

SO SÁNH ĐẶC ĐIỂM HỆ GEN TY THỂ CỦA SÁN LÁ RUỘT NHỎ *HAPLORCHIS TAICHUI* VỚI *METAGONIMUS YOKOGAWAI* VÀ ĐƠN VỊ MÃ HÓA RIBOSOME VỚI *H. PUMILIO* (HỌ HETEROPHYIDAE)

Lê Thị Việt Hà¹, Nguyễn Thị Khuê², Đồng Văn Quyền^{2,3}, Lê Thanh Hòa^{2,3,✉}

¹Học viện Y – Dược học Cổ truyền Việt Nam, Bộ Y tế

²Viện Công nghệ sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

³Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

✉Người chịu trách nhiệm liên lạc. E-mail: imibtvn@gmail.com

Ngày nhận bài: 26.01.2020

Ngày nhận đăng: 10.4.2020

TÓM TẮT

Sán lá ruột nhỏ *Haplorchis taichui* và *H. pumilio* thuộc họ Heterophyidae (Trematoda: Platyhelminthes), được nghiên cứu còn rất hạn chế, đặc biệt là chỉ thị phân tử hệ gen ty thể và đơn vị mã hóa ribosome. Chúng tôi thu nhận toàn bộ hệ gen ty thể (mtDNA) của loài *H. taichui* và toàn bộ phần mã hóa của đơn vị mã hóa ribosome (rTU hay rDNA) loài *H. taichui* và *H. pumilio* của Việt Nam. Dữ liệu nucleotide và amino acid được so sánh giữa *H. taichui* và *Metagonimus yokogawai* về các đặc điểm thành phần kiến tạo gen/hệ gen, đặc điểm sử dụng bộ mã hoặc nucleotide (độ lệch skew) và các cấu trúc lặp liên kề (TRU). Hệ gen ty thể của chúng Htai-QT3-VN có độ dài 15.120 bp và *M. yokogawai* (15.258 bp, Hàn Quốc, KC330755) chứa 36 gen, gồm 12 gen mã hóa protein (*cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4L*, *nad4*, *nad5*, *nad6*, *atp6* và *cob*), 2 gen RNA ribosome (rRNA), 22 gen RNA vận chuyển (tRNA hay *trn*) và một vùng không mã hóa (NCR) giữa *trnE* và *trnG*, chia thành 2 tiểu vùng chứa 5 cấu trúc lặp (182-183 bp/cấu trúc). *H. taichui* (Việt Nam và Lào) sử dụng A = 19,56%, T = 39,71%, G = 28,34%, C = 12,39% (A+T là 59,27% và G+C là 40,73%) cho kiến tạo mtDNA, có giá trị độ lệch (skew/skewness) ở A+T là âm (-0,340) và G+C là dương (0,392); cho 12 gen mã hóa protein (PCG) tương tự; nhưng cho gen ribosome ty thể (MRG, gồm 16S/*rrnL* và 12S/*rrnS*) với A+T ít hơn (57,22%) và với G+C nhiều hơn (42,78%). *M. yokogawai* có tỷ lệ sử dụng A+T thấp hơn (mtDNA/55,68%, PCGs/55,96%, MRGs/54,15%) và G+C cao hơn so với loài *H. taichui*. *H. taichui* của Việt Nam và Lào có 10.164 bp mã hóa cho 3.376 amino acid để kiến tạo 12 PCG với những bộ mã sử dụng nhiều nhất là Phenylalanine (Phe-TTT) và Leucine (Leu-TTG), những bộ mã ít nhất là Glutamine (Gln-CAA), Arginine (Arg-CGC), có thêm Thr-ACA/ACC ở *M. yokogawai*. Phần mã hóa đơn vị mã hóa ribosome (từ 5' 18S đến 3' 28S) của *H. taichui* (7.268 bp) và *H. pumilio* (7.416 bp) được xác định có 5 vùng gen gồm 18S rDNA, ITS1, 5.8S rDNA, ITS2 và 28S rDNA. Các gen 18S và 5.8S của cả 2 loài có độ dài như nhau (1.992 bp/18S, 160 bp/5.8S), gen 28S khác nhau (3.875 bp/*H. taichui* và 3.870 bp/*H. pumilio*). ITS1 ở *H. taichui* (797 bp) và ITS2 ở *H. pumilio* (280 bp) không chứa cấu trúc lặp, trong khi đó ITS1 ở *H. pumilio* (1.106 bp) chứa 5 cấu trúc lặp liên kề TRU gồm 136 bp cho 3 TRU và 116 bp cho 2 TRU và ITS2 ở *H. taichui* (444 bp) chứa 3 TRU (83–85 bp/cấu trúc).

Từ khóa: Đơn vị mã hóa ribosome, *Haplorchis pumilio*, *Haplorchis taichui*, hệ gen ty thể, mtDNA, *Metagonimus yokogawai*, rTU, sán lá ruột nhỏ

ĐẶT VẤN ĐỀ

Họ Heterophyidae Odhner, 1914 thuộc liên

họ Opisthorchioidea Looss, 1899 (Digenea) bao gồm một tập hợp các loài sán lá ruột nhỏ (minute intestinal flukes) có tầm quan trọng về y tế và thú

y với sự phân bố rộng và với hơn 60 loài phổ biến ở hàng chục nước trên thế giới (Chai, Jung, 2017; 2020; Chai, 2019; Santos, Borges, 2020). Hầu hết những người nhiễm bệnh sống ở các nước Châu Á, bao gồm Hàn Quốc, Trung Quốc, Đài Loan, Việt Nam, Lào, Thái Lan, Malaysia, Indonesia, Philippin và Ấn Độ. Thành viên họ sán lá Heterophyidae sử dụng ốc nước ngọt làm vật chủ trung gian thứ nhất, cá làm vật chủ trung gian thứ hai và người/động vật là vật chủ cuối cùng (Chai *et al.*, 2009; Chai, Jung, 2020). Có tầm quan trọng dịch tễ là các loài sán lá ruột nhỏ (heterophyid) trong các chi *Haplorchis*, *Metagonimus*, *Stellantchasmus*, *Procerovum* và *Centrocestus* (Chai *et al.*, 2009; Chai, 2019). Trong số đó, các loài ký sinh và gây bệnh trên người được công nhận có ảnh hưởng lớn nhất là *Haplorchis taichui*, *H. pumilio*, *H. yokogawai*, *Stellantchasmus falcatus*, *Procerovum varium*, *Metagonimus yokogawai* và *Centrocestus formosanus*, phổ biến ở các nước ở Đông Á bao gồm Trung Quốc, Philippin, Hàn Quốc, Đài Loan, Thái Lan, Lào, Campuchia và Việt Nam (Dung *et al.*, 2007; Van *et al.*, 2009; Thaenkham *et al.*, 2011; Chai *et al.*, 2012; Le *et al.*, 2017; Chai, Jung, 2020).

Haplorchis taichui Nishigori, 1924, cùng với *H. pumilio* Looss, 1899, là hai loài gây nhiễm và gây bệnh sán lá ruột nhỏ (haplorchiasis) ở người, được phát hiện phổ biến ở các nước Châu Á, một phần châu Phi và Trung Nam Mỹ (Dung *et al.*, 2007; Chai *et al.*, 2009; Chai, 2019). *Metagonimus yokogawai* là loài có sự phân bố gây bệnh toàn cầu, phổ biến ở Trung Quốc, Châu Âu, Ấn Độ, Nhật Bản, Hàn Quốc, Đài Loan, Nga (Chai, Jung, 2017; 2020). Phân biệt về hình thái học của sán lá ruột nhỏ trưởng thành cũng như các dạng sinh trưởng và phát triển của chúng (trứng (eggs), ấu trùng lông (redia), ấu trùng đuôi (cercariae), ấu trùng gây nhiễm (metacercariae) có những khó khăn và phức tạp nhất định vì hình thái giống nhau, ngay cả sán trưởng thành (Chai, Jung, 2020). Mặc dù vậy, trong hơn 100 năm qua, các loài sán lá ruột nhỏ cơ bản đã được xếp vào các chi riêng biệt tương ứng, nhưng việc chẩn đoán, phân biệt và thẩm định loài đòi hỏi có những khảo sát với kỹ thuật và công nghệ chính

xác hơn (Johansen *et al.*, 2015).

Chỉ thị phân tử (molecular genetic marker) và các công nghệ thao tác và công cụ phân tích phân tử đã góp phần to lớn trong nghiên cứu ứng dụng ở các lĩnh vực chẩn đoán, phân loại, phá hệ và tiên hóa, dịch tễ học phân tử và di truyền quần thể (Hoelzer *et al.*, 2018). Nguồn cung cấp chỉ thị phân tử là trình tự DNA của hệ gen ty thể (mitochondrial genome, mtDNA) và hệ gen nhân chủ yếu là trình tự DNA của đơn vị mã hóa ribosome (ribosomal transcription unit, rTU hay rDNA) (Le *et al.*, 2002; 2020; Hu, Gasser, 2006; Crampton-Platt *et al.*, 2016). Nhu cầu ứng dụng sinh học phân tử rất lớn, đặc biệt là làm sáng tỏ điều kiện và vị trí phân loại, mối quan hệ về loài/chi/họ/liên họ và phân lớp, đặc biệt với các loài “đồng hình”, loài “đa hình”, loài “chị em”, loài “lai” và các dạng phát triển hiện diện trong cùng vật chủ.

