

# PHÂN TÍCH PHẢ HỆ VÀ KHOẢNG CÁCH DI TRUYỀN SÁN LÁ PHỔI *PARAGONIMUS* CỦA VIỆT NAM SỬ DỤNG DỮ LIỆU TOÀN BỘ HỆ GEN TY THỂ

Nguyễn Thị Khuê<sup>1,2</sup>, Đỗ Thị Roan<sup>1,2</sup>, Phạm Thị Khánh Linh<sup>1</sup>,  
Đoàn Thị Thanh Hương<sup>1,2</sup>, Phạm Ngọc Doanh<sup>3</sup>, Lê Thanh Hòa<sup>1,2\*</sup>

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

<sup>2</sup>Học viện Khoa học và Công nghệ Việt Nam, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

<sup>3</sup>Viện Sinh thái và Tài nguyên sinh vật

## TÓM TẮT

Bằng công nghệ Sanger và thế hệ mới (NGS), hệ gen ty thể (mtDNA) của loài sán lá phổi *Paragonimus heterotremus* của Việt Nam (chủng D2YB) và *P. westermani* (chủng QT2) đã được thu nhận, có độ dài là 16.268 bp/D2YB và 14.779 bp/QT2. Có 12 gen mã hóa protein (protein-coding genes, PCGs); 2 gen RNA ribosome ty thể (mitoribosomal genes, MRGs); 22 gen RNA vận chuyển (transfer RNA, tRNAs); và một vùng không mã hóa (non-coding region, NCR), chứa nhiều cấu trúc lặp dài (LRU) và lặp ngắn (SRU). Phân tích phả hệ dựa trên chuỗi amino acid cộng hợp của toàn bộ 12 gen mã hóa protein của 32 chủng/loài bao gồm họ Paragonimidae (15 chuỗi), họ Heterophyidae (4 chuỗi), họ Opisthorchiidae (12 chuỗi) và một chuỗi họ Schistomatidae (làm nhóm ngoại hợp) cho thấy có hai phân nhóm trong nhóm họ Paragonimidae: nhóm chỉ có các chủng của loài *P. westermani* và nhóm của tất cả các chủng của các loài *Paragonimus* khác của thế giới. Tính toán khoảng cách di truyền (KCDT) cho thấy giá trị KCDT ngoại loài cao nhất là giữa loài đồng danh *P. ohirai*/*P. iloktsuenensis* với *P. westermani* nhóm Ấn Độ (19,93-21,23%) và thấp nhất là giữa *P. heterotremus* với *P. skrjabini miyazakii*, chỉ là 12,20-12,47%. Đặc biệt, KCDT nội loài của phân nhóm *P. westermani* Nam Á của các chủng Ấn Độ với các chủng của phân nhóm *P. westermani* Đông Á (Việt Nam, Trung Quốc và Hàn Quốc) rất cao (12,1-12,7%), tương đương với giá trị KCDT ngoại loài giữa *P. heterotremus* với *P. skrjabini miyazakii*. Đặc điểm di truyền, cấu trúc của mtDNA và KCDT của sán lá phổi *Paragonimus*, cho phép sử dụng làm đích khai thác cho nghiên cứu di truyền quần thể, phả hệ, nguồn gốc, tiến hóa, đồng thời dữ liệu gen học sử dụng cho phát triển chẩn đoán, định hướng nghiên cứu tầm định loài/giống/họ ký sinh trùng động vật lây sang người.

**Từ khóa:** Hệ gen ty thể (mtDNA), họ Paragonimidae, khoảng cách di truyền, phả hệ, sán lá phổi *Paragonimus*.

## MỞ ĐẦU

Bệnh sán lá phổi (paragonimiasis) do các loài thuộc chi *Paragonimus*, họ Paragonimidae, phân bộ Troglotremata (Platyhelminthes: Trematoda) gây ra là một trong những bệnh nguy hiểm ở nhiều quốc gia tại châu Á, châu Phi và châu Mỹ, trong đó có Việt Nam (Yoshida *et al.*, 2019; Blair *et al.*, 2022). Hai loài sán lá phổi phổ biến nhất có thể kể đến là *P. heterotremus* và *P. westermani* chủ yếu ở châu Á cùng với *P. ohirai*, *P. skrjabini miyazaki*, *P. skrjabini*, *P. iloktsuenensis*, ngoài ra còn có *P. africanus*, *P. kellicotti* và *P. mexicanus* ở châu Phi và châu Mỹ. Trên 50 loài sán lá phổi được phân chia vào các phức hệ theo nhóm *P. westermani*, nhóm *P. heterotremus*, nhóm *P. s. miyazaki*, nhóm *P. ohirai* ... Nhóm *P. heterotremus*, phân bố ở Nam Trung Quốc, Ấn Độ, Bangladesh, Sri Lanka, Thái Lan, Lào và Tây-Bắc Việt Nam và *P. westermani* có nhiều biến chủng khác nhau ở Đông-Nam Siberi (Nga), Ấn Độ, Sri Lanka, Nepal, Pakistan, Hàn Quốc, Nhật Bản, Trung Quốc, Đài Loan, Philippines, Cambodia, Lào và Việt Nam (Le *et al.*, 2006; Doanh *et al.*, 2013; Blair, 2022). Bệnh sán lá phổi thường xảy ra chủ yếu ở cộng đồng dân tộc thiểu số vùng cao, vùng núi xa xôi, trẻ em tiểu học và trung học do có thói quen ăn cua/giáp xác nhiễm ấu trùng sán lá phổi gây bệnh chưa được nướng/nấu chín (Doanh *et al.*, 2013; Blair, 2022). Ở Việt Nam, loài *P. heterotremus* được phát hiện đầu tiên năm 1993 ở Sìn Hồ, Lai Châu và tiếp theo nhiều địa danh khác ở Yên Bái, Sơn La, Tuyên Quang, Hòa Bình và được nghiên cứu khá chi tiết về dịch tễ học và sinh học phân tử (Kino *et al.*, 1995; Le *et al.*, 2006; Doanh *et al.*, 2013). Loài *P. westermani* mới được ghi nhận gần đây trên vật chủ trung gian là cua suối ở Quảng Trị, chưa phát hiện trên người (Doanh *et al.*, 2009; 2013).

