

ỨNG DỤNG PAN-GENOME TÌM KIẾM CÁC GEN MỤC TIÊU HƯỚNG TỚI PHÁT TRIỂN PHƯƠNG PHÁP PAN-PCR KIỂM SOÁT EDWARDSIELLOSIS Ở THỦY SẢN

Nguyễn Thành Luân

Viện Khoa học ứng dụng HUTECH, Đại học Công nghệ Tp. Hồ Chí Minh (HUTECH)

Sự phát triển bền vững của ngành nuôi trồng thủy sản đóng vai trò rất quan trọng cho an ninh lương thực toàn cầu và phúc lợi của nền kinh tế. Tuy nhiên, sự bùng phát của các dịch bệnh thủy sản do vi khuẩn đang đặt ra những thách thức rất lớn cho sự phát triển các phương pháp kiểm soát sinh học bền vững (bao gồm sự đặc hiệu, sự nhạy và nhanh chóng). Hiện nay, việc phát triển các phương pháp giải trình tự bộ gen kết hợp với so sánh phân tích các bộ gen được áp dụng để tìm ra các phương pháp quản lý hiệu quả cho nhiều bệnh thủy sản (Vincent và ctv., 2016). Do đó, việc phân tích thường quy các bộ gen vi sinh vật sẽ cung cấp những thông tin phân tử thể hiện sự đa dạng thích nghi và các xu hướng tiến hóa của các chủng vi khuẩn gây bệnh thủy sản, làm sáng tỏ các cơ chế gây bệnh cũng như ước tính các mô hình lây truyền bệnh qua các thang dịch tễ học. Dữ liệu giải trình tự bộ gen (whole genome sequencing data) là cơ hội để cách mạng hóa trong nghiên cứu dịch tễ học phân tử của vi sinh vật gây bệnh thủy sản có liên quan đến sức khỏe của con người và cộng đồng (Bayliss và ctv., 2017). Những thách thức trong quản lý bệnh thủy sản là sự thay đổi đa dạng của vi sinh vật gây bệnh, sự thích nghi tương tác với nhiều vật chủ (ví dụ các cơ chế thích nghi và truyền bệnh khác nhau) và thay đổi áp lực chọn lọc từ môi trường, đặc biệt là biến đổi khí hậu, sử dụng kháng sinh quá mức. Vì vậy, phân tích kiểu hình gây bệnh kết hợp với phân tích kiểu gen từ dữ liệu giải trình tự bộ gen hoàn chỉnh là phương pháp rất quan trọng để tái xây dựng các phương thức truyền bệnh ở quy mô địa phương và toàn cầu, từ đó đề xuất giảm thiểu sự phát tán và lây lan của dịch bệnh.

I. TÌNH HÌNH GIẢI TRÌNH TỰ VI KHUẨN *EDWARDSIELLA* GÂY BỆNH THỦY SẢN

1.1. Bệnh thủy sản do *Edwardsiella*

Được biết đến như một bệnh truyền nhiễm mãn tính, Edwardsiellosis gây chết hàng loạt ở nhiều loài cá có giá trị kinh tế quan trọng và luôn là một vấn đề nghiêm trọng trong thâm canh nuôi trồng thủy sản (Griffin và ctv., 2017). Ba loài gây bệnh bao gồm *E. hoshinae*, *E. ictaluri* và *E. tarda* được mô tả trong mối liên hệ với các vật chủ khác nhau, bao gồm chim và bò sát, cá trê nuôi và cá rô phi nuôi, và tồn tại trong các vi môi trường khác nhau như nước hồ, sông, biển và ở trong đường ruột của động vật thủy sinh khỏe mạnh (Buján và ctv., 2018). Trong số đó, *E. tarda* được xem là tác nhân truyền nhiễm toàn thân, gây tử vong hàng loạt ở nhiều loài cá nuôi với tỷ lệ cao, chúng cũng được coi là tác nhân gây bệnh linh hoạt có thể ảnh hưởng đến nhiều vật chủ khác như chim, lưỡng cư và bò sát, động vật có vú biển và con người. Ngoài ra, các hệ vi sinh thái của chúng bao gồm hồ, sông, nước biển và ruột của động vật thủy sinh khỏe mạnh. Gần đây, một tác nhân mới gây dịch bệnh truyền nhiễm cho các loài cá nuôi khác nhau trên toàn cầu là *E. piscicida* (được phát hiện ở hầu hết các chủng cá) trước đây được xác định nhầm là chủng *E. tarda* (Buján và ctv., 2018) vì chúng có chung nhiều đặc điểm kiểu hình với nhau (Shao và ctv., 2015). Trên thực tế, việc phân loại *E. tarda* đã chỉ ra rằng chúng bao gồm các nhóm khác biệt về mặt di truyền dựa trên kết quả phân tích sự phát sinh loài từ các gen (Buján và ctv., 2018) và phân tích sự phát sinh loài từ bộ gen bao gồm gen trung tâm và pan-genomics.