Hệ gen ty thể ở động vật, kể cả ký sinh trùng đa bào, có cấu trúc là một vòng DNA khép kín có độ dài khoảng 13–25 nghìn nucleotide, tùy loài, chứa 12 gen mã hóa protein (*atp6*, *cob*, *cox1–3*, *nad1–6* và *nad4L*), 2 gen RNA ribosome (*rrnL*/(16S) và *rrnS*/(12S)), 22 gen RNA vận chuyển (tRNA hay *trn*) và một vùng không mã hóa (non-coding region, NCR) thường chứa nhiều cấu trúc lặp có kích thước dài ngắn khác nhau (Boore, 1999; Le *et al.*, 2002; Bernt *et al.*, 2013). Đối với ngành Sán dẹt (Platyhelminthes), đã có hàng chục hệ gen ty thể hoàn chỉnh của sán lá (thuộc lớp Trematoda) và sán dây (lớp Cestoda) gây bệnh ở người và động vật đã được thu nhận và phân tích đầy đủ cung cấp nguồn dữ liệu đa dụng (Nguồn: GOBASE: <http://gobase.bcm.umontreal.ca/>) (O'Brien *et al.*, 2009). Có thể kể ra một số hệ gen ty thể của sán lá phổi *Paragonimus ohirai* (KX765277; Le *et al.*, 2019), *P. westermani* ở 3 dạng tồn tại (AF540958; AF219379 và KM280646) (Biswal *et al.*, 2014), *P. heterotremus* (GenBank: KY952166) (Qian *et al.*, 2018; Oeyet *et al.*, 2019), *P. kellicoti* (Wang *et al.*, 2018), sán lá ruột nhỏ *Haplorchis taichui* (GenBank: KF214770, Lee *et al.*, 2013), *Metagonimus yokogawai* (KC330755), sán lá gan nhỏ *Opisthorchis viverrini*, *O. felineus*, *Clonorchis sinensis*

(Shekhovtsov *et al.*, 2010), sán lá gan lớn *Fasciola hepatica*/ *F. gigantica* (KF543342; KF543343) và *F. jacksoni* (Rajapakse *et al.*, 2020), sán lá ruột lớn *Fasciolopsis buski*, *Fascioloides magna* (Ma *et al.*, 2016; 2017), sán lá ruột *Echinochasmus japonicus* (Le *et al.*, 2016); *Echinostoma caproni* và *E. paraensei* (LL250667; KT008005), *E. miyagawai* (Li *et al.*, 2019), *E. revolutum* (Le *et al.*, 2020), *Hypoderaeum conoideum* (Yang *et al.*, 2015) và hàng chục loài khác ở 12 họ khác nhau. Hệ gen ty thể của sán lá đã cung cấp nguồn gen rất giá trị cho xác định loài, chẩn đoán, phân biệt, phá hệ, tiến hóa và di truyền quần thể (Le *et al.*, 2002; 2016; Lee *et al.*, 2013; Liu *et al.*, 2014).

Trong hệ gen nhân ở tế bào động vật còn có hàng trăm rTU sắp xếp nối tiếp nhau thành dãy. Ở ký sinh trùng, mỗi một đơn vị có độ dài khoảng 7–10 kb, chứa 3 gen mã hóa ribosome (18S, 5.8S và 28S) tách biệt nhau bởi 2 vùng giao gen, lần lượt là ITS-1 và ITS-2 (internal transcribed spacer 1 and 2). Các đơn vị được nối với nhau bằng một chuỗi nucleotide, chứa nhiều cấu trúc lặp, gọi là vùng bản lề IGS (non-transcribed intergenic spacer) và vùng ngoại tiếp nối ETS ở đầu 3' của rTU (3' external transcribed spacer). Khung gen đặc trưng của rTU là: IGS-ETS-18S-ITS1-5.8S-ITS2-28S-IGS, tạo nên các cấu trúc nối tiếp nhau (Blair, 2006; McStay, 2016). Bất kỳ gen ribosome nào (18S, 28S) hay vùng giao gen (ITS-1, ITS-2) đều được sử dụng trong phân tích phân loại và quan hệ tiến hóa phân tử của các loài và truy xuất nguồn gốc phá hệ (Weider *et al.*, 2005; Blair, 2006). Với sán lá ruột nhỏ, đã có một số nghiên cứu rTU và kiến tạo dữ liệu ứng dụng chỉ thị rDNA cho phân loại, chẩn đoán và di truyền quần thể (Thaenkham *et al.*, 2011; Le *et al.*, 2017; Chontanarith *et al.*, 2018). Tuy nhiên, nhiều dữ liệu gen học, hoặc chưa có, hoặc chưa cập nhật đầy đủ, của hàng chục loài sán lá kể cả sán lá ruột nhỏ quan trọng có ảnh hưởng y tế cộng đồng.

Trong bài báo này chúng tôi hoàn thiện quá trình giải trình tự và phân tích đặc điểm hệ gen ty thể của loài sán lá ruột nhỏ *H. taichui* của Việt Nam so sánh với *H. taichui* (chủng của Lào) và loài *Metagonimus yokogawai* (Hàn Quốc) nhằm

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 các loài trong họ Heterophyidae.

NGUYÊN LIỆU VÀ PHƯƠNG PHÁP

Mẫu *Haplorchis* spp. trong nghiên cứu

Các mẫu sán lá ruột nhỏ trưởng thành thu từ các bệnh nhân tại Quảng Trị (Việt Nam), gồm *H. taichui* chủng QT3, ký hiệu Htai-QT3-VN, và *H. pumilio*, chủng HPU8, ký hiệu Hpum-HPU8-VN được sử dụng để giải trình tự hệ gen ty thể và đơn vị mã hóa ribosome. Ngoài ra, có một số mẫu khác của *Haplorchis*, *Stellantchasmus*, *Procerovum*, *Centrocestus* cũng được thu thập và nghiên cứu chỉ thị mtDNA và rTU. Mẫu ở dạng tươi đông lạnh, hoặc trong cồn 70%, bảo quản ở -20°C cho đến khi dùng. Mẫu do TS Đỗ Trung Dũng (Viện Sốt rét - Ký sinh trùng và Côn trùng trung ương) cung cấp, đã được xác định loài dựa vào đặc điểm hình thái học (Dung *et al.*, 2007; 2013; 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; Le *et al.*, 2017).

Tách chiết DNA tổng số, thực hiện PCR và giải trình tự

DNA tổng số được tách chiết bằng bộ sinh phẩm GeneJET™ Genomic DNA Purification Kit (Thermo Scientific Inc., MA, USA) theo hướng dẫn của nhà sản xuất. DNA tổng số của mẫu được kiểm tra nồng độ bằng cách đo hấp phụ quang học (OD, optical density) trên máy đo quang phổ kế Nanodrop 1000, bảo quản ở -20°C và pha ở nồng độ 50 ng/ μL , sử dụng 2 μL cho mỗi phản ứng PCR có dung tích 50 μL .

Phản ứng PCR có tổng thể tích 50 μL , gồm: 25 μL PCR Mastermix (Thermo Fisher Scientific, MA, USA), 2 μL mỗi loại mồi (10 pmol/ μL), 2 μL khuôn DNA, 2 μL DMSO (dimethyl sulfoxide) và 17 μL nước khử ion DEPC, được thực hiện trên máy MJ PTC-100 (USA). Chu trình nhiệt gồm: 1 chu kỳ ở $95^{\circ}\text{C}/5$ phút, 35 chu kỳ ở [$94^{\circ}\text{C}/1$ phút, $52^{\circ}\text{C}/4$ phút; $72^{\circ}\text{C}/2$ phút], chu kỳ cuối ở $72^{\circ}\text{C}/10$ phút. Đối với L-PCR, thực hiện bước kéo dài ở $72^{\circ}\text{C}/6-10$

phút. Sản phẩm PCR được điện di kiểm tra trên thạch agarose 1% nhuộm Ethidium bromide, với chỉ thị DNA Lamda (*HindIII*). Sản phẩm PCR đơn băng được tinh sạch bằng GeneJET PCR Purification Kit hoặc phân lập bằng DNA đúng kích thước dự tính (nếu nhiều băng) bằng bộ kit GeneJET Gel Extraction Kit (Thermo Fisher Scientific Inc.). Một số sản phẩm PCR được dòng hoá vào vector pCR2.1 hoặc pCR2.1TOPO TA-cloning Kit (Invitrogen), chuyển nạp vào dòng tế bào IVN α F' để chọn lọc khuẩn lạc. DNA của plasmid tái tổ hợp được tách chiết với bộ hoá chất GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific Inc.). Các sản phẩm DNA được gửi đi giải trình tự tại các cơ sở dịch vụ trong và ngoài nước.

Giải trình tự và xử lý chuỗi nucleotide

Giải trình tự trực tiếp bằng môi xuôi và môi ngược (đối với sản phẩm PCR ngắn); bằng cách giải kế tiếp bằng môi thiết kế lồng vào nhau (primer-walking) với sản phẩm PCR dài. Với đoạn DNA đã chèn vào plasmid tái tổ hợp, sử dụng môi M13F (5' GTAAAACGACGGCCAG 3') và M13R (5' CAGGAAACAGCTATGAC 3'), hoặc môi thích hợp được thiết kế thêm để giải trình tự. Chuỗi nucleotide được biên tập (editing) từ giản đồ trình tự (chromatogram) do cơ sở dịch vụ cung cấp bằng Chromas Pro (<http://technelysium.com.au/wp/chromaspro/>), sắp xếp và thu chuỗi cuối cùng bằng GENEDOC2.7 và MEGA7.0 (Kumar *et al.*, 2016).

Xử lý phân tích chuỗi và thiết lập cấu trúc gen/hệ gen ty thể

Xác định kích thước từng gen ty thể, vùng giao gen, vùng không mã hóa; xác định trật tự sắp xếp gen trong hệ gen ty thể bằng cách so sánh với dữ liệu hiện có của sán đẹt (của chúng tôi và Ngân hàng gen). Mười hai gen mã hóa protein, 2 gen ribosome ty thể được xác định bằng so sánh đối chiếu.

Các gen mã hóa protein (protein-coding genes, PCGs) được xác định bằng cách căn chỉnh đối chiếu với các PCG sẵn có của các loài sán lá đã được công bố hoặc có trong Ngân hàng gen. Xác định mã khởi đầu là ATG/GTG và mã kết

thúc là TAA/TAG (bộ mã TGA là mã kết thúc ở protein của loài bậc cao, nhưng ở ty thể sán đẹt là bộ mã mã hóa Tryptophan). Các PCG đã được dịch mã bằng cách sử dụng bộ mã di truyền ty thể là “Echinoderm Mitochondrial Genetic Code” của sán lá (Bảng số 9 trong GenBank). Thành phần nucleotide và bộ mã (codon) được phân tích với chương trình MEGA 7.0 hoặc MEGA X (Kumar *et al.*, 2016; 2018) và tỷ lệ sử dụng bộ mã (codon usage) cho 12 PCG kết nối được xác định với chương trình GENE INFINITY (Cách sử dụng Codon: http://www.geneinfinity.org/sms/sms_codonusage.html).

Các gen RNA vận chuyển (tRNA hay *trn*), gồm 22 gen được xác định bằng phần mềm chuyên biệt (tRNAscan-SE1.21: www.genetics.wustl.edu/eddy/tRNAscan-SE/ (Lowe, Eddy, 1997); ARWEN tại <http://130.235.244.92/bcgi/arwen.cgi> (Laslett, Canback, 2008); DOGMA tại <http://dogma.cccb.utexas.edu/> và MITOS tại <http://mitos.bioinf.uni-leipzig.de/>).

Cấu trúc của toàn bộ hệ gen ty thể (dạng vòng tròn) và của 22 gen tRNA được mô hình hóa bằng hình vẽ theo qui định cấu trúc gen và hệ gen ty thể, sử dụng các phần mềm chuyên biệt OGDRAW (Greiner *et al.*, 2019).