Phân loại và nghiên cứu phả hệ, tiến hóa và di truyền quần thể, ngoài phân tích đặc điểm hình thái học, đòi hỏi cần có những chỉ thị phân tử và đặc điểm gen/hệ gen của loài sinh vật, trong đó có ký sinh trùng. Hệ gen ty thể (mtDNA) là nguồn cung cấp chỉ thị phân tử tối ưu nhất và được sử dụng trong hơn 40 năm qua (Chan *et al.*, 2021). Ở sán lá, đó là vòng DNA khép kín, khoảng 13,5-23 kb, chứa 12 gen mã hóa protein gồm *cox1-3*, *nad1-6*,

*nad4L*, *atp6* và *cob*; 2 gen RNA ribosome ty thể gồm *rnl* (large ribosomal rRNA hay 16S) và *rns* (small rRNA hay 12S); 22 gen RNA vận chuyển; và một vùng không mã hóa dài ngắn khác nhau tùy loài và thường có cấu trúc lặp khá phức tạp. Đối với loài/chủng mới phát hiện hay mới ghi nhận, chỉ thị mtDNA càng được ứng dụng rộng rãi. Chính vì vậy, đã có hàng chục ngân mtDNA của các loài sinh vật, trong đó có hàng chục mtDNA của sán lá đã được giải trình tự thành công và được ứng dụng hiệu quả trong nghiên cứu bệnh động vật lây sang người (Wey-Fabrizius *et al.*, 2013; Chan *et al.*, 2021).

Trong bài báo này, chúng tôi giới thiệu ứng dụng chỉ thị phân tử từ hệ gen ty thể hoàn chỉnh của một số loài sán lá phổi của Việt Nam, trong đó có *Paragonimus heterotremus*, mẫu Việt Nam (chủng LC ở Lai Châu và D2YB ở Yên Bái) và *P. westermani* (chủng QT2, Quảng Trị) trong nghiên cứu phả hệ và vị trí phân loại của các chủng/loài này trong họ Paragonimidae (phân bộ Troglotremata) và các họ Heterophyidae và Opisthorchiidae (phân bộ Opisthorchiata), cũng như khoảng cách di truyền của 15 chủng/loài *Paragonimus* trong họ Paragonimidae.

## NGUYÊN LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

### Nguyên liệu mẫu sán và tách chiết DNA tổng số

Mẫu *Paragonimus heterotremus* là sán lá phổi trưởng thành (thu ở Yên Bái, Việt Nam), ký hiệu Phet-D2YB-VN và mẫu *P. westermani* (thu ở Quảng Trị), ký hiệu Pwes-QT2-VN. Tất cả các thao tác đều sử dụng các bộ kit của hãng Thermo Fisher Scientific Inc. (MA, USA) và DNA tổng số được tách chiết bằng bộ kit GeneJET™ Genomic DNA Purification Kit theo hướng dẫn của nhà sản xuất, được bảo quản ở -20°C. Khuôn DNA tổng số được sử dụng 2 µL ở nồng độ 50 ng/µL cho mỗi phản ứng PCR và LPCR. Các chuỗi nucleotide được thu nhận từ sản phẩm LPCR bằng công nghệ giải trình tự thế hệ mới NGS (hệ thống PACBIO SEQUEL system)

### Thiết kế môi và thực hiện PCR

Một số cặp môi cơ bản (môi chung) được thiết kế dựa vào so sánh trình tự của các gen chọn lọc lấy từ các mtDNA có trong Ngân hàng gen (Le *et al.*, 2022). LPCR được thực hiện sử dụng bộ sinh phẩm Phusion High-Fidelity PCR Kit 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, 52°C/30 giây và 72°C/1-4 phút; chu kỳ cuối kéo dài 10 phút ở 72°C và đối với LPCR, thực hiện chu kỳ cuối ở 72°C/6-12 phút. Sản phẩm PCR được tinh sạch bằng GeneJET™ Genomic DNA Purification Kit và giải trình tự trực tiếp (hoặc tách dòng) hoặc NGS.

### Giải trình tự bằng phương pháp Sanger và thế hệ mới (NGS)

Chủng Pwes-QT2-VN được giải trình tự bằng phương pháp Sanger; sản phẩm PCR hoặc DNA plasmid tái tổ hợp được gửi giải trình tự tại cơ sở dịch vụ nước ngoài (Macrogen, Hàn Quốc). Trình tự nucleotide từ giải trình tự được xử lý bằng chương trình Chromas 2.6.6 (<http://technelysium.com.au/wp/chromas/>) để thu nhận chuỗi cuối cùng.

Chủng Phet-D2YB-VN bằng NGS sử dụng hệ thống PACBIO SEQUEL có tại Viện Công nghệ sinh học như đã thực hiện với loài *P. skrjabini miyazaki* (Le *et al.*, 2022). Các sản phẩm LPCR (khoảng 4-8 kb/mỗi sản phẩm) bao phủ hết toàn bộ vòng mtDNA được kiểm tra, tinh sạch bằng bộ sinh phẩm GeneJET PCR Purification Kit. Các sản phẩm tinh sạch được gom vào cùng một ống và tiếp tục được tinh sạch lại bằng AMPure XP beads. Thư viện giải trình tự thế hệ mới được chuẩn bị bằng Express Template Prep Kit 2.0 và được đưa vào hệ thống giải trình tự đoạn dài (long-read sequencing). Tín hiệu giải trình tự được xử lý, đánh giá và chuyển đổi thành chuỗi thô, sau đó sắp xếp căn chỉnh thành các lớp chuỗi nucleotide (contig) bằng chương trình PacBio's circular consensus sequencing (CCS), hợp nhất *de novo* bằng phần mềm Canu software v2.0, và kiểm tra chất lượng bằng phần mềm Quast software v5.0.2.

