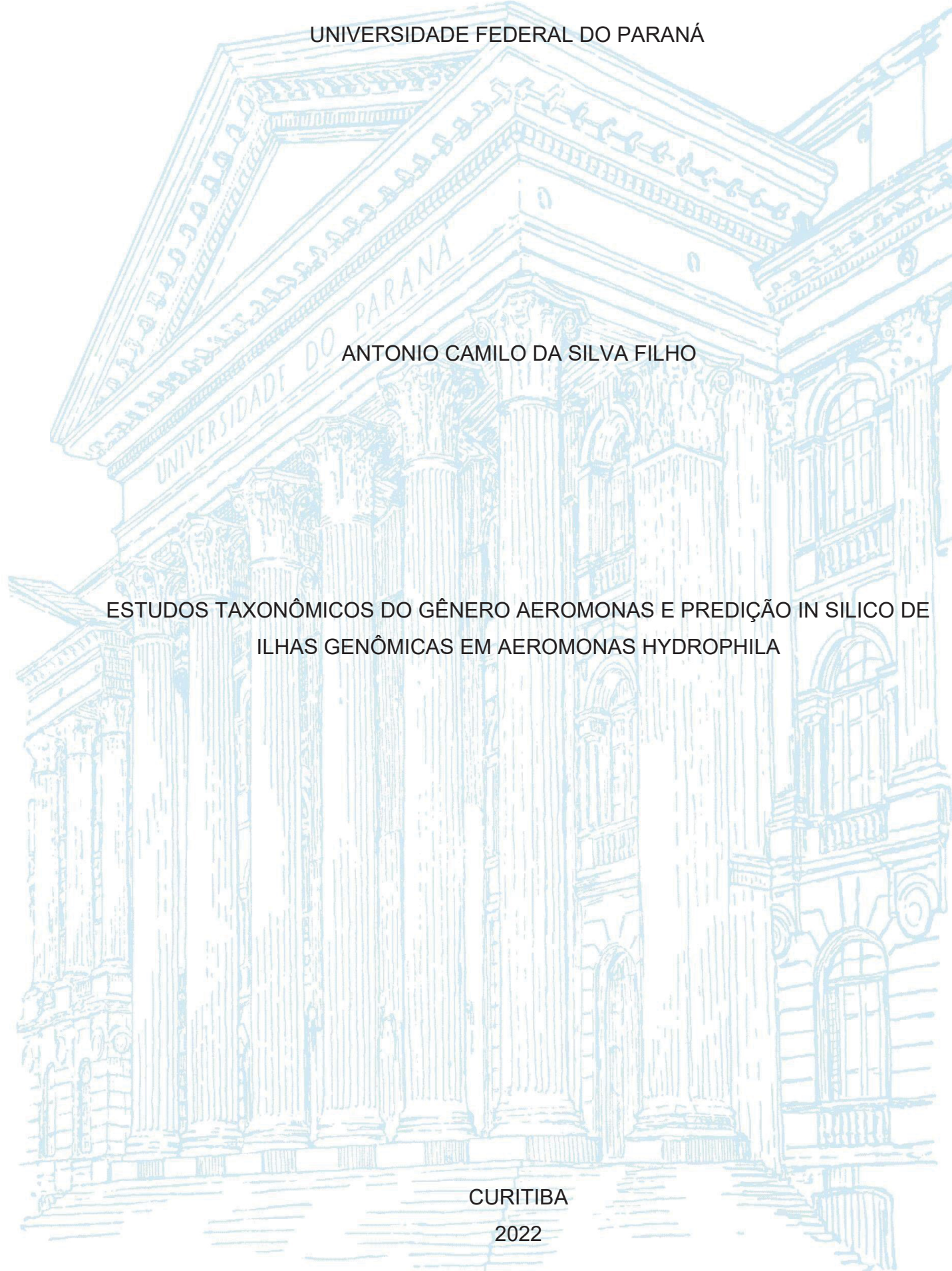




UNIVERSIDADE FEDERAL DO PARANÁ

ANTONIO CAMILO DA SILVA FILHO

ESTUDOS TAXONÔMICOS DO GÊNERO AEROMONAS E PREDIÇÃO IN SILICO DE
ILHAS GENÔMICAS EM AEROMONAS HYDROPHILA

CURITIBA

2022

ANTONIO CAMILO DA SILVA FILHO

ESTUDOS TAXONÔMICOS DO GÊNERO AEROMONAS E PREDIÇÃO IN SILICO DE
ILHAS GENÔMICAS EM AEROMONAS HYDROPHILA

Tese apresentada ao curso de Pós-Graduação em Ciências Farmacêuticas, Setor de Ciências da Saúde, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Doutor em Ciências Farmacêuticas.

Orientador: Prof. Dr. Geraldo Picheth

Coorientadora: Profa. Dra. Cyntia Maria Telles
Fadel-Picheth

Coorientadora: Profa. Dra. Jeroniza Nunes
Marchaukoski

CURITIBA

2022

Silva Filho, Antonio Camilo da
Estudos taxonômicos do gênero *Aeromonas* e predição *in silico* de ilhas
genômicas em *Aeromonas hydrophila* [recurso eletrônico] / Antonio Camilo da
Silva Filho – Curitiba, 2022.
1 recurso online: PDF.

Tese (doutorado) – Programa de Pós-Graduação em Ciências Farmacêuticas.
Setor de Ciências da Saúde, Universidade Federal do Paraná, 2022.

Orientador: Prof. Dr. Geraldo Picheth
Coorientador: Profa. Dra. Cyntia Maria Telles Fadel-Picheth
Coorientador: Profa. Dra. Jeroniza Nunes Marchaukoski

1. *Aeromonas*. 2. Taxonomia. 3. Árvore filogenética. I. Picheth, Geraldo.
II. Fadel-Picheth, Cyntia Maria Telles. III. Marchaukoski, Jeroniza Nunes.
IV. Universidade Federal do Paraná. V. Título.

CDD 579.34

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação CIÊNCIAS FARMACÊUTICAS da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **ANTONIO CAMILO DA SILVA FILHO** intitulada: **Estudos taxonômicos do gênero *Aeromonas* e predição *in silico* de ilhas genômicas em *Aeromonas hydrophila***, sob orientação do Prof. Dr. GERALDO PICHETH, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 04 de Julho de 2022.



GERALDO PICHETH
Presidente da Banca Examinadora



CIBELLE DE BORBA DALLAGASSA
Avaliador Externo (PONTIFÍCIA UNIVERSIDADE CATÓLICA DO
PARANÁ)



ROBERTO TADEU RAITTZ
Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ)



LEONARDO MAGALHÃES CRUZ
Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ)

AGRADECIMENTOS

Aos meus orientadores, Professor Dr. Geraldo Picheth e Professora Dra. Cyntia Maria Telles Fadel-Picheth, a minha coorientadora Professora Dra. Jeroniza Nunes Marchaukoski. Foram quatro anos de estudo, pesquisa, dedicação e amizade. Tive o privilégio de poder aprender muitas coisas das quais sempre vou me recordar, tanto na área da pesquisa, como na vida pessoal. Esse período foi atípico. Devido a pandemia, que ainda estamos vivendo, passamos por várias dificuldades, sendo estas, principalmente emocionais. Eu passei um bom tempo sozinho, mas todos os dias vocês estavam dispostos a conversar sobre qualquer coisa, além da nossa pesquisa, e isso também me ajudou a superar esse momento conturbado. Vocês são exemplos para muitos alunos. Seus ensinamentos irão ser sempre lembrados, muito obrigado por tudo!

À minha família. Meus agradecimentos não cabem em palavras para expressar o quanto vocês foram, e ainda continuam sendo importantes nessa caminhada. Sem vocês, eu não teria conseguido chegar até aqui. Minha mãe Rozeli Smaha, sempre me deu todo o amor e apoio nessa jornada. Minha irmã Emily Smaha, que aguentou todas as minhas frustrações e sempre esteve ao meu lado, não importando a situação. A minha avó materna, Júlia Smaha, meu maior tesouro e maior exemplo que eu tenho na vida. Mesmo estando longe, sinto como você sempre estivesse comigo, me dando forças para continuar. As minhas tias, Tânia Smaha, Ruth Smaha e meu tio João Morozini, que sempre estiverem ao meu lado, ajudando e apoiando em qualquer decisão que eu tenha tomado. Faltam muitas pessoas ainda, mas quero que saibam que cada um foi muito importante na minha trajetória, eu amo cada um de vocês.

À minha esposa, Dra. Camilla Reginatto De Pierri. Lembro como se fosse hoje ... quando nos conhecemos, ainda na faculdade e começamos a trilhar nosso caminho juntos. Somos amigos, colegas de trabalho e de pesquisa. A vida nos proporcionou a oportunidade de sempre estarmos juntos na maioria da nossa trajetória profissional e acadêmica. Eu só tenho a agradecer a cada momento que passamos e por todo o apoio e amor que você sempre me deu, sem você eu não estaria aqui e nada disso seria possível.

À Universidade Federal do Paraná e aos Programas de Pós-graduação em Ciências Farmacêuticas e Pós-graduação em Bioinformática. Foram seis anos frequentando a universidade. Pode parecer clichê, mas o tempo passou muito rápido, parece que foi ontem. Realizei a entrevista para o mestrado e logo depois, fui ao laboratório da Profa, Cyntia Picheth conversar sobre uma possível vaga no doutorado. Foram muitos momentos marcantes, fiz muitas amizades e gostaria de falar sobre todos e agradecer por tudo. Quero deixar meus agradecimentos a todos os professores e colegas que passaram o seu conhecimento para que pudéssemos construir a nossa aprendizagem da melhor maneira possível. Em especial, aos dois professores e amigos Professor Dr. Roberto Tadeu Raittz e Professor Dr. Dieval Guizelini, que sempre estiveram presente, dando suporte, ensinando, apoiando e ajudando a seguir o melhor caminho possível.

Gratidão enorme poder compartilhar tantos momentos juntos, muito obrigado por tudo.

Aos membros da banca avaliadora, Prof. Dr. Roberto Tadeu Raittz, Prof. Dr. Leonardo Magalhães Cruz, Dra. Cibelle de Borba Dallagassa. Os meus agradecimentos por terem aceito o convite e compartilhado seus conhecimentos para melhorar o nosso trabalho, muito obrigado a todos!

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES. Fui bolsista desde o mestrado e no doutorado. Isso só foi possível devido a esta fundação. Hoje, olho para trás e vejo quantos pesquisadores e alunos puderam desenvolver seus trabalhos, projetos, pesquisas através do apoio da CAPES. Agora, eu também faço parte destas pessoas. Cabe a mim, passar o todo o conhecimento adiante para as novas gerações como forma de agradecimento. Os meus mais sinceros agradecimentos.

*Underneath this smile lies everything
All my hopes and anger, pride and shame
Make myself a pact, not to shut doors on the past
Just for today I am free ...*

(Eddie Vedder, Inside Job)

RESUMO

Aeromonas são bacilos gram-negativos de habitat aquático, mas podem ser encontradas em outros ambientes. São oportunistas e podem causar infecções em humanos que variam de diarreia a septicemia. As espécies mais frequentemente recuperadas de amostras clínicas são *A. veronii*, *A. caviae*, *A. hydrophila*. *Aeromonas* também estão relacionadas diretamente com graves infecções em peixes, gerando perdas financeiras na indústria pesqueira. Ilhas genômicas (GIs) são “clusters” que se encontram no genoma bacteriano de determinadas cepas, mas ausentes em outras cepas da mesma espécie. GIs podem contribuir para a evolução e adaptação de bactérias a diferentes ambientes; dependendo dos genes que se encontram nestas regiões, pode-se classificar as GIs, como ilhas de virulência, patogenicidade, metabolismo, resistência à antibiótico ou simbiose. No primeiro artigo, foi realizada a predição *in silico* de GIs, em isolados de *A. hydrophila* com genoma completo. O número de GIs preditas e suas características variaram entre as 17 cepas de *A. hydrophila* que foram classificadas em dois grupos filogenéticos. Poucas GIs foram compartilhadas entre os dois grupos, e contém principalmente genes do “core genoma”. O grupo 1 contém 9 cepas, majoritariamente patógenos de peixes; estas apresentaram maior número de GIs preditas e e com características semelhantes incluindo diversos genes relacionados ao metabolismo do mio-inositol, ácido siálico e L-fucose, que podem estar associados à adaptação e virulência no hospedeiro, e um “cluster” recentemente associado com o metabolismo de DNA. Estas GIs também apresentam “clusters” para biossíntese do antígeno somático OX1 ou OX6. O grupo 2, é constituído por 8 cepas isoladas de diferentes origens, incluindo isolados clínicos e ambientais; estas cepas apresentam GIs únicas e nenhuma característica distintiva do grupo foi identificada. Em resumo, os resultados sugerem que há diversidade entre as GIs presentes em *Aeromonas*, que codificam diversas características associadas ao metabolismo, virulência, resistência aos antibióticos, entre outras, que podem contribuir para a adaptação a hospedeiros e competitividade no ambiente. No segundo artigo a classificação taxonômica de *Aeromonas* com genomas completos foi determinada utilizando os genes *housekeeping gyrB-rpoD-rpoB*, proteoma e Average Nucleotide Identity (ANI). O método Sweep foi utilizada para gerar as árvores filogenéticas. Um total de 177 genomas de *Aeromonas*, depositados no GenBank como membros de 15 espécies ou identificados apenas ao nível de gênero, foram analisados. Algumas divergências foram observadas entre os resultados da análise baseada em genes *housekeeping* e aqueles do proteoma e ANI, que foram mais consistentes. Os resultados mostram que 11 cepas estão erroneamente identificadas ao nível de espécie. As cepas erroneamente identificadas estão classificadas no NCBI como *A. media* (2 cepas), *A. hydrophila* (4 cepas) e *A. veronii* (5 cepas). Destas, 4 provavelmente são *A. rivipollensis* e 3 *A. allosaccharophila*, uma vez que apresentam respectivamente, valores de ANI superiores a 95% com o único representante dessas espécies; 2 pertencem à espécie *A. dhakensis*; 1 *A. jandaei* e 1 cepa permanece como espécie não identificada. Entre as 27 cepas previamente identificadas apenas ao nível de gênero, 22 puderam ser classificadas ao nível de espécie, sendo 3 *A. hydrophila*, 9 *A. caviae*, 5 *A. veronii*, 2 *A. dhakensis*; *A. enteropelogenes*, *A. jandaei* e *A. sanarellii* 1 cepa cada. Três cepas não foram incluídas em nenhum grupo, sugerindo que possam pertencer a espécies distintas das que já tem genoma completo sequenciado, ou até mesmo representar novas espécies de *Aeromonas*.

Palavras-chave: *Aeromonas*; ilha genômica; taxonomia; árvore filogenética; Identidade Nucleotídica Média (ANI).

ABSTRACT

Aeromonas are gram-negative bacilli of aquatic habitat, but can be found in other environments. They are opportunistic and can cause infections in humans ranging from diarrhea to septicemia. The species most frequently recovered from clinical samples are *A. veronii*, *A. caviae*, *A. hydrophila*. *Aeromonas* are also directly related to serious infections in fish, generating financial losses in the fishing industry. Genomic islands (GIs) are “clusters” that are found in the bacterial genome of certain strains, but absent in other strains of the same species. GIs can contribute to the evolution and adaptation of bacteria to different environments; depending on the genes found in these regions, GIs can be classified as islands of virulence, pathogenicity, metabolism, antibiotic resistance or symbiosis. In the first article, the in silico prediction of GIs was performed in *A. hydrophila* isolates with complete genome. The number of predicted GIs and their characteristics varied among the 17 *A. hydrophila* strains that were classified into two phylogenetic groups. Few GIs were shared between the two groups, and they mostly contain core genome genes. Group 1 contains 9 strains, mostly fish pathogens; these showed a greater number of predicted GIs and with similar characteristics including several genes related to the metabolism of myo-inositol, sialic acid and L-fucose, which may be associated with adaptation and virulence in the host, and a cluster recently associated with metabolism of DNA. These GIs also present “clusters” for the biosynthesis of the somatic antigen OX1 or OX6. Group 2 consists of 8 strains isolated from different origins, including clinical and environmental isolates; these strains have unique GIs and no distinctive features of the group have been identified. In summary, the results suggest that there is diversity among the GIs present in *Aeromonas*, which encode several characteristics associated with metabolism, virulence, antibiotic resistance, among others, which may contribute to host adaptation and competitiveness in the environment. In the second article, the taxonomic classification of *Aeromonas* with complete genomes was determined using the housekeeping genes *gyrB-rpoD-rpoB*, proteome and Average Nucleotide Identity (ANI). The Sweep method was used to generate the phylogenetic trees. A total of 177 *Aeromonas* genomes, deposited in GenBank as members of 15 species or identified only at the genus level, were analyzed. Some divergences were observed between the results of the analysis based on housekeeping genes and those of the proteome and ANI, which were more consistent. The results show that 11 strains are misidentified at the species level. The wrongly identified strains are classified in the NCBI as *A. media* (2 strains), *A. hydrophila* (4 strains) and *A. veronii* (5 strains). Of these, 4 are probably *A. rivipollensis* and 3 *A. allosaccharophila*, since they present respectively, ANI values higher than 95% with the only representative of these species; 2 belong to the species *A. dhakensis*; 1 *A. jandaei* and 1 strain remains as an unidentified species. Among the 27 strains previously identified only at the genus level, 22 could be classified at the species level, being 3 *A. hydrophila*, 9 *A. caviae*, 5 *A. veronii*, 2 *A. dhakensis*; *A. enteropelogenes*, *A. jandaei* and *A. sanarellii* 1 strain each. Three strains were not included in any group, suggesting that they may belong to different species from those that already have a complete genome sequenced, or even represent new species of *Aeromonas*.

Keywords: *Aeromonas*; genomic island; taxonomy; phylogenetic tree; Average Nucleotide Identity (ANI).

LISTA DE FIGURAS

CAPÍTULO III: PREDICTION AND ANALYSIS IN SILICO OF GENOMIC ISLANDS IN *Aeromonas hydrophila* – ARTIGO I

FIGURE 1. Phylogenetic tree of the <i>Aeromonas</i> species	51
FIGURE 2. Comparison of O antigen biosynthesis clusters encoded in predicted GIs...	56
FIGURE 3. Sharing of predicted GIs among <i>A. hydrophila</i> strains	62

CAPÍTULO IV: TAXONOMIA de *Aeromonas* – ARTIGO II

FIGURA 1. Descrição das etapas do trabalho.....	88
FIGURA 2. Árvore filogenética (proteoma).....	90
FIGURA 3. Árvore filogenética (genes <i>housekeeping</i>).....	92

LISTA DE TABELAS

CAPÍTULO III: PREDICTION AND ANALYSIS IN SILICO OF GENOMIC ISLANDS IN *Aeromonas hydrophila* – ARTIGO I

TABELA 1. General characteristics of <i>Aeromonas hydrophila</i>	45
TABELA 2. General data of predicted GIs	53
TABELA 3. Main traits encoded in predicted GIs	58

CAPÍTULO IV: TAXONOMIA de *Aeromonas* – ARTIGO II

TABELA 1. Comparação dos valores de ANI entre espécies, <i>A. hydrophila</i> , <i>A. media</i> , <i>A. rivipollensis</i>	94
TABELA 2. Valores de ANI obtidos para <i>A. hydrophila</i> atcc 7966 e <i>Aeromonas</i> sp.....	95
TABELA 3. <i>A. caviae</i> 8LM comparadas com <i>A. caviae</i> SS332 e <i>Aeromonas</i> sp	95
TABELA 4. Comparação de valores de ANI de cepas <i>A. veronii</i> ; <i>Aeromonas</i> sp; <i>A.</i> <i>jandaei</i> ; <i>A. allosaccharophila</i>	97
TABELA 5. <i>A. jandaei</i> em comparação com espécies <i>A. veronii</i> e <i>Aeromonas</i> . sp.....	98
TABELA 6. Comparação ANI entre espécie <i>A. hydrophila</i> e <i>A. dhakensis</i>	98
TABELA 7. Valores de ANI de <i>Aeromonas</i> sp FDAARGOS 1407 comparados com <i>A.</i> <i>enteropelogenes</i>	99
TABELA 8. Bactérias classificadas erroneamente no NCBI	100
TABELA 9. Resultado da nova classificação de <i>Aeromonas</i> sp.....	101

SUMÁRIO

CAPITULO I: INTRODUÇÃO.....	13
1. CONSIDERAÇÕES INICIAIS	13
2. DESCRIÇÃO DO PROBLEMA.....	15
3 OBJETIVOS	16
3.1 <i>Objetivo geral</i>	16
3.2 <i>Objetivos específicos</i>	16
4 ORGANIZAÇÃO DA TESE.....	17
CAPÍTULO II: REVISÃO DA LITERATURA.....	18
1 <i>Aeromonas</i>	18
1.1 <i>Fatores de virulência</i>	21
1.2 <i>Resistência aos antibióticos</i>	24
2 TAXONOMIA.....	25
3 ILHAS GENÔMICAS	28
3.1 <i>História, características e funções</i>	28
3.2 <i>Métodos e ferramenta de predição de ilhas genômicas</i>	31
3.3 <i>Métodos de predição baseados em composição de sequência – Genes</i>	31
3.4 <i>Métodos de predição baseados em composição de sequência – DNA</i>	32
3.5 <i>Métodos de predição baseados em estrutura – Ilha genômica</i>	34
3.6 <i>Métodos de predição baseados em análise comparativa – múltiplos genomas</i>	35
3.7 <i>Métodos de predição Baseados em métodos híbridos</i>	37
CAPÍTULO III: PREDICTION AND ANALYSIS IN SILICO OF GENOMIC ISLANDS IN	
<i>Aeromonas hydrophila</i> – ARTIGO I	39
ABSTRACT	41
1 INTRODUCTION.....	42
2 MATERIALS AND METHODS.....	43
2.1 <i>A. hydrophila</i> genomes	43
2.2 <i>Phylogenetic analysis</i>	47
2.3 <i>GI prediction and analysis</i>	47

2.3	<i>Tools and databases for traits prediction</i>	47
2.4	<i>Phage prediction</i>	48
2.5	<i>Shared GIs</i>	48
2.6	<i>Core Genome and related GIs</i>	48
2.7	<i>Cluster alignments</i>	49
2.8	<i>Cluster alignments</i>	49
3	RESULTS	50
3.1	<i>Phylogenetic analysis and general characteristics of A. hydrophila strains</i>	50
3.2	<i>Characteristics of predicted GIs of A. hydrophila strains</i>	52
3.4	<i>Functions encoded by the predicted GIs</i>	55
3.4	<i>Predicted GIs containing genes from core genome</i>	61
3.5	<i>Shared GIS</i>	61
4	DISCUSSION	63
	REFERENCES	71

CAPÍTULO IV: TAXONOMIA Aeromonas – ARTIGO II	78
RESUMO	78
ABSTRACT	79
1 INTRODUÇÃO	80
2 REVISÃO DA LITERATURA	81
2.1 Métodos baseados no gene 16S rRNA	82
2.2 Métodos baseados em identificação fenotípica	83
2.3 Métodos baseados em genes “housekeeping”	83
2.4 Métodos baseados em genotipagem	84
2.5 Métodos baseados em genômica	85
2.6 Métodos baseados em genoma completo	86
3 MATERIAIS E MÉTODOS	86
3.1 Genomas	86
3.2 Ferramentas para análise e visualização	87
3.3 Tratamento das sequências para análise	88
4 RESULTADOS E DISCUSSÃO	89
4.1 Aeromonas media (12 cepas)	93
4.2 Aeromonas hydrophila (42 cepas)	94

4.3 <i>Aeromonas caviae</i> (29 cepas).....	SUMÁRIO	95
4.4 <i>Aeromonas veronii</i> (25 cepas)		95
4.5 <i>Aeromonas salmonicida</i> (19 cepas)		97
4.6 <i>Aeromonas jandaei</i> (7 cepas)		97
4.7 <i>Aeromonas dhakensis</i> (4 cepas).....		98
4.8 <i>Aeromonas enteropelogenes</i> (4 cepas)		98
4.9 <i>Aeromonas schubertii</i> (2 cepas).....		99
4.10 <i>Aeromonas simiae</i> (1 cepa)		99
4.11 <i>Aeromonas allosaccharophila</i> (1 cepa)		99
4.12 <i>Aeromonas encheleia</i> (1 cepa)		99
4.13 <i>Aeromonas rivipollensis</i> (1 cepa).....		99
4.14 <i>Aeromonas sanarellii</i> (1 cepa).....		99
4.15 <i>Aeromonas rivuli</i> (1 cepa).....		100
5 CONCLUSÕES PRELIMINARES.....		100
REFERÊNCIA		103

CAPÍTULO I: INTRODUÇÃO

1 CONSIDERAÇÕES INICIAIS

Com um maior número de genomas sequenciados foi possível obter diversas informações genéticas dos organismos em grande escala, demonstrando que, além da presença dos genes do “core” genoma, que codificam as funções metabólicas essenciais, estão presentes também uma grande quantidade de genes acessórios adquiridos por transferência horizontal (Schmidt; Hensel, 2004); (Bello-López et al., 2019).