Vùng không mã hóa (non-coding region, NCR) được xác định bởi ranh giới của gen tRNA cuối cùng và gen kế tiếp sau đó. Các loài *Haplorchis* spp. có NCR nằm giữa hai RNA vận chuyển là tRNA^{Glu} (*trnE*) và tRNA^{Gly} (*trnG*). NCR ở mtDNA của đa số các sán đẹt có chứa các chuỗi/cấu trúc lặp liên kề (tandem repeat unit, thường ký hiệu TR hay RU hay TRU). Cấu trúc lặp liên kề được xác định bằng phần mềm Tandem Repeat Finder v3.01 (Benson, 1999).

Thu nhận và phân tích đơn vị mã hóa ribosome (rTU) của *H. taichui* và *H. pumilio*

Môi được thiết kế dựa trên một số công bố về rDNA đã có trên Ngân hàng gen hoặc dữ liệu của chúng tôi đã công bố và đang có (Le *et al.*, 2017; 2020) để thu nhận toàn phần mỗi đơn vị rTU (5'18S-ITS1-5.8S-ITS2-28S-IGS 3') có độ dài

khoảng từ 7–8 cho đến 9–10 kb. Phần mã hóa được tính từ đầu 5' của gen 18S cho đến hết 3' của gen 28S, độ dài thường là khoảng 6–7 kb.

Các chương trình ChromasPro, BioEdit, GENEDOC2.7 đã được sử dụng để thu nhận, xử lý, phân tích các chuỗi nucleotide, đặc điểm các gen ribosome (18S, 5.8S, 28S) và vùng giao gen (internal transcribed spacer, ITS) gồm ITS1 và ITS2 và vùng bản lề IGS. Hầu hết các loài chỉ thu nhận được đoạn gen 28S rRNA của vùng D1–D3 khoảng 1.200–1.250 bp để phân tích so sánh và xây dựng cây phả hệ. Chương trình Tandem Repeat Finder v3.01 (Benson, 1999) được sử dụng để tìm kiếm chuỗi (cấu trúc) lặp có thể có ở vùng giao gen ITS1 hay ITS2.

Tính toán giá trị độ lệch (skew/skewness)

Giá trị độ lệch (skew/skewness value) là sự chênh lệch sử dụng A, T, G, C để kiến tạo, có thể là chuỗi nucleotide của toàn bộ hay một phần mtDNA, hoặc một gen hay đa gen cộng hợp PCG hoặc chuỗi nucleotide không mã hóa; hoặc các gen rRNA (gen 18S hay 28S ribosome hay ITS1/ITS2) hoặc toàn bộ đơn vị mã hóa ribosome (rTU); hoặc bất kỳ chuỗi nucleotide/amino acid nào. Chương trình GENEDOC2.7 được sử dụng tính tỷ lệ từng nucleotide (%) và tính toán giá trị A+T và G+C. Giá trị độ lệch (skew) dao động theo lệch âm và lệch dương (từ -1 đến +1) được tính theo công thức mô tả bởi Perna, Kocher (1995): $[AT\ skew = (A + T)/(A - T)]$ và $[GC\ skew = (G + C)/(G - C)]$. Nếu độ lệch bằng 0 thì tần số sử dụng A và T cũng như G và C như nhau trong chuỗi DNA; nếu AT skew lệch âm thì được hiểu là T được sử dụng nhiều hơn A (Le *et al.*, 2004). Đối với sán lá (trematode), do tỷ lệ sử dụng A và T nhiều hơn G và C nên giá trị độ lệch của AT skew luôn có giá trị âm và của GC skew luôn có giá dương (Le *et al.*, 2002; 2004; 2019; 2020; Rajapakse *et al.*, 2020).

KẾT QUẢ VÀ THẢO LUẬN

Phân tích và so sánh đặc điểm hệ gen ty thể

Tóm lược sắp xếp hệ gen và cấu trúc gen ty thể của sán lá ruột nhỏ *H. taichui*

Toàn bộ chuỗi nucleotide hệ gen ty thể hoàn

chỉnh của sán lá ruột nhỏ *H. taichui* mẫu Việt Nam (Htai-QT3-VN) có độ dài 15.120 bp, đã được thu nhận và phân tích chi tiết. Các chủng của loài *H. taichui* đã được thu thập ở Quảng Trị (Việt Nam), trong số đó một hệ gen ty thể đã được thu nhận và phân tích cấu trúc gen nhưng chưa đầy đủ về đặc điểm các gen mã hóa protein (Lê Thanh Hòa và cs., 2016). Chúng tôi tiếp tục hoàn thiện phân tích và so sánh với các chủng khác trên thế giới, kết quả được trình bày trong bài báo này.

Toàn bộ mtDNA của *H. taichui* có cấu trúc là vòng DNA khép kín chứa 36 gen, gồm 12 gen mã hóa protein (PCG, protein-coding genes), 2 gen RNA ribosome (*rrnL* và *rrnS*), 22 gen RNA vận chuyển amino acid (tRNA) và một vùng DNA không mã hóa (NCR). Bảng 1 liệt kê sắp xếp thứ tự các gen của *H. taichui* (mẫu Việt Nam) so sánh với *H. taichui* (mẫu của Lào) (Lee *et al.*, 2013). Sắp xếp dạng vòng tròn (circular map) trình bày minh họa ở Hình 1A với các cấu trúc lặp TRU (tandem repeat unit) được xác định chính xác và đầy đủ số lượng. Các gen cách nhau bởi vùng giao gen (là chuỗi nucleotide nối giữa gen trước với gen liền kề) là 0–25 nucleotide (Bảng 1). Một số gen sử dụng 1–3 nucleotide của nhau, lồng vào nhau, để kết thúc hoặc khởi đầu gen. 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 và của tất cả các loài sán dẹt (Le *et al.*, 2002).

Thứ tự sắp xếp dạng thẳng (linear gene arrangement) của 36 gen và vùng không mã hóa, được mở vòng tại *cox3*, như sau: *cox3-H-cob-nad4L-nad4-QFM-atp6-nad2-VAD-nad1-NPIK-nad3-S₁W-cox1-T-rrnL-C-rrnS-cox2-nad6-YL₁S₂L₂R-nad5-E-NCR[TRU1-5]-G* (Hình 1B). Vùng DNA không mã hóa (NCR) có độ dài 1.692 bp (mẫu Việt Nam) và 1.705 bp (mẫu Lào), được chia thành các tiểu vùng: tiểu vùng chuỗi giao gen dài số 1 (246 bp) và số 2 (487 bp), các tiểu vùng chứa 5 đơn vị lặp liền kề TRU (182 và 183 bp/mỗi một chuỗi) và một tiểu vùng không mã hóa (49 bp) nối tRNA^{Gly} với *cox3*. Hai gen *nad4L* (264 bp) và *nad4* (1.278 bp) sắp xếp kế tiếp và lồng vào nhau trên hai khung gen khác nhau, trong đó gen *nad4* gói 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' của *cox2* gối vào đầu 5' của *nad6* (Bảng 1), đây là một trường hợp hiếm gặp ở sán dẹt được biết cho đến nay. Chênh lệch độ dài khoảng 11–13 bp giữa *H. taichui* Htai-QT3-VN và Htai-LA là ở vùng NCR.

Phương thức sử dụng nucleotide

Đê kiến tạo hệ gen ty thể (15.120 bp), *H. taichui* (Việt Nam) sử dụng A = 19,56%, T = 39,71%, G = 28,34%, C = 12,39%, với tổng thể thành phần A+T là 59,27% và G+C là 40,73%, giá trị độ lệch (skew/skewness) ở A+T là âm (-0,340) và G+C là dương (0,392). Tổng số 10.164 nucleotide được sử dụng để xây dựng 12 PCG trong đó có 17,02% (A), 43,06 (T), 27,99% (G) và 11,92% C, tổng thành phần A+T là 60,08% và G+C là 39,92%, giá trị độ lệch cũng tương tự mtDNA, skew = -0,433 (A+T) và skew = 0,403 (G+C) (Bảng 2). Hai gen ribosome ty thể (MRG) là 16S (*rrnL*) và 12S (*rrnS*) sử dụng A+T ít hơn (57,22%) và G+C nhiều hơn (42,78%) so với mtDNA và PCG, do vậy giá trị độ lệch skew cũng thấp hơn.

Phương thức sử dụng nucleotide A, T, G, C của *H. taichui* của Lào cũng tương tự đồng loài ở Việt Nam. Trong khi đó loài sán lá ruột nhỏ *M. yokogawai* có tỷ lệ sử dụng A+T thấp hơn (*M. yokogawai*: mtDNA/55,68%; PCGs/55,96%; MRGs/54,15%) và G+C cao hơn so với loài *H. taichui* (Bảng 2).

Tổng thể thành phần A+T sử dụng ở mtDNA của *H. taichui* cũng tương tự như ở một số loài sán lá gan lớn *Fasciola* spp. (họ Fasciolidae) và sán lá gan nhỏ (họ Opisthorchiidae), vào khoảng 60%, nhưng khác xa và ít hơn nhiều so với các loài sán máng *Schistosoma* spp. thuộc họ

Schistosomatidae (thiên vị về A+T, thường là khoảng 70% A+T) (Le *et al.*, 2002).

Đặc điểm gen RNA ribosome ty thể và RNA vận chuyển

RNA ribosome ty thể (gồm hai tiểu phần, *rrnL* và *rrnS*) và RNA vận chuyển (gồm 22 tRNA), 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 1). Chuỗi gen của tRNA^{Cys} (*trnC*) nằm xen vào giữa hai gen RNA ribosome, *rrnL* và *rrnS* (Bảng 1). Cấu trúc bậc hai suy diễn điển hình của *rrnL* và *rrnS* ở *H. taichui* tương tự của các loài sán dẹt và giống như ở sán lá gan lớn *F. hepatica* (Le *et al.*, 2002).

Hệ gen ty thể *H. taichui* mã hóa cho 22 RNA vận chuyển amino acid, có độ dài của tRNA biến động từ 60 bp (ở tRNA^{Val}) đến 73 bp (ở tRNA^{Glu}), đa phần 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 “đối mã” (anticodon) và vùng cánh tay TΨC (TΨC-stems), chỉ chứa 4 bp (ít khi 5 bp, chỉ có ở *trnC*). Về cấu trúc, ngoại trừ tRNA của Serine 1 (*trnS*₁), hầu như tất cả tRNA (21 gen RNA còn lại) của mtDNA *H. taichui* đều có cấu trúc bậc hai suy diễn hình “lá sồi” (clover-leaf) hoàn toàn đặc trưng của sán dẹt, kể cả tRNA của Serine 2 (*trnS*₂), của 2 tRNA vận chuyển Leucine (*trnL*₁; *trnL*₂) và của Cystein (*trnC*). Chỉ duy nhất cấu trúc của *trnS*₁ của *H. taichui* bị khuyết cánh tay dihydrouridine (DHU-arm), tạo nên một vòm cung nucleotide. Khuyết thiếu cánh tay DHU ở tRNA của Serine và Leucine cũng thường gặp ở mtDNA của một số loài sán dẹt khác (Le *et al.*, 2002; Liu *et al.*, 2014; Ma *et al.*, 2016; 2017).