### Phân tích chuỗi gen và hệ gen

Sắp xếp căn chỉnh, xác định cấu trúc gen và toàn bộ hệ gen ty thể được thực hiện trực tuyến (online) qua chương trình Gene Infinity (<http://www.geneinfinity.org/>) hoặc các chương trình so sánh có tại ExPasy (<https://www.expasy.org/>). Thành phần amino acid của mã hóa protein được suy diễn bằng bộ mã của mtDNA của sán đẹt (echinoderm mitochondrial genetic code, translation table 9 trong GenBank) sử dụng GENEDOC2.7 (<http://w.w.w.psc.edu/biomed/genedoc>). Độ dài và cấu trúc tRNA dựa vào tRNAscan-SE 1.21 (<http://lowelab.ucsc.edu/tRNAscan-SE/>) và ARWEN (<http://130.235.244.92/bcgi/arwen.cgi>); các cấu trúc lặp (repeat unit, RU) bằng phần mềm Tandem Repeats Finder v3.01 có tại (<https://tandem.bu.edu/trf/trf.html>).

### Phân tích xây dựng cây phả hệ và tính toán khoảng cách di truyền

Trình tự amino acid cộng hợp của 12 gen mã hóa protein (khoảng 3354 aa đến 3406 aa) từ 32 chủng/loài của 3 họ sán lá (Paragonimidae, Heterophyidae và Opisthorchiidae) liệt kê trong Bảng 1 được sử dụng làm thành phần nội hợp (ingroup) trong phân tích phả hệ. *Schistosoma haematobium* (họ Schistosomatidae) được sử dụng làm nhóm ngoại hợp (outgroup). Họ Paragonimidae gồm 15 chủng thuộc các loài *P. heterotremus*, *P. westermani*, *P. ohirai*, *P. iloktsuenensis*, *P. s. miyazakii* và *P. kellycotti* ngoài tham gia thành phần phân tích phả hệ còn được sử dụng tính toán khoảng cách di truyền. Sắp xếp căn chỉnh của 32 chuỗi amino acid được thực hiện bằng

HỘI NGHỊ KHOA HỌC TOÀN QUỐC VỀ CÔNG NGHỆ SINH HỌC 2024

GENEDOC2.7 và MAFFT 7.122. Phân tích phả hệ thực hiện trực tuyến bằng PhyML tại <https://ngphylogeny.fr/>, cây phả hệ được xây dựng bằng chương trình MEGA 11 sử dụng phương pháp “*tiếp cận cực đại*” (ML, maximum-likelihood), thuật toán Jones-Taylor-Thornton (JTT) và Nearest-Neighbor-Interchange (NNI), với hệ số tin tưởng bootstrap 1000 lần lặp lại. Khoảng cách di truyền song chuỗi (pairwise genetic distance) được tính toán giữa 14 chủng/loài *Paragonimus* sử dụng chương trình MEGA 11, thông số Kimura-2, phương pháp p-distance và hệ số tin tưởng bootstrap 1000 lần lặp lại (Tamura *et al.*, 2021).

**Bảng 1. Danh sách và thông tin loài/chủng cung cấp toàn bộ chuỗi nucleotide hệ gen và 12 gen mã hóa protein (nucleotide và amino acid) sử dụng phân tích phả hệ của các loài *Paragonimus* của Việt Nam với các chủng/loài trong các họ Paragonimidae, Heterophyidae, Opisthorchiidae và Schistosomatidae trong nghiên cứu này**