1.2. Đặc điểm kiểu hình và phân tử *Edwardsiella*

Trong kết quả thử nghiệm sinh hóa, loài *E. anguillarum* có khả năng sản xuất acetoin từ glucose (VP dương tính) và có thể lên men arabinose phân biệt từ các loài khác. Loài này có khả năng gây bệnh cho lươn. Trên thực tế, các phân loại trong chi *Edwardsiella* rất khó phân biệt với nhau theo đặc điểm của trình tự gen 16S rDNA và đặc điểm hình thái, sinh lý hoặc sinh hóa. Thực tế, *E. piscicida* đã chia sẻ nhiều đặc điểm kiểu hình giống với *E. tarda* và thậm chí trước đây còn bị nhầm lẫn với nhau. Đối với động vật thủy sản, *E. piscicida* đã được chứng minh với nhiều đặc điểm di truyền khác biệt dựa trên các kỹ thuật phân tử và phương pháp phân tích sự phát sinh loài. Quan trọng hơn, các nghiên cứu gần đây dựa trên các phương pháp so sánh nhiều gen cho thấy rằng loài *E. tarda* mang các dấu hiệu khác biệt về mặt di truyền; hầu hết các phân lập từ cá thực sự thuộc về loài *E. piscicida*, không phải *E. tarda*. Do đó, phương pháp phân tích bộ gen trong các nghiên cứu mới là cần thiết để làm rõ vị trí phân loại của các loài *Edwardsiella*.

II. CÔNG CỤ SO SÁNH BỘ GEN VÀ SỰ KẾT HỢP

2.1. Pan-genome

Pan-genome là kết quả của phép so sánh bộ gen của các sinh vật với nhau để tìm ra các gen trung tâm (core gen) hiện diện ở tất cả các bộ gen trong dữ liệu phân tích, và một nhóm các gen phân tán (dispensable gen) có mặt một trong các bộ gen hoặc đặc trưng cho sinh vật đơn lẻ (Medini và ctv., 2005). Sự phân tích pan-genome sẽ cho biết khả năng thu hoặc mất gen của một loài sinh vật trong quá trình tiến hóa và cũng được dùng so sánh sự khác biệt về cấu trúc protein giữa các loài. Các gen đặc trưng có thể được sử dụng cho các ứng dụng khác nhau như nghiên cứu định danh, chức năng của gen trong môi trường sống.

Mục đích chính của việc phân tích pan-genome là so sánh bộ gen của các chủng vi

khuẩn khác nhau trong một loài (intra-species) hoặc chi/giống (inter-species) (Snipen và ctv., 2009). Các nghiên cứu về bộ gen mang lại những hiểu biết đáng kể về sự tiến hóa của vi khuẩn, sự thích nghi thích hợp, cấu trúc quần thể và tương tác vật chủ. Những nghiên cứu này cũng có thể được áp dụng cho các vấn đề như xác định gen độc lực, chế tạo vacxin và thuốc (Chaplin và ctv., 2015).

2.2. Sự kết hợp các công cụ trong phân tích gen chức năng

Phân tích pan-genome cho phép hiểu về sự tiến hóa của loài gây bệnh đa kháng thuốc. Phương pháp này cũng cho phép hiểu rõ hơn về yếu tố di truyền mang khả năng gây bệnh ở nhiều loại vi khuẩn bằng cách so sánh các chủng độc lực và không độc lực khi chỉ sử dụng kết quả phân tích các gen trung tâm. Đặc biệt, sự phân tích các gen/cụm gen (core, dispensable, singleton) kết hợp hệ thống phân tích chức năng như RAST (Rapid Annotation using Subsystem Technology), KEGG, WebMGA (một máy chủ web tùy biến để phân tích các chuỗi trình tự) có thể giúp xác định các yếu tố quan trọng xuất hiện trong khả năng gây độc của chủng độc lực. Ngoài ra, khả năng chuyển hóa đường phức tạp có thể được hiển thị bằng các so sánh sự phân bố của các enzyme sử dụng đường (cấu hình Cazy, carbohydrate-active enzymes) giữa các phân lập. Thông tin Cazy đặc trưng là một lợi thế của quá trình chọn lọc cho phép vi sinh vật thích nghi với hệ sinh thái của chúng.