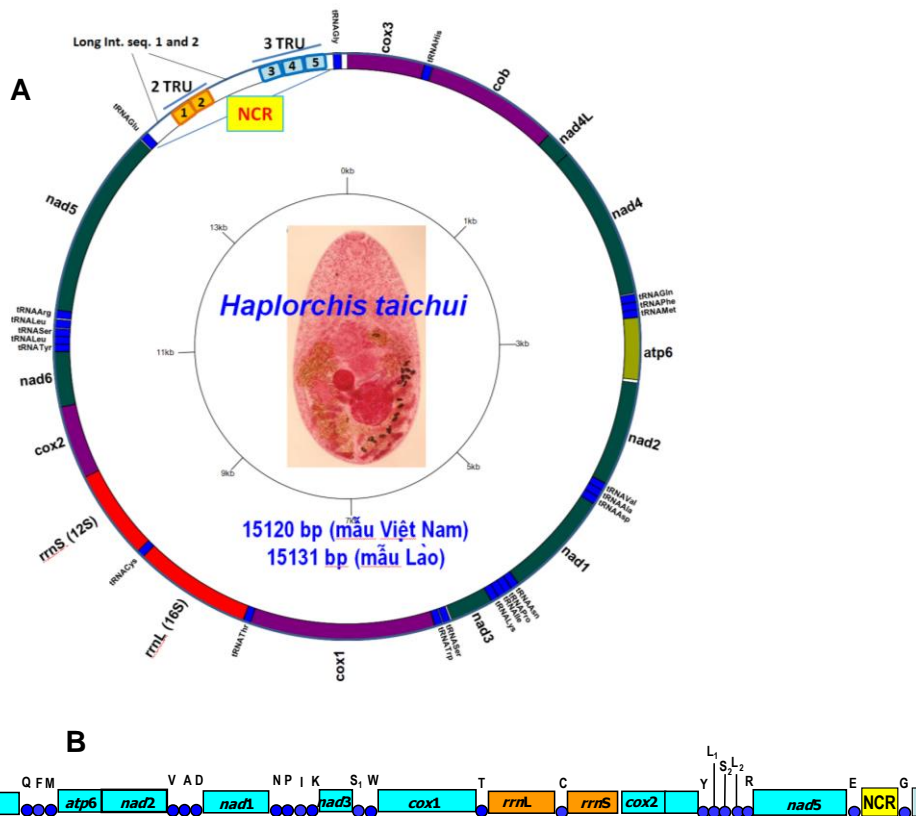
Bảng 1. Vị trí của gen và các đặc điểm gen trong hệ gen ty thể hoàn chỉnh của loài sán lá ruột nhỏ *H. taichui* (chủng QT3, Việt Nam, 15.120 bp; GenBank: MG972809) và *H. taichui* (LA, Lào, 15.131 bp; GenBank: KF214770).

| Gen/vùng giao gen | Vị trí (5' > 3') | Đặc điểm [bp/aa(start/stop)] và các vùng gen | Đối mã của tRNA | Chuỗi giao gen (bp) | Vị trí (5' > 3') | Đặc điểm [bp/aa(start/stop)] và các vùng gen | Đối mã của tRNA | Chuỗi giao gen (bp) |
|-------------------|------------------|--|-----------------|---------------------|--------------------------|--|-----------------|---------------------|
| | | Htai-QT3-VN (MG972809) (Việt Nam) | | | Htai-LA (KF214770) (Lào) | | | |

| | | | | | | | | |
|---------------------------|-------------|--------------------|-----|-----|-------------|--------------------|-----|-----|
| cox3 | 1-645 | 645/214/(ATG/TAA) | | +2 | 1-645 | 645/214/(ATG/TAA) | | +2 |
| tRNA ^{His} | 648-713 | 66 | GTG | +6 | 648-713 | 66 | GTG | +6 |
| cob | 720-1829 | 1110/369/(ATG/TAG) | | +1 | 720-1829 | 1110/369/(ATG/TAG) | | +1 |
| nad4L | 1831-2094 | 264/87/(ATG/TAG) | | -40 | 1831-2094 | 264/87/(ATG/TAG) | | -40 |
| nad4 | 2055-3335 | 1281/426/(GTG/TAA) | | +9 | 2055-3335 | 1281/426/(GTG/TAA) | | +9 |
| tRNA ^{Gln} | 3345-3410 | 66 | TTG | +3 | 3345-3410 | 66 | TTG | +3 |
| tRNA ^{Phe} | 3414-3477 | 64 | GAA | -1 | 3414-3477 | 64 | GAA | -1 |
| tRNA ^{Met} | 3477-3540 | 64 | CAT | 0 | 3477-3540 | 64 | CAT | 0 |
| atp6 | 3541-4056 | 516/171/(ATG/TAG) | | +25 | 3541-4056 | 516/171/(ATG/TAG) | | +25 |
| nad2 | 4082-4951 | 870/289/(ATG/TAA) | | +4 | 4082-4951 | 870/289/(ATG/TAA) | | +4 |
| tRNA ^{Val} | 4956-5015 | 60 | TAC | +1 | 4956-5015 | 60 | TAC | +1 |
| tRNA ^{Ala} | 5017-5079 | 63 | TGC | +2 | 5017-5079 | 63 | TGC | +2 |
| tRNA ^{Asp} | 5082-5148 | 67 | GTC | 0 | 5082-5148 | 67 | GTC | 0 |
| nad1 | 5149-6054 | 906/302/(GTG/TAG) | | +2 | 5149-6054 | 906/302/(GTG/TAG) | | +2 |
| tRNA ^{Asn} | 6057-6122 | 66 | GTT | +3 | 6057-6122 | 66 | GTT | +3 |
| tRNA ^{Pro} | 6126-6190 | 65 | TGG | -3 | 6126-6190 | 65 | TGG | -3 |
| tRNA ^{Ile} | 6187-6251 | 65 | GAT | +3 | 6187-6251 | 65 | GAT | +3 |
| tRNA ^{Lys} | 6255-6319 | 65 | CTT | 0 | 6255-6319 | 65 | CTT | 0 |
| nad3 | 6320-6679 | 360/119/(ATG/TAG) | | +13 | 6320-6679 | 360/119/(ATG/TAG) | | +13 |
| tRNA ^{Ser1(AGN)} | 6693-6754 | 62 | GCT | +7 | 6693-6754 | 62 | GCT | +7 |
| tRNA ^{Trp} | 6762-6825 | 64 | TCA | +4 | 6762-6825 | 64 | TCA | +4 |
| cox1 | 6830-8371 | 1542/513/(ATG/TAG) | | -1 | 6830-8371 | 1542/513/(ATG/TAG) | | -1 |
| tRNA ^{Thr} | 8371-8434 | 64 | TGT | 0 | 8371-8434 | 64 | TGT | 0 |
| rrL (16S) | 8435-9413 | 979 | | 0 | 8435-9413 | 979 | | 0 |
| tRNA ^{Cys} | 9414-9477 | 64 | GCA | 0 | 9414-9477 | 64 | GCA | 0 |
| rrS (12S) | 9478-10226 | 749 | | 0 | 9478-10224 | 747 | | 0 |
| cox2 | 10227-10850 | 624/201/(ATG/TAG) | | -14 | 10225-10848 | 624/201/(ATG/TAG) | | -14 |
| nad6 | 10837-11295 | 459/152/(ATG/TAG) | | +1 | 10835-11293 | 459/152/(ATG/TAG) | | +1 |
| tRNA ^{Tyr} | 11297-11359 | 63 | GTA | 0 | 11295-11357 | 63 | GTA | 0 |
| tRNA ^{Leu1(CUN)} | 11360-11425 | 66 | TAG | -2 | 11358-11423 | 66 | TAG | -2 |
| tRNA ^{Ser2(UCN)} | 11424-11487 | 64 | TGA | +8 | 11422-11485 | 64 | TGA | +8 |
| tRNA ^{Leu2(UUR)} | 11496-11563 | 68 | TAA | +11 | 11494-11561 | 68 | TAA | +11 |
| tRNA ^{Arg} | 11575-11639 | 64 | TCG | +1 | 11573-11637 | 64 | TCG | +1 |
| nad5 | 11641-13227 | 1587/528/(GTG/TAA) | | +11 | 11639-13225 | 1587/528/(GTG/TAA) | | +11 |
| tRNA ^{Glu} | 13239-13311 | 73 | TCC | 0 | 13237-13309 | 73 | TCC | 0 |
| NCR | 13312-15003 | 1692 | | | 13310-15014 | 1705 | | |
| Long Int seq. 1 | 13312-13557 | 246 | | 0 | 13310-13568 | 259 | | 0 |
| TRU1 (182) | 13558-13739 | 182 | | 0 | 13569-13750 | 182 | | 0 |
| TRU2 (182) | 13740-13921 | 182 | | 0 | 13751-13931 | 182 | | 0 |
| Long Int seq. 2 | 13922-14408 | 487 | | 0 | 13932-14419 | 487 | | 0 |
| TRU3 (183) | 14409-14595 | 183 | | 0 | 14420-14606 | 183 | | 0 |
| TRU4 (183) | 14596-14781 | 183 | | 0 | 14607-14789 | 183 | | 0 |
| TRU5 (183) | 14782-14961 | 183 | | +42 | 14790-14972 | 183 | | +42 |
| tRNA ^{Gly} | 15004-15071 | 68 | TTC | 0 | 15015-15082 | 68 | TTC | 0 |
| unique seq. | 15072-15120 | 49 | | 0 | 15083-15131 | 49 | | 0 |

Ghi chú: bp: base pair; aa: amino acid; start: mã khởi đầu; stop: mã kết thúc; tRNA: RNA vận chuyển; Int. seq.: Chuỗi giao gen (+, Số nucleotide trước gen kề cận sau đó; -, Số nucleotide lồng vào gen kề cận sau đó); *Dấu sao: RNA vận chuyển (tRNA) khuyết thiếu cánh tay DHU-arm; NCR: non-coding region (vùng không mã hóa); Long Int seq. 1 và 2: chuỗi giao gen dài 1 và 2; TRU: đơn vị (chuỗi) lặp liên kề (182 và 183 bp/mỗi một chuỗi); unique seq.: chuỗi nucleotide nối gen vận chuyển tRNA^{Gly} và cox3; VN: Việt Nam; LA: Lào.