| Họ/Loài/Chủng và nước phân lập                            | Số ngân hàng gen | Ký hiệu chủng      | mtDNA (bp) | PCGs (bp) | PCGs (aa) | Nước thu thập | Tài liệu (nếu có)                |
|---|------------------|--------------------|------------|-----------|-----------|---------------|----------------------------------|
| <b>Paragonimidae</b>                                      |                  |                    |            |           |           |               |                                  |
| 1 <i>Paragonimus heterotremus</i> -(D2YB)-Vietnam         | ==               | Phet-(D2YB)-VN     | 16.268     | 10,101    | 3355      | Vietnam       | Nghiên cứu này                   |
| 2 <i>Paragonimus heterotremus</i> -(LC)-Vietnam           | KY952166         | Phet-(LC)-VN       | 15.898     | 10,101    | 3355      | Vietnam       | GenBank                          |
| 3 <i>Paragonimus heterotremus</i> -(GX)-China             | MH059809         | Phet-(GX)-CN       | 13.927     | 10,101    | 3355      | China         | Qian <i>et al.</i> (2018)        |
| 4 <i>Paragonimus westermani</i> -(QT2)-Vietnam            | ==               | Pwes-(QT2)-VN      | 14.779     | 10,104    | 3356      | Vietnam       | Nghiên cứu này                   |
| 5 <i>Paragonimus westermani</i> -(dog1)-China             | MN412705         | Pwes-(dog1)-CN     | 14,790     | 10,104    | 3356      | China         | GenBank                          |
| 6 <i>Paragonimus westermani</i> -(dog2)-China             | MN412706         | Pwes-(dog2)-CN     | 14.774     | 10,104    | 3356      | China         | GenBank                          |
| 7 <i>Paragonimus westermani</i> (3n)-(Bogil)-Korea        | AF219379         | Pwes3n-(Bogil)-KR  | 14,965     | 10,101    | 3355      | South Korea   | GenBank                          |
| 8 <i>Paragonimus westermani</i> (2n)-(Haenam)-Korea       | AF540958         | Pwes2n-(Haenam)-KR | 14,244     | 10,101    | 3355      | South Korea   | GenBank                          |
| 9 <i>Paragonimus westermani</i> -(Type I)-India           | KM280646         | Pwes-(Type 1)-IN   | 14,015     | 10,104    | 3356      | India         | GenBank                          |
| 10 <i>Paragonimus westermani</i> -(AP)-India              | KX943544         | Pwes-(AP)-IN       | 14,975     | 10,104    | 3356      | India         | Biswal <i>et al.</i> (2014)      |
| 11 <i>Paragonimus westermani</i> -(IND2009)-India         | CM017921         | Pwes-(IN2009)-IN   | 20,273     | 10,104    | 3356      | India         | Oye <i>et al.</i> (2019b)        |
| 12 <i>Paragonimus ohirai</i> (Kino)-Japan                 | KX765277         | Pohi-(Kino)-JP     | 14,818     | 10,104    | 3356      | Japan         | Le <i>et al.</i> (2019)          |
| 13 <i>Paragonimus iloktsuensis</i> -(Amami)-Japan         | ON961029         | Pilo-(Amami)-JP    | 14,827     | 10,104    | 3356      | Japan         | Le <i>et al.</i> (2023)          |
| 14 <i>Paragonimus skrjabini miyazakii</i> -(OkuST1)-Japan | ON782295         | Psmiy-(OkuST1)-JP  | 17,591     | 10,098    | 3354      | Japan         | Le <i>et al.</i> (2022)          |
| 15 <i>Paragonimus kelicotti</i> -(Ozark)-United States    | MH322000         | Pkel-(Ozark)-US    | 13,786     | 10,098    | 3354      | United States | Wang <i>et al.</i> (2018)        |
| <b>Heterophyidae</b>                                      |                  |                    |            |           |           |               |                                  |
| 16 <i>Cryptocotyle lingua</i> -(66766)-Norway             | OL853496         | Clin-(66766)-NO    | 13,983     | 10173     | 3379      | Norway        |                                  |
| 17 <i>Haplorchis taichui</i> -(QT3)-Vietnam               | MG972809         | Htai-(QT3)-VN      | 15,120     | 10,164    | 3379      | Vietnam       | GenBank                          |
| 18 <i>Haplorchis taichui</i> -(Laos)-Laos                 | KF214770         | Htai-(Laos)-LA     | 15,131     | 10,164    | 3392      | Laos          | Lee <i>et al.</i> (2013)         |
| 19 <i>Metagonimus yokogawai</i> -Korea                    | KC330755         | Myok--KR           | 15,258     | 10,245    | 3405      | South Korea   | GenBank                          |
| <b>Opisthorchiidae</b>                                    |                  |                    |            |           |           |               |                                  |
| 20 <i>Clonorchis sinensis</i> -(Amur)-Russia              | FJ381664         | Csin-(Amur)-RU     | 13,875     | 10,209    | 3403      | Russia        | Shekhovtsov <i>et al.</i> (2010) |
| 21 <i>Clonorchis sinensis</i> -(ND)-Vietnam               | ==               | Csin-(ND)-VN       | 14,461     | 10,209    | 3394      | Vietnam       | Nghiên cứu này                   |
| 22 <i>Clonorchis sinensis</i> -(GD)-China                 | JF729303         | Csin-(GD)-CN       | 13,879     | 10,209    | 3394      | China         | Cai <i>et al.</i> (2012)         |
| 23 <i>Clonorchis sinensis</i> -Korea                      | JF729304         | Csin-KR            | 13,877     | 10,209    | 3394      | South Korea   | Cai <i>et al.</i> (2012)         |
| 24 <i>Clonorchis sinensis</i> -(Jinju)-Korea              | MT607652         | Csin-(Jinju)-KR    | 18,304     | 10,209    | 3391      | South Korea   | Kinkar <i>et al.</i> (2020)      |
| 25 <i>Opisthorchis viverrini</i> -Laos                    | JF739555         | Oviv-LA            | 13,510     | 10,206    | 3391      | Laos          | Cai <i>et al.</i> (2012)         |
| 26 <i>Opisthorchis viverrini</i> -(KK)-Thailand           | ==               | Oviv-(KK)-TH       | 14,046     | 10,200    | 3400      | Thailand      | Nghiên cứu này                   |
| 27 <i>Opisthorchis felineus</i> -(UstTula)-Russia         | EU921260         | Ofel-(UstTula)-RU  | 14,277     | 10,218    | 3406      | Russia        | Shekhovtsov <i>et al.</i> (2010) |
| 28 <i>Opisthorchis sudarikovi</i> -(Swabi)-Pakistan       | MK033132         | Osud-(Swabi)-PK    | 13,641     | 10,197    | 3397      | Pakistan      | Suleman <i>et al.</i> (2019)     |
| 29 <i>Opisthorchis parageminus</i> -(PC6)-Vietnam         | ==               | Opar-(PC6)-VN      | 13,540     | 10,224    | 3396      | Vietnam       | Nghiên cứu này                   |
| 30 <i>Amphimerus</i> sp.-(JM2019)-Ecuador                 | MK238506         | AmpSP-(JM2019)-EC  | 15,151     | 10,179    | 3381      | Ecuador       | Ma <i>et al.</i> (2019)          |
| 31 <i>Metorchis orientalis</i> -(HLJ)-China               | KT239342         | Mori-(HLJ)-CN      | 13,834     | 10,176    | 3392      | China         | Na <i>et al.</i> (2016)          |
| <b>Schistosomatidae outgroup</b>                          |                  |                    |            |           |           |               |                                  |
| 32 <i>Schistosoma haematobium</i> -(N10)-Mali             | DQ157222         | Shae-(N10)-ML      | 15,003     | 10,116    | 3365      | Mali          | Littlewood <i>et al.</i> (2006)  |

mtDNA: Mitochondrial genome; PCGs: Gen mã hóa protein (protein-coding genes); (=): Đang đăng ký.

**KẾT QUẢ VÀ THẢO LUẬN**

**Hệ gen ty thể của sán lá phổi *Paragonimus heterotremus* (D2YB) và *P. westermani* (QT2) của Việt Nam**

Hệ gen ty thể hoàn chỉnh của sán lá phổi *P. heterotremus* chủng D2YB (Phet-D2YB-VN) là vòng DNA khép kín có độ dài 16.268 bp và của *P. westermani* chủng QT2 (Pwes-QT2-VN) là 14.779 bp. Cả hai hệ gen đều có 12 gen mã hóa protein gồm *cox1-3*, *nad1-6*, *nad4L*, *atp6* và *cob*; 2 gen RNA ribosome ty thể (MRG) là *rrnL* (16S) và *rrnS* (12S); 22 gen RNA vận chuyển (tRNA); và một vùng không mã hóa (NCR), chứa nhiều cấu trúc lặp dài (LRU, long repeat unit) và lặp ngắn (SRU, short repeat unit). Bản đồ hệ gen mtDNA mạch thẳng mở vòng tại 5' gen *cox3*, của Phet-D2YB-VN và Pwes-QT2-VN đều có trật tự sắp xếp: 5'-*cox3*-H-*cob*-*nad4L*-*nad4*-QFM-*atp6*-*nad2*-VAD-*nad1*-NPIK-*nad3*-S1W-*cox1*-T-*rrnL*-C-*rrnS*-*cox2*-*nad6*-YL1S2L2R-*nad5*-GE-NCR[LRU-SRU]-3'. Độ dài và sắp xếp gen trong vùng mã hóa của mtDNA ở hai loài Phet-D2YB-VN và Pwes-QT2-VN cơ bản giống nhau. Tuy nhiên, vùng NCR của 2 chủng/loài này có khác nhau và thể hiện tính đa hình (polymorphism) của sán lá phổi bởi số lượng và kích thước khác nhau của các cấu trúc lặp dài (*P. heterotremus* D2YB có 6 LRU/370 bp mỗi đơn vị và 5 SRU/114 bp mỗi đơn vị và *P. westermani* QT2 có 8 LRU/119 bp mỗi đơn vị; 2 SRU/118 bp mỗi đơn vị). Đặc biệt, cấu trúc lặp dài LRU của *P. heterotremus* (D2YB) đều chứa toàn bộ chuỗi nucleotide của gen RNA vận chuyển amino acid glutamine (tRNA<sup>Glu</sup>) định vị ở giữa của mỗi một LRU, và đó cũng là một đặc điểm cấu trúc mới phát hiện gần đây ở vùng NCR của loài sán lá phổi *P. s. miyazakii* (Le et al., 2022).