Trong kết quả phân tích pan-genome trước đây của chúng tôi (Nguyễn và Kim, 2018; Nguyen và ctv., 2018) đã chỉ ra rằng chủng WFLU12 phân lập từ cá biển biểu hiện các đặc tính đặc thù cho việc sản xuất, chuyển đổi năng lượng, vận chuyển và chuyển hóa đường bằng cách phân tách các gen đặc trưng cho từng chủng. Bên cạnh đó, dựa vào các gen phân tách từ pan-genome, các gen được chọn hoàn toàn có thể được kiểm tra chức năng lại trên cơ sở dữ liệu các gen độc lực (VFDB), các nhóm gen chức năng (COG), Từ điển bách khoa về gen của Kyoto (KEGG) và cơ sở dữ liệu gen kháng kháng sinh (ARGB).

III. TRƯỜNG HỢP PHÂN TÍCH BỘ GEN CỦA CÁC *EDWARDSIELLA* BẰNG PAN-GENOME

Chúng tôi thực hiện phân tích pan-genome của các loài *Edwardsiella* phân lập từ các nguồn khác nhau (bảng 1), sau đó tiến hành so sánh với mục tiêu tiếp cận một phương pháp khác dùng phân loại chính xác các vi khuẩn thuộc giống *Edwardsiella* gây bệnh thủy sản. Quan trọng hơn, việc so sánh các gen phân tán sẽ giúp tìm

ra nhiều gen chỉ thị cho loài và hướng tới sử dụng các gen trong kỹ thuật chẩn đoán sớm như Pan-PCR cho các mẫu lâm sàng.

3.1. Các bước thực hiện trong phân tích pan-genome

3.1.1. Chọn lọc dữ liệu

Trong nghiên cứu này, tổng cộng có 15 bộ gen hoàn chỉnh của các loài *Edwardsiella* phân lập từ các hệ sinh thái khác nhau (bảng 1) được thu thập từ Ngân hàng gen NCBI.

Bảng 1. Thông tin về bộ gen của các loài *Edwardsiella* từ nguồn NCBI dùng trong nghiên cứu này

Tên chủng	Mã nhận diện (Assession no.*)	Nguồn gốc phân lập (Ecological niche)
<i>Edwardsiella tarda</i>		
EIB202	CP001135.1	the outbreak in farmed turbot in Shandong province of China, 2009
FL6-60	CP002154.1	catfish
FL95-01	CP011359.1	recovered from channel catfish 1995, USA
KC-Pc-HB1	CP023706.1	blood of <i>Pseudorca crassidens</i> , 2017, Korea
ET-1	NZ_LC127084.1	host: <i>Paralichthys olivaceus</i>
<i>Edwardsiella anguillarum</i>		
ET080813	CP006664.1	<i>Anguilla japonica</i> , 2008, China
<i>Edwardsiella sp</i>		
EA181011	CP011364.1	<i>Epinephelus aeneus</i> , organism modifier note <i>Edwardsiella cf. piscicida</i>
LADL05-105	CP011516.1	<i>Oreochromis sp.</i> , Tilapia, 2005, USA
<i>Edwardsiella ictaluri</i>		
93-146	CP001600.2	a natural channel catfish outbreak of enteric septicemia of catfish, 1993, in Louisiana
RUSVM-1	CP020466.1	<i>Oreochromis niloticus</i> , 2012
MS-17-156	CP028813.1	diseased catfish from East Mississippi, 2017, USA
<i>Edwardsiella piscicida</i>		
C07-087	CP004141.1	Channel catfish gastrointestinal septicemia isolate, Mississippi, 2013
S11-285	CP016044.1	<i>Ictalurus punctatus</i> , 2011, USA:Mississippi
ETW41	CP019440.1	Eel pond water, 2010, Korea: Naju
<i>Edwardsiella hoshinae</i>		
ATCC 35051	CP016043.1	<i>Varanus sp.</i> , 1980, Chad: N'Djamena

*, <ftp://ftp.ncbi.nih.gov/genomes/>

3.1.2. Định danh và phân tích sự phát sinh loài

Sự phát sinh loài dựa trên toàn bộ thông tin bộ gen (phylogenomic) của các chủng *Edwardsiella* được thiết lập dựa trên sự phân tích giá trị nucleotide trung bình (ANI) bằng phần mềm JSpecies v1.2.1 (Richter và ctv., 2009) và được hiển thị dưới dạng bản đồ nhiệt bằng phần mềm Gene-E (<http://www.broadinstitute.org/cancer/software/GENE-E/>).