Hệ gen ty thể *Haplorchis taichui* (mạch vòng và mạch thẳng)



Hình 1. Sơ đồ hệ gen ty thể của sán lá ruột nhỏ *H. taichui* dạng vòng tròn (A) và dạng mạch thẳng (B) mở vòng tại 5' của gen *cox3* (Ngân hàng gen: MG972809 (mẫu Việt Nam) và KF214770 (mẫu Lào)). Các gen mã hóa protein và rDNA ribosome được viết tắt theo Boore (1999) và các công bố trước đây của chúng tôi (Le *et al.*, 2002; 2019; 2020; Lee *et al.*, 2013). Các gen RNA vận chuyển amino acid (tRNA) được đánh dấu bằng viết tắt ba chữ của amino acid hoặc một chữ cái mà tRNA đó vận chuyển (xem Bảng 1). Vùng không mã hóa (NCR) nằm giữa tRNA^{Glu} và tRNA^{Gly} bao gồm 5 đơn vị cấu trúc lặp liên kề (TRU1–5) và 2 chuỗi nối (Long Int. Seq. 1 và 2) xen kẽ.

Bảng 2. Tỷ lệ sử dụng thành phần nucleotide A, T, G, C và giá trị độ lệch (skew/skewness) sử dụng của các gen mã hóa protein (PCGs), các gen ribosome ty thể (MRGs) và toàn bộ hệ gen ty thể (mtDNA) của các loài sán lá ruột nhỏ trong họ Heterophyidae.

| Loài | | Độ dài (bp) | A (%) | T (%) | G (%) | C (%) | A+T (%) | AT-skew | G+C (%) | GC-skew | |
|----------------------|---|-------------|--------|-------|-------|-------|---------|---------|---------|---------|-------|
| HETEROPHYIDAE | | | | | | | | | | | |
| 1 | <i>Haplorchis taichui</i> (VN) (MG972809) | mtDNA | 15.120 | 19,56 | 39,71 | 28,35 | 12,38 | 59,27 | -0,340 | 40,73 | 0,392 |
| | | PCGs | 10.164 | 17,02 | 43,06 | 27,99 | 11,92 | 60,08 | -0,433 | 39,92 | 0,403 |
| | | MRGs | 1.730 | 22,25 | 34,97 | 29,36 | 13,42 | 57,22 | -0,222 | 42,78 | 0,373 |
| 2 | <i>Haplorchis taichui</i> (LA) (KF214770) | mtDNA | 15.131 | 19,56 | 39,67 | 28,34 | 12,42 | 59,23 | -0,340 | 40,97 | 0,390 |
| | | PCGs | 10.164 | 17,02 | 43,08 | 27,96 | 11,93 | 60,11 | -0,434 | 39,89 | 0,402 |
| | | MRGs | 1.728 | 22,34 | 35,01 | 29,28 | 13,37 | 57,35 | -0,221 | 42,65 | 0,373 |

| | | | | | | | | | | | |
|---|---|-------|--------|-------|-------|-------|-------|-------|--------|-------|-------|
| 3 | <i>Metagonimus yokogawai</i> (KR) (KC330755) | mtDNA | 15.258 | 17,79 | 37,89 | 30,11 | 14,21 | 55,68 | -0,361 | 44,32 | 0,359 |
| | | PCGs | 10.245 | 15,31 | 40,65 | 29,68 | 14,35 | 55,96 | -0,453 | 44,04 | 0,348 |
| | | MRGs | 1.736 | 19,93 | 34,22 | 30,82 | 15,03 | 54,15 | -0,264 | 45,85 | 0,344 |

Ghi chú: A: Adenine; T: Thymine; G: Guanine; C: Cytosine. PCGs: protein-coding genes (các gen mã hóa protein); MRGs: mito-ribosomal genes (các gen ribosome ty thể); mtDNA: toàn bộ hệ gen ty thể; VN: Việt Nam; LA: Lào; KR: Hàn Quốc.

Bảng 3. Thống kê số lượng và tỷ lệ sử dụng bộ mã kiến tạo các gen mã hóa protein (PCG) trong hệ gen ty thể của loài *H. taichui* (Việt Nam), *H. taichui* (Lào) và loài *M. yokogawai* (Hàn Quốc) thuộc họ Heterophyidae.

| aa | Codon | Htai (Việt Nam) | | Htai (Lào) | | Myok (Hàn Quốc) | | aa | Codon | Htai (Việt Nam) | | Htai (Lào) | | Myok (Hàn Quốc) | | |
|------------|-------|-----------------|------|------------|------|-----------------|------|-------------|------------|-----------------|------|------------|------|-----------------|------|------|
| | | No | % | No | % | No | % | | | No | % | No | % | No | % | |
| Ala | GCG | 36 | 1,06 | 37 | 1,09 | 33 | 0,97 | Pro | CCG | 16 | 0,47 | 16 | 0,47 | 27 | 0,79 | |
| | GCA | 9 | 0,27 | 10 | 0,30 | 20 | 0,59 | | CCA | 15 | 0,44 | 15 | 0,44 | 14 | 0,41 | |
| | GCT | 82 | 2,42 | 79 | 2,33 | 77 | 2,25 | | CCT | 45 | 1,33 | 45 | 1,33 | 44 | 1,29 | |
| | GCC | 21 | 0,62 | 20 | 0,59 | 19 | 0,56 | | CCC | 16 | 0,47 | 16 | 0,47 | 17 | 0,50 | |
| Cys | TGT | 89 | 2,63 | 90 | 2,66 | 80 | 2,34 | Gln | CAG | 19 | 0,56 | 19 | 0,56 | 21 | 0,61 | |
| | TGC | 12 | 0,35 | 11 | 0,33 | 33 | 0,97 | | CAA | 5 | 0,15 | 5 | 0,15 | 11 | 0,32 | |
| Asp | GAT | 61 | 1,80 | 61 | 1,80 | 52 | 1,52 | Arg | CGG | 24 | 0,71 | 24 | 0,71 | 23 | 0,67 | |
| | GAC | 10 | 0,30 | 10 | 0,30 | 18 | 0,53 | | CGA | 7 | 0,21 | 7 | 0,21 | 8 | 0,23 | |
| Glu | GAG | 63 | 1,86 | 63 | 1,86 | 64 | 1,87 | CGC | CGT | 30 | 0,89 | 31 | 0,92 | 32 | 0,94 | |
| | GAA | 15 | 0,44 | 15 | 0,44 | 13 | 0,38 | | CGC | 5 | 0,15 | 5 | 0,15 | 7 | 0,21 | |
| Phe | TTT | 301 | 8,88 | 295 | 8,71 | 234 | 6,85 | Ser | AGG | 70 | 2,07 | 71 | 2,10 | 71 | 2,08 | |
| | TTC | 37 | 1,09 | 41 | 1,21 | 88 | 2,58 | | AGA | 26 | 0,77 | 27 | 0,80 | 24 | 0,70 | |
| Gly | GGG | 136 | 4,01 | 136 | 4,01 | 142 | 4,16 | AGT | AGT | 61 | 1,80 | 61 | 1,80 | 53 | 1,55 | |
| | GGA | 42 | 1,24 | 42 | 1,24 | 29 | 0,85 | | AGC | 21 | 0,62 | 20 | 0,59 | 13 | 0,38 | |
| | GGT | 92 | 2,72 | 92 | 2,72 | 100 | 2,93 | | TCG | 39 | 1,15 | 39 | 1,15 | 54 | 1,58 | |
| | GGC | 32 | 0,94 | 32 | 0,95 | 29 | 0,85 | | TCA | 20 | 0,60 | 18 | 0,53 | 21 | 0,62 | |
| His | CAT | 47 | 1,39 | 46 | 1,36 | 46 | 1,35 | TCT | TCT | 100 | 2,95 | 102 | 3,01 | 96 | 2,81 | |
| | CAC | 7 | 0,21 | 8 | 0,24 | 14 | 0,41 | | TCC | 12 | 0,35 | 13 | 0,38 | 26 | 0,76 | |
| Ile | ATA | 69 | 2,04 | 69 | 2,04 | 53 | 1,55 | Thr | ACG | 20 | 0,59 | 21 | 0,62 | 25 | 0,72 | |
| | ATT | 95 | 2,80 | 93 | 2,75 | 88 | 2,58 | | ACA | 10 | 0,23 | 8 | 0,24 | 7 | 0,21 | |
| | ATC | 13 | 0,38 | 14 | 0,41 | 32 | 0,94 | | ACT | 47 | 1,39 | 49 | 1,45 | 50 | 1,46 | |
| Lys | AAG | 45 | 1,33 | 44 | 1,30 | 56 | 1,64 | ACC | ACC | 10 | 0,30 | 10 | 0,30 | 7 | 0,21 | |
| Leu | TTG | 224 | 6,61 | 226 | 6,67 | 255 | 7,47 | Val | GTG | 132 | 3,90 | 132 | 3,90 | 183 | 5,36 | |
| | TTA | 151 | 4,46 | 150 | 4,43 | 73 | 2,14 | | GTA | 64 | 1,89 | 63 | 1,86 | 37 | 1,08 | |
| | CTG | 56 | 1,65 | 54 | 1,59 | 101 | 2,96 | | GTT | 171 | 5,05 | 173 | 5,11 | 146 | 4,28 | |
| | CTA | 19 | 0,56 | 20 | 0,59 | 8 | 0,23 | | GTC | 37 | 1,09 | 37 | 1,09 | 39 | 1,14 | |
| | CTT | 98 | 2,90 | 99 | 2,92 | 102 | 2,99 | | Trp | TGG | 71 | 2,10 | 71 | 2,10 | 92 | 2,69 |
| | CTC | 16 | 0,47 | 16 | 0,47 | 27 | 0,79 | | | TGA | 46 | 1,36 | 46 | 1,36 | 36 | 1,05 |
| Met | ATG | 113 | 3,34 | 112 | 3,31 | 111 | 3,25 | Tyr | TAT | 145 | 4,28 | 145 | 4,28 | 99 | 2,90 | |
| Asn | AAA | 30 | 0,89 | 31 | 0,92 | 24 | 0,70 | Stop | TAC | 30 | 0,89 | 30 | 0,89 | 53 | 1,55 | |
| | AAT | 35 | 1,03 | 35 | 1,03 | 37 | 1,08 | | TAG | 9 | 0,27 | 9 | 0,27 | 10 | 0,29 | |
| | AAC | 6 | 0,18 | 6 | 0,18 | 9 | 0,26 | | TAA | 3 | 0,09 | 3 | 0,09 | 2 | 0,06 | |

Ghi chú: aa: amino acid ký hiệu bằng ba chữ theo Cơ sở dữ liệu DDBJ (<http://www.ddbj.nig.ac.jp/sub/ref2-e.html>). Ký hiệu loài: Htai (Việt Nam): *H. taichui*, chủng QT3 của Việt Nam trong nghiên cứu này (GenBank: MG972809); Htai (Lào): *H. taichui* của Lào (GenBank: KF214770); Myok: *M. yokogawai* của Hàn Quốc (GenBank: KC330755). Stop: mã kết thúc gen mã hóa protein. Các bộ mã được sử dụng nhiều nhất (Phe-TTT; Leu-TTG) và ít nhất (Gln-CAA; Arg-CGC; Thr-ACA; Thr-ACC) ở *H. taichui* và *M. yokogawai* được bôi màu.