Genomas bacterianos possuem característica de fluxo constante de perda e ganho de material genético (Soucy; Huang; Gogarten; 2015). Estes eventos estão relacionados a diversos fatores, como por exemplo, a presença de elementos genéticos móveis auto transmissíveis (MGEs), profagos e elementos conjugativos, que demonstraram ser um recurso muito importante (Frost et al., 2005); (Tekedar et al., 2019). Com isso, as bactérias podem adquirir características adaptativas benéficas sob certas condições ambientais ou de crescimento (Koonin, 2016); (Liu et al., 2018).

A evolução e transformação dos genomas bacterianos está relacionado diretamente com a transferência de genes entre as diferentes espécies bacterianas. Isso ocorre principalmente devido as Ilhas Genômicas, conhecidas também como “GIs”. Estas GIs, são caracterizados como blocos de genes que podem atribuir determinadas funções acessórias para a bactéria. As GIs podem estar presentes em genomas procarióticos patogênicos e não patogênicos (Hacker; Kaper, 2000); (Selvaraj et al., 2022).

As GIs foram descritas pela primeira vez no ano de 1990, por Hacker e colaboradores. Eles identificaram clusters de genes virais que (codificam genes de hemolisina) estavam presentes em genomas de várias cepas de *Escherichia coli*, e ausentes em outras cepas. Estes clusters foram denominados como Ilhas de Patogenicidade (PAIs) (Hacker et al; 1990). Desde então, outras GIs foram caracterizadas e de acordo com o seu conteúdo genético identificadas como GIs de Resistência a Antibióticos, Metabólicas, de Simbiose (Hacker; Kaper, 2000); (Fu et al., 2022).

Como exemplo de funções codificadas em GIs e relacionadas com a capacidade de sobrevivência de patógenos no corpo do hospedeiro, levando a doenças, estão o sistema de secreção do tipo III, a capacidade de absorção de ferro, a produção e secreção de toxinas e as adesinas (Dobrindt et al, 2004); (Gal-Mor; Finlay, 2006); (Pang et al., 2015). Existem estudos que sugerem que a patogenicidade pode ser regulada pela perda

seletiva ou recuperação de GIs específicas (Lawrence, 2005); (Manson; Gilmore, 2006); (Gonçalves et al., 2020) e que as PAIs, podem ser transferidas ou deslocadas dos cromossomos a uma taxa detectável, ocasionando diferentes fenótipos patogênicos. A detecção de diversas GIs, com suas características distintas e diferentes funções codificadas, tornou-se um fator importante no estudo da evolução microbiana (Finlay; Falkow, 1997); (Middendorf et al., 2004); (Bertelli; Tilley; Brinkman, 2019).

O gênero *Aeromonas* consiste de bacilos gram-negativos, anaeróbios facultativos que pertencem à família *Aeromonadaceae*. São compostos por organismos mesófilos, capazes de desenvolver-se melhor em condições de temperatura moderada, e psicrófilos, que se reproduzem em baixas temperaturas. Algumas espécies podem causar uma série de doenças em animais de sangue quente e frio (Igbinosa et al., 2012); (Guz et al., 2021). Nos últimos anos, *Aeromonas* mesófilas tem chamado atenção por serem associadas com doenças transmitidas por alimentos (Pablos; Remacha; Rodriguez-Calleja, 2010); (Hoel; Vadstein; Jakobsen, 2019). As gastroenterites e infecções de feridas são as doenças mais frequentemente associadas com *Aeromonas* em humanos. No entanto outras infecções tais como septicemia, infecções do trato urinário, infecções de feridas, infecções do trato hepatobiliar, pneumonia, infecções oculares, fascíte necrosante entre outras também podem ser causadas por essas bactérias (Janda; Abbott, 2010). As quatro espécies mais frequentemente relacionadas com infecção em humanos são: *A. hydrophila*, *A. caviae*, *A. veronii biovar sobria*, *A. dhakensis* (Janda; Abbott, 2010), (Fernandez-Bravo; Figueras, 2020)

Aeromonas são descritas como resistentes a ampicilina e a cefalosporinas de 1ª geração, exceto a espécie *A. trota* (Fernandez-Bravo; Figueras, 2020). Os antibióticos fluoroquinolona e aminoglicosídeos geralmente apresentaram boa atividade antimicrobiana para espécies de *Aeromonas* isoladas de fontes clínicas. No entanto foi observado resistência para cefalosporina de 3ª geração. O aumento do uso inapropriado dos antibióticos na aquicultura e no tratamento humano, ocasionou um aumento significativo da resistência das bactérias aos antimicrobianos (Tena et al., 2007); (Li et al., 2015); (Ghenghesh et al., 2015).

Com o uso indevido e generalizado dos antimicrobianos, patógenos bacterianos resistentes estão surgindo rapidamente (Andersson; Hughes, 2014); (Lamy; Baron; Barraud, 2022). O impacto no meio ambiente devido à pressão seletiva exercida pelos antibióticos, torna-se muito preocupante, e não apenas em ambientes restritos, como hospitais por exemplo. A presença destes compostos pode-se acumular e estar presente

em diferentes comunidades, levando ao crescimento do resistoma ambiental e gerando diversos depósitos de novos genes de resistência (Tyrrell et al., 2019). Por exemplo, uma prática tradicional na área da agricultura é a utilização de esterco para devolver os nutrientes ao solo. Entretanto, não é possível mensurar o impacto do uso contínuo de estrume para microbioma ambiental e no resistoma. A pressão seletiva exercida pelos antibióticos, utilizados como fator de crescimento em animais, em doses terapêuticas e subterapêuticas, no microbioma intestinal do animal, excretado nas fezes. Com isso, a aplicação de esterco em terras agrícolas pode se tornar um grande problema devido a possível transmissão de bactérias resistentes a antibióticos do gado para as plantações, animais e humanos (Tyrrell et al., 2019). Diante disso, as *Aeromonas* que se encontram em grandes proporções, distribuídas no solo e presentes também em ecossistemas aquáticos, possuem uma atribuição relevante como reservatório de genes de resistência antimicrobiana (Kümmerer, 2009); (Chen et al., 2016).

A virulência em *Aeromonas* é multifatorial e está relacionada a expressão de genes como *exoA*, *alt*, *act*, entre outros que codificam diferentes toxinas, componentes estruturais como flagelos e adesinas, sistemas de secreção (T3SS, T6SS, etc), e metaloproteases (Tomás, 2012); (Beaz-Hidalgo; Figueras, 2013); (Kim et al., 2019). Estes fatores de virulência permitem que a bactéria seja capaz de colonizar, invadir e superar os mecanismos de resposta imune do hospedeiro, ocasionando a infecção e doenças (Janda; Abbott, 2010); (Tomás, 2012) (Fernández-Bravo et al., 2020).

2. DESCRIÇÃO DO PROBLEMA

Aeromonas é um bacilo gram-negativo de habitat aquático. Se encontra amplamente distribuído no ambiente e algumas espécies, entre elas *A. hydrophila*, podem causar infecções em humanos e animais. Este gênero de bactérias possui uma grande diversidade de fatores de virulência, tornando-se um organismo muito versátil. *Aeromonas* podem causar gastroenterites e infecções em diversos órgãos e são consideradas um patógeno emergente. Entretanto, existem dificuldades na taxonomia do gênero, na identificação laboratorial errônea dessas bactérias como membros de outras espécies, e mais estudos são necessários para esclarecer a patogênese de *Aeromonas*.

Ilhas Genômicas (GIs), são clusters de genes que estão presentes em algumas estirpes do mesmo gênero, mas ausentes em outras. O interesse em ilhas genômicas cresceu exponencialmente nos últimos anos devido a capacidade de sequenciamento e

análise de genoma microbianos em larga escala. As predições das ilhas genômicas se tornaram essenciais para conduzir investigações de populações microbianas e estudos detalhados sobre genômica comparativa. Na área da microbiologia clínica e epidemiologia por exemplo, identificar regiões genômicas adquiridas se torna cada vez mais necessário. A descobertas destas regiões pode ajudar a identificar ilhas genômicas associadas a um uma cepa patogênica, relacionada a um surto de doença, ou que tenha um fenótipo particular de interesse, como maior virulência, transmissibilidade ou resistência aos antibióticos.

3 OBJETIVOS

3.1 Objetivo geral

- Definir a taxonomia do gênero *Aeromonas* e realizar a predição de ilhas genômicas na espécie *A. hydrophila*, caracterizando sua diversidade genética em relação as propriedades de virulência, metabolismo e resistência a antibióticos.

3.2 Objetivos específicos

- Analisar a classificação taxonômica do gênero *Aeromonas*, através de genomas completos.
- Identificar possíveis classificações errôneas das espécies depositadas no NCBI.
- Realizar a predição de ilhas genômicas *in silico* em genomas completos da espécie *A. hydrophila*.
- Identificar as possíveis funções biológicas codificadas nas ilhas genômicas relacionando as informações com bancos de dados de referência em virulência, metabolismo e resistência a antibióticos.
- Analisar e determinar a relação das ilhas genômicas entre as cepas e seus possíveis impactos na área clínica e aquicultura.

4 ORGANIZAÇÃO DA TESE

- O capítulo 1, apresenta as considerações iniciais, descrição do problema e os objetivos.
- O capítulo 2, descreve a revisão da literatura das *Aeromonas*, sua taxonomia e contexto geral sobre Ilhas Genômicas.
- O capítulo 3, refere-se ao artigo publicado na *Frontiers in Microbiology* (Qualis A1 – Fator de impacto 5.6).
- O capítulo 4, corresponde ao segundo artigo, estudos taxonômicos do gênero *Aeromonas*. O artigo inclui: resumo, introdução, revisão bibliográfica, materiais e métodos, resultados, discussão e conclusão.

Pelo desenvolvimento de muitos materiais suplementares de grande extensão dos artigos, os arquivos foram disponibilizados em plataformas digitais de compartilhamento, com o objetivo de reduzir o tamanho do documento da tese sem omitir informações que fazem parte do trabalho. Os arquivos consistem em Supplementary_Tables_artigo_1, Supplementary_figures_artigo_1, Tabelas_Suplementares_artigo_2 depositados no Google Drive com link público (<https://drive.google.com/drive/folders/1td4wfckp9H65-nydjLLy7PqbVNO3e-k7?usp=sharing>).

CAPÍTULO II: REVISAO DA LITERATURA

1 *Aeromonas*

Os primeiros isolados de *Aeromonas* foram reportados em 1891, por Sanarelli e colaboradores que descreveram a bactéria então designada como *Bacillus hydrophilus fuscus*. Entretanto, o Comitê Internacional de Bacteriologia Sistemática estabeleceu a autoria do gênero para Stanier em 1943. Em 1965, *Aeromonas* foram incluídas na família *Vibrionaceae*. Somente após os estudos de Colwell e colaboradores em 1986, sobre sequência dos genes 16S rRNA, 5S rRNA e hibridização DNA-DNA, mostrando que *Aeromonas* formavam um grupo filogenético distinto, foi criada a família *Aeromonadaceae* (Colwell et al., 1986; citado por Fernandez-Bravo e Figueras, 2020).

Com base em suas características, as espécies do gênero *Aeromonas* foram divididas em dois grupos: crescimento ótimo a 35-37 °C, com isolados móveis, responsáveis por várias infecções em humanos e definido como “*A. hydrophila*” nomeado como “mesofílico”; crescimento ótimo 22–28 °C, com cepas não móveis, ligadas a doenças de peixes e identificadas em “*A. salmonicida*”, sendo nomeado como “psicrofílico”. As *Aeromonas* são ubíquas no ambiente aquático e amplamente distribuídas no solo e alimentos. São bacilos gram-negativos aeróbias facultativos, não ácido-resistentes, crescem em 0 a 4% de NaCl (Abott, 2010); (Hoel; Vadstein; Jakobsen, 2019). Quando estes patógenos entéricos emergentes entram em contato com o sistema de distribuição de água formam biofilmes (Igbinosa et al., 2012); (Thomas et al., 2020). Esta bactéria pode causar várias complicações, como, a doença da perna vermelha em rãs e ascite em carpas, tendo impacto direto na economia (Slevin et al., 1988); (Liu et al., 2020).

Aeromonas podem ser encontradas e isoladas em vários ambientes aquáticos, tais como águas de superfície, subterrâneas, água potável, engarrafada, residual, água do mar e de irrigação (Janda; Abbott, 2010); (Guz et al., 2021). Diversos estudos relataram que a incidência de *Aeromonas* em águas residuais é alta (Kampfer et al., 1996); (McLellan et al., 2010); (Al-Jassim et al., 2015); (Varela; Nunes; Manaia, 2016); (Gomes et al., 2021). A utilização de tratamento primário, secundário ou biológico para recuperação de água em estações de tratamento de efluentes, não reduzem muito a concentração de *Aeromonas* (Salvat; Ashbolt, 2019). Procedimentos adicionais, como tratamento químicos (ozônio, cloração) e físicos (radiação ultravioleta), podem resultar na eliminação por completo das *Aeromonas* (Latif-Eugenín; Beaz-Hidalgo; Figueras, 2016a),

(Latif-Eugenín; Beaz-Hidalgo; Figueras, 2016b); (Fernandez-Cassi et al., 2016); (Latif-Eugenín et al., 2017).

Em relação a tratamentos naturais, o procedimento é realizado através do armazenando temporário de águas residuais em lagoas ou lagoas rasas, com objetivo de purificação para reduzir os níveis de *Aeromonas* e dependendo da qualidade microbiológica da água recuperada, ela pode ser usada para irrigação (Jjemba et al., 2010); (Aravena-Román et al., 2013); (Cui; Liang, 2019). Por outro lado, a utilização da água da irrigação contaminada por *Aeromonas* pode afetar frutas e vegetais, causando infecções em humanos (Pianetti et al. 2004); (Latif-Eugenín; Beaz-Hidalgo; Figueras, 2016a). Resultados epidemiológicos confirmam que o veículo de disseminação dessa bactéria é a água. Os dados mostraram que foi possível identificar as mesmas cepas de *Aeromonas* em amostras de água potável e nas fezes de indivíduos com diarreia (Khajanchi et al., 2010); (Pablos et al., 2011); (Liu et al., 2022). Cepas isoladas de alface e tomate foram recuperadas da água de irrigação (Latif-Eugenín et al., 2017).

Aeromonas também foram identificadas em água potável. A quantidade de nutrientes disponíveis, temperatura e concentração de cloro residual influência as concentrações desta bactéria (Borrel; Figueras; Guarro, 1998); (Figueras et al., 2005); (Salvat; Ashbolt, 2019). Latif-Eugenín e colaboradores também mostraram a importância das condições de cultivo para a detecção de *Aeromonas* em água (Latif-Eugenín; Beaz-Hidalgo; Figueras, 2016a). Devido à sua capacidade de crescer no sistema de esgoto, estudos metagenômicos de águas residuais revelaram que *Aeromonas* é uma das bactérias mais predominantes (McLellan et al., 2010); (McLellan; Roguet, 2019).

Diversas cepas de *Aeromonas* foram isoladas de diferentes alimentos, como, frutas, vegetais, laticínios, carnes e salsichas, peixes e mariscos (Abbott et al., 2003); (Chang et al., 2008); (Hoel; Vadstein; Jakobsen, 2017); (Hammad et al., 2018); (Hoel; Vadstein; Jakobsen, 2019). Os fatores que determinam o número de *Aeromonas* nestes alimentos são a temperatura, salinidade ou pH (Martin-Carnahan; Joseph, 2005); (John et al., 2019). Em relação a temperatura, diversos estudos mostram que as *Aeromonas* conseguem sobreviver em temperaturas mais baixas (2–10 °C). Conservante comum para produtos como carnes e peixes é o sal ou o cloreto de sódio (NaCl). Entretanto, as *Aeromonas* são capazes de crescer em concentrações de NaCl de até 4%. Um critério para o crescimento bacteriano é o pH, que foi demonstrado a capacidade das *Aeromonas* sobreviver em pH = 5 (John; Vidyalakshmi; Hatha, 2019).

As espécies mais prevalentes em carnes são *A. veronii*, *A. salmonicida*, *A. jandaei*, *A. simiae*, *A. hydrophila* e *A. caviae* com base no sequenciamento dos genes “rpoD” e “gyrB” (Fontes et al., 2010); (Nagar; Shashidhar; Bandekar, 2013); (Elbehiry et al., 2019). A tilápia e os salmonídeos, peixes destinados ao consumo humano, apresentam *Aeromonas* spp. que após identificação molecular correspondiam a *A. bivalvium*, *A. bestiarum*, *A. encheleia*, *A. veronii*, *A. allosaccharophila*, *A. salmonicida* e *A. hydrophila* (Castro-Escarpulli et al., 2003); (Figueras; Beaz-Hidalgo, 2014). Com base em métodos de identificação molecular, a diversidade de espécies de *Aeromonas* isoladas de mexilhões, ostras, camarões e berbigões é semelhante às espécies encontradas em peixes (Figueras; Beaz-Hidalgo, 2014); (Woodring et al., 2012); (Yano et al., 2015).

A identificação de *Aeromonas* em frutas e vegetais, como tomate, alface, pepino, batata, cebola, cenoura, foram descritas em diversos estudos (Chang et al., 2008); (Nagar; Shashidhar; Bandekar, 2011); (Nagar; Shashidhar; Bandekar, 2013); (Balali et al., 2020). A água contaminada, em conjunto com estes alimentos, se torna a principal fonte de infecção por *Aeromonas* em casos de diarreia (Figueras; Borrego, 2010); (Figueras; Beaz-Hidalgo, 2014). As espécies clínicas mais prevalentes detectadas foram, *A. hydrophila* e *A. caviae* (Chang et al., 2008); (Nagar; Shashidhar; Bandekar, 2011); (Nagar; Shashidhar; Bandekar, 2013) (Surek et al., 2010); (Prediger et al., 2012); (Assis et al., 2014); (Rossato et al., 2015).

Com estudos descrevendo septicemias em rãs e peixes doentes, o gênero *Aeromonas* tem sido considerado um patógeno animal. As espécies *A. salmonicida* e *A. hydrophila* são espécies patogênicas que afetam principalmente os salmonídeos. Podem causar diversas complicações, como úlceras, hemorragias, furunculose e septicemias (Farmer; Arduino; Hickman-Brenner, 2006); (Beaz-Hidalgo et al., 2009); (Beaz-Hidalgo et al., 2010); (Hayatgheib et al., 2020). Surtos mundiais em peixes de águas quentes começaram a aparecer, sendo o agente responsável uma cepa hipervirulenta de *A. hydrophila* de acordo com os estudos de (Rasmussen-Ivey et al., 2016). Devido a estas causas, as indústrias da aquicultura sofreram grandes perdas financeiras Rasmussen-Ivey et al. (2016); (Vincent et al., 2019).

Existem vários estudos que determinaram a incidência de *Aeromonas* em infecções humanas ao redor do mundo. Um estudo na Califórnia de 1988, identificou que cerca de 10,5 casos por milhão de pessoas tiveram infecções por *Aeromonas* anualmente (King; Werner; Kizer, 1992). Em 2004 na Inglaterra, os casos de bacteremia por *Aeromonas* chegaram a 1,5 casos por milhão de pessoas de acordo com (Jand; Abbott,

2010). Na França, no ano de 2006, foi estimado 0,66 casos por milhão de pessoas (Lamy; Kodjo; Laurent, 2009). Durante os anos de 2008 e 2010 em Taiwan, um estudo demonstrou que a incidência de bacteremia por *Aeromonas* foi de 76 casos por milhão de pessoas (Wu et al., 2014). A incidência de doenças causadas por *Aeromonas* varia, e pode estar relacionada a localização geográfica e hábitos de higiene em regiões de baixo desenvolvimento (Wu et al., 2014); (Tsai et al., 2015); (Baltra; Mathur; Misra, 2016);

Somente quatro espécies de *Aeromonas* são responsáveis por 96,5% das infecções humanas causadas por bactérias dessa espécie. Entre elas, *Aeromonas caviae* (29,9%), *Aeromonas dhakensis* (26,3%), *Aeromonas veronii* (24,8%) e *Aeromonas hydrophila* (15,5 %) (Figueras; Beaz-Hidalgo, 2015). A espécie *Aeromonas salmonicida*, geralmente associadas a doenças de peixes, também foram relatadas em infecções humanas (Vincent et al., 2019).

1.1 Fatores de virulência

A patogênese de *Aeromonas* é multifatorial (Tomás, 2012); (Beaz-Hidalgo; Figueras, 2013) (Grim et al., 2013); (Reyes-Rodriguez et al., 2019). Diversos fatores de virulência participam na patogênese, facilitando a colonização bacteriana e superando a resposta imune do hospedeiro (Figueras; Beaz-Hidalgo, 2015). Na fase inicial de infecções, a etapa crítica ocorre na adesão de bactérias aos tecidos do hospedeiro. A colonização ocorre quando as bactérias começam a aderir nas células e tecidos de seu hospedeiro, afetando assim os seus meios de defesa. A camada S, pili, lipopolissacarídeos (LPS), cápsula e os flagelos são os componentes estruturais mais estudados em *Aeromonas* envolvidos no processo de colonização (Tomás, 2012); (Pessoa et al., 2019).

A camada S é composta por subunidades de uma única proteína ou glicoproteína, e forma o envelope celular mais externo sendo produzida por uma ampla gama de bactérias. Devido a sua exposição na superfície celular, a camada S está relacionada a várias funções de patogenicidade tais como adesão, proteção contra o complemento e ataque por fagócitos, entre outras (Tomás, 2012; Beaz-Hidalgo; Figueras, 2013; Fernández-Bravo; Figueras, 2020). Em *Aeromonas* mesofílicas a camada S foi descrita em cepas pertencentes ao sorotipo O:11, e em *A. hydrophila* patogênicas sorotipos O:14 e O:81 (Tomás, 2012); (Rasmussen-Ivey, 2019).

Pili são estruturas proteicas filamentosas relacionadas com adesão encontradas na superfície bacteriana. Podem apresentar diferentes funções, incluindo invasão de

células hospedeiras, agregação celular, transferência de DNA, formação de biofilme (Beaz-Hidalgo; Figueras, 2013).

Os lipopolissacarídeos são constituídos por três subunidades: o polissacarídeo O (antígeno O), o núcleo do LPS (polissacarídeo central) e o lipídio A, que compõe a membrana externa das bactérias gram-negativas (Heiss et al., 2019). Antígeno O tem a função de proteger as bactérias infectantes da morte do complemento sérico e da fagocitose, mediando a sua patogenicidade. Se encontra na parte mais exposta à superfície do lipopolissacarídeo, um componente da membrana externa das bactérias gram-negativas. Integra repetição de subunidades de oligossacarídeos, fornecendo especificidade imunológica devido a sua variabilidade (Tomás, 2012). Existem vários tipos de agrupamentos de genes do antígeno O descritos em *Aeromonas hydrophila* (Hossain et al., 2013; Pang et al., 2015; Cao et al., 2018) e previamente associadas a Glis (Hossain et al., 2013; Pang et al. al., 2015).

As cápsulas revestem a membrana externa das bactérias, são estruturas polissacarídicas que tem a capacidade de interagir com o meio ambiente. Sua característica mais importante é a capacidade de resistir à fagocitose e ao sistema complemento, sendo um importante fator de virulência (Merino; Tomás, 2010); (Heiss et al., 2019).

Aeromonas mesofílicas possuem flagelo polar e muitas cepas também expressam múltiplos flagelos laterais, de modo induzível. Estes flagelos são organelas que tem como função a mobilidade bacteriana. A sua estrutura é composta por um filamento, gancho e corpo basal (Beaz-Hidalgo; Figueras, 2013); (Mendoza-Barberá; Merino; Tomás, 2021). As funções dos flagelos polar e lateral são distintas, o primeiro está associado com a motilidade da bactéria em meios líquidos, enquanto a capacidade das bactérias se moverem por superfícies sólidas ou viscosas é devido ao flagelo lateral (Tomás 2012); (Lowry et al., 2014). Ambos participam na formação de biofilme e estão associados com adesão em enterócitos humanos (Kirov et al., 2004) (Janda e Abbott, 2010); (Prediger et al., 2020) e na adesão e invasão de linhagens de células de peixes (Beaz-Hidalgo, Figueras, 2013).