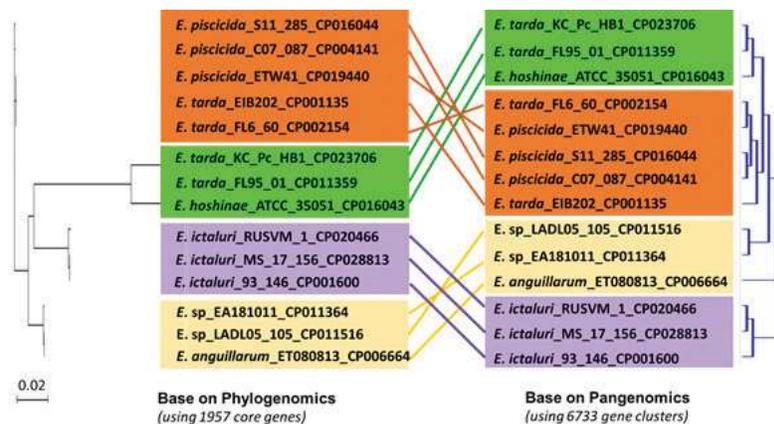
3.1.3. Phân tích pan-genome

Pan-genome của 15 loài *Edwardsiella* được phân tích bằng hệ thống cơ sở dữ liệu của EDGAR v2.2 (Blom và ctv., 2016). Theo đó, nhóm gen bao gồm số lượng gen thuộc gen trung tâm (core gen), gen phân tán (accessory gen) và gen đơn lẻ (singleton gen) được trích xuất theo các thông số mặc định của chương trình. Để tìm sự khác biệt trong việc sử dụng các gen làm chỉ thị phân tích các mẫu lâm sàng, các gen phân tán được phân tích lại bằng phần mềm Gene-E.

3.2. Mối quan hệ phát sinh loài từ dữ liệu *Edwardsiella* pan-genome

Kết quả phân tích pan-genome cho thấy có 6733 gen mã hóa protein của 15 chủng *Edwardsiella* có 29,07% (1957 gen) là gen trung

tâm và 70,93% còn lại là gen phân tán và gen đơn lẻ. Sự phát sinh loài từ dữ liệu gen trung tâm và pan-genome cho thấy rằng chủng *E. tarda* EIB202, FL6_60 và ET-1 là thuộc loài *E. piscicida* (hình 1). ET-1 vẫn đang được sử dụng dưới tên loài là *E. tarda* trong dữ liệu của NCBI gần đây nhất (2018) vì có kết quả tương đồng cao với chủng EIB202. Kết quả này cho thấy bằng cách sử dụng gen trung tâm và pan-genome, sự phát sinh loài của các chủng *Edwardsiella* cũng được phân biệt rõ ràng. Phù hợp với các nghiên cứu trước đây (Buján và ctv., 2018; Shao và ctv., 2015), phân tích của chúng tôi cho thấy pan-genome có thể ứng dụng trong việc xác định mối quan hệ loài giữa các chủng *E. tarda*, *E. piscicida* và 4 loài trong chi *Edwardsiella* có thể phân biệt rõ ràng (hình 1) và hướng tới phân tích các chủng có nguồn gốc thủy sản khác như *Aeromonas*. Hai chủng *Edwardsiella* sp. còn lại (EA181011 và LADL05_105) có giá trị ANIs tương ứng là 99,65 và 99,58 (dữ liệu không được hiển thị) và phân cụm rất tốt với chủng *E. anguillarum* ET080813 (hình 2). Do đó, hai bộ chủng *Edwardsiella* sp. EA181011 và *Edwardsiella* sp. LADL05_105 có định danh thuộc loài *E. anguillarum* dựa trên cơ sở của cả hai phân tích ANI và phân tích các gen phân tán (hình 2).