Đặc điểm gen mã hóa protein

H. taichui sử dụng bảng mã di truyền tương tự như ở các loài sán dẹt (trematode), đó là bảng mã “Echinoderm Mitochondrial Code”/Bảng 9 trong Ngân hàng gen được nghiên cứu cho đến nay (Le *et al.*, 2002; 2004; Lee *et al.*, 2013). Tất cả 64 bộ mã đều được sử dụng, 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. Bộ mã TGA, thông thường là bộ mã kết thúc ở các loài sinh vật nhân thật (metazoan), nhưng ở sán dẹt nói chung và sán lá ruột nhỏ *H. taichui*, mã hóa cho amino acid Tryptophan (W) (Le *et al.*, 2002). Loài *M. yokogawai* (của Hàn Quốc) có một số protein có mã khởi đầu và kết thúc khác với *H. taichui*. Không có bất kỳ trường hợp nào có bộ mã kết thúc khiếm khuyết (chỉ có T hay TA) 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 sán lá *H. taichui* đang nghiên cứu ở đây (Bảng 1).

Đặc điểm sử dụng bộ mã trong các gen mã hóa protein

H. taichui sử dụng 10.164 bp mã hóa cho 3.376 amino acid để kiến tạo 12 gen mã hóa protein (trong đó 36 nucleotide cho 12 bộ mã kết thúc). *H. taichui* mtDNA sử dụng nhiều nhất là 301 bộ mã cho Phenylalanine (Phe-TTT), 224 cho Leucine (Leu-TTG) và 171 cho Valine (Val-GTT) và sử dụng ít nhất, chỉ có 5 bộ mã lần lượt cho Glutamine (Gln-CAA) và Arginine (Arg-CGC) và 6 bộ mã cho Asparagine (Asn-AAC). Tương tự, *H. taichui* (Lào) và *M. yokogawai* (Hàn Quốc) sử dụng nhiều nhất cho Phe-TTT và Leu-TTG và ít nhất cho Gln-CAA/Arg-CGC. Ngoài ra, *M. yokogawai* còn sử dụng ít nhất các bộ mã khác của Thr-ACA và Thr-ACC (Bảng 3).

Như vậy, những bộ mã sử dụng nhiều nhất của sán lá ruột nhỏ thuộc về Phenylalanine (Phe-TTT) và Leucine (Leu-TTG), những bộ mã ít sử dụng nhất là Glutamine (Gln-CAA), Arginine

(Arg-CGC), và có thể thêm Thr-ACA/ACC ở *M. yokogawai*, đó cũng là phương thức sử dụng mã nhiều nhất và ít nhất thường gặp ở hệ gen ty thể sán lá đã công bố trước đây (Le *et al.*, 2004; 2020; Chen *et al.*, 2013) (Bảng 3).

Đặc điểm vùng không mã hóa với những cấu trúc lặp liên kề

Vùng không mã hóa protein (non-coding region, NCR) ở mtDNA của sán lá ruột nhỏ *H. taichui*, được xác định giới hạn giữa 2 tRNA vận chuyển là tRNA^{Glu} và tRNA^{Gly}, độ dài 1.692 bp (mẫu Việt Nam) và 1.705 bp (mẫu Lào), được chia làm hai tiểu vùng chứa 5 cấu trúc lặp liên kề. Tiểu vùng thứ nhất chứa 2 cấu trúc lặp TRU1 và TRU2 (182 bp/mỗi một cấu trúc), cách tRNA^{Glu} bởi chuỗi giao gen dài 1 (Long Int seq. 1) có độ dài là 246 bp ở mẫu Việt Nam và 259 bp ở mẫu Lào, chênh lệch 13 bp (Bảng 2) (182 và 183 bp). Tiếp theo là chuỗi giao gen dài 2 (Long Int seq. 2) có độ dài là 487 bp ở cả chủng *H. taichui* của Việt Nam và chủng *H. taichui* của Lào. Tiểu vùng thứ hai chứa 3 cấu trúc lặp TRU3–TRU5 (183 bp/mỗi một cấu trúc), cách tRNA^{Gly} kế tiếp sau đó là 42 bp. Sự có mặt của các cấu trúc lặp liên kề hay còn gọi là các minisatellite trong NCR của hệ gen ty thể sán lá ruột nhỏ nói riêng và sán lá nói chung thể hiện đặc tính đa hình (polymorphism) chưa rõ chức năng. Vùng không mã hóa có cấu trúc lặp liên kề là nét đặc trưng ở mtDNA của một số loài sán lá, như bắt gặp trong mtDNA của sán lá ở họ Fasciolidae (sán lá gan lớn *Fasciola hepatica*, *F. gigantica*, *F. jacksoni*; sán lá ruột lớn *Fascioloides magna*, *Fasciolopsis buski*), ở họ Echinochasmidae và Echinostomatidae (sán lá ruột *Echinostoma revolutum*, *E. miyagawai*, *E. paraensei*, *E. hortense*, *Echinochasmus japonicus*, *Artyfechinostomum sufrartyfex*), ở họ Opisthorchiidae (sán lá gan nhỏ *Clonorchis sinensis*, *Opisthorchis viverrini*, *O. felinus*) và ở họ Schistosomatidae (sán máng *Schistosomaspp.*) được biết cho đến nay, tạo nên cấu trúc đa hình mang tính di truyền quần thể đang rất được quan tâm nghiên cứu (Le *et al.*, 2002; 2020; Shekhovtsov *et al.*, 2010; Cai *et al.*, 2012; Liu *et al.*, 2014; Li *et al.*, 2019; Ma *et al.*, 2016; 2017; Rajapakse *et al.*, 2020; Kinkar *et al.*, 2020).

Phân tích và so sánh đơn vị mã hóa ribosome

Cấu trúc đơn vị mã hóa ribosome của các loài *H. taichui* và *H. pumilio*

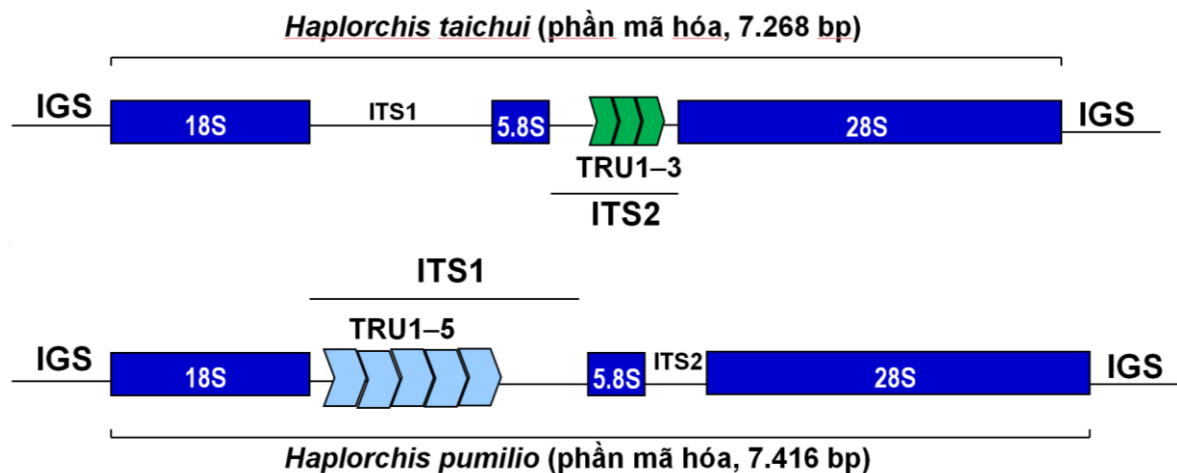
Các môi chung (liệt kê tại Le *et al.* (2017)) được sử dụng để thực hiện PCR thu nhận sản phẩm có độ dài khác nhau của rDNA từng loài *H. taichui* và *H. pumilio* và giải trình tự. Sau khi xử lý chuỗi, chúng tôi thu được toàn bộ phần mã hóa ribosome (từ 5' 18S đến 3' 28S) của mỗi loài (Hình 2; Bảng 4), bao gồm:

i) Chuỗi nucleotide của phần mã hóa ribosome (coding rDNA) của loài *H. taichui* (Htai-QT3-VN) có độ dài là 7.268 bp, còn thiếu

vùng IGS chưa thu nhận được;

ii) Chuỗi nucleotide phần mã hóa ribosome (coding rDNA) của loài *H. pumilio* (Hpum-HPU8-VN) có độ dài là 7.416 bp, còn thiếu vùng IGS chưa thu nhận được.

Trong phần mã hóa ribosome, cấu trúc của rDNA của từng loài và 5 phân đoạn DNA của 5 gen/vùng giao gen đã được xác định bao gồm: 18S rDNA, ITS1, 5.8S rDNA, ITS2 và 28S rDNA. Trình tự thu được đã hoàn tất toàn bộ phần mã hóa ribosome của cả hai loài *H. taichui* và *H. pumilio* mà trước đây chúng tôi mới công bố một phần và chưa phân tích một cách chính xác các cấu trúc lặp (Le *et al.*, 2017).



Hình 2. Sắp xếp trật tự phần mã hóa (coding rDNA) của đơn vị mã hóa ribosome của sản lá ruột nhỏ *H. taichui* và *H. pumilio* (mẫu Việt Nam). TRU: tandem repeat unit (các cấu trúc chuỗi lặp liền kề).

Độ dài gen và vùng giao gen của các loài *H. taichui* và *H. pumilio*

Dựa trên một số công bố trước đây về rDNA của một số loài sán dẹt (Briscoe *et al.*, 2016; Su *et al.*, 2018; Le *et al.*, 2017; 2020) và dữ liệu rDNA đã được phân tích công bố trong Ngân hàng gen, chúng tôi đã xác định được kích thước và vị trí của từng gen/vùng giao gen trong phần

mã hóa rTU của từng loài, bao gồm gen 18S rRNA, vùng giao gen ITS1, 5.8S rRNA, vùng giao gen ITS2, gen 28S rRNA (Bảng 4). Kết quả cho thấy, các gen 18S và 5.8S của cả 2 loài đều có độ dài hoàn toàn giống nhau, cụ thể: 1.992 bp ở gen 18S, 160 bp ở gen 5.8S, nhưng gen 28S có độ dài khác nhau, 3.875 bp ở loài *H. taichui* và 3.870 ở loài *H. pumilio*.

