Chinese Sci Bull* 57(1): 54-67.
- Kim Văn Vạn, Lê Thanh Hòa, Nguyễn Văn Đê, Nguyễn Thị Bích Nga, Nguyễn Thị Tuyết Nhung, Anders Dalgaard (2007) Phân biệt sản lá ruột nhỏ *Haplorchis taichui* và *H. pumilio* với các loài sản lá khác sử dụng chỉ thị ITS-2 (internal transcribed spacer). *Tap chí Khoa học và Phát triển* 1(5): 36-42.
- Le TH, Blair D, McManus DP (2002) Mitochondrial genomes of parasitic flatworms. *Trends Parasitol* 18: 206--213.
- Le TH, Blair D, McManus DP (2004) Codon usage and bias in mitochondrial genome of platyhelminths. *Korean J Parasitol* 42(4): 159-167.
- Lê Thanh Hòa, Nguyễn Thị Bích Nga, Vũ Thị Tiên, Nguyễn Văn Đê (2013). Phân tích gen học hệ gen ty thể của hai loài sản lá gan nhỏ *Clonorchis sinensis* và *Opisthorchis viverrini*, kí sinh và gây bệnh ở người và động vật tại Việt Nam. Hội nghị khoa học Công nghệ sinh học toàn quốc 2013. *Nhà XB Tự nhiên và Công nghệ*, 2013: 97-101.
- Lee D, Choe S, Park H, Jeon HK, Chai JY, Sohn WM, Yong TS, Min DY, Rim HJ, Eom KS (2013) Complete mitochondrial genome of *Haplorchis taichui* and comparative analysis with other trematodes. *Korean J Parasitol* 51(6): 719-726.
- Nguyễn Thị Khuê, Đỗ Trung Dũng, Lê Thanh Hòa (2015) Phân tích tương đồng gen ty thể *cox1* và thẩm định loài sản lá ruột nhỏ *Haplorchis taichui* và *H. pumilio* (Platyhelminthes: Trematoda: Heterophyidae) thu thập ở các tỉnh phía bắc Việt Nam. Báo cáo Khoa học toàn văn Hội nghị Kí sinh trùng học toàn quốc lần thứ 42 (Nghệ An, 2-3/4/2015). *Nhà xuất bản Khoa học tự nhiên và Công nghệ*, 2015: 64-71.
- O'Brien EA, Zhang Y, Yang L, Wang E, Marie V, Lang BF, Burger G (2009) GOBASE: an organelle genome database. *Nucleic Acids Res* 37(Database issue): D946-D950.
- Shekhovtsov SV, Katokhin AV, Kolchanov NA, Mordvinov VA (2010) The complete mitochondrial genomes of the liver flukes *Opisthorchis felineus* and *Clonorchis sinensis* (Trematoda). *Parasitol Int* 59(1): 100-103.
- Sripa B, Brindley PJ, Mulvenna J, Laha T, Smout MJ, Mairiang E, Bethony JM, Loukas A (2012) The tumorigenic liver fluke *Opisthorchis viverrini*-multiple pathways to cancer. *Trends Parasitol* 28(10): 395-407.

## GENETIC CHARACTERIZATION OF MITOCHONDRIAL GENOME OF THE SMALL INTESTINAL FLUKE, *HAPLORCHIS TAICHUI* (TREMATODA: HETEROPHYIDAE), VIETNAMESE SAMPLE

Thanh Hoa Le<sup>1,✉</sup>, Nguyen Thi Khue<sup>1</sup>, Nguyen Thi Bich Nga<sup>1</sup>, Do Thi Roan<sup>1</sup>, Do Trung Dung<sup>2</sup>, Le Thi Kim Xuyen<sup>1</sup>, Doan Thi Thanh Huong<sup>1</sup>

<sup>1</sup>*Institute of Biotechnology, Vietnam Academy of Science and Technology*

<sup>2</sup>*National Institute of Malariaology, Entomology and Parasitology, Ministry of Health*

### SUMMARY

The small intestinal fluke, *Haplorchis taichui* Nishigori, 1924, belonging to genus *Haplorchis* (family Heterophyidae, class Trematoda, phylum Platyhelminthes), is a zoonotic pathogen causing disease in humans and animals. Complete mitochondrial genome (mtDNA) of *H. taichui* (strain HTAQT, collected from Quang Tri) was obtained and characterized for structural genomics providing valuable data for studies on epidemiology, species identification, diagnosis, classification, molecular phylogenetic relationships and prevention of the disease. The entire nucleotide mtDNA sequence of *H. taichui* (HTAQT) is 15,119 bp in length, containing 36 genes, including 12 protein-coding genes (*cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4L*, *nad4*, *nad5*, *nad6*, *atp6* and *cob*); 2 ribosomal

✉ Author for correspondence: E-mail: [imibtvn@gmail.com](mailto:imibtvn@gmail.com)

RNA genes, *rrnL* (16S) and *rrnS* (12S); 22 transfer RNA genes (tRNA or *trn*), and a non-coding region (NR), divided into two sub-regions of short non-coding (short, SNR) and long non-coding (long, LNR). LNR region, 1.692 bp in length, located between the position of *trnG* (transfer RNA-Glycine) and *trnE* (Glutamic acid), contains 6 tandem repeats (TR), arranged as TR1A, TR2A, TR1B, TR2B, TR3A, TR3B, respectively. Each protein coding gene (overall, 12 genes), ribosomal rRNA (2 genes) and tRNA (22 genes) were analyzed, in particular, protein-coding genes were defined in length, start and stop codons, and rRNA and tRNA genes for secondary structure.

**Keywords:** *H. taichui*, mitochondrial genome, mtDNA, small intestinal fluke.