**Phân tích phả hệ của sán lá phổi *Paragonimus***

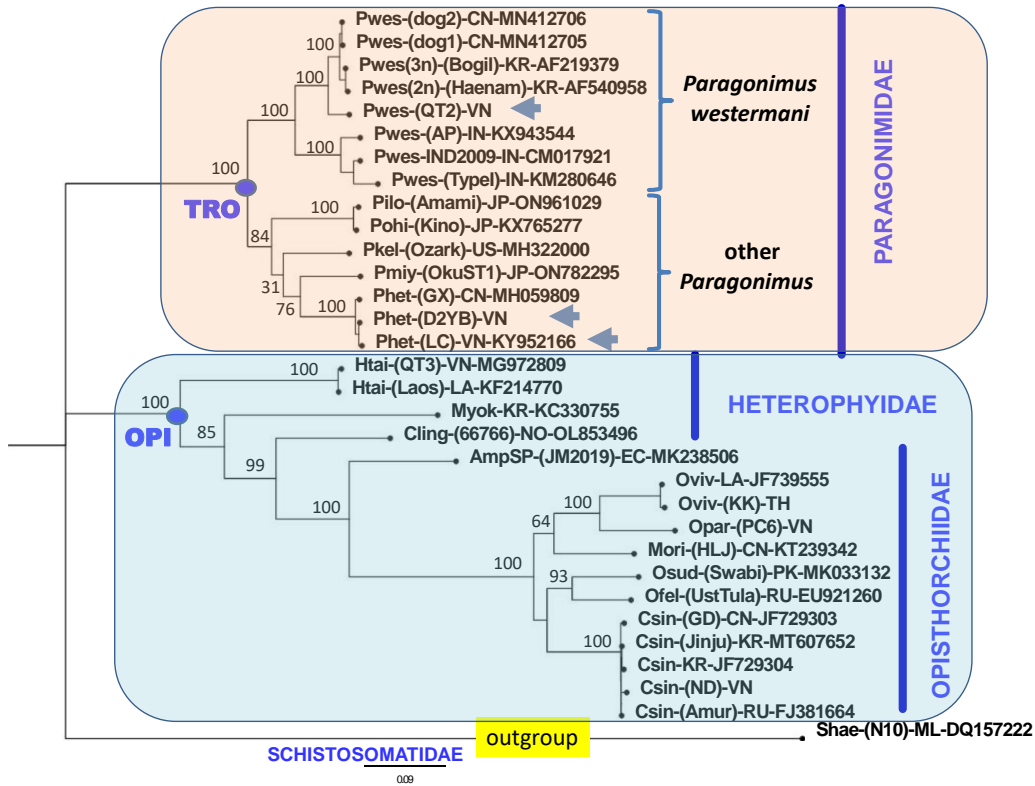
Cây phả hệ của 32 chủng/loài (liệt kê ở **Bảng 1**) dựa trên phân tích chuỗi aa cộng hợp của 12 gen mã hóa protein bằng MEGA X được trình bày ở **Hình 1**. Trên cây phả hệ có nhóm phân định tách biệt nhau: i) nhóm thứ nhất là Troglotremata (ký hiệu là TRO), gồm 15 chủng *Paragonimus* (họ Paragonimidae, phân bộ Troglotremata) và nhóm thứ hai là Opisthorchiata (ký hiệu là OPI) gồm 16 chủng, trong đó có 4 chủng của *Cryptocotyle*, *Haplorchis*, *Metagonimus* (họ Heterophyidae) và 12 chủng của *Clonorchis*, *Opisthorchis*, *Amphimerus* và *Metorchis* (họ Opisthorchiidae) thuộc phân bộ Opisthorchiata; và iii) nhóm thứ ba là chuỗi ngoại hợp của loài *S. haematobium* (họ Schistosomatidae). Kết quả cũng cho thấy các chủng sán lá phổi của Việt Nam và thế giới được phân thành hai phân nhóm trong họ Paragonimidae với hệ số tin tưởng 100%, gồm toàn bộ các chủng của loài *P. westermani* (kể cả *P. westermani* chủng QT2 của Việt Nam) ở một phân nhóm và phân nhóm kia là tất cả các chủng của tất cả các loài *Paragonimus* còn lại của thế giới, kể cả *P. heterotremus* chủng D2YB của Việt Nam. Hai phân nhóm của *Paragonimus* trong nhánh phân loại họ Paragonimidae được sắp xếp ở vị trí “chị em” (sister) rất rõ ràng.

**Khoảng cách di truyền giữa các chủng/loài sán lá phổi *Paragonimus***

Trong nghiên cứu này, khoảng cách di truyền (pairwise genetic distance) (KCDT) được tính toán dựa trên phân tích chuỗi amino acid cộng hợp của 12 gen mã hóa protein (3354-3406 aa) của 15 chủng/6 loài sán lá phổi trong họ Paragonimidae (liệt kê ở **Bảng 1**). Sáu loài bao gồm: *P. heterotremus* các chủng của Việt Nam và Trung Quốc; *P. westermani* các chủng của Việt Nam, Trung Quốc, Hàn Quốc và Ấn Độ; *P. ohirai* chủng của Nhật Bản; *P. iloktsuenensis* chủng của Nhật Bản; *P. skrjabini miyazakii* chủng của Nhật Bản; và *P. kellycotti* Bắc Mỹ chủng của Mỹ.