Em bactérias gram-negativas foram descritos seis sistemas de secreção designados como Sistema de Secreção tipo I a VI. Estes sistemas diferem quanto a sua estrutura, substratos secretados e modo de secreção, no meio extracelular ou diretamente nas células hospedeiras (Tseng; Tyler; Setubal, 2009); (Moriel et al., 2021). Três deles foram associados com virulência em *Aeromonas* (Tomás, 2012); (Marcoux et al., 2020).

O Sistema de Secreção Tipo II (SSTII), considerado como via geral de secreção, desempenha funções importantes em organismos ambientais e patógenos oportunistas. A função do SSTII e seus efetores na patogenicidade e virulência foi estudado em *A. hydrophila* ML09-119, um patógeno de peixes. Foi mostrado que um mutante deficiente na produção de SSTII teve a virulência completamente atenuada em ensaios *in vivo* em *catfish*. Além disso, verificou-se, no mutante, a ausência de grande parte das proteínas secretadas no meio de cultura incluindo proteínas de virulência degradativas como elastases. A atividade hemolítica foi eliminada, sugerindo que a vasta maioria das hemolisinas são secretadas via SSTII. Portanto, o SSTII e seus efetores desempenham uma função essencial na patogenicidade da ML09-119 (Barger; Liles; Newton, 2020).

Diferentemente do SSTII, que secreta proteínas no ambiente, os sistemas de secreção tipo III (SSTIII) e VI (SSTVI), secretam seus efetores (proteínas tóxicas) diretamente no citoplasma das células alvo (Tomás, 2012).

Cepas de *A. hydrophila* e *A. salmonicida* mutantes para SSTIII apresentaram virulência reduzida em relação às cepas selvagens (Yu et al., 2004), (Fernández-Bravo, Figueras, 2020). Distintos modos de ação foram identificados para estes efetores SSTIII; AexT é uma toxina bifuncional com atividade de ADP ribosiltransferase e GTPase; AopP interfere com as vias de sinalização das proteínas quinases mitógeno-ativadas (MAPK) e/ou do fator nuclear kappaB (NF-kB); AopH tem atividade de tirosina fosfatase; AopO serina/treonina quinase e AexU similar à toxina AexT (Tomás, 2019) (Fernández-Bravo, Figueras, 2020). Estudo recente utilizando ferramentas de bioinformática e citotoxicidade em levedura, identificou 23 possíveis efetores do SSTIII em *Aeromonas* dos quais apenas 8 (AexT, AextU, AopP, AopH, AopO, Ati2 e AopX e AopS inicialmente identificados como pseudogenes em *A. salmonicida*) eram conhecidos (Rangel et al., 2019). Estudos sobre estes efetores poderão esclarecer os seus mecanismos de ação e ampliar o conhecimento sobre a participação do SSTIII na virulência de *Aeromonas*.

O SSTVI pode translocar proteínas efetoras diretamente nas células alvo bacterianas (SSTVI antibacteriano) ou eucarióticas (SSTVI anti-eucariótico). A translocação dos efetores é realizada através de uma estrutura tubular, formada por hexâmetros da proteína Hcp, em cuja extremidade se encontra o dispositivo perfurante, formado pelas proteínas VgrG e PAAR, responsável pela perfuração da membrana da célula alvo (Cianfanelli, Monlezun, Coulthurst, 2016). Além da função como componentes estruturais do SSTVI, Hcp e VgrG também apresentam atividade como efetores (Suarez et al., 2008, 2010) (Sha et al., 2013). Diferentes proteínas efetoras podem ser

translocadas fundidas ou por interação não covalente com um dos componentes do tubo expelido (Cianfanelli, Monlezun, Coulthurst, 2016).

A participação na virulência, induzindo apoptose em células Hela, foi demonstrada para Hcp e VgrG-1 (apresenta uma extensão C-terminal com atividade de actina ADP-ribosilase); Hcp inibe a fagocitose por macrófagos, desempenhando um papel na modulação da resposta imune; mutantes para Hcp e VgrG tiveram sua virulência em camundongos reduzida (Suarez et al., 2008, 2010 a,b) (Sha et al., 2013).

Posteriormente foi observado que o SSTVI pode apresentar atividade antibacteriana, eliminando bactérias competidoras nas infecções ou no meio ambiente (Cianfanelli, Monlezun, Coulthurst, 2016) (Fernández-Bravo, 2020). Estudo recente, baseado em análise genômica, mostrou que a presença de SSTVI é comum entre *Aeromonas* e foi encontrado em 22 espécies do gênero incluindo isolados clínicos e ambientais (Moriel et al., 2020). Possíveis efetores foram localizados, entre eles genes codificando proteínas contendo domínios DUF1910 e DUF 1911 que podem representar um par toxina/antitoxina, e o domínio lisozima (Moriel et al., 2020).

1.2 Resistência aos antibióticos

A resistência aos antibióticos é uma decorrência da resposta genética evolutiva das bactérias, mediada pela existência de genes que podem ser encontrados em plasmídeos, integrons ou no genoma da bactéria (Fernandez-Bravo; Figueras, 2020). *Aeromonas* são descritas como resistentes à ampicilina com exceção de algumas cepas e da espécie *A. trota* (Carnahan et al., 1991); (Janda; Abbott, 2010). Estas bactérias também podem apresentar resistência a cefalosporinas de primeira geração e outras penicilinas (Aravena-Romnán et al., 2012). Porém, são suscetíveis a aminoglicosídeos, fluoroquinolonas, monobactâmicos, carbapenêmicos, fenicóis, tetraciclinas, macrolídeos, penicilinas de espectro estendido, e cefalosporinas de terceira e quarta geração (Janda, Abbott, 2010) (Codjoe; Donkor, 2017).

O aumento da resistência aos antimicrobianos beta-lactâmicos (cefalosporinas, carbapenêmicos, monobactâmicos, penicilinas e derivados) foi ocasionado devido a presença de genes que codificam a produção de beta-lactamases (Janda; Abbott, 2010); (Ndi; Barton, 2011). As metalo- β -lactamase de classe B, cefalosporinase de classe C e penicilinase de classe D, foram descritas como as três principais β -lactamases em *Aeromonas* (Dias et al., 2014). Estudos caracterizaram as β -lactamases relacionadas às diferentes espécies de *Aeromonas* em cinco grupos (I-V): (I) *A. trota* com β -lactamases

de classe C; (II) *A. veronii* contendo lactamases de classe B e D; (III) *A. schubertii* contendo lactamases de classe D; (IV) *A. caviae* expressando β -lactamases de classe C e D; (V) *A. hydrophila* expressando β -lactamases de classe B, C e D (Fosse; Giraud-Morin; Madinier, 2003; Fosse, 2010). No entanto, há relatos de *A. veronii* bv. sobria produtora de cefalosporinase de classe C (Janda; Abbott, 2010) e resistente ao imipenem (Sánchez-Céspedes et al., 2009).

A disseminação e aumento da resistência aos antibióticos, podem estar relacionadas com diversos fatores, como, elementos genéticos móveis, plasmídeos, elementos de inserção, cassetes associados a integrons e ilhas de patogenicidade (Piotrowska; Popowska, 2015). Isolados de *Aeromonas* de truta arco-íris, mostraram a presença de β -lactamases e integron de classe 1 (Vega-Sánchez et al., 2014). Estudo de (Pitowska et al., 2017) sugerem que cepas de *Aeromonas* resistentes a antibióticos podem se disseminar de águas residuais para outros ambientes. Encontraram uma grande variedade de genes de beta-lactamases presentes em plasmídeos. Hossain e colaboradores identificaram a ocorrência de vários genes de resistência antimicrobiana em integrons classe 1 em isolados resistentes de *Aeromonas* (Hossain et al., 2019).

2 TAXONOMIA

Na microbiologia clínica, uma das responsabilidades é atribuir nomes científicos para os isolados microbianos. Com esta identificação, é possível obter informações sobre o agente etiológico, suas relações patológicas como causador de uma doença infecciosa e possível terapia eficaz antimicrobiana (Zhou et al., 2019). Um dos primeiros métodos para realizar a identificação do isolado dependia da comparação de uma descrição morfológica e fenótipo. A partir do Manual de Bacteriologia Sistemática de Bergey ou o Manual de Microbiologia Clínica, os pesquisadores criam referencias padrões para estes isolados (Galusko; Kuever, 2020). Podem ainda compilar os resultados das características de cada espécie de bactéria a partir do Centro de Controle e Prevenção de Doenças ou “American Type Culture Collection”. (Krieg; Holt, 1984); (Murray et al., 1999). Entretanto, não haveria correspondência perfeita quando comparado o isolado clínico, deixando para o microbiologista o julgamento para a identificação. Programas de computadores ajudam nestes processos, mas isso pode variar entre os laboratórios (Sussman et al., 1986); (Simon et al., 2019).

Um novo padrão para identificação de bactérias começou a ser desenvolvido na década de 1980. Este método foi baseado na demonstração das relações filogenéticas

das bactérias. Relacionaram uma parte estável do código genético para identificação taxonômica. Para isto, os genes utilizados foram os que codificam para o 5S, 16S e rRNA 23S. Para fins taxonômicos bacterianos, o gene 16S rRNA é o mais utilizado (Böttger, 1989); (Bukin et al., 2019).

É muito difícil atribuir valores numéricos para essa proporção, devido que a ocorrência desta mudança na sequência do gene 16 rRNA pode não ocorrer da mesma forma para todos os organismos (grupos taxonômicos distintos podem ter proporções de mudanças diferentes) (Murali et al., 2018). Além disso, em alguns momentos durante a evolução, esta proporção pode ser diferente entre várias posições ao longo do gene. As regiões chamadas de “*hot spot*” apresentam um maior número de mutações, porém, estas regiões não são as mesmas para todas as espécies (Ueda et al., 1999); (Bukin et al., 2019). Vários agentes antimicrobianos também tem como alvo o gene 16S rRNA. Por consequência, as mutações presentes neste gene podem chegar a interferir na suscetibilidade dos organismos a estes agentes (Pfister et al., 2003); (Belete, 2019). As sequências do gene 16S rRNA no gênero *Aeromonas* demonstram que estes organismos são compostos por um grupo muito restrito de espécies, a diferença entre algumas delas é de poucos nucleotídeos com a similaridade variando entre 96,7 a 100% (Martinez-Murcia; Benlloch; Collins, 1992), (Alperi et al., 2008); (Pessoa et al., 2019).

Foram descobertas algumas diferenças na comparação entre o sequenciamento do gene 16S rRNA e os resultados de hibridização DNA-DNA, (Sneath, 1993); (Martinez-Murcia, 1999); (Awan et al, 2018). Foi relatado a presença de polimorfismo deste gene especialmente em cepas de *Aeromonas media* e *Aeromonas veronii*. Com estes indícios, outros genes foram utilizados como ferramentas para análise taxonômica e filogenética deste gênero (Morandi; Zhaxybayeva; Gogartenj, 2005); (El-Bahar et al., 2019).

Genes “*housekeeping*” são classificados como um dos melhores marcadores moleculares, em conjunto com o gene 16S rRNA, para análises filogenéticas e taxonômicas em nível de espécie. A utilização destes genes melhorou os resultados das filogenias, sendo mais confiáveis e ao mesmo tempo ajudando na compreensão da estrutura populacional bacteriana e também da epidemiologia (Stackebrandt; Frederiksen; Garrity, 2002); (Navarro; Martínez-Murcia, 2018).

Em 1995 Yamamoto & Harayama construíram um conjunto de “*primers*” que possibilitou a amplificação e também o sequenciamento dos nucleotídeos do gene *gyrB* (responsável pela codificação da subunidade B da DNA girase) de múltiplas bactérias. Os resultados evidenciaram que as análises filogenéticas da sequência desse gene

retratavam as relações evolutivas das espécies estreitamente relacionadas (Yamamoto; Harayama, 1996); (Yamamoto; Bouvet; Harayama, 1999). As análises do gene *gyrB* já foram utilizadas para responder questões filogenéticas das espécies de *Aeromonas* (Yáñez et al., 2003); (Soler et al., 2004) e também para identificar novas espécies (Pidíyar et al., 2004). Outro gene *housekeeping*, gene *rpoD*, que codifica a subunidade beta da RNA polimerase, também foi aplicada ao estudo taxonômico das *Aeromonas* (Soler et al., 2004), a inclusão destes genes na análise de *Aeromonas* ajudou a aperfeiçoar a taxonomia do gênero (Navarro; Martínez-Murcia, 2018).

O sequenciamento do gene 16S rRNA se tornou a técnica molecular mais utilizada no laboratório clínico para identificação de gênero e espécie das bactérias (Janda; Abbott, 2007); (Church et al., 2020). Entretanto, para o gênero *Aeromonas*, esta técnica molecular possui suas limitações (Fernández-Bravo et al., 2020). Em alguns relatos, o sequenciamento deste gene pode ser usado para identificar de forma definitiva a espécie (Hua et al., 2004) mas em outros casos não (Al-Benwan et al., 2007); (Fernández-Bravo et al., 2020). As causas dessas diferenças podem estar associadas pela aparente evolução em mosaico dos operons *rrn* das *Aeromonas* (Morandi et al., 2005); (Navarro; Martínez-Murcia, 2018).

A diversidade intragenômica devido a polimorfismos de nucleotídeos *rrn* foi identificada em grande parte das espécies de *Aeromonas*, com uma variação do mínimo de 0,06% a um máximo de 1,5% (Morandi et al., 2005); (Alperi et al., 2008). Um exemplo está em *Aeromonas veronii*, contendo cerca de 6 cópias do gene 16S rRNA. Estas cópias são diferentes entre si, o que impossibilita o seu uso para identificação de espécie no gênero *Aeromonas*. De acordo com Alperi e colaboradores (2008) a existência de microheterogeneidade nos genes 16S rRNA de *Aeromonas* interfere com a correta identificação dessas bactérias, e sugerem que outros genes sejam utilizados para este fim (Alperi et al., 2008); (Navarro; Martínez-Murcia, 2018).

Os valores de hibridização DNA-DNA, que foi por muito tempo o padrão ouro utilizado para validação das espécies bacterianas, podem não ser muito úteis para correlacionar as semelhanças da sequência do gene 16S rRNA no gênero *Aeromonas* (Christensen; Olsen, 2018). Tendo como exemplo *Aeromonas caviae* e *Aeromonas trota*, possuem apenas 30% de parentesco a nível de DNA, enquanto suas sequências de 16S rRNA diferem em apenas 3 nucleotídeos (Morandi et al., 2005). *Aeromonas sobria* junto com *Aeromonas veronii* possui cerca de 60% a 65% de relação em análises de pareamento de DNA. Suas diferenças de nucleotídeos chegam a 14 bases em sua

sequência do gene 16S rRNA (Martinez-Murcia; Benlloch; Collins, 1992). No estudo de Martínez-Murcia e colaboradores, *Aeromonas salmocida* e *Aeromonas bestiarum* não puderam ser diferenciadas pelo sequenciamento do gene 16S rRNA com uma variação de apenas 2 nucleotídeos (Martínez-Murcia et al., 2005). O uso do sequenciamento do gene 16S rRNA para identificação taxonômica para o gênero *Aeromonas* possui várias restrições, não sendo uma técnica muito útil para uso. Os genes “housekeeping” tem se mostrado uma opção para ajudar a solucionar essa problemática, incluindo os genes *gyrB*, *rpoD*, *rpoB* (Morandi et al., 2005); (Al-Benwan et al., 2007); (Alperi et al., 2008); (Fernández-Bravo; Figueras, 2020).

3 ILHAS GENÔMICAS

3.1 História, características e funções

Com o avanço computacional e das tecnologias de sequenciamento, é possível sequenciar genomas completos em larga escala. Isso proporciona aos pesquisadores realizar estudos genômicos comparativos (Chen et al., 2019). Devido ao aumento dos depósitos de genomas completos e com a capacidade dos métodos de bioinformática, é possível estudar e analisar as características associadas as bactérias e suas relações entre gêneros e espécies. Estas análises se tornam os primeiros passos para decifrar a genômica microbiana, e ajudam a decifrar as relações evolutivas (Bentley; Parkhill, 2004); (Dobrindt et al., 2004); (Zhu et al., 2019)

Um dos mecanismos que proporciona a diversidade de genes bacterianos entre as espécies é chamado de transferência horizontal de genes, conhecido também como (HGT) (Niehus et al., 2015). Esse processo evolutivo tem a capacidade de fornecer e distribuir genes entre linhagens bacterianas próximas ou distantes, sendo recurso significativo de variabilidade genética (Treangen; Rocha, 2011). Os clusters de genes adquiridos por estes processos, são comumente descritos como ilhas genômicas (GIs) (Hacker; Kaper, 2000); (Bertelli; Tilley; Brinkman, 2019).

GIs são clusters/conjunto de genes que se encontram no genoma bacteriano de determinadas cepas, mas ausentes em outras cepas da mesma espécie. O conceito de GIs foi originalmente criado em 1990, por Hacker e colaboradores. Eles estudavam isolados uropatogênicos de *Escherichia coli* (UPEC), analisando a sua base genética da virulência. Descobriram clusters de genes (codificam hemolisina), contendo fatores de

virulência, que estavam presentes em determinadas cepas e ausentes em outras. Assim, designaram esse cluster como Ilhas de Patogenicidade (PAIs) (Hacker et al., 1990). Depois dessa descoberta, as GIs representam agora um grupo muito mais extenso e variado do que somente as PAIs, presentes em vários genomas bacterianos (Dobrindt et al., 2004); (Banerjee, 2021).

A classificação das GIs pode ser feita de acordo com os seus genes funcionais. As ilhas de patogenicidade (PAIs), codificam genes importantes para patogenicidade/virulência bacteriana; ilhas de resistência (RIs), genes que codificam proteínas de resistência a antibióticos; ilhas metabólicas (MIs), possuem genes relacionados ao metabolismo; ilhas de simbiose (SIs), genes capazes de sustentar uma relação simbiótica hospedeiro-bactéria (Hacker et al., 1990); (Barcellos et al., 2007); (Bertelli; Tilley; Brinkman, 2019). Suas características principais são:

- GIs são segmentos de DNA relativamente grandes possuindo cerca de 10 a 200 mil pares de bases. Existem também regiões contendo menos que 10 kb que são denominadas de Ilhotas genômicas (Hacker; Kaper, 2000); (Juhas, 2019).
- Possui vários padrões que se diferenciam do restante do genoma. Estas regiões possuem sua porcentagem GC% menor do que a do genoma, diferenças nas frequências de tetranucleotídeos ou uso de códons (Hacker; Kaper, 2000); (Lu; Leong, 2018).
- Estão geralmente localizadas adjacentes a genes de tRNA, e são flanqueadas por repetições diretas (DRs), com 16 a 20 pares de bases. As DRs podem ser originárias devido a integração das GIs no genoma e podem ser utilizadas como sequências de reconhecimento para sua excisão (Schmidt; Hensel, 2004); (Bertelli; Tilley; Brinkman, 2019).
- Algumas GIs podem possuir genes funcionais, no quais estão relacionados a sistemas de conjugação de plasmídeos ou fagos (Schmidt; Hensel, 2004). Elementos de inserção e mobilização também podem estar presentes, como as integrases e transposons, responsáveis pela integração/mobilização do material genético das GIs (Gal-Mor; Finlay, 2006); (Juhas, 2019).

- Dependendo das funções relacionadas às GIs, ilhas de patogenicidade, simbiose, metabólica, fitness ou resistência, elas possuem genes que podem oferecer vantagem seletiva para as bactérias hospedeiras (Dobrindt et al., 2004; Schmidt; Hensel, 2004); (Lu; Leong, 2018)..

Muitas das GIs descritas na literatura seguem esta definição de classificação e características. Entretanto, um número significativo não possui uma ou mais das particularidades citadas acima. Isso pode ocorrer devido a várias GIs estarem em um estado evolutivo de regressão, como foi sugerido anteriormente por Dobrindt et al. (2004). Alguns estudos propõem que as GIs tiveram múltiplas origens evolutivas, devido algumas delas possuírem fagos e outros genes relacionados a plasmídeos conjugativos (Toussaint; Merlin 2002); (Frost et al., 2005); (Leplae, Lima-Mendez; Toussaint, 2006); (Juhas, 2019). As GIs podem apresentar alguns padrões em suas características e elementos, mas não é possível afirmar um mecanismo único para o estilo de vida, ou seja, as funções necessárias para manutenção, excisão, transferência ou integração nos genomas bacterianos (Burrus; Waldor, 2004); (Carraro et al., 2017).

Diversas GIs foram estudadas e analisadas, buscando compreender melhor os seus estilos de vida (Ravatt et al., 1998); (Larbig et al., 2002); (Williams, 2002); (Lu; Leong, 2018). As integrases tem a função de realizar a inserção, catalisada por recombinases do tipo fago específicas e frequentemente são codificadas pela própria GI. Além disso, estas integrases não são conservadas em todas as GIs (Williams, 2002); (Al-Jabri et al., 2018). Algumas integrases que são codificadas pelas GIs, estão associadas as famílias lambda, P4 ou XerD (Burrus et al., 2002); (He et al., 2015). Além disso, as integrases estão relacionadas também com a excisão das GIs. Isto ocorre devido ao auxílio das excisionase, uma proteína de bacteriófago codificada pelo gene Xis (Lesic et al., 2004); (Ramsay et al., 2006); (Novick; Ram, 2016).

De acordo com um estudo sobre ilhas de patogenicidade, descritas como GIs de fagos, evidenciaram a capacidade de existirem em um estado semelhante a um plasmídeo auto-replicante (Ubeda et al., 2007). Como comentado anteriormente, as GIs estão relacionadas aos fagos e também plasmídeos conjugativos, devido suas origens evolutivas. Devido a isso, além da transformação, sua transferência geralmente ocorre via conjugação e transdução (Chen; Christie; Dubnau, 2005); (Brow-Jaque; Calero-

Cáceres; Muniesa, 2015).

3.2 Métodos e ferramenta de predição de ilhas genômicas

Devido a várias características já conhecidas e analisadas em diversos estudos, parece fácil realizar a predição dessas GIs nos genomas bacterianos. Entretanto, estas regiões candidatas a GIs tem se mostrado cada vez mais como um mosaico, aumentando assim a complexidade para realizar as suas predições (Vernikos, 2008); (Lu; Leong, 2016a). Todos os elementos que constituem uma GI podem ter sido adquiridos por eventos de HGT, passando por evoluções subsequentes, como perda de genes e rearranjo genômico (Dobrindt, 2004); (Bertelli; Tilley; Brinkman, 2019). Devido a isso, a variedade das GIs dificulta em achar uma maneira eficaz de integrar vários métodos para predição das regiões candidatas. Optar por apenas um método pode ocasionar perda de muita informação, consequentemente, na perda do número de GIs preditas (Ravenhall et al., 2015). Apesar das adversidades, os métodos de predição já existentes estão contribuindo para o entendimento do comportamento das GIs, esses métodos se dividem em dois principais grupos. Os métodos de composição de sequência, analisam gene/DNA/Estrutura da GI, baseado em um único genoma. Métodos baseados em análise comparativa, utilizam múltiplos genomas para suas análises, realizando principalmente alinhamentos de regiões candidatas com organismos estreitamente relacionados (Langille; Hsiao; Brinkman, 2010); (Lu; Leong, 2016a).

3.3 Métodos de predição baseados em composição de sequência – Genes

A composição de uma região transferida por HGT pode ser diferente da média do genoma receptor. As ferramentas de predição, atuam relacionando essas composições e características, medindo suas diferenças em relação ao genoma e determinando critérios para buscas. Estes critérios podem incluir a variação do conteúdo de guanina e citosina (GC%), uso de códons, uso de aminoácidos e frequências de oligonucleotídeos (Bertelli; Tilley; Brinkman, 2019). Após identificar estas características, é realizado então, uma medida de pontuação. Esta medida vai ser realizada a partir dos genes ou regiões genômicas com os critérios citados anteriormente, em comparação com o genoma completo. Pontuações abaixo ou acima de um determinado limite de corte (predefinido ou calculado dinamicamente) são considerados atípicos. Os genes ou regiões genômicas são então mesclados, obtendo assim as GIs candidatas (Lu; Leong, 2016a).

Algumas ferramentas de predição utilizam um ou múltiplos critérios para obter seus resultados. O método de Karlin (Azad; Lawrence, 2011) e o PAI-IDA (Tu; Ding, 2003) detectam GIs e PAIs buscando vários critérios, como conteúdo de GC%, frequências de dinucleotídeos, uso de códons e uso de aminoácidos. Para isso, os dois preditores utilizam “janela deslizante” em todo o genoma. Esta janela funciona como se fosse um “*scanner*”, calculando a média dos critérios, preservando a subsequência de maior qualidade ao longo do genoma (Lu; Leong, 2016a). Entretanto, ferramentas como IslandPath-DINUC (Hsiao et al., 2003) e SIGI-HMM (Waack et al., 2006), utilizam apenas um critério para predição. O IslandPath-DINUC identifica somente o viés de dinucleotídeo, analisando vários genes consecutivos. Por outro lado, SIGI-HMM reconhece possíveis genes adquiridos por HGT, com base apenas no viés de uso de códons de cada gene. Este método substituiu o modelo heurístico utilizado pelo Hidden Markov Model (HMM) (Hsiao et al., 2005).