Hình 1. Sự phân cấp của các chủng *Edwardsiella* phân lập từ các hệ sinh thái khác nhau
Cây phân loài (bên phải) dựa trên sự chia sẻ các gen giữa các chủng và cây phát sinh loài (bên trái) dựa trên các core gen của 15 chủng khảo sát. Thanh chỉ thị cùng một chủng của cả hai cây phân loài nhằm làm nổi bật mức độ tương đồng giữa hai phương pháp.

3.3. Phân tích sự đa hình của gen phân tán của *Edwardsiella*

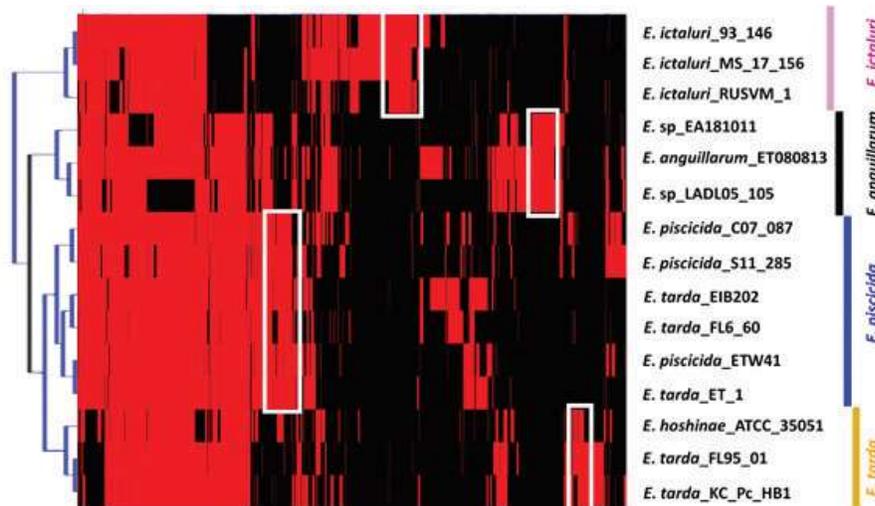
Cụ thể, tính đa hình của các gen phân tán sẽ cung cấp thông tin rất có giá trị về các biện pháp kiểm soát bệnh Edwardsiellosis. Phân tích sự hiện diện/vắng mặt của gen trong bộ gen giữa các loài (hình 2) có thể chỉ ra các dấu hiệu cho việc nhận diện phân tử và có thể được sử dụng để thiết kế các markers và xác định chính xác các loài trong chi *Edwardsiella*, đặc biệt là sử dụng để phân biệt các loài mới với *E. tarda* (Buján và ctv., 2018; Fogelson và ctv., 2016).

IV. ỨNG DỤNG PAN-GENOME TRONG PHÁT HIỆN BỆNH EDWARDSIELLOSIS

Kết quả phân tích sự đa hình (hình 2) cho thấy rằng pan-genome có thể sử dụng như một công cụ mạnh mẽ để phân biệt sự khác nhau về các gen chức năng của các loài trong chi

Edwardsiella. Đặc biệt, chúng tôi tin rằng pan-PCR, một phương pháp PCR dùng định danh phân tử có độ tin cậy cao dựa trên các gen mục tiêu được chọn lọc từ gen phân tán, sẽ là một công cụ thường quy trong phòng thí nghiệm dùng phân biệt tất cả các chủng *Edwardsiella* từ mẫu bệnh lâm sàng.

Quan trọng hơn, số lượng các trình tự bộ gen vi khuẩn phân lập từ thủy sản trên cơ sở dữ liệu của NCBI đang tăng theo cấp số nhân. Dữ liệu này cung cấp một tiềm năng lớn cho việc kiểm tra tổng hợp về dịch tễ bệnh và các tương tác giữa mầm bệnh và vật chủ. Dễ thấy rằng, pan-genome là một công cụ hiệu quả có thể được mở rộng để phân tích các vi sinh vật thủy sinh và hiểu được các đặc tính phân tử giúp chúng thích nghi với các vật chủ và môi trường khác nhau. Ví dụ như *E. tarda* phân lập từ cá bệnh có thể được chia thành nhóm nước ngọt và nhóm biển/di cư (Shao và ctv., 2018).