Dựa vào giá trị KCDT, nhóm của loài *P. westermani* (8 chủng) chia thành 3 phân loài chính: nhóm phân loài *P. westermani* Việt Nam (1 chủng Việt Nam), nhóm phân loài *P. westermani* Đông Á (2 chủng Trung Quốc và 2 chủng Hàn Quốc), nhóm *P. westermani* Nam Á/Ấn Độ (3 chủng Ấn Độ). Kết quả phân tích cho thấy, KCDT nội loài của loài *P. heterotremus* giữa 2 chủng Việt Nam (D2YB và LC) và Trung Quốc (GX) là rất thấp (0,48-1,09%) cũng giống như KCDT nội loài *P. westermani* của nhóm Đông Á giữa 4 chủng của Trung Quốc và Hàn Quốc (dog1, dog2, Bogil và Haenam) là (0,15-1,09%), trong khi đó, KCDT nội loài *P. westermani* của nhóm Nam Á giữa 3 chủng của Ấn Độ (Type I, AP và IND2009) là rất cao (3,25-6,23%). Chủng QT2 (Việt Nam) tuy cùng loài *P. westermani*, nhưng giá trị KCDT nội loài của chủng này so với nhóm Đông Á là 3,94-4,51% và so với nhóm Nam Á (Ấn Độ) là rất cao (12,1-12,7%) (**Bảng 2**). Giá trị KCDT nội loài của *P. westermani* trong nhóm 3 chủng của Ấn Độ có biên thiên mạnh, đạt 6,23% giữa chủng Type I và AP và 4,16% giữa chủng AP và IND2009. Tuy được coi là hai loài, nhưng giá trị KCDT giữa loài *P. ohirai* và *P. iloktsuenensis* chỉ là 0.79%, mức tương đương nội loài vốn có của các chủng trong một loài *P. heterotremus*. Hai loài *P. ohirai* và *P. iloktsuenensis* được xác định lại là đồng danh (synonymy) trên cơ sở phân tích mtDNA và đơn vị mã hóa ribosome (Le et al., 2023).

Giá trị KCDT ngoại loài cao nhất là giữa loài đồng danh *P. ohirai/P. iloktsuenensis* với *P. westermani* nhóm Ấn Độ (19,93-21,23%) và thấp nhất là giữa *P. heterotremus* với *P. skrjabini miyazakii*, chỉ là 12,20-12,47%. Điều đặc biệt chú ý là, KCDT nội loài của *P. westermani* giữa nhóm 3 chủng của Ấn Độ (nhóm Nam Á) với nhóm 5 chủng Đông Á (Trung Quốc và Hàn Quốc) và cả Việt Nam là xấp xỉ 11,5% đến 12%, một giá trị tương đương với KCDT ngoại loài giữa *P. heterotremus* với *P. skrjabini miyazakii* (**Bảng 2**).



**Hình 1.** Cây phả hệ ML của 32 chủng/loài của các họ Paragonimidae, Heterophyidae, Opisthorchiidae và ngoại hợp Schistosomatidae. Phân tích phả hệ dựa trên sắp xếp cộng hợp amino acid của 12 gen mã hóa protein bằng phương pháp ML (“*tiếp cận cực đại*”, maximum-likelihood) với 1000 lần lặp lại. *Schistosoma haematobium* (họ Schistosomatidae) là nhóm ngoại hợp (outgroup). Hiện thị trên mỗi nhánh/nhóm là giá trị hệ số tin tưởng (bootstrap). Nút cơ bản của phân bộ Troglotremata (TRO) và Opisthorchiata (OPI) được đánh dấu bằng vòng tròn nhỏ tại nút phân nhánh. Mũi lên chỉ các chủng/loài *Paragonimus* của Việt Nam, đối tượng chính của nghiên cứu này. Ở mỗi chuỗi, viết tắt 4 chữ đầu tiên là ký hiệu loài, tiếp đến là chủng (nếu có, trong ngoặc), tiếp theo là ký hiệu hai chữ cái của quốc gia, nơi mẫu được phân lập và công bố, cuối chuỗi là số đăng ký Ngân hàng gen (Bảng 1). Thanh tỷ lệ đại diện cho số lượng thay thế trên mỗi nhánh (0.09 = 9/100).

**Bảng 2.** Khoảng cách di truyền giữa các chủng của Việt Nam với các chủng/loài của *Paragonimus* trong họ Paragonimidae

| Chủng/Loài            | 1     | 2     | 2     | 4     | 5     | 6     | 7     | 8     | 9     | 10    | 11    | 12    | 13    | 14    |
|-----------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1 Phet-(D2YB)-VN      |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 2 Phet-(LC)-VN        | 0,48  |       |       |       |       |       |       |       |       |       |       |       |       |       |
| 3 Phet-GX-CN          | 0,85  | 1,09  |       |       |       |       |       |       |       |       |       |       |       |       |
| 4 Pwes-QT2-VN         | 17,13 | 17,42 | 17,06 |       |       |       |       |       |       |       |       |       |       |       |
| 5 Pwes-(dog1)-CN      | 17,16 | 17,41 | 17,01 | 4,00  |       |       |       |       |       |       |       |       |       |       |
| 6 Pwes-(dog2)-CN      | 17,16 | 17,42 | 17,02 | 3,94  | 0,15  |       |       |       |       |       |       |       |       |       |
| 7 Pwes(3n)-Bogil-KR   | 17,35 | 17,60 | 17,21 | 4,30  | 0,88  | 0,88  |       |       |       |       |       |       |       |       |
| 8 Pwes(2n)-Haenam-KR  | 17,73 | 17,98 | 17,59 | 4,51  | 1,09  | 1,09  | 0,21  |       |       |       |       |       |       |       |
| 9 Pwes-(Typel)-IN     | 18,76 | 19,01 | 18,65 | 12,10 | 11,21 | 11,12 | 11,20 | 11,44 |       |       |       |       |       |       |
| 10 Pwes-(AP)-IN       | 19,19 | 19,44 | 19,19 | 12,70 | 11,53 | 11,51 | 11,49 | 11,69 | 6,23  |       |       |       |       |       |
| 11 Pwes-(IND 2009)-IN | 17,74 | 17,99 | 17,70 | 12,13 | 11,91 | 11,90 | 12,09 | 12,32 | 3,25  | 4,16  |       |       |       |       |
| 12 Pohi-(Kino)-JP     | 14,72 | 14,93 | 14,51 | 19,01 | 18,90 | 18,90 | 19,10 | 19,45 | 21,05 | 21,04 | 19,93 |       |       |       |
| 13 Pilo-(Amami)-JP    | 14,89 | 15,10 | 14,68 | 19,16 | 19,05 | 19,05 | 19,25 | 19,56 | 21,16 | 21,23 | 20,08 | 0,79  |       |       |
| 14 Pmiy-(Oku ST1)-JP  | 12,24 | 12,47 | 12,20 | 17,42 | 17,45 | 17,49 | 17,57 | 17,95 | 19,17 | 19,26 | 18,07 | 14,79 | 14,90 |       |
| 15 Pkel-(Ozark)-US    | 13,22 | 13,40 | 13,09 | 17,63 | 17,01 | 16,95 | 17,21 | 17,55 | 18,94 | 19,26 | 18,10 | 15,03 | 15,17 | 13,33 |