Em geral, os métodos baseados na composição da sequência de genes não são difíceis de serem utilizados. Entretanto, ocorre muitos falsos positivos e falsos negativos em seus resultados. As regiões que contêm genes altamente expressos por exemplo, devido a sua composição podem ser detectados como falsos positivos (Merkl, 2004). GIs derivadas de outros genomas, contendo composição similar podem não ser identificadas, levando a um falso negativo (Lee et al., 2013). Estes métodos de análise em nível de composição de gene, requerem também anotações confiáveis. Sendo assim, não podem ser aplicados a genomas incompletos ou recém-sequenciados. Para isso, os métodos de composição de sequência de DNA são utilizados. (Lu; Leong, 2016a).

3.4 Métodos de predição baseados em composição de sequência – DNA

Com o aumento de genomas recém-sequenciados, não contendo anotações completas, as bases para predições de GIs são realizadas apenas com análises das sequências de DNA, em vez de analisar genes e regiões individuais. Para realizar as predições, as ferramentas segmentam todo o genoma. Isto ocorre através de dois métodos principais: baseados em janelas deslizantes e sem janelas deslizantes (Bertelli; Tilley; Brinkman, 2019).

As predições a partir dos métodos em janelas deslizantes são muito utilizados para predição de GIs. Toda a sequência do genoma é segmentada em conjuntos de regiões menores, com um limite de corte único. Vários programas são utilizados para este propósito, como, Alien Hunter (Vernikos; Parkhill, 2006), Centroid (Rajan et al., 2007),

Design-Island (Chatterjee et al., 2008), INDeGenIUS (Shrivastava et al., 2010), e GI-SVM (Lu; Leong, 2016b). Entre as diferenças entre estas ferramentas, as principais estão no tamanho da janela deslizante utilizada, medida de similaridade e na determinação do limite de corte (Lu; Leong, 2016a).

Alien Hunter foi a primeira ferramenta para predição de GIs em sequências genômicas com dados brutos. Suas análises buscam segmentos por meio de entropia relativa, com base em motivos de ordem variável interpolados (IVOM). O limite de corte pode ser determinado através dos clusters de k-mers ou desvio padrão (Vernikos; Parkhill, 2006).

Centroid funciona a partir das análises de janela não sobreposta e de tamanho fixo. Estas análises identificam a média dos vetores de frequência k-mer para todas as janelas, e com base nas distâncias de Manhattan, as janelas de outlier são selecionadas por um limite de corte derivado do desvio padrão (Rajan et al., 2007).

Design-Island faz uso de duas fases para suas análises de predições. Durante a primeira fase, uma janela de tamanho variável é usada para obter GIs iniciais. Enquanto na segunda fase, uma janela menor de tamanho fixo é usada para refinamento para obter as regiões finais candidatas a GIs (Chatterjee et al., 2008).

INDeGenIUS possui o seu método de predição muito semelhante ao Centroid. Entretanto, esta ferramenta utiliza janelas sobrepostas de tamanho fixo, que são calculados por agrupamento hierárquico (Shrivastava et al., 2010).

GI-SVM utiliza método mais recente. Faz uso de frequências k-mer de ordem fixa ou variável. O limite de corte pode ser atribuído de acordo com agrupamento k-means unidimensional (Lu; Leong, 2016b).

Poucas ferramentas utilizam métodos sem janelas para realizar as suas predições, os principais são GC Profile (Elhai; Taton, 2012) e MJSD (Arvey et al., 2009).

O GC Profile é um método para analisar a distribuição e a variação global do conteúdo GC do genoma. As regiões candidatas a GIs possuem uma porcentagem menor de GC% quando comparadas com a do genoma total (Elhai; Taton, 2012)

MJSD é um método de segmentação recursiva baseado na medida de divergência de Markov Jensen-Shannon (MJSD). Para realizar as suas análises, esta ferramenta separa o genoma em dois segmentos (esquerda e direita) quando encontra uma região que possui diferenças estatisticamente significativas. Após isso, cada segmento identificado é comparado com o genoma total com um limite de corte predefinido, identificando assim, as regiões candidatas a GIs (Arvey et al., 2009).

Vantagens e desvantagens estão presentes em relação aos dois métodos, predição por sequência de DNA e composição de sequência (Lu; Leong, 2016a). Os métodos baseados em janela são altamente sensíveis em sua identificação e regiões candidatas. A ferramenta Alien Hunter por exemplo, possui o maior recall, demonstrado no estudo de (Langille; Hsiao; Brinkman, 2008). GI-SVM constatou uma sensibilidade maior em um estudo mais recente (Lu; Leong, 2016b). Contudo, a precisão das duas ferramentas é baixa. Isto pode ocorrer devido a limitação das informações de entradas para análise (Lu; Leong, 2016a)

Entretanto, os métodos que não utilizam janelas podem identificar regiões candidatas a GIs e não GIs com mais precisão (Zhang et al., 2014). A ferramenta GC *Profile*, identificou várias GIs positivas em diversos genomas. Mas possui limitações de apenas identificar regiões com porcentagem GC muito abaixo do genoma (Elhai; Taton, 2012). O preditor MJSD apresentou resultados melhores em GIs com tamanho maior que 10kb, mas não obteve resultados significativos em regiões menores (Arvey et al., 2009).

3.5 Métodos de predição baseados em estrutura – Ilha genômica

Os métodos baseados na estrutura da GI são construídos a partir das variadas características presentes na região candidata (tRNAs, integrases, repetições diretas, transposases, fagos, etc). Dependendo da abordagem para realizar a predição, esses métodos podem ser divididos em integração direta e aprendizado de máquina (Lu; Leong, 2016a).

Na integração direta, as ferramentas utilizam vários filtros para identificar GIs com mais exatidão. Alguns desses recursos integrados funcionam somente para validação. Isso é devido a grande variação na estrutura da GI. Os dois principais preditores que entram nessa categoria são *IslandPath* (Hsiao et al., 2003) e *Islander* (Mantri; Williams, 2004).

IslandPath foi o primeiro programa desenvolvido para identificar várias características relacionadas as GIs ao longo do genoma (viés de GC, viés de dinucleotídeos, genes relacionados à mobilidade). Entretanto, esta ferramenta somente localiza esses elementos, deixando para o pesquisador tomar a decisão se aquela região é uma GI ou não (Hsiao et al., 2003).

Islander identifica GIs principalmente pela presença de tRNAs, buscando a presença deste gene nas regiões candidatas. Vários critérios são usados para tentar

excluir resultados falsos positivos. Principal filtro são as regiões que não possuem integrases (Mantri; Williams, 2004).

Nos métodos de aprendizado de máquina, várias ferramentas para predição de GIs foram construídas. As principais são: *Relevance Vector Machine* (RVM), Alien Hunter (Vernikos; Parkhill, 2008), GI-Detector (Che et al., 2010) e GI-Hunter (Che et al., 2014). As principais distinções entre os métodos destas ferramentas são a escolha dos conjuntos de dados de treinamento e os algoritmos de aprendizado (Lu; Leong, 2016a).

O RVM foi o primeiro aprendizado de máquina a ser construído para realizar predição de GIs. Os seus conjuntos de dados para treinamento se baseiam em características presentes nas regiões candidatas a GIs, como por exemplo, ponto de inserção (tRNA), tamanho da GI, expressão genica, repetições diretas, proteínas relacionadas a fagos, integrases (Vernikos; Parkhill, 2008).

GI-Detector é muito semelhante a ferramenta RVM. Utiliza os mesmos recursos para montar o seu conjunto de treinamento, entretanto, também implementa o algoritmo de aprendizado baseado em árvore de decisão (Lu; Leong, 2016a).

GI-Hunter é parecido com o algoritmo do GI-Detector, mas possui algumas mudanças principalmente na utilização do seu conjunto de dados que é proveniente do *IslandPick* (Langille; Hsiao; Brinkman, 2008). Suas diferenças também estão relacionadas ao modo como analisam o tamanho e as repetições do GI, sendo substituídas por genes altamente expressos e distância intergênica média (Che et al., 2014).

Os métodos de integração direta que utilizam a estrutura das GIs para realizar as suas predições, podem gerar resultados robustos. Por exemplo, a presença de genes de tRNAs e integrases nas regiões candidatas é um forte indício de uma GI positiva. Entretanto, dependem muito de vários recursos relacionados interligados, fazendo com que as GIs mais “antigas”, que não possuam estes genes, sejam descartadas (Lu; Leong, 2016a).

Por outro lado, os métodos baseados em aprendizado de máquina têm a capacidade de integrar e sistematizar vários recursos para suas análises, melhorando a capacidade de predição das GIs. Porém, o desempenho da ferramenta é muito dependente da qualidade do seu conjunto de treinamento (Lu; Leong, 2016a).

3.6 Métodos de predição baseados em análise comparativa – múltiplos genomas

Os métodos de análise comparativa utilizam múltiplos genomas para realizar as

suas predições. Este método compara vários genomas de espécies estreitamente relacionadas, um contra o outro, para identificar regiões candidatas a GIs. Os resultados esperados são regiões presentes em um ou mais genomas que não são encontradas no restante do conjunto. Esta comparação é realizada através de ferramentas de alinhamento de sequências, como BLAST (Altschul et al., 1997) para alinhamento local, MAUVE (Darling et al., 2004) alinhamento de genoma completo. Estas duas ferramentas dependem de análise manual. Os principais preditores com capacidade de análises semi-automáticas (escolha do genoma para comparação pelo pesquisador) são tRNAcc (Ou et al., 2006), MobilomeFINDER (Ou; He, 2007). IslandPick (Langille; Hsiao; Brinkman, 2008) automatizado.

BLAST e MAUVE são capazes de identificar regiões específicas em cepas únicas, sendo estas candidatas a GIs, além disso, MAUVE ainda pode ser usado para localizar regiões conservadas. Vernikos e Parkhill em um estudo, realizaram comparações de todo o genoma através da ferramenta BLAST. Após isso, uma curadoria manual foi feita para encontrar as GIs “confiáveis”, utilizando então, estas regiões para treinar o modelo de estruturas. Com o MAUVE, eles foram capazes de identificar o ganho e perda de gene por um modelo de máxima parcimônia obtido a partir de alinhamentos (Vernikos; Parkhill, 2008).

tRNAcc faz uso dos alinhamentos do MAUVE para realizar a predição das GIs. Isso ocorre pela identificação dos genes tRNAs conservados juntos com uma região flangeadora a jusante nos genomas selecionados. Algum tempo depois, este método foi integrado no MobilomeFINDER, ferramenta de predição web. Análises complementares também foram integradas, como, conteúdo GC%, sequências de codificação específicas da cepa derivadas da análise por BLAST e diferenças de dinucleotídeos. Entretanto, os genomas para comparação devem ser selecionados manualmente (Ou et al., 2006); (Ou; He, 2007).

IslandPick consegue realizar a seleção do genoma para comparação automaticamente. Para fazer isto, o método constrói uma matriz de distância utilizando a comparação dos genomas todos contra todos. Construindo vários pontos de corte para identificar o genoma mais adequado. O MAUVE é utilizado para se obter os alinhamentos dos genomas completos, resultando em grandes regiões únicas no genoma de consulta. Após essa etapa, o BLAST é aplicado como um filtro, para retirar as duplicações do genoma, resultando nas GIs candidatas (Langille; Hsiao; Brinkman, 2008).

Os métodos baseados em análise comparativa trazem algumas vantagens quando

comparamos com os métodos de composição de sequência. Mas ainda assim, existem incertezas em suas predições (Langille; Hsiao; Brinkman, 2008); (Lu; Leong, 2016a). Isto ocorre devido à dificuldade de conseguir separar os genes adquiridos por HGT ou pela perda de genes (Ravenhall et al., 2015). Além disso, os rearranjos genômicos podem influenciar nos alinhamentos das sequências. As aplicações de métodos baseados em diversos genomas são limitadas. Devido a necessidade de utilizar um genoma de referência para comparação, muitas vezes estes organismos relacionados podem não estar disponíveis (Darling et al., 2004); (Bertelli; Tilley; Brinkman, 2019).

3.7 Métodos de predição Baseados em métodos híbridos

Alguns métodos compostos, ou chamados também de híbridos, foram desenvolvidos para tentar melhorar as predições das GIs nos genomas bacterianos. Estas ferramentas trabalham com várias metodologias integradas para a tomada de decisão (Bertelli; Tilley; Brinkman, 2019). Algumas das principais ferramentas são, EGID (Che et al, 2011), GIST (Hasan et al., 2012), PIPs (Soares et al., 2012), GIPSY (Soares et al., 2016), IslandViewer4 (Bertelli et al., 2017).

O preditor EGID foi primeira versão da ferramenta GIST, sendo ela uma atualização. Utiliza cinco abordagens/métodos para realizar as suas predições: INDeGenIUS, IslandPath, PAI-IDA, AlienHunter, COLOMBO SIGI-HMM. Para delimitar os resultados de cada método é agregado valores sobre um limite de corte. Resultados individuais são separados para não formarem regiões sobrepostas uma nas outras. Então, é gerado um consenso das regiões candidatas, mesclando as GIs com as melhores pontuações EGID (Che et al, 2011), GIST (Hasan et al., 2012).

IslandViewer 4 possui métodos baseados em composição (IslandPath-DIMOB, SIGI-HMM e Islander) e também método de genômica comparativa IslandPick. Esta ferramenta também atua como um banco de dados de GIs. Para escolha do genoma de referência, o método seleciona o organismo disponível no banco de dados. Entretanto, os usuários podem selecionar manualmente um genoma de interesse para comparação, realizando novamente a análise. Se este não estiver disponível, somente outros métodos iram ser computados no organismo de interesse (Bertelli et al., 2017).

PIPs foi o primeiro preditor a identificar e classificar funções das GIs, sendo GIPSY sua atualização mais recente. Para suas predições utiliza scripts personalizados, SIGI-HMM, presença de genes de transposase e tRNA. As GIs são classificadas em quatro

subtipos com base na presença de fatores de virulência, metabolismo, resistência a antibióticos ou simbiose utilizando bancos de dados integrados para comparação. Os usuários são responsáveis pelos genomas de referência para comparação, estes não podem ser determinados automaticamente (Soares et al., 2012), GIPSY (Soares et al., 2016).

Com uma ampla diversidade de preditores de GIs e seus diversos métodos e características, a escolha de uma ferramenta para uso dependerá do genoma a ser estudado. Os métodos baseados em genômica comparativa propõem definições mais robustas para definir os limites das GIs, mas dependem exclusivamente de genomas de referência para suas análises. Para estudos em um único genoma, os preditores baseados na composição de sequências são comumente usados. Possuem uma grande variedade de metodologias utilizadas, mas possuem baixa precisão para regiões que não possuam um ou mais critérios envolvidos para realizar a predição. Combinar métodos e ferramentas ainda é a melhor forma de diminuir os riscos de perder GIs chaves para obter resultados significativos nas análises e pesquisas (Bertelli; Tilley; Brinkman, 2019)

As GIs podem proporcionar adaptações para as bactérias em novos ambientes, principalmente em locais adversos. Com a aquisição destas GIs, as bactérias podem adquirir novas capacidades e funções, de acordo com o seu material genético presente. Em microrganismos patogênicos e ambientais as GIs estão amplamente distribuídas. Isso faz com que ocorra um aumento na relevância para os pesquisadores em tentar entender o papel destas GIs na evolução, e o impacto funcional dentro das populações bacterianas (Hacker; Carniel, 2001); (Lu; Leong, 2016a); (Juhas, 2019); (Bertelli; Tilley; Brinkman, 2019).

CAPÍTULO III: PREDICTION AND ANALYSIS IN SILICO OF GENOMIC ISLANDS IN *Aeromonas hydrophila* – ARTIGO I

Frontiers. Microbiology., 29 November 2021. <http://doi.org.10.3389/fmicb.2021.769380>



frontiers
in Microbiology

Evolutionary and Genomic
Microbiology

SECTION ABOUT ARTICLES RESEARCH TOPICS FOR AUTHORS EDITORIAL BOARD

< Articles

ORIGINAL RESEARCH article

Front. Microbiol., 29 November 2021 | <https://doi.org/10.3389/fmicb.2021.769380>



Prediction and Analysis *in silico* of Genomic Islands in *Aeromonas hydrophila*

Antonio Camilo da Silva Filho^{1*}, Jeroniza Nunes Marchaukoski², Roberto Tadeu Raittz², Camilla Reginatto De Pierri³, Diogo de Jesus Soares Machado², Cyntia Maria Telles Fadel-Picheth¹ and Geraldo Picheth¹

¹Department of Clinical Analysis, Federal University of Parana, Curitiba, Brazil

²Department of Bioinformatics, Professional and Technical Education Sector, Federal University of Parana, Curitiba, Brazil

³Department of Biochemistry and Molecular Biology, Federal University of Parana, Curitiba, Brazil

Prediction and analysis *in silico* of genomic islands in *Aeromonas hydrophila*

Antonio Camilo da Silva Filho¹, Jeroniza Nunes Marchaukoski², Roberto Tadeu Raittz², Camilla Reginatto De Pierri³, Diogo de Jesus Soares Machado², Cyntia Maria Telles Fadel-Picheth¹, Geraldo Picheth¹

¹Department of Clinical Analysis, Federal University of Parana, Curitiba, Brazil.

²Department of Bioinformatics, Professional and Technical Education Sector, Federal University of Parana, Curitiba, Brazil.

³Department of Biochemistry and Molecular Biology, Federal University of Parana, Curitiba, Brazil.

* **Correspondence:** Antonio Camilo da Silva Filho. antonio.camilofilho@gmail.com

Keywords: *Aeromonas hydrophila*¹, Genomic Island², Virulence³, Metabolism⁴, Antibiotic Resistances⁵.

ABSTRACT

Aeromonas are Gram-negative rods widely distributed in the environment. They can cause severe infections in fish related to financial losses in the fish industry, and are considered opportunistic pathogens of humans causing infections ranging from diarrhea to septicemia. The objective of this study was to determine in silico the contribution of genomic islands to *A. hydrophila*. The complete genomes of 17 *A. hydrophila* isolates, which were separated into two phylogenetic groups, were analyzed using a genomic island (GI) predictor. The number of predicted GIs and their characteristics varied among strains. Strains from group 1, which contains mainly fish pathogens, generally have a higher number of predicted GIs, and with larger size, than strains from group 2 constituted by strains recovered from distinct sources. Only a few predicted GIs were shared among them and contained mostly genes from the core genome. Features related to virulence, metabolism, and resistance were found in the predicted GIs, but strains varied in relation to their gene content. In strains from group 1, O Ag biosynthesis clusters OX1 and OX6 were identified, while strains from group 2 each had unique clusters. Metabolic pathways for myo-inositol, L-fucose, sialic acid, and a cluster encoding QueDEC, tgtA5, and proteins related to DNA metabolism were identified in strains of group 1, which share a high number of predicted GIs. No distinctive features of group 2 strains were identified in their predicted GIs, which are more diverse and possibly better represent GIs in this species. However, some strains have several resistance attributes encoded by their predicted GIs. Several predicted GIs encode hypothetical proteins and phage proteins whose functions have not been identified but may contribute to *Aeromonas* fitness. In summary, features with functions identified on predicted GIs may confer advantages to host colonization and competitiveness in the environment.

1 INTRODUCTION

The bacterial genome is composed of a core genome containing the genetic information required for essential functions, and a flexible gene pool, which encodes additional traits that can be beneficial under certain circumstances. The flexible gene pool represents variable chromosomal regions and includes mobile and accessory genetic elements, such as bacteriophages, plasmids, insertion sequences, and genomic islands (GIs) (Dobrindt et al., 2004).

GIs are syntenic blocks of accessory genes acquired by horizontal gene transfer (HGT) which contribute to the diversification and adaptation of microorganisms and offer a selective advantage for host bacteria. According to their gene content, GIs are described as pathogenicity, resistance, symbiosis, metabolic or fitness islands (Weinstock et al., 2000; Juhas et al., 2009; Rao et al., 2020). Most GIs are relatively large segments of DNA, usually between 10 and 200 kb, detected by comparisons among closely related strains; however, regions with sizes < 10 kb have also been identified and named genomic islets (Hacker et al., 1990; Juhas et al., 2009; Li and Wang, 2021). GIs usually differ from chromosomes in terms of GC content, tetranucleotide frequency, and codon usage. They are often flanked by small directly repeated (DR) sequences and are often inserted into tRNA genes. They can carry functional or cryptic genes encoding integrases, factors related to plasmid conjugation, phages involved in GI transfer, and transposons or insertion sequence (IS) elements, which may be implicated in mobilizing genetic material onto or deleting DNA from GI (Buchrieser et al., 1998; Juhas et al., 2009; Partridge et al., 2018). Currently, bioinformatics tools developed for GI prediction (Vernikos et al., 2006; Pundhir et al., 2008; Che et al., 2014; Soares et al., 2016; Wei et al., 2016; Bertelli et al., 2017) are available for the study of these mobile elements.

Aeromonas are Gram-negative bacteria ubiquitous in aquatic environments and can be isolated from virtually every environmental niche where bacterial ecosystems exist (Janda and Abbott, 2010). There are 36 species recognized of *Aeromonas* (Fernández-Bravo and Figueras, 2020), of which *A. hydrophila* is the most studied and the first species of the genus to have a sequenced genome (Seshadri et al., 2006; Grim et al., 2014; Rasmussen-Ivey et al., 2016a). *A. hydrophila* has been isolated from fresh and marine waters, diseased fish, poikilothermic aquatic animals, and warm-blooded animals (Martin-Carnahan and Joseph, 2005). In fish, *A. hydrophila* can cause a variety of diseases, including septicemia, red sore disease, and ulcerative infections, and has been linked to

fish death worldwide, resulting in great economic losses (Janda and Abbott, 2010; Hossain et al., 2013, 2014; Rasmussen-Ivey et al., 2016b). In humans, *A. hydrophila* is associated with both intestinal and extraintestinal diseases (Martin-Carnahan and Joseph, 2005), such as gastroenteritis, wound infection, septicemia, pneumonia, soft tissue infections, peritonitis, hepatobiliary tract infections, and necrotizing fasciitis (Janda and Abbott, 1998, 2010; Grim et al., 2014).

Genome sequencing and comparative genomic analysis have allowed major advances in the study of *Aeromonas* virulence. Highly virulent strains recovered from humans or fish have been identified, several genes have been associated with pathogenicity and some are related to GIs in some strains (Seshadri et al., 2006; Hossain et al., 2013, Grim et al., 2014; Pang et al., 2015; Rasmussen-Ivey et al., 2016b). The objective of this study was to analyze *A. hydrophila* genomes using a GI predictor, determine the distribution of predicted GIs, their characteristics, functions encoded, core genome content, and sharing among strains.

2 MATERIALS AND METHODS

2.1 *A. hydrophila* genomes

The complete genomes of *A. hydrophila* available at the National Center for Biotechnology Information database (Geer et al., 2010) as at 12/01/2019 were used in this study. The source and genome characteristics of the *A. hydrophila* strains are shown in (Table 1).