Bảng 4. Vị trí gen và vùng giao gen của đơn vị sao chép ribosome (rTU) của loài sán lá ruột nhỏ của Việt Nam, *H. taichui* (chủng QT3) (7.268 bp, Htai-QT3-VN) và *H. pumilio* (chủng HPU8) (7.416 bp, Hpum-HPU8-VN).

| Gen/Vùng giao gen | Vị trí (5'→3') | Cấu trúc lặp | Kích thước (bp) | Vùng giao gen (bp) | Ghi chú |
|--------------------------|----------------|--------------|-----------------|--------------------|-----------------------|
| <i>H. taichui</i> | | | | | GenBank: KX815126 |
| 7.268 bp | | | | | |
| 18S | 1–1992 | | 1992 | 0 | Gen rRNA |
| ITS1 | 1993–2779 | | 797 | 0 | Không có cấu trúc lặp |
| 5.8S | 2790–2949 | | 160 | 0 | Gen rRNA |
| ITS2 | 2950–3393 | | 444 | +121 | 121 bp đến TR1 |
| | 3071–3155 | TRU1 | 85 | 0 | Cấu trúc lặp liền kề |
| | 3156–3238 | TRU2 | 83 | 0 | Cấu trúc lặp liền kề |
| | 3239–3323 | TRU3 | 85 | +70 | Cấu trúc lặp liền kề |
| 28S | 3394–7268 | | 3875 | | Gen rRNA |
| IGS | // | | // | | Chưa hoàn thành |
| <i>H. pumilio</i> | | | | | GenBank: KX815125 |
| 7.416 bp | | | | | |
| 18S | 1–1992 | | 1992 | 0 | Gen rRNA |
| ITS1 | 1993–3098 | | 1106 | +66 | 66 bp đến TRA1 |
| | 2059–2194 | TRU1 | 136 | 0 | Cấu trúc lặp liền kề |
| | 2195–2330 | TRU2 | 136 | 0 | Cấu trúc lặp liền kề |
| | 2331–2466 | TRU3 | 136 | 0 | Cấu trúc lặp liền kề |
| | 2467–2582 | TRU4 | 116 | +7 | Cấu trúc lặp liền kề |
| | 2590–2705 | TRU5 | 116 | 0 | Cấu trúc lặp liền kề |
| | 2706–3098 | Int seq. | 393 | 0 | Chuỗi nối |
| 5.8S | 3099–3258 | | 160 | 0 | Gen rRNA |
| ITS2 | 3259–3546 | | 280 | 0 | Không có cấu trúc lặp |
| 28S | 3547–7416 | | 3870 | | Gen rRNA |
| IGS | // | | // | | Chưa hoàn thành |

Ghi chú: TRU: tandem repeat unit (các cấu trúc lặp liền kề).

Mặt khác, độ dài vùng giao gen ITS ở mỗi loài là khác nhau và cấu trúc chuỗi nucleotide trong mỗi vùng giao gen ở hai loài cũng khác nhau. Cụ thể:

- Vùng giao gen ITS1 ở *H. taichui* có độ dài là 797 bp, không chứa cấu trúc lặp (tandem repeat unit), trong khi đó ITS1 ở *H. pumilio* có

độ dài 1.106 bp chứa 5 cấu trúc lặp liền kề (TRU), trong đó 3 TRU (TRU1–3) có độ dài 136 bp/mỗi một cấu trúc và 2 TRU (TRU4–5) có độ dài 116 bp/mỗi một cấu trúc.

- Vùng giao gen ITS2 ở *H. taichui* có độ dài là 444 bp chứa 3 cấu trúc lặp liền kề (TRU), trong đó 2 TRU (TRU1 và 3) có độ dài 85 bp/mỗi một

cấu trúc, còn TRU2 chỉ có 83 bp. Trong khi đó, ITS2 ở *H. pumilio* có độ dài 280 bp không chứa cấu trúc lặp.

- Vùng IGS nằm ở đầu cuối 3' của rDNA của cả hai loài đều chưa xác định được.

KẾT LUẬN

Hệ gen ty thể và đơn vị mã hóa ribosome (phần mã hóa) của sán lá ruột nhỏ *H. taichui* mẫu của Việt Nam (chủng Htai-QT3-VN) và của *H. pumilio* mẫu của Việt Nam (chủng Hpum-HPU8-VN) đã được giải mã và phân tích đầy đủ. Toàn bộ mtDNA của *H. taichui* có kích thước 15.120 bp ngắn hơn so với một chủng của Lào 15.131 bp (GenBank: KF214770), 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 gồm *rrnL* (16S) và *rrnS* (12S); 22 gen RNA vận chuyển; và một vùng không mã hóa có các cấu trúc lặp liên kề (5 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 và với loài *M. yokogawai* (Hàn Quốc) cùng họ Heterophyidae. Phần mã hóa ribosome (5' 18S đến 3' 28S) của *H. taichui* và *H. pumilio* đã được thu nhận còn thiếu vùng IGS, trong đó loài *H. taichui* (Htai-QT3-VN) có độ dài là 7.268 bp, *H. pumilio* (Hpum-HPU8-VN) có độ dài là 7.416 bp; 5 phân đoạn DNA của 5 gen/vùng giao gen đã được xác định bao gồm: 18S rRNA, ITS1, 5.8S rRNA, ITS2 và 28S rRNA. Các gen 18S và 5.8S của cả 2 loài đều có độ dài giống nhau (1.992 bp/gen 18S, 160 bp/gen 5.8S), gen 28S có độ dài khác nhau (3.875 bp/*H. taichui* và 3.870 bp/*H. pumilio*). ITS1 ở *H. taichui* (797 bp) và ITS2 ở *H. pumilio* (280 bp) không chứa cấu trúc lặp, trong khi đó ITS1 ở *H. pumilio* (1.106 bp) chứa 5 cấu trúc lặp liên kề TRU, 136 bp cho 3 TRU và 116 bp cho 2 TRU; và ITS2 ở *H. taichui* (444 bp) chứa 3 TRU (83 – 85 bp/cấu trúc).

Lời cảm ơn: Nghiên cứu này được tài trợ bởi Quỹ Phát triển khoa học và công nghệ Quốc gia (NAFOSTED) trong đề tài mã số 106.06-2012.05

và 108.02-2017.09 do GS.TS Lê Thanh Hòa chủ nhiệm và nghiên cứu sinh (Lê Thị Việt Hà) tham gia là thành viên thực hiện đề tài.

TÀI LIỆU THAM KHẢO

- Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27: 573–580.
- Bernt M, Brabant A, Schierwater B, Stadler PF (2013) Genetic aspects of mitochondrial genome evolution. *Mol Phylogenet Evol* 69(2): 328–338.
- Biswal DK, Chatterjee A, Bhattacharya A, Tandon V (2014) The mitochondrial genome of *Paragonimus westermani* (Kerbert, 1878), the Indian isolate of the lung fluke representative of the family Paragonimidae (Trematoda). *Peer J* 2: e484.
- Blair D (2006) Ribosomal DNA variation in parasitic flatworms. In: Maule A, editor. Parasitic Flatworms: Molecular Biology, Biochemistry, Immunology and Control. CAB International pp 96–123.
- Boore JL (1999) Animal mitochondrial genomes. *Nucleic Acids Res* 27: 1767–1780.
- Briscoe AG, Bray RA, Brabec J, Littlewood DTJ (2016) The mitochondrial genome and ribosomal operon of *Brachycladium goliath* (Digenea: Brachycladiidae) recovered from a stranded minke whale. *Parasitol Int* 65(3): 271–275.
- 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 (2019) Chapter 1. Heterophyids in Human Intestinal Flukes: From Discovery to Treatment and Control. Springer, Dordrecht 2019. ISBN: 978-94-024-1702-9.
- Chai JY, Jung BK (2017) Fishborne zoonotic heterophyid infections: An update. *Food and Waterborne Parasitology* 8-9: 33–63.
- Chai JY, Jung BK (2019) Epidemiology of Trematode Infections: An Update. In: Toledo R., Fried B. (eds) Digenetic Trematodes. *Advances in Experimental Medicine and Biology*, vol 1154. Springer, Cham. https://doi.org/10.1007/978-3-030-18616-6_12.
- Chai JY, Jung BK (2020) Foodborne intestinal flukes:

- A brief review of epidemiology and geographical distribution. *Acta Trop* 201: 105210.
- Chai JY, Shin EH, Lee SH, Rim HJ (2009) Foodborne intestinal flukes in Southeast Asia. *Korean J Parasitol* 47 Suppl: S69–102. Review
- 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 plathelminth mitochondrial genomes. *Genet Mol Res* 12(1): 587–596.
- Chontanarith T, Anucherngchai S, Tejangkura T (2018) The rapid detection method by polymerase chain reaction for minute intestinal trematodes: *Haplorchis taichui* in intermediate snail hosts based on 18S ribosomal DNA. *J Parasit Dis* 42(3): 423–432.
- Crampton-Platt A, Yu DW, Zhou X, Vogler AP (2016) Mitochondrial metagenomics: letting the genes out of the bottle. *Gigascience* 5: 15. Review.
- 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: 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.
- Dung DT, Hop NT, Thaenkham U, Waikagul J (2013) Genetic differences among Vietnamese *Haplorchis taichui* populations using the COI genetic marker. *J Helminthol* 87(1): 66–70.
- Greiner S, Lehwerk P, Bock R (2019) Organellar Genome DRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res* 47(W1): W59–W64.
- Hoelzer K, Moreno Switt AI, Wiedmann M, Boor KJ (2018) Emerging needs and opportunities in foodborne disease detection and prevention: From tools to people. *Food Microbiol* 75: 65–71.
- Hu M, Gasser RB (2006) Mitochondrial genomes of parasitic nematodes - progress and perspectives. *Trends Parasitol* 22(2): 78–84.
- Johansen MV, Lier T, Sithithaworn P (2015) Towards improved diagnosis of neglected zoonotic trematodes using a One Health approach. *Acta Trop* 141(B): 161–169.
- Kinkar L, Young ND, Sohn W-M, Stroehlein AJ, Korhonen PK, Gasser RB (2020) First record of a tandem-repeat region within the mitochondrial genome of *Clonorchis sinensis* using a long-read sequencing approach. *PLoS Negl Trop Dis* 14(8): e0008552.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol Biol Evol* 35: 1547–1549.
- Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33: 1870–1874.
- Laslett D, Canback B (2008) ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* 24: 172–175
- 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 plathelminths. *Korean J Parasitol* 42(4): 159–167.
- Le TH, Nguyen KT, Nguyen NT, Doan HT, Dung DT, Blair D (2017) The ribosomal transcription units of *Haplorchis pumilio* and *H. taichui* and the use of 28S sequences for phylogenetic identification of common heterophyids in Vietnam. *Parasit Vectors* 10: 17.
- Le TH, Nguyen KT, Nguyen NTB, Doan HTT, Agatsuma T, Blair D (2019) The complete mitochondrial genome of *Paragonimusohirai* (Paragonimidae: Trematoda: Platyhelminthes) and its comparison with *P. westermani* congeners and other trematodes. *Peer J* 7: e7031.
- Le TH, Nguyen NT, Nguyen KT, Doan HT, Dung DT, Blair D (2016) A complete mitochondrial genome from *Echinochasmus japonicus* supports the elevation of Echinochasmidae Odhner, 1910 to family rank (Trematoda: Platyhelminthes). *Infect Genet Evol* 45: 369–377.
- Le TH, Pham LTK, Doan HTT, Le XTK, Saijuntha W, Rajapakse RPVJ, Lawton SP (2020) Comparative mitogenomics of the zoonotic parasite *Echinostoma revolutum* resolves taxonomic relationships within the ‘*E. revolutum*’ species group and the Echinostomata (Platyhelminthes: Digenea). *Parasitology* 147: 566–576.
- Lê Thanh Hòa, Nguyễn Thị Khuê, Nguyễn Thị Bích Nga, Đỗ Thị Roan, Đỗ Trung Dũng, Lê Thị Kim