Khoảng cách di truyền nội loài/phân loài của *P. heterotremus* và *P. westermani* (phân nhóm Đông Á và phân nhóm Nam Á/Ấn Độ) được đóng khung; KCDT nội loài của *P. westermani* của chủng Qt2 Việt Nam với phân nhóm Đông Á và phân nhóm Nam Á/Ấn Độ được bôi đậm con số; và KCDT nội loài của *P. westermani* của phân nhóm Nam Á/Ấn Độ với phân nhóm Đông Á và Việt Nam được bôi nền. Một số giá trị khác được bôi màu để theo dõi.

Nghiên cứu giải trình tự toàn bộ mtDNA và trật tự gen, cấu trúc gen của sán lá phổi *Paragonimus* rất quan trọng nhằm cung cấp hiểu biết gen/hệ gen, tiến hóa, chỉ thị phân tử cho nghiên cứu phân biệt loài, di truyền quần thể và phả hệ tiến hóa. Với dữ liệu mtDNA của *P. heterotremus* chủng D2YB (Việt Nam) và *P. westermani* chủng QT2 (Việt Nam), hiện nay, họ Paragonimidae đã có 15 mtDNA hoàn chỉnh được công bố và sử dụng dữ liệu (Bảng 1). Cũng giống như một số loài sán đẹt trước đây, mtDNA của *P. heterotremus* chủng D2YB và *P. westermani* chủng QT2 có trật tự sắp xếp gen, sử dụng gen, bộ mã ATG/GTG để khởi đầu và TAA/TAG để kết thúc gen mã hóa protein được biết cho đến nay. Tuy nhiên vùng NCR của mtDNA có một số khác biệt đặc trưng. Do khai thác công nghệ NGS nên NCR của cả hai loài được thu nhận và đó là độ dài và cấu trúc đích thực của chúng. Các cấu trúc lặp dài và ngắn được phát hiện trong vùng NCR của cả *P. heterotremus* (D2YB) và *P. westermani* (QT2) trong nghiên cứu này, cũng được thấy ở *P. s. miyazakii* (Le et al., 2022) và *P. ohirai/P. iloktsuenensis* (Le et al., 2023) và nhiều loài sán đẹt khác, tạo nên đa hình (polymorphism) cấu trúc và kích thước mtDNA dài ngắn khác nhau. Nghiên cứu phá hệ các loài họ Paragonimidae và các họ liên quan (Heterophyidae và Opisthorchiidae) theo công bố của Le et al. (2022; 2023) và ở nghiên cứu này đều đã chỉ ra là có hai nhóm loài có quan hệ “đồng vị” (monophyly) đó là phân nhóm *P. westermani* riêng biệt sắp xếp ở vị trí “chị em” với phân nhóm các loài *Paragonimus* khác tạo nên của thế giới. Cây phả hệ cũng cho thấy sự sắp xếp “lệch vị” (paraphyly) và “đa vị” (polyphyly) của các loài và các chi có trong họ Heterophyidae và Opisthorchiidae của phân bộ Opisthorchiata.

Khoảng cách di truyền của các chủng/loài sán lá phổi *Paragonimus* trong họ Paragonimidae phản ánh một số nét đặc trưng của riêng họ này. Đó là KCDT nội loài của phân nhóm *P. westermani* Nam Á của các chủng Ấn Độ với các chủng của phân nhóm *P. westermani* Đông Á (Việt Nam, Trung Quốc và Hàn Quốc) rất cao, tương đương với giá trị KCDT ngoại loài giữa *P. heterotremus* và *P. heterotremus* với *P. skrjabini miyazakii*. Sự biến động đặc biệt trong di truyền của các chủng ở Ấn Độ của loài *P. westermani* đã được một số nghiên cứu đề cập (Devi et al., 2013; Blair et al., 2022) và có thể cho phép tính đến các phân dòng tiến hóa của quần thể *P. westermani* ở Ấn Độ sâu sắc hơn so với quần thể *P. westermani* ở Đông và Đông-Nam Á. Những nét đặc trưng về di truyền và cấu trúc của mtDNA của sán lá phổi *Paragonimus* và của họ Paragonimidae, KCDT và biến đổi di truyền nội loài và ngoại loài của chúng, cho phép làm đích khai thác cho nghiên cứu di truyền quần thể, phá hệ, nguồn gốc, tiến hóa của sán lá phổi *Paragonimus*, đồng thời dữ liệu gen học ty thể của chúng được sử dụng cho phát triển chẩn đoán, định hướng nghiên cứu thẩm định loài/giống/họ kí sinh trùng động vật lây sang người.

## KẾT LUẬN

Toàn bộ mtDNA của sán lá phổi *P. heterotremus* (D2YB, Việt Nam) và *P. westermani* (Qt2, Việt Nam) được thu nhận bằng phương pháp giải trình tự thông thường và thế hệ mới (NGS). Phả hệ của 32 chủng loài, trong đó có 15 chủng/loài *Paragonimus* được xây dựng và xác định có hai phân nhóm riêng biệt trong chi *Paragonimus*. Khoảng cách di truyền của các chủng/loài có những nét riêng biệt tạo nên sự đặc trưng di truyền của nhóm sán lá phổi trong lớp Sán lá (Trematoda).