TABLE 1. General characteristics of *Aeromonas hydrophila*

Strain	RefSeq ID	Genome Size (Mb)	Genome GC%	Gene number	Source	Collection Year	Country
<i>A. hydrophila</i> ATCC 7966	NC_008570.1	4.74	61.50	4,283	Canned milk	No info	USA
<i>A. hydrophila</i> AL06-06	NZ_CP010947.1	4.88	61.40	4,572	Goldfish	2006	USA
<i>A. hydrophila</i> AHNIH1	NZ_CP016380.1	4.91	61.20	4,551	Perirectal swab	2013	USA
<i>A. hydrophila</i> ML09-119	NC_021290.1	5.02	60.80	4,646	Catfish	2009	USA
<i>A. hydrophila</i> AL09-71	NZ_CP007566.1	5.02	60.80	4,649	Catfish	2009	USA
<i>A. hydrophila</i> pc104A	NZ_CP007576.1	5.02	60.80	4,648	Pond soil	2010	USA
<i>A. hydrophila</i> NJ-35	NZ_CP006870.1	5.28	60.50	4,902	Diseased carp	2010	China
<i>A. hydrophila</i> JBN2301	NZ_CP013178.1	5.13	60.80	4,767	Crucian carp	2009	China
<i>A. hydrophila</i> J-1	NZ_CP006883.1	5.00	60.90	4,616	Diseased carp	1989	China
<i>A. hydrophila</i> GYK1	NZ_CP016392.1	4.95	60.80	4,534	Siniperca chuatsi	2001	China
<i>A. hydrophila</i> D4	NZ_CP013965.1	5.10	60.80	4,729	Megalobrama amblycephal	2012	China
<i>A. hydrophila</i> strain ZYAH75	NZ_CP016990.1	4.96	61.30	4,598	Wound secretion	2015	China
<i>A. hydrophila</i> strain ZYAH72	NZ_CP016989.1	5.16	60.70	4,797	Crucian carp	2015	China
<i>A. hydrophila</i> AH10	NZ_CP011100.1	4.91	61.10	4,561	Grass carp	2011	China
<i>A. hydrophila</i> WCHAH045096	NZ_CP028568.1	5.02	61.10	4,691	Sewage	2015	China

<i>A. hydrophila</i> MX16A	NZ_CP018201.1	4.78	61.60	4,445	Water	2012	China
<i>A. hydrophila</i> 4AK4	NZ_CP006579.1	4.52	62.00	4,164	Raw Sewage	1999	China
<i>A. hydrophila</i> YL17	NZ_CP007518.2	4.80	61.60	4,392	Compost	No info	Malaysia
<i>A. hydrophila</i> KN-Mc-1R2	NZ_CP027804.1	4.91	61.00	4,629	<i>Myocastor coypus</i>	2016	South Korea

2.2 Phylogenetic analysis

A phylogenetic tree was generated using the software Spaced Words Projection – SweeP (De Pierri et al., 2020) with the complete genomes of all *Aeromonas* species available from the NCBI database. Default SweeP criteria were used in the analysis. Data regarding the organisms used are shown in (Supplementary Table 1).

2.3 GI prediction and analysis

IslandViewer 4 (Bertelli et al., 2017) was used to predict GIs in *A. hydrophila* strains. GIs < 10 kb were considered genomic islands unless otherwise stated. Some GIs predicted by SIGI-HMM (Waack et al., 2006) and IslandPath-DIMOB (Hsiao et al., 2003) two components of IslandViewer 4, were partially superimposed, increasing the number of predictions. Then, a verification of these regions was conducted to determine the start and end of the GIs manually, avoiding misidentification in the number of predicted GIs. After the identification of each predicted GI by IslandViewer 4, the regions were selected and visualized using Artemis software (Rutherford et al., 2000) to generate.GBK and .FASTA files for each GI. These files contained the amino acid sequence, the name, and the protein ID of each product of each GI, and were used to create multi-fasta files for each microorganism. These files were submitted to RAST (Brettin et al., 2015) for standardization of gene nomenclature. The IslandViewer version 4.0 (Web Server) was used.

2.3 Tools and databases for traits prediction

To determine the traits encoded by the predicted GIs, datasets containing the amino acid sequences of the GIs of each *A. hydrophila* strain were compared with the amino acid sequences of the databases indicated below using the online web server CD-HIT Suite: Biological Sequence Clustering and Comparison (CD-HIT-2D; version v4.8.1-2019-0228) (Huang et al., 2010). The minimum threshold for sequence identity cut-off was 70%. All other parameters of CD-HIT-2D are indicated in Supplementary Table 2.

The following databases were used to identify traits related to virulence, antibiotic resistance, or drug targets encoded on the predicted GIs:

Pathosystems Resource Integration Center – PATRIC (Wattam et al., 2017). The virulence factor database – VFDB (Liu et al., 2019). PHIDIAS: Pathogen-host interaction data integration and analysis system VICTORS (Sayers et al., 2019). The Comprehensive

Antibiotic Resistance Database CARD (Jia et al., 2017). National Database of Antibiotic Resistant Organisms (<https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance>). DrugBank (Wishart et al., 2018). Therapeutic Target Database (TTD) (Li et al., 2018). For the identification of metabolic traits, KEGG Mapper (Kanehisa and Sato 2020) was used with family/genus 642 ID. The data obtained from the tests were manually curated.

2.4 Phage prediction

Phage regions were analyzed using PHASTER Phage Search Tool Enhanced Release (Arndt et al., 2016). The PHASTER search phages regions using BLAST through file analysis in .GBK or .FASTA format. The sequences are compared with the NCBI Phage database and a phage database developed by Srividhya et al. (2006). Phage-like genes are then grouped into phage regions using DBSCAN (Ester et al., 1996). Regions corresponding to phages are assigned scores and colors depending on the level of identity and coverage of the region. Intact phage (> 90 score) are colored green; questionable phage (70-90 score) labelled blue; incomplete phage (<70 score) are colored red. For this analysis, complete genomes of *A. hydrophila* were submitted to PHASTER to search for phage regions, and the results were compared with predicted GIs for identification of phage-containing GIs.

2.5 Shared GIs

To compare the predict GIs, genes of each GI were concatenated and used to create individual multi-fast files for each strain. After this step, BLASTp was used to infer homology, considering a minimum of 75% of identity and sequence coverage for comparisons between single GIs against all other predicted GIs in all strains. Power Bi, a software for generating interactive reports with data modeling tools, and chord diagrams were used to represent the proportion of GIs shared among the strains. An image was generated using the data obtained from the total number of shared GIs. The nodes were organized around a circle, with the proportion between points connected to each other, represented proportionally by the size of each arc. (<https://github.com/Microsoft/PowerBI-visuals-chord>). Power Bi version 2.96.1061.0 (Windows 64-bit) was used in this study.

2.6 Core Genome and related GIs

The core genome of *A. hydrophila* strains was determined using the web server

EDGAR version 3.0 (Web Server) <https://edgar3.computational.bio.uni-giessen.de/> (Dieckmann et al., 2021). This bioinformatics tool is based on 'score ratio values' (SRVs) methodology for orthology estimation. To calculate the SRVs, all BLAST scores are normalized against the maximum score, which is the self-hit of a query sequence. Then, the distribution of the SRVs is analyzed to derive a threshold that is tailored to the data, usually around an SRV of 0.3 (30%). Furthermore, an initial e-value cutoff of 1e-05 is used. This orthologous approach is such that only one-to-one pairs are found. For duplicated genes and paralogs, a single hit is found, and additional copies are missed. Pseudogenes are excluded from these analyses. Core genome calculation requires one reference genome, for which *A. hydrophila* ATCC 7966 was chosen. Results from the core genome (EDGAR) are presented with a locus tag for each gene and strain with the amino acid sequence. After identifying the genes from the core genome, they were compared with predicted GIs using CD-HIT, a tool developed for clustering very long microbial genomic sequences. In this analysis the comparison was performed gene-by-gene for all GIs of each strain using amino acid sequences. The cut-off rate used was 90% identity.

2.7 Cluster alignments

Mauve, a tool for multiple genome alignments that can be used to view genomic rearrangements, inversions, and large insertions in related genomes, was used to analyze O antigen clusters. The alignment is shown as an image formed by colored blocks, each one of them representing genome region aligned and internally free of genomic rearrangements (Darling et al., 2004). Mauve version 2.4.0 (Windows 64-bit) was used in this study.

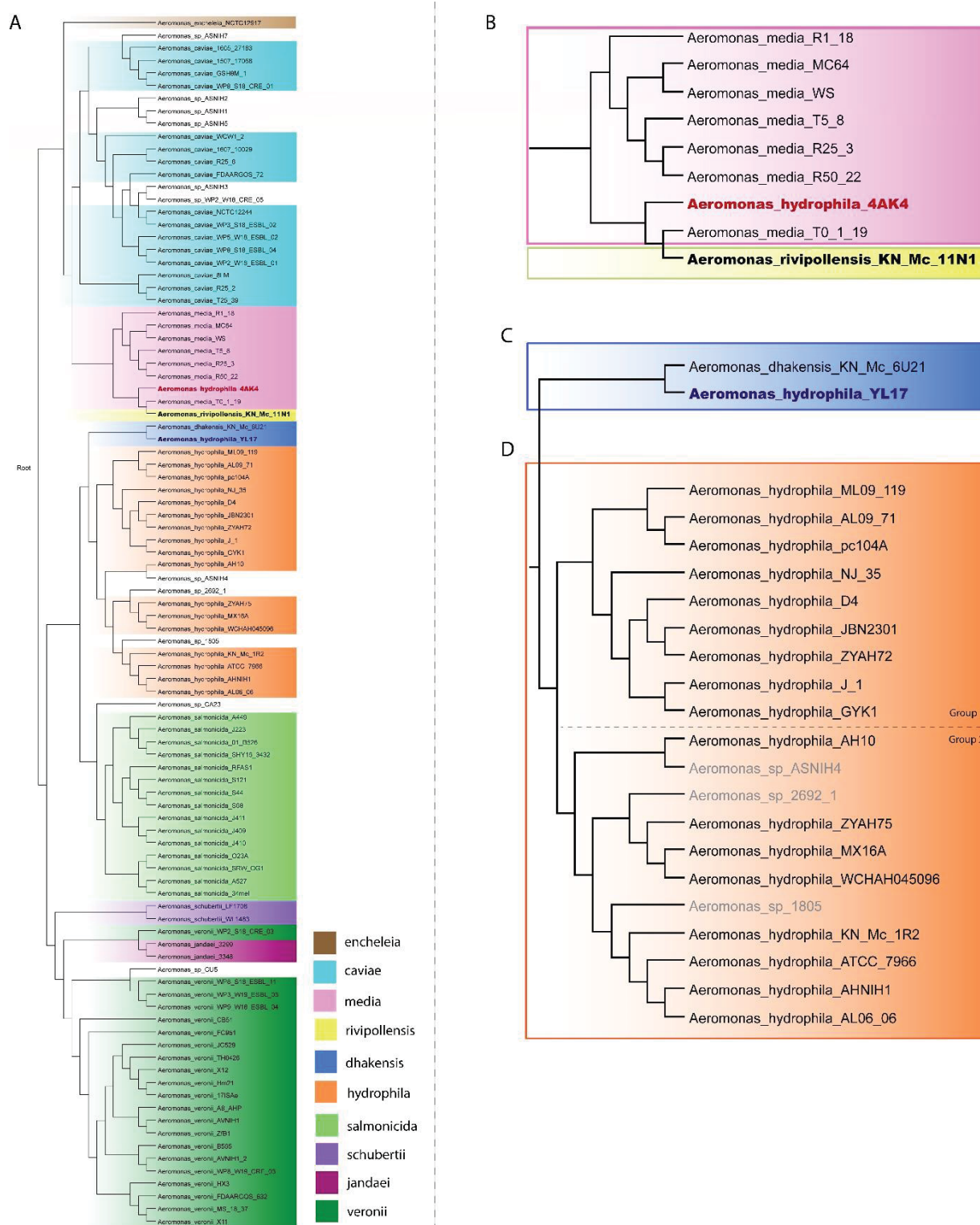
2.8 Cluster alignments

Variables with normal distributions (Kolmogorov-Smirnov) were reported as mean $\pm$ 1-standard deviation. Comparison among groups were tested with Student t-test (2-tailed) or by one-way ANOVA for multiple groups comparisons. A P-value <0.05 was considered significant. All calculations were performed using MedCalc version 17.6 (MedCalc Statistical Software bvba, Ostend, Belgium).

3 RESULTS

3.1 Phylogenetic analysis and general characteristics of *A. hydrophila* strains

A phylogenetic tree based on the complete genome of *Aeromonas* species was generated to verify the relationships among *A. hydrophila* isolates. A total of 96 genomes from 10 *Aeromonas* species and some strains identified only to the genus level were included in this analysis, the results of which are shown in Figure 1. Strains of *A. schubertii*, *A. veronii*, *A. salmonicida*, *A. dhakensis*, *A. hydrophila*, *A. encheleia*, *A. media*, *A. rivipollensis*, *A. jandaei*, and *A. caviae* were separated into distinct branches of the phylogenetic tree according to their species. However, there were some exceptions, with *A. veronii* WP2_S18_CRE03 grouped with *A. jandaei* strains 3299 and 3348; *A. hydrophila* YL17 grouped together *A. dhakensis* KN_Mc_6U21, and *A. hydrophila* 4AK4 with *A. media* T0-1-19 and *A. rivipollensis* KN_Mc_11N1, suggesting that these strains were misidentified at the species level. Thus, the genomes of strains YL17 and 4AK4 were removed from this study. Additionally, a small number of strains previously identified only at the genus level clustered together with certain species (Figure 1).

FIGURE 1. Phylogenetic tree of the *Aeromonas* species.

LEGEND: (A) Phylogenetic tree generated from the whole genome of 96 *Aeromonas* strains. (B) Highlighted branch of "*A. hydrophila* 4AK4." (C) Highlighted branch of "*A. hydrophila* YL17." (D) Highlighted branch of *A. hydrophila*; strains analyzed are subgrouped into two main subclades, separated by dotted lines: ML09-119, AL09-71, pc104A, NJ-35, D4, JBN2301, ZYAH72, J-1, and GYK1 (group 1); AH10, ZYAH75, MX16A, WCHAH045096, KN-Mc-1R2, ATCC 7966, AHNIH1, and AL06-06 (group 2).

The 17 complete *A. hydrophila* genomes that were analyzed for predicted GIs differed in genome size, which varies by up to 0.54 Mb, as observed between strains ATCC 7966, which has the smallest genome (4.74 Mb) and NJ-35 (5.28 Mb) the largest; the GC content varied from 60.50% in strain NJ-35 to 61.60% in strain MX16A (Table 1).

Phylogenetic analysis classified the 17 *A. hydrophila* strains into two main groups (Figure 1). Group 1 contained nine strains, isolated in China or the USA, recovered from diseased fish, except for pc104A, which was recovered from the soil of a catfish pond (Table 1). The Chinese and American *A. hydrophila* strains were separated into different subclades of group 1. Group 2 contained strains recovered from different sources, including the environment (ATCC 7966, MX16A, and WCHAH045096), human clinical samples (ZYAH75, AHNIH1), *Myocastor* (KN-Mc-1R2), and diseased fish (AH10 and AL06-06). They were also separated into subclades according to their geographical origins.

Strains from groups 1 and 2 differ in relation to the mean genome size; 5.08 and 4.89 Mb, respectively. Interestingly, all strains of group 1 presented genomic GC contents below 61% (60.50% to 60.90%), while those of group 2 were 61% or higher (61.00 to 61.60%; Table 1), a significant difference ($P < 0.001$; t-test), however stats were underpowered.

3.2 Characteristics of predicted GIs of *A. hydrophila* strains

The number of predicted GIs varied widely among the strains, from 13 in *A. hydrophila* ATCC 7966 and MX16A to 33 in strain NJ-35, and the average size ranged from 12402 bp in AL06-06 to 29317 in JBN2301. A significant difference was observed in terms of the ratio between the total GI sequence size and host genome size, which varied from 4% in strain ATCC 7966 to 12% in NJ-35. The variation in GC content within GIs ranged from 34% to 66%, both in *A. hydrophila* isolate AH10 (Table 2). In general, the genomes of group 1 strains presented a higher average number of predicted GIs, genes in GIs, larger mean GI size, and ratio of GIs relative to genome size than those of group 2 (Table 2).

Table 2. General data of predicted GIs

Strain	GIs ¹	>10 kb ²	< 10 kb ³	A.S GIs ⁴	Genes ⁵	> GC% ⁶	< GC% ⁷	Av. GC% ⁸	Gen.Siz ⁹	GI Total Size ¹⁰	Ratio % ¹¹
ATCC 7966	13	6	7	16279	204	63%	43%	51%	4744448	211630	4%
MX16A	13	7	6	17185	246	63%	39%	51%	4783504	223399	5%
ZYAH75	15	10	5	19549	401	59%	37%	51%	4955171	293237	6%
AH10	15	11	4	19391	314	66%	34%	51%	4908265	290871	6%
KN-Mc-1R2	16	7	9	20054	379	60%	39%	50%	4911246	320867	7%
AL06-06	19	9	10	12402	252	63%	38%	51%	4884823	235642	5%
AHNIH1	20	17	3	20042	404	65%	38%	51%	4906118	400837	8%
WCHAH04509 6	27	18	9	19387	556	63%	37%	52%	5022867	523460	10%
<u>J-1*</u>	17	15	2	28323	504	58%	35%	48%	5000814	481498	10%
<u>GYK1</u>	19	16	3	25238	479	58%	35%	47%	4951765	479521	10%
<u>JBN2301*</u>	19	17	2	29317	597	58%	35%	49%	5127362	557016	11%
<u>ML09-119*</u>	21	18	3	24627	531	58%	35%	49%	5024500	517159	10%
<u>AL09-71*</u>	22	18	4	23889	541	58%	35%	49%	5023861	525556	10%
<u>ZYAH72</u>	22	19	3	26601	626	58%	35%	49%	5159182	585215	11%

<u>pc104A*</u>	22	19	3	24085	546	58%	35%	49%	5023829	529868	11%
<u>D4</u>	23	18	5	23873	579	62%	35%	50%	5100520	549072	11%
<u>NJ-35*</u>	33	20	13	19477	661	65%	35%	52%	5279644	642740	12%

LEGEND: Strains from phylogenetic group 1 are underlined; *hypervirulent *A. hydrophila* (vAh). ¹Total number of predicted GIs. ²Total GIs > 10,000 base pairs. ³Total GIs < 10,000 base pairs. ⁴Average size of GIs. ⁵Total genes in predicted GIs. ⁶Higher percentage of guanine and cytosine in GIs. ⁷Lower percentage of guanine and cytosine in GIs. ⁸Average percentage of guanine and cytosine in GIs. ⁹Genome size. ¹⁰Total size of GIs. ¹¹Percentage of GI size in relation to genome size.

Altogether, the data indicated differences in relation to genome size and GC content, distribution, and characteristics of GIs of strains from groups 1 and 2. Genes identified on the predicted GIs are listed in Supplementary Table 3.

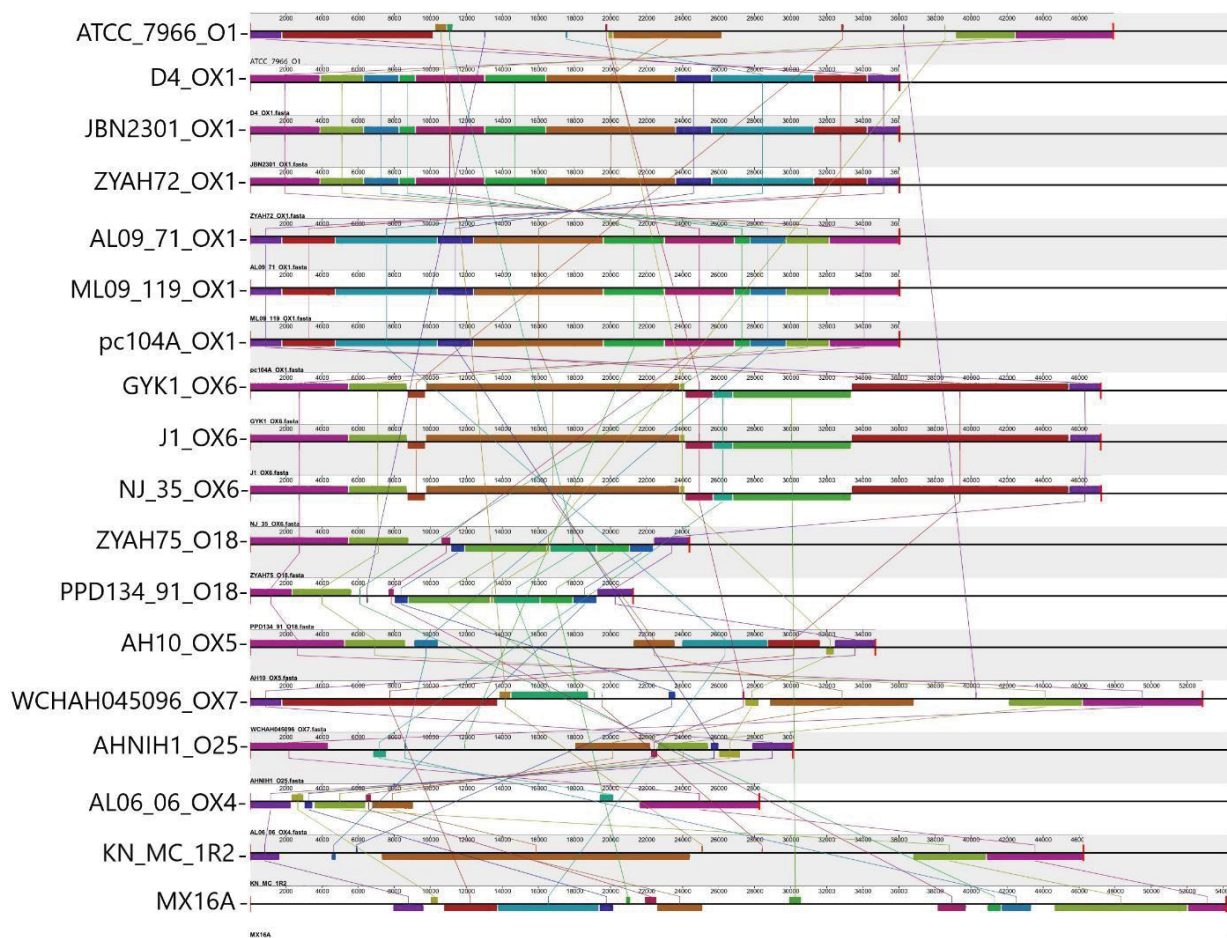
3.4 Functions encoded by the predicted GIs

Determination of the functions of genes in the predicted GIs was based on databases related to virulence, resistance, metabolism, drug targets, and RAST annotation. A complete list of genes with identified functions and the corresponding databases is shown in Supplementary Table 4. The main traits identified are described below.

3.3.1 Traits related to virulence

O antigen. Gene clusters encoding proteins associated with the biosynthesis of O antigen (O Ag) were identified in the GIs of all *A. hydrophila* strains. Ten distinct O Ag clusters were found among the 17 strains analyzed in this study, including serogroups O1, O25, and putative serogroups OX1, OX4, OX5, OX6, and OX7 (Figure 2). The O Ag cluster of strain ZYAHA75 presented identity and query coverage of 89.69% and 97%, respectively, with serogroup O18 of *A. hydrophila* strain PPD134/91 compared in Figure 2. The O Ag cluster of KN-Mc-1R2 contains *cysN* and *cysD* genes encoding enzymes related to sulfur metabolism, and that of strain MX16A is interrupted by genes encoding transposases, suggesting that HGT played a role in the formation of this cluster. The comparison of the OAg cluster sequence of strain MX16A showed higher homology with that of *Aeromonas veroni* A8-AHP, with 77% of coverage and 93% of identity. When compared to the same species, better results were observed with *A. hydrophila* G5380, with 50% of coverage and 94% of identity. For KN-MC-1R2, better results were obtained with *Aeromonas hydrophila* 3924, with 44% of coverage and 98% of identity. When compared with another species, 41% of cover and 89% of identity was found with the cluster of *Aeromonas caviae* KAM345.

FIGURE 2. Comparison of O antigen biosynthesis clusters encoded in predicted GIs



LEGEND: Alignment of the O Ag cluster from 17 *A. hydrophila* strains analyzed and strain PPD134-91. Blocks with the same color indicate homologous regions. The scale indicates the size of the cluster. Strain designations and respective serogroups are indicated below the representation of the cluster. Serogroups identified OX1 (D4, JBN2301, ZYAH72, AL09-71, ML09-119, pc104A,); OX6 (GYK1, J-1, NJ-35); O18 (comparison between PPD134/91 serogroup O18, and ZYAH75 cluster); OX5 (AH10); OX7 (WCHAH045096); O25 (AHNIN1); OX4 (AL06-06); O1 (ATCC 7966); not determined (KN-MC-1R2, MX16A). The O Ag cluster of strains pc104a, AL09-71, and ML09-119 are in the reverse orientation in the genome of the strains. The name of each strain and the respective serogroup are indicated in front of the blocks representing the O Ag cluster.

Flagella. Genes from polar flagella region 2 (*flaA flaBGHJ maf-1*) were identified in the predicted GIs of most strains, with additional genes found in several of them. Three types of predicted GI were observed according to the identity and number of genes encoded. One, approximately 7 kb encoding *flaA flaBGHJ maf-1* and a hypothetical protein was observed in strains ATCC 7966, AL06-06, and AHNIN1. Predicted GIs of approximately 19 kb encoding the polar flagella region 2 together with other genes, including a flagellin-like, transposase, *neuB*, and *flmD* (related to polar flagella glycosylation), were found in

all strains from group 1 and in AH-10. Finally, GIs of approximately 15 kb containing the polar flagella region 2 genes, *neuB*, *flmD*, and other genes distinct from those of the GI described above, were identified in strains WCHAH045096 and MX16A. In strains KN-Mc-1R2 and ZYAH75, no GIs containing flagellar region 2 genes were predicted, although these regions are present in their genomes.