- Xuyến, Đoàn Thị Thanh Hương (2016) 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. *Tạp chí Công nghệ Sinh học* 14(2): 215–224.
- Lee D, Choe S, Park H, Jeon HK, Chai JY, Sohn WM, Yong TS, Min DY, Rim HJ, EomKS (2013) Complete mitochondrial genome of *Haplorchis taichui* and comparative analysis with other trematodes. *Korean J Parasitol* 51(6): 719–726.
- Li Y, Qiu YY, Zeng MH, Diao PW, Chang QC, Gao Y, Zhang Y, Wang CR (2019) The complete mitochondrial genome of *Echinostoma miyagawai*: Comparisons with closely related species and phylogenetic implications. *Infect Genet Evol* 75: 103961.
- Liu GH, Gasser RB, Young ND, Song HQ, Ai L, Zhu XQ (2014) Complete mitochondrial genomes of the ‘intermediate form’ of *Fasciola* and *Fasciola gigantica*, and their comparison with *F. hepatica*. *Parasit Vectors* 7: 150.
- Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* 25: 955–964.
- Ma J, He JJ, Liu GH, Leontovyc R, Kašný M, Zhu XQ (2016) Complete mitochondrial genome of the giant liver fluke *Fascioloides magna* (Digenea: Fasciolidae) and its comparison with selected trematodes. *Parasit Vectors* 9: 429.
- Ma J, Sun MM, He JJ, Liu GH, Ai L, Chen MX, Zhu XQ (2017) *Fasciolopsis buski* (Digenea: Fasciolidae) from China and India may represent distinct taxa based on mitochondrial and nuclear ribosomal DNA sequences. *Parasit Vectors* 10(1): 101.
- McStay B (2016) Nucleolar organizer regions: genomic ‘dark matter’ requiring illumination. *Genes & Development* 30(14): 1598–1610.
- Nguyễn Thị Khuê, Đỗ Trung Dũng, Lê Thanh Hòa (2015) Phân tích tương đồng gen ty thể *cox1* và thẩm định loài sán 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 Ký 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.
- Oey H, Zakrzewski M, Narain K, Devi KR, Agatsuma T, Nawaratna S, Gobert GN, Jones MK, Ragan MA, McManus DP, Krause L (2019) Whole-genome sequence of the oriental lung fluke *Paragonimus westermani*. *Gigascience* 1: 8(1).
- Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *Journal of Molecular Evolution* 41: 353–358.
- Qian L, Zhou P, Li W, Wang H, Miao T, Hu L (2018) Characterization of the complete mitochondrial genome of the lung fluke, *Paragonimus heterotremus*. *Mitochondrial DNA Part B* 3(2): 560–561.
- Rajapakse RPVJ, Lawton SP, Karunathilake KJK, Perera BVP, Nguyen NTB, Le TH (2019) Molecular characterization of *Fasciola jacksoni* from wild elephants (*Elephas maximus maximus*) of Sri Lanka; a taxonomic evaluation. *Parasitology* 146(10): 1247–1255.
- Santos CP, Borges JN (2020) Current Knowledge of Small Flukes (Digenea: Heterophyidae) from South America. *Korean J Parasitol* 58(4): 373–386.
- Shekhovtsov SV, Katokhin AV, Kolchanov NA and Mordvinov VA (2010) The complete mitochondrial genomes of the liver flukes *Opisthorchis felineus* and *Clonorchis sinensis* (Trematoda). *Parasitol Int* 59(1): 100–103.
- Su X, Zhang Y, Zheng X, Wang XX, Li Y, Li Q, Wang CR (2018) Characterization of the complete nuclear ribosomal DNA sequences of *Eurytrema pancreaticum*. *Journal of Helminthology* 92: 484–490.
- Thaenkham U, Nawa Y, Blair D, Pakdee W (2011) Confirmation of the paraphyletic relationship between families Opisthorchiidae and Heterophyidae using small and large subunit ribosomal DNA sequences. *Parasitol Int* 60(4): 521–523.
- Van KV, Dalgaard A, Blair D, Le TH (2009) *Haplorchis pumilio* and *H. taichui* in Vietnam discriminated using ITS-2 DNA sequence data from adults and larvae. *Exp Parasitol* 123: 146–151.
- Wang T, Wang Y, Xu F, Li X, Qu R, Song L, Tang Y, Lin P (2018) Characterization of the complete mitochondrial genome of the lung fluke, *Paragonimus kellicotti*. *Mitochondrial DNA Part B* 3(2): 715–716.

Weider LJ, Elser JJ, Crease TJ, Mateos M, Cotner JB, Markow TA (2005) The functional significance of ribosomal rDNA variation: Impacts on the evolutionary ecology of organisms. *Annu Rev Ecol Evol Syst* 36: 219–242.

Yang X, Gasser RB, Koehler AV, Wang L, Zhu K, Chen L, Feng H, Hu M, Fang R (2015) Mitochondrial genome of *Hypoderaeum conoideum* - comparison with selected trematodes. *Parasit Vectors* 8: 97.

COMPARATIVE ANALYSIS OF THE CHARACTERISTICS OF *HAPLORCHIS TAICHUI* MITOCHONDRIAL GENOME WITH *METAGONIMUS YOKOGAWAI* AND ITS RIBOSOMAL TRANSCRIPTION UNIT WITH *H. PUMILIO* (FAMILY HETEROPHYIDAE)

Le Thi Viet Ha¹, Nguyen Thi Khue², Dong Van Quyen^{2,3}, Le Thanh Hoa^{2,3}

¹Vietnam University of Traditional Medicine

²Institute of Biotechnology, Vietnam Academy of Science and Technology

³Graduate University of Science and Technology, Vietnam Academy of Science and Technology

SUMMARY

Minute intestinal flukes, *Haplorchis taichui* and *H. pumilio*, belong to the family Heterophyidae (Trematoda: Platyhelminthes), which have been studied very limited, especially the molecular markers of the mitochondrial genomes (mtDNA) and the ribosome transcription units (rTU or rDNA). We have obtained the complete mitochondrial genome of *H. taichui* and the coding part of ribosome transcription unit of *H. taichui* and *H. pumilio* of Vietnam. Nucleotide and amino acid data were compared between *H. taichui* and *Metagonimus yokogawai* for genomic/gene composition, codon usage (skew/skewness), and tandem repeat units (TRU). The complete mtDNA of *H. taichui* (strain Htai-QT3-VN) with the length of 15,120 bp and *M. yokogawai* (15,258 bp; Korea; KC330755) contains 36 genes, including 12 protein-coding genes (*cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4L*, *nad4*, *nad5*, *nad6*, *atp6* and *cob*), 2 ribosomal RNA genes (rRNA); 22 transfer RNA (tRNA or *trn*) and a noncoding region (NCR) between *trnE* and *trnG*, divided into 2 sub-regions containing 5 TRUs (182–183 bp/TRU). *H. taichui* (Vietnam and Laos) uses A = 19.56%, T = 39.71%, G = 28.34%, C = 12.39% (A + T is 59.27% and G + C is 40.73%) for mtDNA construction, whose skew/skewness value at A+T is negative (−0.340) and G+C is positive (0.392); for 12 protein-coding genes (PCGs) is similar; but for the mito-ribosomal genes (MRGs, of 16S/*rrnL* and 12S/*rrnS*) it is less for A+T (57.22%) and more for G+C (42.78%). *M. yokogawai* had lower A+T (mtDNA/55.68%; PCGs/55.96%; MRGs/54.15%) and higher G+C usage rate than *H. taichui*. *H. taichui* of Vietnam and Laos has 10,164 bp encoding for 3,376 amino acids to construct 12 PCGs with the mostly used codons as Phenylalanine (Phe-TTT) and Leucine (Leu-TTG), and the leastly used codons as Glutamine (Gln-CAA), Arginine (Arg-CGC). Additional codon, Thr-ACA/ACC can be added as the least used in *M. yokogawai*. The rTU (from 5' 18S to 3' 28S) of *H. taichui* (7,268 bp) and *H. pumilio* (7,416 bp) were identified with 5 genomic regions including 18S rDNA, ITS1, 5.8S rDNA, ITS2 and 28S rDNA. The 18S and 5.8S genes of both species were of the same length (1,992 bp for 18S, 160 bp for 5.8S), but different for 28S genes (3,875 bp for *H. taichui* and 3,870 bp for *H. pumilio*). ITS1 in *H. taichui* (797 bp) and ITS2 in *H. pumilio* (280 bp) do not contain TRUs, whilst ITS1 in *H. pumilio* (1,106 bp) contains 5 TRUs (136 bp for 3 TRU and 116 bp for 2 TRUs); and ITS2 in *H. taichui* (444 bp) contain 3 TRUs (83–85 bp/each).

Keywords: Ribosome transcription unit, *Haplorchis pumilio*, *Haplorchis taichui*, mitochondrial genome, mtDNA, *Metagonimus yokogawai*, rTU, minute intestinal flukes.