**Lời cảm ơn:** Nguyễn Thị Khuê được tài trợ bởi Chương trình học bổng sau tiến sĩ trong nước của Quỹ Đổi mới sáng tạo Vingroup (VINIF), mã số VINIF.2023.STS.85.

## TÀI LIỆU THAM KHẢO

- Blair D (2022). Lung flukes of the genus *Paragonimus*: ancient and re-emerging pathogens. *Parasitol* 149: 1286-1295.
- Chan AHE, Chaisiri K, Saralamba S, Morand S, Thaenkham U (2021) Assessing the suitability of mitochondrial and nuclear DNA genetic markers for molecular systematics and species identification of helminths. *Parasit Vectors* 14: 233.
- Devi KR, Narain K, Mahanta J, Nirmolia T, Blair D, Saikia SP, Agatsuma T (2013). Presence of three distinct genotypes within the *Paragonimus westermani* complex in northeastern India. *Parasitol* 140(1): 76-86.
- Doanh PN, Horii Y, Nawa Y (2013). *Paragonimus* and paragonimiasis in Vietnam: an update. *Korean J Parasitol* 51(6): 621-627.
- Doanh PN, Shinohara A, Horii Y, Habe S, Nawa Y (2009). Discovery of *Paragonimus westermani* in Vietnam and its molecular phylogenetic status in *P. westermani* complex. *Parasitol Res* 104(5): 1149-1155.
- Kino H, De NV, Vien HV, Chuyen LT, Sano M (1995). *Paragonimus heterotremus* Chen et Hsia, 1964 found from a dog in Vietnam. *Jpn J Parasitol* 44: 470-472.
- Le TH, De NV, Blair D, McManus DP, Kino H, Agatsuma T (2006). *Paragonimus heterotremus* Chen and Hsia, 1964, in Vietnam: a molecular identification and relationships of isolates from different hosts and geographical origins. *Acta Trop* 98(1): 25-33.
- Le TH, Nguyen KT, Pham LTK, Doan HTT, Agatsuma T, Blair D (2022). The complete mitogenome of the Asian lung fluke *Paragonimus skrjabini miyazakii* and its implications for the family Paragonimidae (Trematoda: Platyhelminthes). *Parasitology* 149:1709-1719.
- Le TH, Nguyen KT, Pham LTK, Doan HTT, Do RT, Le XTK, Agatsuma T, Blair D (2023). Mitogenomic and nuclear ribosomal transcription unit datasets support the synonymy of *Paragonimus iloktsuenensis* and *P. ohirai* (Paragonimidae: Platyhelminthes). *Parasitol Res* 122(7): 1531-1544.
- Tamura K, Stecher G, Kumar S (2021). MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Mol Biol Evol* 38: 3022-3027.
- Wey-Fabrizius AR, Podsiadlowski L, Herlyn H, Hankeln T (2013). Platyzoan mitochondrial genomes. *Mol Phylogenet Evol* 69: 365-375.
- Yoshida A, Doanh PN, Maruyama H (2019). *Paragonimus* and paragonimiasis in Asia: An update. *Acta Trop* 199: 105074.

## PHYLOGENETIC ANALYSIS AND GENETIC DISTANCE ESTIMATION OF *PARAGONIMUS* LUNG FLUKES OF VIETNAM USING DATA FROM THE COMPLETE MITOGENOMES

Nguyễn Thị Khuê<sup>1,2</sup>, Đỗ Thị Roan<sup>1,2</sup>, Phạm Thị Khánh Linh<sup>1</sup>,  
Đoàn Thị Thanh Hương<sup>1,2</sup>, Phạm Ngọc Doanh<sup>3</sup>, Lê Thanh Hòa<sup>1,2\*</sup>

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

<sup>2</sup>*Graduated University of Science and Technology, Vietnam Academy of Science and Technology*

<sup>3</sup>*Institute of Ecology and Biological Resources*

### SUMMARY

Using Sanger and next-generation sequencing (NGS), the mitochondrial genome (mtDNA) of the Vietnamese lung flukes *Paragonimus heterotremus* (strain D2YB) and *P. westermani* (strain QT2) was determined and are 16,268 bp/D2YB and 14,779 bp/QT2 in length, respectively. There are 12 protein-coding genes (PCG); 2 mitoribosomal genes (MRG); 22 transfer RNA genes (tRNAs); and a non-coding region (NCR) with multiple long (LRU) and short repeats (SRU) structures. Phylogenetic analysis was performed using the concatenated amino acid sequences of all 12 protein-coding genes from 32 strains/species, including families Paragonimidae (15 sequences), Heterophyidae (4 sequences), Opisthorchiidae (12 sequences) and a sequence of Schistomatidae (as an outgroup). The phylogenetic tree clearly identified two subgroups among the Paragonimidae: one having only *P. westermani* species and another comprising all other worldwide *Paragonimus* species. Pairwise genetic distance (GD) estimation reveals that the highest GD value is between the synonymous *P. ohirai*/*P. iloktsuenensis* and *P. westermani* of the Indian group (19.93-21.23%), while the lowest between *P. heterotremus* and *P. skrjabini miyazakii*, only 12.20-12.47%. The intraspecific GD value, between the South Asian *P. westermani* subgroup (Indian strains) and the East Asian *P. westermani* subgroup (Vietnam, China, and Korea strains) is particularly high (12.1-12.7%), matching the interspecific GD value between *P. heterotremus* and *P. heterotremus* and *P. skrjabini miyazakii*. *Paragonimus*' genetic characteristics, mtDNA structure, and genetic distance make it an exploitable target for population genetics, phylogenetic and evolutionary studies, as well as the development of diagnostics and research orientation to evaluate species/genus/families of zoonotic parasites.

**Keywords:** Mitogenome (mtDNA), Paragonimidae, genetic distance, phylogeny, lung fluke *Paragonimus*.

---

\* Author for correspondence: Tel: +0912336855; Email: imibtvn@gmail.com