Components of type VI secretion system (T6SS). Genes encoding VgrG and Hcp, components of the expelled puncturing structure of T6SS were found in the predicted GIs of strains AH10, AL06-06, and KN-Mc-1R2. In the first two strains, VgrG-encoding genes are present in two predicted GIs. Additionally, certain features that may potentially be associated with virulence are encoded in predicted GIs of some strains, including ankyrin proteins (AnkB), which are encoded together with catalase (KatE) and, in most of them, with super oxide dismutase; Fic (filamentation induced by cyclic AMP) proteins, of which two types PA0574 and KPN03553 were identified; pilin Flp; the zona occludens toxin and accessory cholera enterotoxin, were sequentially encoded (Table 3). Of the genes with identified functions, some were found in genomic islets, which contain low numbers of genes and are generally poorly explored or studied. Among the genes identified in genomic islets are those encoding Fic domain protein PA0574 (strain ZYAH72, GI-18); zone occludens toxin and accessory cholera enterotoxin (strains KN-Mc-1R2, GI-11, AHNIH1, GI-17), and VgrG (strains AH10, GI15; AL06-06, GI-16).

3.3.2 Traits related to resistance

Genes associated with antibiotic resistance were restricted to the predicted GIs from strains AL06-06, ZYAH75, MX16A, and WCHAH045096. These genes were associated with resistance to sulfonamide, aminoglycoside, tetracycline, rifampin, chloramphenicol, and macrolides, in addition to TEM beta-lactamase, and distinct extended-spectrum beta-lactamases that are able to hydrolyze different beta-lactams. Additionally, genes related to mercury, copper, and chromium compounds, and ammonium quaternary resistance were identified in the GIs of some strains (Table 3, Supplementary Table 4).

Traits\Strains	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
O-antigen cluster	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
Polar flagellum (region 2)	Green	Green	Green	Green	Green	Green	Green	Green	Green	White	Green	White	Green	Green	Green	Green	Green
Lateral flagella	White	White	White	White	White	White	White	White	White	Green	White	White	White	Green	White	White	White
VgrG	White	White	White	White	White	White	White	White	White	Green	White	White	White	Green	White	Green	White
HcP	White	White	White	White	White	White	White	White	White	Green	White	White	White	Green	White	Green	White
AnkB*	Green	Green	Green	Green	White	White	Green	Green	White	White	White	White	White	Green	White	Green	White
Flp pilus*	White	Green	Green	White	Green	Green	White	White	Green	White	White	White	White	White	White	White	White
Fic domain protein, PA0574 type*	Green	Green	Green	Green	Green	Green	Green	Green	Green	White	White	White	White	White	Green	White	White
Fic domain protein, KPN03553 type*	White	White	White	White	White	White	White	White	White	White	White	Green	Green	White	White	White	White
Inhibitor of vertebrate c-type lysozyme (Ivy)*	Green	Green	Green	Green	Green	Green	Green	White	White	Green	White	White	Green	Green	White	White	White
Zona occludens toxin*/Accessory cholera enterotoxin*	White	White	White	White	White	White	White	White	White	Green	White	White	Green	Green	White	White	White
L-threonine biosynthesis#	White	White	White	White	White	White	White	White	White	White	White	White	Blue	White	Blue	White	White
Sulfate adenylyltransferase/adenylylsulfate kinase#	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
Tryptophan biosynthesis#	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
Dha / GldA#	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
IMP dehydrogenase/GMP synthase#	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
QueDEC / tgtA5	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
Inositol metabolism	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
L-fucose/Sialic acid metabolism	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
Aminoglycoside	White	White	White	White	White	White	White	White	White	White	Orange	Orange	Orange	White	White	Orange	White
Beta-lactam	White	White	White	White	White	White	White	White	White	White	Orange	Orange	White	White	White	Orange	White
Chloramphenicol	White	White	White	White	White	White	White	White	White	White	White	Orange	White	White	White	White	White
Macrolide	White	White	White	White	White	White	White	White	White	White	Orange	Orange	White	White	White	White	White
Rifampin	White	White	White	White	White	White	White	White	White	White	White	Orange	White	White	White	White	White
Sulfonamide	White	White	White	White	White	White	White	White	White	White	Orange	Orange	Orange	White	White	Orange	White
Tetracycline	White	White	White	White	White	White	White	White	White	White	Orange	Orange	Orange	White	White	Orange	White
Copper resistance	Orange	Orange	Orange	Orange	Orange	Orange	Orange	Orange	Orange	Orange	Orange	Orange	Orange	Orange	Orange	Orange	Orange
Mercuric resistance	White	White	White	White	White	White	White	White	White	White	Orange	White	Orange	White	White	White	White
Quaternary ammonium compounds	White	White	White	White	White	White	White	White	White	White	White	Orange	Orange	White	White	Orange	White

LEGEND: Colors indicate distinct classes of traits; green corresponds to virulence or *putative virulence genes; blue indicates metabolism; orange corresponds to resistance. #Indicates core genome. Numbers in the header correspond to the strains (¹ZYAH72, ²JBN2301, ³D4, ⁴ML09-119, ⁵AL09-71, ⁶pc104A, ⁷NJ-35, ⁸J-1, ⁹GYK1, ¹⁰KN-Mc-1R2, ¹¹MX16A, ¹²ZYAH75, ¹³WCHAH045096, ¹⁴AH10, ¹⁵AHNIH1, ¹⁶AL06-06, ¹⁷ATCC 7966).

AL06-06: Aminoglycoside 3"-nucleotidyltransferase - ANT(3'')-Ia; chloramphenicol O-acetyltransferase, CatB family; dihydropteroate synthase type-2 sulfonamide resistance protein; small multidrug resistance (SMR) efflux transporter - QacE delta 1, quaternary ammonium compounds; tetracycline resistance, MFS efflux pump – TetC.

MX16A: Aminoglycoside 3"-nucleotidyltransferase - ANT(3'')-Ia; aminoglycoside N(3)-acetyltransferase - AAC(3)-II,III,IV,VI,VIII,IX,X; class A beta-lactamase - VEB family, extended-spectrum; dihydropteroate synthase type-2 sulfonamide resistance protein; macrolide 2'-phosphotransferase - Mph(A) family; SMR efflux transporter - QacE delta 1, quaternary ammonium compounds; tetracycline resistance, MFS efflux pump Tet(A); mercuric resistance operon (regulatory protein MerR; mercuric transport protein, MerT; periplasmic mercury(+2) binding protein, MerP; mercuric ion reductase; mercuric resistance transcriptional repressor, MerD; mercuric transport protein, MerE).

WCHAH045096: Aminoglycoside 3"-nucleotidyltransferase - ANT(3'')-Ia; dihydropteroate synthase type-2 - sulfonamide resistance protein; small multidrug resistance (SMR) efflux transporter - QacE delta 1, quaternary ammonium compounds; tetracycline resistance, MFS efflux pump - Tet(A); mercuric resistance operon (regulatory protein MerR; mercuric transport protein, MerT; periplasmic mercury(+2) binding protein, MerP; mercuric ion reductase; mercuric resistance transcriptional repressor, MerD; mercuric transport protein, MerE).

ZYAH75: Aminoglycoside 3"-nucleotidyltransferase ANT(3'')-Ia; aminoglycoside 3'-phosphotransferase APH(3')-I; aminoglycoside 3"-phosphotransferase - APH(3'')-I; aminoglycoside N(6')-acetyltransferase - AAC(6')-Ib/AAC(6')-II; aminoglycoside 6-phosphotransferase - APH(6)-Ic/APH(6)-Id; chloramphenicol resistance, MFS efflux pump - CmlA family; chloramphenicol/florfenicol resistance, MFS efflux pump - FloR family; class A beta-lactamase - CTX-M family, extended-spectrum; class A beta-lactamase - TEM family; class D beta-lactamase - OXA-10 family; dihydropteroate synthase type-2 - sulfonamide resistance protein; macrolide 2'-phosphotransferase - Mph(A) family; rifampin ADP-ribosyl transferase; SMR, efflux transporter - QacE, quaternary ammonium compounds.

3.3.3 Traits related to metabolism

Genes or clusters of genes associated with metabolism were identified in the predicted GIs (Table 3). Among them, the gene cluster *dhaKLM*, encoding subunits of dihydroxyacetone kinase (Dha), *gldA*, glycerol dehydrogenase (GldA), and *dhaR*, encoding the transcriptional regulator DhaR, are associated with glycerol metabolism.

Genes encoding the enzymes inosine-5'-monophosphate dehydrogenase (IMP dehydrogenase) and GMP synthase were also identified in the predicted GIs of several strains. They participate in purine metabolism in the synthesis of GMP from IMP.

Other predicted GIs contained genes related to amino acid biosynthesis. Enzymes that participate in the biosynthesis of tryptophan from chorismate are encoded by the predicted GIs from several strains. Additionally, genes encoding the bifunctional aspartate kinase/homoserine dehydrogenase I, homoserine kinase, and threonine synthase, enzymes that are part of the aspartate pathway of amino acid biosynthesis related to threonine biosynthesis, were identified in predicted GIs in a few strains.

The *cysN* and *cysD* genes encoding subunits of the enzyme sulfate adenylyltransferase, and *cysC* encoding adenylylsulfate kinase were found to be part of the predicted GIs of a certain strains. These enzymes participate in some steps in the process of assimilatory sulfate reduction, in which sulfate is reduced to hydrogen sulfide and then incorporated in cysteine and methionine biosynthesis.

Genes encoding 6-carboxytetrahydropterin synthase (QueD), 7-carboxy-7-deazaguanine synthase (QueE), and 7-cyano-7-deazaguanine synthase (QueC) were identified in several of the predicted GIs, and participate in the synthesis of queuosine.

Genes associated with carbohydrate metabolism, including sialic acid, fucose, and inositol metabolism, were also identified in the predicted GIs (Table 3).

3.3.4 Phages

PHASTER analysis indicated the presence of intact and questionable phages in strain genomes. For intact phages, the most common were *Aeromonas* phage phiO18P, *Salmonella* phage SEN8, *Escherichia* phage Lys12581Vzw, *Shigella* phage POCJ13, and for questionable phages *Escherichia* phage 520873 and *Shigella* phage POCJ13. Only strains ATCC 7966, GYK1, and MX16A did not have any phage-related regions with a score > 70. However, no intact phages were found on the predicted GIs, but several of them contained partial phage sequences

(Supplementary Table 5).

3.3.5 Other features encoded in predicted GIs

Several other features are encoded in predicted GIs, including proteins related to chemotaxis signal sensing/transduction, response regulators, and a two-component system; a type I restriction modification system, DNA repair proteins, transporters, toxin-antitoxin modules, cytochrome associated proteins, an inhibitor of vertebrate c-type lysozyme (Ivy), and flagella and pili biosynthesis-associated proteins. Finally, GIs encoding transposon-associated proteins, conjugation-related proteins, or almost exclusively hypothetical proteins, were also observed in *A. hydrophila*. The distribution of these features varied among strains (Supplementary Table 3).

3.4 Predicted GIs containing genes from core genome

The presence of genes from the core genome was detected in several of the predicted GIs from all strains (Supplementary Table 3), which reached 13% (strain AH10) to 27% (strain AL06-06), with an average of 18% among the *A. hydrophila* isolates.

However, in a few islands, a high number of genes from the core genome were detected. This was observed mainly in the predicted GIs containing the gene cluster *dhaKLM gldA*, which were identified as part of the core genome. In addition, genes encoding ribosomal proteins and transcription-related factors, rRNAs, RNA polymerase subunit $\beta\beta'$, and some tRNAs, among others, are present in the predicted GIs. The presence of repeat regions and transposases between these genes was also observed, which may have contributed to its identification as a GI. The predicted GI containing the cluster *scsABCD*, associated with copper resistance, also contains a high proportion of genes from the core genome.

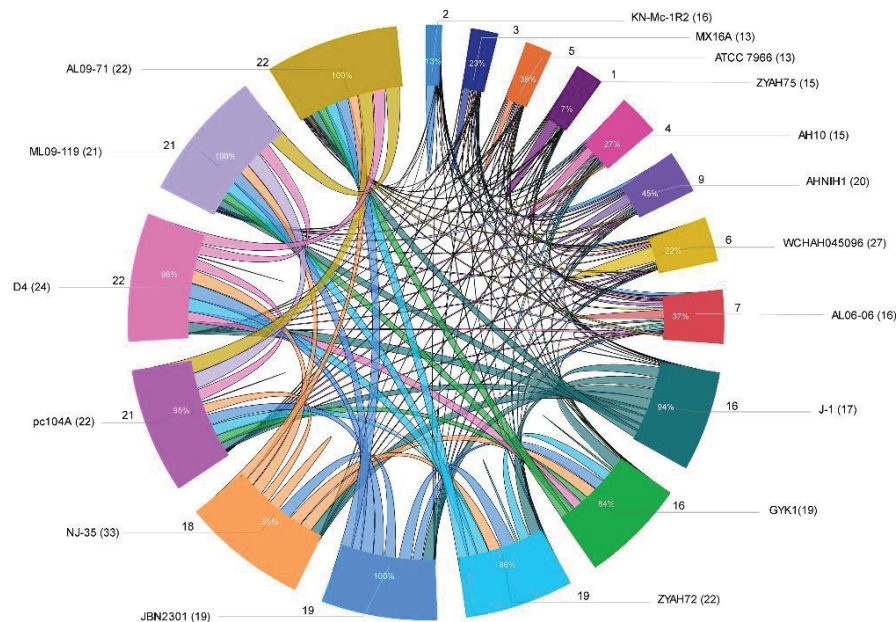
Other genes from the core genome in predicted GIs encode the enzyme IMP dehydrogenase, enzymes associated with biosynthesis of tryptophan or threonine, and the enzyme adenylylsulfate kinase. These and other genes related to the core genome are highlighted in blue in Supplementary Table 3.

3.5 Shared GIS

Several strains share predicted GIs, as indicated by comparative analysis performed considering at least 75% amino acid sequence identity and coverage as criteria (Figure 3; Supplementary Table 6). Some of the predicted GIs were shared among most of the strains, while others were shared only among those of phylogenetic group 1 or group 2. Strains from group 1

shared all or most of their predicted GIs, while those from group 2 shared only a few (Figure 3; Supplementary Table 6).

FIGURE 3. Sharing of predicted GIs among *A. hydrophila* strains



LEGEND: Ratio of GIs shared among strains. Distribution of the data is presented as the total number of predicted GIs, the total number of shared GIs, and their related percentages. The circle indicates the predicted GIs shared proportionally by the size of each arc; the names of the strains are indicated outside the circle with the number of predicted GIs in parentheses; the number of GIs shared is indicated by the lines and the proportion inside the arc; lines connecting strains indicate sharing of predicted GIs.

The main predicted GIs with identified functions shared among two or more strains from both groups included those encoding ScsABCD, DhAKLM/GIdA, and enzymes associated with tryptophan biosynthesis/lvy. Of these, the first is the most widespread and is present in all strains except KN-Mc-1R2. This strain also contains the cluster ScsABCD; however, it is encoded in a larger GI that did not meet the criteria used in this study to define a shared island with other strains. The predicted GIs with characterized functions shared only among strains from group 1 included those encoding inositol and sialic acid/L-fucose metabolism; QueDEC, the O antigen cluster for putative OX1 or OX6 serogroups, and Fic PA0574 (Supplementary Table 3;6). Additionally, predicted GIs encoding toxin/antitoxin systems, phages or hypothetical proteins are shared among them. In group 2, predicted GIs encoding the zona occludens toxin/accessory cholera enterotoxin and flagellar regions 1 and 2 were the most common shared GIs (Supplementary Table 3;6).

4 DISCUSSION

In this study, complete genomes of *A. hydrophila* strains were analyzed using the Island Viewer 4 bioinformatics tool for GI prediction in order to determine the distribution of GIs and their contributions in such strains, however, while some of the features encoded have functions defined, others are putative and have to be experimentally demonstrated in *Aeromonas*. A phylogenetic analysis performed with Sweep software showed that two strains identified as *A. hydrophila* (4AK4 and YL17) were misidentified at species level. These findings are in agreement with the studies by Beaz-Hidalgo et al., (2015) and Moriel et al. (2021) which indicated several problems with the taxonomic affiliation of *Aeromonas* genomes deposited at the NCBI database. These findings show the relevance of using complete genomes for taxonomy of *Aeromonas*. The *A. hydrophila* strains were classified into two distinct phylogenetic groups (Figure 1). Group 1 contained strains isolated from diseased fish, except pc104A, which was recovered from the soil of a catfish pond that experienced an epidemic outbreak of septicemia caused by *Aeromonas* (Pang et al., 2015). Most of the strains were related to epidemic outbreaks of motile *Aeromonas* septicemia in catfish that occurred in the USA (ML09-119, AL09-71, and pc104A) or carp in China (NJ-35, JBN2301, and J1), and share a recent common ancestor, as shown by comparative genomic analyses (Hossain et al., 2014). These strains were previously classified as hypervirulent *A. hydrophila* (vAh) based on phenotypic and genotypic tests or *in silico* analysis for pathotype-specific PCR of vAh isolates (Pang et al., 2015; Rasmussen-Ivey et al., 2016b).

Strains ZYAH72, D4 and GYK1, isolated from diseased fish in China, are also part of phylogenetic group 1. These US and Chinese strains were separated into distinct subclades of group 1 (Figure 1), what is in agreement with the data of Pang and collaborators (2015).

Strains from phylogenetic group 2 were recovered from different sources and were also separated into two subclades based on their geographical origins (Figure 1), and the South Korean strain grouped together with the US strains. Three of these strains (AH10, ATCC 7966, and AL06-06) were previously studied and considered to be non-virulent *A. hydrophila* (non-vAh) in relation to the hypervirulent vAh strains recovered from fish disease outbreaks (Pang et al., 2015; Rasmussen-Ivey et al., 2016b). Our data confirmed the results of other phylogenetic studies indicating the separation of vAh from non-epidemic strains (Pang et al., 2015; Rasmussen-Ivey et al., 2016b).

Strains from groups 1 and 2 differed in several characteristics, including genome size, number of predicted GIs, GI average size, and ratio of GIs relative to genome size (Table 2), in which strains from group 1 generally present greater numbers. Genomic GC content also varied

among these strains, from 60.50% to 60.90% in group 1 and from 61% to 61.60% in group 2. Strains from groups 1 and 2 also differed in terms of the functions encoded on the predicted GIs, although some features were common among them (Table 3).

The distribution of predicted GIs containing the main functions identified, and some features whose function have yet to be demonstrated in *A. hydrophila*, varied among the strains (Table 3), while some were common and other were encoded in a few strains. Among the identified characteristics are clusters for the synthesis of O Ag. O Ag, the most surface-exposed part of the lipopolysaccharide, a component of the outer membrane of Gram-negative bacteria, mediates pathogenicity by protecting infecting bacteria from serum complement killing and phagocytosis. It consists of repeating oligosaccharide subunits whose variability confers immunological specificity (Tomás, 2012). Several O Ag gene cluster types have been described in *A. hydrophila* (Hossain et al., 2013; Pang et al., 2015; Cao et al., 2018) and previously associated with GIs (Hossain et al., 2013; Pang et al., 2015). Gene clusters for O Ag were found in the predicted GIs of all strains; however, they differed in their gene content and were related to 10 distinct serogroups (Figure 2).

Thirteen of the strains analyzed in this study had their O Ag gene clusters previously determined (Hossain et al., 2013; Pang et al., 2015; Cao et al., 2018).

The putative serogroup OX1 was proposed for strains ML09-119, AL09-71, pc104a, JBN2301, and D4 strains (Cao et al., 2018). Strain ZYAH72, which had no previously analyzed O Ag cluster, shared the same cluster with the strains above (Figure 2). Thus, it should also be included in the putative OX1 serogroup. Strains NJ-35 and GYK1 were classified as putative serogroups OX6 (Cao et al., 2018). According to our data, strain J-1 should also be included in this serogroup because it shares the same O Ag biosynthesis cluster. This is in agreement with Pang et al. (2015), who showed that J-1 and NJ-35 have identical O Ag gene clusters.

All other strains had unique O Ag clusters. In contrast to the vAh fish pathogens that presented putative OX1 or OX6 O Ag clusters, strains AL06-06 and AH10 were also recovered from diseased fish but were considered as non-vAh (Rasmussen-Ivey et al., 2016b) and presented clusters OX4 and OX5 respectively (Cao et al., 2018). Strain ATCC 7966 belongs to the O1 serogroup (Seshadri et al., 2006), AHNIH1 to O25, WCHAH045096 to OX7 (Cao et al., 2018). The ZYAHA75 O cluster is similar to O18, and those of MX16A and KN-MC-1R2 were not determined (Figure 2).

Thus, no O Ag gene cluster was shared among strains from phylogenetic groups 1 and 2. All strains from group 2 exhibited unique O Ag clusters. Of the 10 distinct serogroups detected, only OX1 and OX6 were shared, and the identification of these clusters might be an alert for

potential virulent strains related to outbreaks of epidemic diseases in fish.

Several genes associated with polar flagella biogenesis and chemotaxis were identified in the predicted GIs of most strains (Table 3; Supplementary Table 3). Mesophilic *Aeromonas* have a polar flagellum for swimming in liquid, and some strains can also express multiple inducible lateral flagella for swarming over viscous environments or surfaces (Janda and Abbott, 2010). The polar flagellum genes are clustered in five distinct regions, whereas lateral flagella genes are organized in only one cluster (Tomás, 2012). Beyond their role in motility, in *Aeromonas* flagella act as adhesins for human intestinal enterocytes and in biofilm formation (Kirov et al., 2004), and are essential for adhesion and the ability to invade fish cell lines (Beaz-Hidalgo and Figueras, 2013). In the current study, GIs containing gene clusters for polar flagella region 2 (*flaA flaBGHJ maf-1*) (Canals et al., 2006) were predicted. However, the data indicated that gene content downstream of polar flagella region 2 is heterogeneous in different strains (Supplementary Table 3) and that genes associated with flagella glycosylation, such as *flmD* and *neuB* (Tabei et al., 2009; Wilhelms et al., 2012; Fulton et al., 2015) are adjacent to region 2 in certain *A. hydrophila* isolates. These data are in agreement with a recent study (Forn-Cuní et al., 2021) that showed that *pse/flm* genes (*flmA flmB neuA flmD neuB*) related to polar flagellin glycosylation are clustered in highly polymorphic GIs with three main genomic glycosylation islands identified.

Groups I and III contain almost exclusively the *pse* genes (*flmA flmB neuA flmD neuB*). *Aeromonas* belonging to these groups have flagellins modified with a single monosaccharide pseudaminic acid derivative (Forn-Cuní et al., 2021). Strains from phylogenetic group 1 and strain AH-10 belong to the glycosylation genomic island group I, while MX16A and WCHA045096 were included in group III according to Forn-Cuní and collaborators (2021).

The glycosylation GI group II contains other genes between the *pse/flm* genes, and are present in strains that modify polar flagellins with heterogeneous glycan moieties. Strains ATCC 7966, AL06-06, and AHNIHI are part of this group together with KN-Mc-1R2 and ZYAHA75 (Forn-Cuní et al., 2021). Flagellin glycosylation is essential for flagellar function (Tabei et al., 2009) and polar flagellar glycan composition can differ between *A. hydrophila* strains (Wilhelms et al., 2012; Fulton et al., 2015; Forn-Cuní et al., 2021). Because *Aeromonas* flagella have several roles in pathogenesis (Kirov et al., 2004; Beaz-Hidalgo and Figueras, 2013), based on data showing that strains ML09-119 and AL09-71 (pathogens of catfish), AH10, J-1, JBN2301, and NJ 35 (pathogen from carp) share the same predicted GIs related to flagellar region 2, and results from Forn-Cuní and collaborators (2021), showing that they belong to the glycosylation GI group I, we speculate that flagellin glycosylation patterns may favor interactions with determined hosts.

VgrG and Hcp play roles in *Aeromonas* virulence, influencing bacterial motility, protease production, and biofilm formation, inducing host cell toxicity, apoptosis, and activation of macrophages (Suarez et al., 2008, 2010 a,b; Sha et al., 2013). They are components of the T6SS which in addition to virulence plays roles in inter-bacterial competition. T6SS is encoded within gene clusters containing the core genes of the secretion machinery however, additional copies of *hcp* and *vgrG* can be found outside the main cluster (Cianfanelli et al., 2016; Moriel et al., 2021), which explains the presence of these genes in two distinct predicted GIs in certain strains (Supplementary Table 3). However, a search in the genome showed that strain AL06-06 does not have the complete T6SS cluster, differently from strains AH10 and KN-Mc-1R2. This is in agreement with Tekedar et al. (2019) which showed that this strain have only three components of T6SS, *hcp*, *tssH* and *vgrG*. In their study they also showed that virulent strains including ML09-119, AL09-71 and pc104A also have only these T6SS components, and that *hcp1* and *vgrG1* contribute for virulence of the ML09-119 in catfish.