Hình 2. Tính đa hình của các gen phân tán chọn lọc từ pan-genome của *Edwardsiella*
Sự hiện diện/vắng mặt của 2578 gen được hiển thị bằng màu đỏ/đen.

V. KẾT LUẬN

Công cụ so sánh pan-genome giúp định danh lại các loài *Edwardsiella*, cụ thể chủng *E. tarda* EIB202, FL6_60 và ET-1 thuộc loài *E. piscicida*.

Chúng tôi đề xuất có những chỉnh sửa định danh bổ sung trên hệ thống NCBI để dữ liệu được sử dụng đúng cho các nghiên cứu tiếp theo. Dựa vào các phân tích chi tiết về thành phần/chức năng các gen trong hệ gen của *Edwardsiella*,

chúng tôi cung cấp những thông tin về các gen có thể sử dụng trong thiết kế quy trình định danh phân tử các loài *Edwardsiella* từ mẫu lâm sàng. Ngoài ra, các phương pháp phân tích pan-genome có thể dùng để khám phá các gen được chia sẻ liên quan tới khả năng thích nghi và phát triển các loại vaccin tiềm năng nhằm cải thiện hiệu quả ngăn chặn dịch bệnh thủy sản.

Lời cảm ơn: Quỹ nghiên cứu Khoa học và công nghệ, Trường đại học Công nghệ, TP. HCM (HUTECH) đã tài trợ cho nghiên cứu này.

TÀI LIỆU THAM KHẢO

- Blom, J., Kreis, S., Spänig, S., Juhre, T., Bertelli, C., Ernst, C., Goesmann A., 2016. EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic Acids Res.* 44, 22–28.
- Buján, N., Mohammed, H., Balboa, S., Romalde, J.L., Toranzo, A.E., Arias CR4, Magariños, B., 2018. Genetic studies to re-affiliate *Edwardsiella tarda* fish isolates to *E. piscicida* and *E. anguillarum* species *Systematic and Applied Microbiology*, 41, 30–37.
- Fogelson, S.B., Petty, B.D., Reichley, S.R., Ware, C., Bowser, P.R., Crim, M.J., Getchell, R.G., Sams, K.L., Marquis, H., Griffin, M.J., 2016. Histologic and molecular characterization of *Edwardsiella piscicida* infection in large-mouth bass (*Micropterus salmoides*). *J. Vet. Diagn. Invest.* 28, 338–344.
- Griffin, M.J., Greenway, T.E., Wise, D.J., 2017. *Edwardsiella* spp. In: Woo, P.T.K. Cipriano, R.C. (Eds.). *CAB International, Boston*. 190–210.
- Nguyen, T.L., Kim, D.H., 2018. Genome-wide comparison reveals a probiotic strain *Lactococcus lactis* WFLU12 isolated from the gastrointestinal tract of olive flounder (*Paralichthys olivaceus*) harboring genes supporting probiotic action, *Mar. Drugs* 16 E140. doi: 10.3390/md16050140.
- Nguyen, T.L., Chun, W-K., Kim, A., Kim, N., Roh, H.J., Lee, Y., Yi, M., Kim, S., Park, C-II., Kim, D-H., 2018. Dietary probiotic effect of *Lactococcus lactis* WFLU12 on low-molecular-weight metabolites and growth of olive flounder (*Paralichthys olivaceus*), *Front. Microbiol.* 9 2059.
- Richter, M., Rosselló-Móra, R., 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A.* 106, 19126–31.
- Shao, S., Lai, Q., Liu, Q., Wu, H., Xiao J., Shao, Z., Wang, Q., Zhang, Y., 2015. Phylogenomics characterization of a highly virulent *Edwardsiella* strain ET080813(T) encoding two distinct T3SS and three T6SS gene clusters: Propose a novel species as *E. anguillarum* sp. nov. *Syst. Appl. Microbiol.* 38, 36–47.
- Shao, J., Guo, Q., Hu, R., Zemao, G., 2018. Comparative genomic insights into the taxonomy of *Edwardsiella tarda* isolated from different hosts: Marine, freshwater and migratory fish. *Aquac Res.* 49, 197–204.
- Vincent, A.T., Trudel, M.V., Freschi, L., Nagar, V. Gagné-Thivierge, C. Levesque, R.C. *et al.*, 2016. Increasing genomic diversity and evidence of constrained lifestyle evolution due to insertion sequences in *Aeromonas salmonicida*, *BMC Genomics* 17 - 44 ./.