Some other genes with functions identified in predicted GIs (Supplementary Table 3) are associated with virulence in other bacteria, but their roles in *Aeromonas* have yet to be determined. Among them are the zonula occludens toxin and the accessory cholera enterotoxin, which have been described in *Vibrio cholerae* and have the ability to affect intestinal function in rabbits (Fasano et al., 1997; Trucksis et al., 1993). Another is the ankyrin protein, AnkB. In plant-pathogens *Xanthomonas oryzae* pv. *oryzae* and *X. oryzae* pv. *oryzicola*, AnkB participates in biofilm formation, swimming ability, exopolysaccharide production, and defense against oxidative stress. In these bacteria, the *ankB* gene is located 58 bp downstream of *catB*, which encodes the enzyme catalase. AnkB affects *catB* gene expression, catalase activity, and sensitivity to H₂O₂ (Pan et al., 2018). In *A. hydrophila*, the *ankB* gene is in tandem with *kate*, which encodes catalase, and in most strains with a gene encoding a superoxide dismutase precursor. Although *A. hydrophila* is an opportunistic pathogen in animals, this led us to speculate regarding its roles in defense against oxidative stress, as shown in *X. oryzae*.

Fic proteins are widespread in bacteria and are often found encoded in GIs. They may play a role as toxins secreted from pathogens, or as the toxin component of toxin/antitoxin modules. However, the great majority of Fic proteins have unknown functions (Veyron et al., 2018).

Several genes from the core genome were found in the predicted GIs (Supplementary Table 3), which should be further considered. Except for the *scsABCD* locus associated with copper resistance, which was present in the predicted GIs of all strains analyzed, the distribution of GIs encoding resistance-associated features was limited to four strains in phylogenetic group 2

(Table 3). The full *scs* operon appears to be limited to some Enterobacteriaceae (*Salmonella enterica*, *Serratia proteamaculans*, *Citrobacter koseri*, and *Klebsiella pneumoniae*), *Aeromonas hydrophila*, and *Photobacterium profundum* (Subedi et al., 2019). The *scsABCD* locus was originally identified in *S. enterica* Typhimurium, and is required to deal with copper and H₂O₂ stress, helping to increase *Salmonella* survival under severe copper and oxidative stress, hostile conditions encountered by the pathogen during its intracellular survival (López et al., 2018) thus, contributing to virulence. However, copper may also be found in water (Domek et al., 1984). Considering that *Aeromonas* are ubiquitous in aquatic habitats and are considered as opportunistic pathogens (Janda and Abbott, 2010), potential roles of *scsABCD* may contribute to fitness in environment and virulence in the host.

In *A. hydrophila*, *scsSBCD* is located in the predicted GIs distinct from those encoding features associated with resistance (Supplementary Table 3)

Aeromonas have intrinsic resistance to some β -lactams due to the production of one to three inducible β -lactamases. The resistance phenotypes vary in relation to the species according the production of one to three inducible β -lactamases, with five basic phenotypes described (Fosse, 2010). In *A. hydrophila* three classes of β -lactamases, a class B metallo- β -lactamase, a class C cephalosporinase, and a class D oxacillinase were described (Fosse, 2010; Janda and Abbott, 2010), conferring to the species the penicillinase-cephalosporinase-carbapenemase phenotype exhibiting resistance to ampicillin, amoxicillin, ticarcillin, cephalothin, and profile susceptible or intermediate to imipenem (Fosse, 2010). We looked for genes encoding these beta lactamases and subclass B2 metallo- β -lactamase (CphA family, or ImiH or ImiS), cephalosporin-hydrolyzing class C β -lactamase (CepS or CepH) and OXA-12 family class D β -lactamase (AmpH/OXA-724 or AmpS/OXA-725 or OXA-726) were identified in the chromosome of all 17 strains. Additionally, a tetracycline efflux MFS transporter TetE was identified in the chromosome of 2 strains.

Aeromonas are usually susceptible to aminoglycosides, tetracyclines, chloramphenicol, sulfonamides, trimethoprim, nitrofurans, nalidixic acid, fluoroquinolones (Fosse, 2010), carbapenems, third- and fourth-generation cephalosporins, macrolides, monobactams and extended spectrum penicillins (Janda and Abbott, 2010; Fernández-Bravo and Figueras, 2020). However, strains presenting resistance to one or more of these antibiotics are common in clinical and environmental strains (Aravena-Román et al., 2012; Esteve et al., 2015). Studies on molecular antibiotic resistance have shown that class 1 integrons, involved in intraspecific and interspecific dissemination of resistance, and large plasmids are associated to resistance to

several antibiotics in *Aeromonas* (Nguyen et al., 2014; Hughes et al., 2016).

A. hydrophila multidrug-resistant strains are frequent, mainly among clinical isolates (Esteve et al., 2015, Hughes et al., 2016). Among the strains analyzed, the following contains plasmids, as informed at the NCBI database: JBN2301 (3 plasmids), D4 (4 plasmids), AHNIH1 (1 plasmid), AL06-06 (3 plasmids) and WCHAH045096 (6 plasmids). Plasmids from strains JBN2301 and AL06-06 do not contain resistance genes. In strain D4, only one of the plasmids contain a gene encoding a quinolone resistance pentapeptide repeat protein QnrS2. The plasmid of strain AHNIH1, contains several genes encoding for resistance, as described (Hughes et al., 2016). Includes three class A beta-lactamases (KPC-2, SHV-12 and CARB-12) and genes encoding for resistance to aminoglycosides, chloramphenicol, fluoroquinolones, macrolides, sulfonamide and trimethoprim. Three of the plasmids of WCHAH045096 carry resistance genes. They are associated to the resistance to several classes of antibiotics including aminoglycosides, sulfonamide, trimethoprim, quinolone, tetracycline and carbapenem.

Predicted GIs containing antibiotic resistance associated genes had heterogeneous distribution among the *A. hydrophila* (Supplementary Table 4), possibly reflecting horizontal gene transfer and selection with the exposure to antibiotics in their environments. These predicted GIs were identified in four strains, recovered from water, sewage, fish or human, indicating that antibiotic resistance associated with GIs is spread in clinical and environmental strains. Most of these predicted GIs also encode transposases, integrases, and mobile elements, and one of them also includes proteins related to conjugation (Supplementary Table 3). In the strain AL06-06, a predicted GI encodes for aminoglycoside, sulfonamide, chloramphenicol and tetracycline resistance. The predicted GI of WCHAH045096 encodes for resistance to the same antibiotics described above, and contains genes for several proteins associated with mercury resistance. The predicted GI of strain MX16A encodes antibiotic resistance for tetracycline, aminoglycoside, chloramphenicol, sulfonamide, macrolide and beta-lactams e proteins related to mercury resistance. Strain ZYAH75 has three predicted GIs containing resistance genes. One of them contain genes for resistance to aminoglycoside and sulfonamide together with proteins for mercury resistance. This predicted GI is a mosaic of proteins associated with mobile elements, it encodes conjugative transfer proteins TrbJ, TrbK and TrbK, IncP-type oriT binding proteins TraJ and TraK, together with integrase, transposase, IS and mobile element proteins. Genes for chloramphenicol and aminoglycoside resistance are encoded in two other islands, which contains additionally genes encoding class A extended-spectrum beta-lactamase CTM-3, class A broad-spectrum beta-lactamases TEM-1, and gene for sulfonamide resistance; or genes for resistance to

macrolides and quinolone, oxacillin-hydrolyzing class D beta-lactamase OXA-10, and also proteins associated with mercury resistance. Integron integrase *IntI1* are encoded in two of the predicted GI of ZYAH75, and those from strains WCHAH045096 and MX16A. In several of these predicted GIs there is more than one gene encoding for aminoglycoside modifying enzymes, and for sulfonamide resistance. Therefore, in addition to the intrinsic resistance characteristic of the species, some strains analyzed contain determinants for resistance to tetracycline in the chromosome (2 strains); plasmids carrying genes for resistance to one and up to six distinct classes of antibiotics, and even more than one plasmid related to resistance, and predicted GIs in which several genes of resistance are encoded, the most common are those related to the aminoglycoside and sulfonamide resistance. *Aeromonas* are ubiquitous in aquatic environments and widely distributed, being found in every environmental niche where bacterial ecosystems exist, facilitating the contact with humans (Janda and Abbott, 2010). Thus, the presence of GIs carrying resistance genes in *Aeromonas* may impact the fitness of the bacteria, facilitate the resistance dissemination, and spread of resistant strains in several environments.

Some of the predicted GIs shared among strains from phylogenetic group 1 encode metabolic pathways for utilization of myo-inositol and L-fucose/sialic acid, which were previously described in GIs of *A. hydrophila* strains NJ-35, J-1, ML09-119, AL09-71, and pc104A (Pang et al., 2015). In the current study, these features were also found in the predicted GIs of other strains of group 1. These pathways may be linked to the full virulence of fish disease epidemics *A. hydrophila* (Pang et al., 2015), suggesting that these GIs may be related to fitness or enhancing adaptation to hosts and competitiveness.

Other predicted GIs common to vAh contain genes encoding QueD, QueE, and QueC (Supplementary Table 3), which participate in the synthesis of PreQ₀, a precursor of queuosine-(Q) tRNA modification in bacteria. PreQ₀ is reduced to PreQ₁ and inserted in the G residue at position 34 of tRNAs with GUN anticodons by tRNA guanosine transglycosylase (bTGT); however, the enzyme can also use PreQ₀ when PreQ₁ is absent (Thiaville et al., 2016). However, predicted *A. hydrophila* GIs encoding QueDEC also encode TgtA5, a variant of TGT that has divergent features from bTGT, helicases, and proteins involved in DNA repair (Supplementary Table 3). Similar clusters have been described in a GI of *Salmonella enterica* serovar Montevideo and by comparative genomic analysis in several other bacteria, including *A. hydrophila* ML09-119. In *S. enterica* serovar Montevideo, *Kineococcus radiotolerans*, *Comamonas testosteroni*, and *Sphingopyxis alaskensis*, clustering of *tgtA5* and *preQ₀* synthesis genes is involved in inserting 7-deazapurine derivatives into DNA, suggesting a new and complex system of DNA modification

(Thiaville et al., 2016). The SAM-dependent methyltransferase HI0095 (UbiE paralog), which is associated with virulence in vAh (Rasmussen-Ivey et al., 2016b) is also encoded in this predicted GI. Thus, the roles of these genes in *A. hydrophila* DNA modification and/or virulence have to be determined.

No distinctive characteristics of strains from group 2 were identified, and most isolates shared only a small number of predicted GIs (Supplementary Table 6), suggesting a high degree of strain diversity. Of note, several of these strains have multiple determinants of antimicrobial resistance encoded by their predicted GIs. Group 2 contains only eight strains; however, they may better reflect the diversity of GIs in *A. hydrophila* because previous studies have shown that most strains from group 1 are highly related and constitute a clonal group (Hossain et al., 2014; Pang et al., 2015; Rasmussen-Ivey et al., 2016b).

Addendum. At the end of this study, new complete genomes of *A. hydrophila* were available in the NCBI database reaching a total of 40, of which 19 were previously analyzed. Information on the new 21 strains is showed in (Supplementary table NST1). A complementary phylogenetic analysis including all 40 genomes indicated that four of the strains classified as *A. hydrophila* were misidentified at species level; two are the new strains NEB724 and B11, and YL17 and 4AK4 which were previously excluded from our study (Supplementary figure NSF1). The 36 *A. hydrophila* were separated in two main branches (Supplemental figure NSF1). However, one branch contains only new strains isolated from chicken (4 strains). The other branch includes subgroups containing the strains previously classified in group 1 (Vah) plus new isolates from snakes (2), fish (1) and without info (1); and the strains from previous group 2, in which new isolates from water (4), fish (4), human (1), and no source info (2) were included (Supplemental figure NSF1). The raw data of Island Viewer4 analysis performed with the new genomes are showed in (Supplemental table NST2). Among the 19 new genomes of *A. hydrophila*, 735 GIs were predicted, however these GIs were not manually curated. Gene clusters for O Ag biosynthesis were identified in predicted GIs. The relationship among the O antigen gene clusters of the 36 *A. hydrophila* was analyzed through nucleotide sequences alignment performed using the Clustal Omega Web Server (<https://www.ebi.ac.uk/Tools/msa/clustalo/>), with “default” parameters and the software IQ-TREE - Efficient Tree Reconstruction (version 1.6.12 – Windows 64; Nguyen et al., 2015, Mol. Biol. Evol., 32:268–274, 2015), which was used to build the phylogenetic tree of the clusters. The parameters used were “default” for ModelFinder, with bootstrap and maximum likelihood values. Results are indicated in the (Supplementary Figure

NSF2), which indicates that two of the new strains present O Ag cluster already identified among the strains previously analyzed. Strain LHW39 presents the OX1 cluster which was observed among some of the VAh, and strain HX-3 the OX-5 cluster, identified in strain AH10. Fourteen O Ag cluster, apparently distinct of those identified among strains previously analyzed, were found in GIs of 17 new genomes (Supplementary figure NSF2); similar O Ag clusters were found in GIs of strains 23-C-23 and WCX23-1; Brac6 and ONP3-1; GSH8-2 and WP8-S18.ESBL02. Predicted GIs containing resistance genes were found in 5 strains (Supplementary table NSF3), and are related with resistance to tetracycline (4 strains), beta-lactam (class A extended-spectrum beta-lactamase CTX-M-14; 1 strain), macrolide (2 strains), aminoglycosides and sulfonamide (1 strain). The number of resistance features found in the predicted GIs of the new genomes varies from 1 to 3, in contrast with those observed in the previous analyzes which range from 3 to 9. These preliminary results confirm the previous analyzes indicating that some features, such as O Ag clusters, are associated to GIs in all strains. Features encoded on predicted GIs of the new genomes confirms the contribution of these elements for the diversity and antimicrobial resistance of *A. hydrophila*.

REFERENCES

- Aravena-Román, M., Inglis, T. J. J., Henderson, B., Riley, T. V., and Changa, B. J. (2012). Antimicrobial susceptibilities of *Aeromonas* strains isolated from clinical and environmental sources to 26 antimicrobial agents. *Antimicrob. Agents Chemother.* 56, 1110–1112. doi: 10.1128/AAC.05387-11.
- Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387.
- Beaz-Hidalgo, R., and Figueras, M. J. (2013). *Aeromonas* spp. whole genomes and virulence factors implicated in fish disease. *J. Fish Dis.* 36, 371–388. doi: 10.1111/jfd.12025.
- Beaz-Hidalgo, R., Hossain, M. J., Liles, M. R., and Figueras, M.-J. (2015). Strategies to avoid wrongly labelled genomes using as example the detected wrong taxonomic affiliation for *Aeromonas* genomes in the GenBank database. *PLoS One* 10:e0115813. doi: 10.1371/journal.pone.0115813.
- Bertelli, C., Laird, M. R., Williams, K. P., Simon Fraser University Research Computing Group, Lau, B. Y., Hoad, G., et al. (2017). IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* 45, W30–W35. doi: 10.1093/nar/gkx343.
- Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., et al. (2015). RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation

pipelines and annotating batches of genomes. *Sci. Rep.* 10:8365. doi: 10.1038/srep08365.

Buchrieser, C., Prentice, M., and Carniel, E. (1998). The 102-kilobase unstable region of *Yersinia pestis* comprises a high-pathogenicity island linked to a pigmentation segment which undergoes internal rearrangement. *J. Bacteriol.* 180, 2321–2329.

Canals, R., Ramirez, S., Vilches, S., Horsburgh, G., Shaw, J. G., Tomás, J. M., et al. (2006). Polar flagellum biogenesis in *Aeromonas hydrophila*. *J. Bacteriol.* 188, 542–555. doi: 10.1128/JB.188.2.542-555.2006.

Cao, H., Wang, M., Wang, Q., Xu, T., Du, Y., Li, H., et al. (2018). Identifying genetic diversity of O antigens in *Aeromonas hydrophila* for molecular serotype detection. *PLoS One* 13:e0203445. doi: 10.1371/journal.pone.0203445.

Che, D., Wang, H., Fazekas, J., and Chen, B. (2014). An accurate genomic island prediction method for sequenced bacterial and archaeal genomes. *J. Proteomics Bioinform.* 7, 214–221. doi: 10.4172/jpb.1000322.

Cianfanelli, F. R., Monlezun, L., and Coulthurst, S. J. (2016). Aim, load, fire: the type VI secretion system, a bacterial Nanoweapon. *Trends Microbiol.* 24, 51–62.

Darling, A. C., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704.

De Pierri, C. R., Voyceik, R., Santos de Mattos, L. G. C., Kulik, M. G., Camargo, J. O., Repula de Oliveira, A. M., et al. (2020). SWEEP: representing large biological sequences datasets in compact vectors. *Sci. Rep.* 10:91.

Dieckmann, M. A., Beyvers, S., Nkouamedjo-Fankep, R. C., Hanel, P. H. G., Jelonek, L., Blom, J., et al. (2021). EDGAR3.0: comparative genomics and phylogenomics on a scalable infrastructure. *Nucleic Acids Res.* 49, W185–W192.

Dobrindt, U., Hochhut, B., Hentschel, U., and Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.* 2, 414–424. doi: 10.1038/nrmicro884.

Domek, M. J., LeChevallier, M. W., Cameron, S. C., and McFeters, G. A. (1984). Evidence for the role of copper in the injury process of coliform bacteria in drinking water. *Appl. Environ. Microbiol.* 48, 289–293. doi: 10.1128/aem.48.2.289-293.1984.

Ester, M., Kriegel, H., Sander, J., and Xu, X. (1996). “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *Proceedings of the KDD-1996 Second International Conference on Knowledge Discovery and Data Mining*, (Menlo Park, CA: AAAI Press), 226–231.

Esteve, C., Alcaide, E., and Giménez, M. J. (2015). Multidrug-resistant (MDR) *Aeromonas* recovered from the metropolitan area of Valencia (Spain): diseases spectrum and prevalence in the environment. *Eur. J. Clin. Microbiol. Infect. Dis.* 34, 137–145. doi: 10.1007/s10096-014-2210-z.

- Fasano, A., Uzzau, S., Fiore, C., and Margaretten, K. (1997). The enterotoxic effect of Zonula occludens Toxin on rabbit small intestine involves the paracellular pathway. *Gastroenterology* 112, 839–846.
- Fernández-Bravo, A., and Figueras, M. J. (2020). An update on the genus *Aeromonas*: taxonomy, epidemiology, and pathogenicity. *Microorganisms* 8:129. doi: 10.3390/microorganisms8010129.
- Forn-Cuní, G., Fulton, K. M., Smith, J. C., Twine, S. M., Mendoza-Barberà, E., Tomás, J. M., et al. (2021). Polar flagella glycosylation in *Aeromonas*: genomic characterization and involvement of a specific glycosyltransferase (Fgi-1) in heterogeneous flagella glycosylation. *Front. Microbiol.* 11:595697. doi: 10.3389/fmicb.2020.595697.
- Fosse, T. (2010). “Chapter 42. *Aeromonas*, vibrio and plesiomonas,” in *Antibiogram*, eds P. Courvalin, R. Leclercq, and L. B. Rice (Paris: ESKA Publishing, ASM Press), 509–518.
- Fulton, K. M., Mendoza-Barberá, E., Twine, S. M., Tomás, J. M., and Merino, S. (2015). Polar glycosylated and lateral non-glycosylated flagella from *Aeromonas hydrophila* strain AH-1 (Serotype O11). *Int. J. Mol. Sci.* 16, 28255–28269. doi: 10.3390/ijms161226097.
- Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., et al. (2010). The NCBI BioSystems database. *Nucleic Acids Res.* 38, D492–D496. doi: 10.1093/nar/gkp858.
- Grim, C. J., Kozlova, E. V., Ponnusamy, D., Fitts, E. C., Sha, J., Kirtley, M. L., et al. (2014). Functional genomic characterization of virulence factors from necrotizing fasciitis-causing strains of *Aeromonas hydrophila*. *Appl. Environ. Microbiol.* 80, 4162–4183. doi: 10.1128/AEM.00486-14.
- Hacker, J., Bender, L., Ott, M., Wingender, J., Lund, B., Marre, R., et al. (1990). Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extra intestinal *Escherichia coli* isolates. *Microb. Pathog.* 8, 213–225. doi: 10.1016/0882-4010(90)90048-U.
- Hossain, M. J., Sun, D., McGarey, D. J., Wrenn, S., Alexander, L. M., Martino, M. E., et al. (2014). An Asian origin of virulent *Aeromonas hydrophila* responsible for disease epidemics in United States-farmed catfish. *mBio* 5:e00848-14. doi: 10.1128/mBio.00848-14.
- Hossain, M. J., Waldbieser, G. C., Sun, D., Capps, N. K., Hemstreet, W. B., et al. (2013). Implication of lateral genetic transfer in the emergence of *Aeromonas hydrophila* isolates of epidemic outbreaks in channel catfish. *PLoS One* 8:e80943. doi: 10.1371/journal.pone.0080943.
- Hsiao, W., Wan, I., Jones, S. J., and Brinkman, F. S. L. (2003). IslandPath: aiding detection of genomic islands in prokaryotes. *Bioinformatics* 19, 418–420. doi: 10.1093/bioinformatics/btg004.
- Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics (Oxf. Engl.)* 26, 680–682. doi: 10.1093/bioinformatics/btq003.
- Hughes, H. Y., Conlan, S. P., Lau, A. F., Dekker, J. P., Michelin, A. V., Youn, J. H., et al. (2016). Detection and whole-genome sequencing of carbapenemase-producing *Aeromonas hydrophila* isolates from routine perirectal surveillance culture. *J. Clin. Microbiol.* 54, 1167–1170. doi: 10.1128/JCM.03229-15.

- Janda, J. M., and Abbott, S. L. (1998). Evolving concepts regarding the genus *Aeromonas*: an expanding panorama of species, disease presentations, and unanswered questions. *Clin. Infect. Dis.* 27, 332–344.
- Janda, J. M., and Abbott, S. L. (2010). The genus *Aeromonas*: taxonomy, pathogenicity, and infection. *Clin. Microbiol. Rev.* 23, 35–73. doi: 10.1128/CMR.00039-09.
- Jia, B., Raphenya, A. R., Alcock, B., Waglechner, N., Guo, P., Tsang, K. K., et al. (2017). CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 45, D566–D573. doi: 10.1093/nar/gkw1004.
- Juhas, M., van der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W., and Crook, D. W. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.* 33, 376–393. doi: 10.1111/j.1574-6976.2008.00136.
- Kanehisa, M., and Sato, Y. (2020). KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci.* 29, 28–35. doi: 10.1002/pro.3711.
- Kirov, S. M., Castrisios, M., and Shaw, J. G. (2004). *Aeromonas* flagella (Polar and Lateral) are enterocyte Adhesins that contribute to biofilm formation on surfaces. *Infect. Immun.* 72, 1939–1945.
- Li, W., and Wang, A. (2021). Genomic islands mediate environmental adaptation and the spread of antibiotic resistance in multiresistant *Enterococci* – evidence from genomic sequences. *BMC Microbiol.* 21:55. doi: 10.1186/s12866-021-02114-4.
- Li, Y. H., Yu, C. Y., Li, X. X., Zhang, P., Tang, J., Yang, Q., et al. (2018). Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res.* 46, D1121–D1127. doi: 10.1093/nar/gkx1076.
- Liu, B., Zheng, D., Jin, Q., Chen, L., and Yang, J. (2019). VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res.* 47, D687–D692. doi: 10.1093/nar/gky1080.
- López, C., Checa, S. K., and Soncini, F. C. (2018). CpxR/CpxA controls *scsABCD* transcription to counteract copper and oxidative stress in *Salmonella enterica* serovar typhimurium. *J. Bacteriol.* 200: e00126-18. doi: 10.1128/JB.00126-18.
- Martin-Carnahan, M., and Joseph, S. W. (2005). “*Aeromonadales* ord. nov,” in *Bergey’s Manual of Systematic Bacteriology: The Proteobacteria, Part B: The gammaproteobacteria*, Vol. 2, eds D. J. Brenner, N. R. Krieg, J. T. Staley, G. M. Garrity, D. R. Boone, P. de Vos, et al. (New York, NY: Springer), 556–587. doi: 10.1007/0-387-28022-712.
- Moriel, B., Prediger, K. C., Souza, E. M., Pedrosa, F. O., Fadel-Picheth, C. M. T., and Cruz, L. M. (2021). In silico comparative analysis of *Aeromonas* type VI secretion system. *Braz. J. Microbiol.* 52, 229–243. doi: 10.1007/s42770-020-00405-y.
- Nguyen, H. N. K., Van, T. T. H., Nguyen, H. T., Smooker, P. M., Shimeta, J., and Coloe, P. J. (2014). Molecular characterization of antibiotic resistance in *Pseudomonas* and *Aeromonas*

isolates from catfish of the Mekong Delta, Vietnam. *Vet. Microbiol.* 171, 397–405. doi: 10.1016/j.vetmic.2014.01.028.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300.

Pan, X., Xu, S., Wu, J., Duan, Y., Zheng, Z., Wang, J., et al. (2018). Ankyrin-like protein AnkB interacts with CatB, affects catalase activity, and enhances resistance of *Xanthomonas oryzae* pv. *oryzae* and *Xanthomonas oryzae* pv. *oryzicola* to phenazine-1-carboxylic acid. *Appl. Environ. Microbiol.* 84: e02145-17. doi: 10.1128/AEM.02145-17.

Pang, M., Jiang, J., Xie, X., Wu, Y., Dong, Y., Kwok, A. H. Y., et al. (2015). Novel insights into the pathogenicity of epidemic *Aeromonas hydrophila* ST251 clones from comparative genomics. *Sci. Rep.* 5:09833. doi: 10.1038/srep09833.

Partridge, S. R., Kwong, S. M., Firth, N., and Jensen, S. O. (2018). Mobile genetic elements associated with antimicrobial resistance. *Clin. Microbiol. Rev.* 31:e00088-17. doi: 10.1128/CMR.00088-17.

Pundhir, S., Vijayvargiya, H., and Kumar, A. (2008). PredictBias: a server for the identification of genomic and pathogenicity islands in prokaryotes. *In Silico Biol.* 8, 223–234.

Rao, R. T., Sharma, S., Sivakumar, N., and Jayakumar, K. (2020). Genomic islands and the evolution of livestock-associated *Staphylococcus aureus* genomes. *Biosci. Rep.* 40:BSR20202287. doi: 10.1042/BSR20202287.

Rasmussen-Ivey, C. R., Hossain, M. J., Odom, S. E., Terhune, J. S., Hemstreet, W. G., Shoemaker, C. A., et al. (2016b). Classification of a hypervirulent *Aeromonas hydrophila* pathotype responsible for epidemic outbreaks in warm-water fishes. *Front. Microbiol.* 7:1615. doi: 10.3389/fmicb.2016.01615.

Rasmussen-Ivey, C. R., Figueras, M. J., McGarey, D., and Liles, M. R. (2016a). Virulence factors of *Aeromonas hydrophila*: in the wake of reclassification. *Front. Microbiol.* 7:1337. doi: 10.3389/fmicb.2016.01337.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M., et al. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* 16, 944–945. doi: 10.1093/bioinformatics/16.10.944.

Sayers, S., Li, L., Ong, E., Deng, S., Fu, G., Lin, Y., et al. (2019). Victors: a web-based knowledge base of virulence factors in human and animal pathogens. *Nucleic Acids Res.* 47, D693–D700. doi: 10.1093/nar/gky999.

Seshadri, R., Joseph, S. W., Chopra, A. K., Sha, J., Shaw, J., Graf, J., et al. (2006). Genome sequence of *Aeromonas hydrophila* ATCC 7966T: Jack of all trades. *J. Bacteriol.* 188, 8272–8282. doi: 10.1128/JB.00621-06.

Sha, J., Rosenzweig, J. A., Kozlova, E. V., Wang, S., Erova, T. E., Kirtley, M. L., et al. (2013). Evaluation of the roles played by Hcp and VgrG type 6 secretion system effectors in *Aeromonas*

hydrophila SSU pathogenesis. *Microbiology* 159, 1120–1135. doi: 10.1099/mic.0.063495-0.

Soares, S. C., Geyik, H., Ramos, R. T., de Sá, P. H., Barbosa, E. G., Baumbach, J., et al. (2016). GIPSy: genomic island prediction software. *J. Biotechnol.* 232, 2–11. doi: 10.1016/j.jbiotec.2015.09.008.

Srividhya, K. V., Rao, G. V., Raghavenderan, L., Mehta, P., Prilusky, J., Manicka, S., et al. (2006). “Database and comparative identification of prophages,” in *Intelligent Control and Automation, Lecture Notes in Control and Information Sciences*, Vol. 344, eds D.-S. Huang, K. Li, and G. W. Irwin (Berlin: Springer), 863–868.

Suarez, G., Sierra, J. C., Erova, T. E., Sha, J., Horneman, A. J., and Chopra, A. K. (2010a). A type VI secretion system effector protein, VgrG1, from *Aeromonas hydrophila* that induces host cell toxicity by ADP ribosylation of actin. *J. Bacteriol.* 192, 155–168. doi: 10.1128/JB.01260-09.

Suarez, G., Sierra, J. C., Kirtley, M. L., and Chopra, A. K. (2010b). Role of Hcp, a type 6 secretion system effector, of *Aeromonas hydrophila* in modulating activation of host immune cells. *Microbiology* 156, 3678–3688. doi: 10.1099/mic.0.041277-0.

Suarez, G., Sierra, J. C., Sha, J., Wang, S., Erova, T. E., Fadl, A. A., et al. (2008). Molecular characterization of a functional type VI secretion system from a clinical isolate of *Aeromonas hydrophila*. *Microb. Pathog.* 44, 344–361. doi: 10.1016/j.micpath.2007.10.005.

Subedi, P., Paxman, J. J., Wang, G., Ukuwela, A. A., Xiao, Z., and Heras, B. (2019). The Scs disulfide reductase system cooperates with the metallochaperone CueP in *Salmonella* copper resistance. *J. Biol. Chem.* 294, 15876–15888. doi: 10.1074/jbc.RA119.010164.

Tabei, S. M. B., Hitchen, P. G., Day-Williams, M. J., Merino, S., Vart, R., Pang, P. C., et al. (2009). An *Aeromonas caviae* genomic Island is required for both O-antigen lipopolysaccharide biosynthesis and flagellin glycosylation. *J. Bacteriol.* 191, 2851–2863. doi: 10.1128/JB.01406-08.

Tekedar, H. C., Abdelhamed, H., Kumru, S., Blom, J., Karsi, A., and Lawrence, M. L. (2019). Comparative genomics of *Aeromonas hydrophila* secretion systems and mutational analysis of hcp1 and vgrG1 genes from T6SS. *Front. Microbiol.* 9:3216. doi: 10.3389/fmicb.2018.03216.

Thiaville, J. J., Kellner, S. M., Yuan, Y., Hutinet, G., Thiaville, P. C., Jumpathong, W., et al. (2016). Novel genomic island modifies DNA with 7-deazaguanine derivatives. *Proc. Natl. Acad. Sci. U.S.A.* 113, E1452–E1459. doi: 10.1073/pnas.1518570113.

Tomás, J. M. (2012). The main *Aeromonas* pathogenic factors. *Int. Sch. Res. Netw. ISRN Microbiol.* 2012:256261. doi: 10.5402/2012/256261.

Trucksis, M., Galen, J. E., Michalski, J., Fasano, A., and Kaper, J. B. (1993). Accessory cholera enterotoxin (Ace), the third toxin of a *Vibrio cholerae* virulence cassette. *Proc. Natl. Acad. Sci. U.S.A.* 90, 5267–5271.

Vernikos, G. S., and Parkhill, J. (2006). Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics* 22, 2196–2203. doi: 10.1093/bioinformatics/btl369.

- Veyron, S., Peyroche, G., and Cherfils, J. (2018). FIC proteins: from bacteria to humans and back again. *Pathog. Dis.* 76:fty012. doi: 10.1093/femspd/fty012.
- Waack, S., Keller, O., Asper, R., Brodag, T., Damm, C., Fricke, W. F., et al. (2006). Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics* 7:142. doi: 10.1186/1471-2105-7-142.
- Wattam, A. R., Davis, J. J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., et al. (2017). Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Res.* 45, D535–D542. doi: 10.1093/nar/gkw1017.
- Wei, W., Gao, F., Du, M. Z., Hua, H. L., Wang, J., Guo, F. B., et al. (2016). Zisland Explorer: detect genomic Islands by combining homogeneity and heterogeneity properties. *Brief. Bioinform.* 18, 357–366. doi: 10.1093/bib/bbw019.
- Weinstock, G. M., Sodergren, E. J., Smajs, D., and Norris, S. J. (2000). “Identification of virulence genes in silico: infectious disease genomics,” in *Virulence Mechanisms of Bacterial Pathogens*, 3rd Edn, eds K. A. Brogden, J. A. Roth, T. B. Stanton, C. A. Bolin, F. C. Minion, and M. J. Wannemuehler (Washington, DC: American Society of Microbiology), 251–261.
- Wilhelms, M., Fulton, K. M., Twine, S. M., Tomás, J. M., and Merino, S. (2012). Differential glycosylation of polar and lateral flagellins in *Aeromonas hydrophila* AH-3. *J. Biol. Chem.* 287, 27851–27862. doi: 10.1074/jbc.M112.376525.
- Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46, D1074–D1082. doi: 10.1093/nar/gkx1037.

CAPÍTULO IV: TAXONOMIA *Aeromonas* – ARTIGO II

RESUMO

Pertencente à família *Aeromonadaceae*, o gênero *Aeromonas* compreende bacilos gram-negativos, que se encontram amplamente distribuídos em ambientes aquáticos. São microrganismos oportunistas, sendo que algumas espécies são capazes de causar infecções em humanos e animais. Estas bactérias podem ainda ser encontradas em diferentes habitats, ambientes e produtos alimentares. A taxonomia deste gênero é complexa, requerendo diversos testes para a sua identificação. Os genomas completos de 177 cepas de *Aeromonas* pertencentes a 15 espécies ou identificadas apenas ao nível de gênero, foram utilizadas neste estudo para analisar a sua classificação taxonômica que está descrita e depositada nos bancos de dados de referência (NCBI). As análises foram realizadas através da construção de duas árvores filogenéticas, baseadas nas sequências de genomas completos (proteomas); genes “housekeeping” *gyrB*, *rpoD*, *rpoB*, e de uma matriz com valores de identidade média de nucleotídeos - ANI. Para a construção da primeira árvore, utilizamos as sequências de genomas completos com a ferramenta SWeeP (em formato de amino ácido). Na segunda, usamos apenas os genes *housekeeping gyrB*, *rpoD*, *rpoB* com suas sequências concatenadas através do SWeeP. Para a matriz com os valores, usamos as sequências de genomas completos, com a ferramenta fastANI em formato de nucleotídeos. A comparação dos resultados das árvores filogenéticas e dos índices de valores do ANI, que indicam a classificação taxonômica das cepas utilizando diferentes elementos, mostram maior concordância entre as análises do proteoma e ANI. Os resultados indicam que 11 cepas estão erroneamente identificadas ao nível de espécie. 22 das 27 cepas identificadas apenas ao nível de gênero puderam ser classificadas ao nível de espécie. E 3 cepas não foram incluídas em nenhum grupo, sugerindo que possam pertencer a espécies distintas das que já tem genoma completo sequenciado, ou até mesmo representar novas espécies de *Aeromonas*.

Palavras-chave: *Aeromonas*; taxonomia; árvore filogenética; identidade nucleotídica média (ANI); genes *housekeeping*.

ABSTRACT

Belonging to the family Aeromonadaceae, the genus *Aeromonas* comprises gram-negative bacilli, which are widely distributed in aquatic environments. They are opportunistic microorganisms, and some species are capable of causing infections in humans and animals. These bacteria can still be found in different habitats, environments and food products. The taxonomy of this genus is complex, requiring several tests for its identification. The complete genomes of 177 *Aeromonas* strains belonging to 15 species or identified only at the genus level were used in this study to analyze their taxonomic classification, which is described and deposited in the reference databases (NCBI). The analyzes were performed through the construction of two phylogenetic trees, based on the sequences of complete genomes (proteomes); the “housekeeping” genes *gyrB*, *rpoD*, *rpoB*, and a matrix with values of mean nucleotide identity - ANI. For the construction of the first tree, we used the complete genome sequences with the SWeeP tool (in amino acid format). In the second, we used only the housekeeping genes *gyrB*, *rpoD*, *rpoB* with their sequences concatenated through SWeeP. For the matrix with the values, we used the whole genome sequences with the fastANI tool in nucleotide format. The comparison of the results of the phylogenetic trees and the indices of ANI values, which indicate the taxonomic classification of the strains using different elements, show greater agreement between the proteome and ANI analyses. The results indicate that 11 strains are misidentified at the species level. 22 of the 27 strains identified only at the genus level could be classified at the species level. And 3 strains were not included in any group, suggesting that they may belong to different species from those that already have a complete genome sequenced, or even represent new species of *Aeromonas*.

Keywords: *Aeromonas*; taxonomy; phylogenetic tree; average nucleotide identity (ANI); housekeeping genes.

1 INTRODUÇÃO

A história do gênero *Aeromonas* se baseia nos relatos da bacteriologia médica moderna, a partir do seu nascimento como uma ciência laboratorial no final do século XIX e início do século XX, até a era pós-genômica molecular (Park et al., 2020). A princípio, *Aeromonas* foram caracterizadas apenas como responsáveis por doenças sistêmicas em animais poiquilotérmicos (Bhowmick; Bhattacharjee, 2018). Nos dias de hoje, essa bactéria é reconhecida não apenas como um relevante patógeno, causando doenças em peixes e outras espécies de animais de sangue frio, mas também como um agente etiológico que é capaz de causar uma série de infecções em humanos imunocompetentes, mas principalmente em pessoas debilitadas e imunocomprometidas (Janda; Abbott, 2010); (Yamasaki et al., 2020).

O Código Internacional de Nomenclatura de Bactérias (Código Bacteriológico) rege as regras para a base da nomenclatura bacteriana juntamente com alterações aprovadas pelo Comitê Internacional de Sistemática de Procariotos (Sneath, 1992); (Pallen; Telatin; Oren, 2021). Os responsáveis por questões relacionadas a taxonomia são a União Internacional de Sociedade Microbiológicas. Porém, o encaminhamento para uma proposta de publicação identificando uma nova espécie, não resulta necessariamente em validade ou exatidão (Euzéby; Tindall, 2004); (Sanford et al., 2021).

No momento da publicação, significa apenas que os requerimentos mínimos foram cumpridos no que corresponde as exigências de nomenclatura para descrever um novo táxon. Dentro destes requisitos está o nome da espécie caracterizado de forma direta, novo nome dado a etimologia, descrição das propriedades do táxon atribuídas, e uma intitulação de estirpe (Euzéby; Tindall, 2004). Estas normas de nomenclatura se aplicam somente para categorias taxonômicas, até o nível de subespécie. Não incluem por exemplo, organizações abaixo desse nível, como biovar, biogrupo, biótipo e sorotipo (Janda; Abbott, 2010); (Rosseló-Móra; Whitman, 2019).

Os microbiologistas clínicos enfrentam grandes problemas em determinar o papel de *Aeromonas* diante das doenças infecciosas e também como identificá-las. Parte destas dificuldades se devem a um legado confuso de nomes de espécies, ao desenvolvimento relativamente recente de um método para identificação fenotípica desses organismos, e o reconhecimento tardio de que várias espécies de *Aeromonas* podem causar infecções em humanos (Abbott et al., 2003); (Fernández-Bravo; Figueras, 2020). Uma das alternativas com mais chance de sucesso para tentar solucionar estes problemas é o uso de sequenciamento do genoma completo e análise de microarray com MALDI-TOF (Nash et al., 2006); (Lamy et al.,

2011); (Pessoa et al., 2019).

O sequenciamento do gene 16S rRNA é frequentemente utilizado para identificar bactérias ao nível de gênero e espécie (Janda; Abbott, 2007); (Church et al., 2020). Em relação ao gênero *Aeromonas*, esta técnica tem suas limitações. Em alguns estudos, determinaram que esta técnica é válida para identificação definitiva de algumas espécies (Hua et al., 2004) mas de outras não (Al-Benwan et al., 2007); (Fernández-Bravo et al., 2020). Os motivos desta problemática do gene 16S rRNA para o gênero das *Aeromonas* está relacionada a provável evolução dos operons *rrn* em mosaico (Morandi et al., 2005); (Navarro; Martínez-Murcia, 2018).

Foram identificadas, em várias espécies de *Aeromonas*, diferenças intragenômicas, representadas pelos polimorfismos dos nucleotídeos *rrn*. Estas alterações tiveram uma variação mínima de 0,06% e um máximo de 1,5% (Alperi et al., 2008); (Morandi et al., 2005). *Aeromonas veronii* possui cerca de 6 cópias do gene 16S rRNA, entre eles ocorre uma diferença de até 1,5%. Devido a essas diversidades presentes nas sequências, o seu uso para identificação de espécies se torna inviável. No estudo de Alperi e colaboradores, foram analisadas 999 cepas de *Aeromonas*, utilizando 16S DNA RFLP (polimorfismo de comprimento de fragmento de restrição do gene 16S rRNA). Os resultados demonstraram que cerca de 8,1% dos isolados apresentaram perfis atípicos, e não poderiam ser atribuídas a nenhuma espécie, devido a polimorfismo no gene 16S rRNA (Alperi et al., 2008).

Outra dificuldade do 16S para identificação de *Aeromonas*, pode ser exemplificada por *Aeromonas caviae* e *Aeromonas trota*, que possuem apenas 30% de parentesco ao nível de hibridização de DNA. Entretanto, suas sequências do gene 16S rRNA se diferenciam em apenas 3 nucleotídeos (Morandi et al., 2005). Também para *Aeromonas sobria* e *Aeromonas veronii* que chegam entre 60% a 65% na hibridização de DNA, porém, nas sequências do gene 16S rRNA diferem em 14 nucleotídeos (Martinez-Murcia; Benlloch; Collins, 1992). Como indicado acima, o método de identificação de espécies através do do 16S rRNA apresenta limitações para o gênero *Aeromonas*. Entretanto, a utilização dos genes “housekeeping” *rpoD*, *rpoB* e *gyrB*, são promissores para tentar resolver esta adversidade (Morandi et al., 2005); (Al-Benwan et al., 2007); (Alperi et al., 2008); (Fernández-Bravo; Figueras, 2020).

2 REVISÃO DA LITERATURA

Aeromonas são bacilos gram-negativos, citocromo-oxidase positivos, anaeróbios facultativos, fermentam glucose e não formam esporos (Parker; Shaw, 2011); (Erdem et al.,

2011); (Percival; Williams, 2014). O gênero *Aeromonas* tem sido discutido há mais de cem anos, e acredita-se que os primeiros isolados tenham sido identificados e relatados no ano de 1890 (Shubert, 1967). No decorrer dos anos, essas bactérias passaram por várias classificações e reclassificações, entre os mais diversos gêneros, como *Aerobacter*, *Pseudomonas*, *Escherichia*, *Proteus*, entre outros (Percival; Williams, 2014); (Pessoa et al., 2019).

Compondo parte da classe *Gammaproteobacteria*, ordem *Aeromonadales* e compartilhando a família *Aeromonadaceae* com outros seis gêneros (*Zobellella*; *Tolumonas*; *Pseudaeromonas*; *Oceanisphaera*; *Oceanimonas Dongshaea*), o gênero *Aeromonas*, contém atualmente 32 espécies reconhecidas: *Aeromonas allosaccharophila*; *aquática*; *australiensis*; *bestiarum*; *bivalvium*; *cavernícola*; *caviae*; *dhakensis*; *diversa*; *encheleia*; *enteropelogenes*; *eucrenophila*; *finlandiensis*; *fluvialis*; *hydrophila*; *jandaei*; *lacus*; *lusitana*; *media*; *molluscorum*; *piscícola*; *popoffii*; *rivipollensis*; *rivuli*; *salmonicida*; *sanarellii*; *schubertii*; *simiae*; *sóbria*; *taiwanensis*; *tecta*; *veronii*.
(<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=84642>).

(<https://www.ncbi.nlm.nih.gov/genome/?term=aeromonas>) último acesso 17/03/2022.

A classificação taxonômica nada mais é que o arranjo ordenado de forma hierárquica (domínio, filo, classe, ordem, família, gênero, espécie) das bactérias, com base no parentesco, analisada por várias plataformas, incluindo métodos fenotípicos e genéticos (Rosseló-Móra; Whitman, 2019). Suas relações são apresentadas visualmente pelas árvores filogenéticas, com galhos (*branches*) conectando os grupos de parentescos por linhagens ou espécies e em clados por grupos taxonômicos de ancestralidade comum (Pessoa et al., 2019).

Os processos podem passar por vários tipos de análise, até chegar uma conclusão definitiva, como análises de genes de interesse, utilização de ferramentas filogenéticas e métodos empregados para a seleção de cepas de referência para estudo e comparação. Os principais métodos para classificação são baseados no gene 16S rRNA, na identificação fenotípica, genes “*housekeeping*”, em genotipagem e análises genômicas (Fernández-bravo; Figueras, 2020).

2.1 Métodos baseados no gene 16S rRNA

Para identificação de espécies bacterinas, o gene 16S rRNA é utilizado como um marcador estável. A sua distribuição é universal permitindo assim a comparação entre os microrganismos (Figueras et al., 2011); (Soriano et al., 2014). A estrutura das regiões do gene 16S rRNA são variadas, contribuindo assim para a diferenciação de organismos estreitamente relacionados.

Além disso, essas regiões conservadas ajudam na comparação de organismos distantes, permitindo assim a construção de *primers* universais (Martínez-Murcia; Lamy, 2015).

O gene 16S de *Aeromonas* foram sequenciados primeiramente em 1992 (Martinez-Murcia; Benlloch; Collins, 1992). Isso ocorreu com as cepas das espécies que estavam descritas até aquele momento e analisadas com a hibridização DNA-DNA (Martinez-Murcia; Benlloch; Collins, 1992). O intervalo de similaridade interespecies para o gene 16S do gênero *Aeromonas* chegou a 96-100%, e que a maioria das espécies apresentavam assinaturas característica na região V3 (Martinez-Murcia; Benlloch; Collins, 1992).

O gene 16S rRNA possui limitações para identificar de forma correta as espécies do gênero *Aeromonas*. Isto ocorre devido a presença de algumas mutações em algumas das diversas cópias do gene, em posições específicas (Alperi et al., 2008); (Roger et al., 2012); (Martínez-Murcia; Lamy, 2015). Outra técnica muito utilizada para a identificação e comparação de isolados microbianos é a ionização/dessorção a laser assistida por matriz acoplada a espectrometria de massa por tempo de voo (MALDI-TOF MS), (Lamy et al., 2011); (Vávrová et al., 2015); (Fernández, 2019). Este método é capaz de identificar principalmente as proteínas que estão relacionadas ao gene 16S rRNA (Latif-Eugenín, 2015).

2.2 Métodos baseados em identificação fenotípica

Este método é baseado em características fisiológicas, morfológicas e bioquímicas (Stackebrandt et al., 2002); (Abbott; Cheung; Janda, 2003); (Chen et al., 2019). O fenótipo do gênero *Aeromonas* inclui coloração de Gram como gram-negativa, anaeróbio facultativo, citocromo oxidase positiva, fermentação da glucose (Abbott; Cheung; Janda, 2003); (Janda; Abbott, 2010). Vários testes adicionais são necessários para a sua caracterização ao nível de gênero e espécie.

Devido ao comportamento versátil das linhagens, é difícil realizar a caracterização em nível de espécie usando este método. No estudo de Beaz-Hidalgo e colaboradores, 119 cepas de *Aeromonas* previamente caracterizadas fenotipicamente foram re-identificadas com base em métodos moleculares (sequências 16S rRNA-RFLP e *rpoD*). Os resultados apontaram que somente 35,5% das cepas foram identificadas corretamente a nível de espécie (Beaz-Hidalgo et al., 2010), (Fernández-Bravo; Figueras, 2020).

2.3 Métodos baseados em genes “housekeeping”

A manutenção das funções celulares basais, sendo estas, essenciais para a sobrevivência

de um organismo é realizada pelos genes *housekeeping*, também conhecidos como genes de manutenção. Devido a esta característica, estes genes são amplamente utilizados como controles internos para estudos experimentais e computacionais (Biddanda et al., 2021). Como por exemplo, usados para análises taxonômicas, apresentando algumas características principais: devem estar presentes em todas as bactérias; não devem conter cópias no genoma; e apresentar duas regiões conservadas para poder realizar o desenho de primers (Harayama; Kasai, 2006); (Navarro; Martínez-Murcia, 2018).

Em *Aeromonas*, os genes “housekeeping” utilizados para a classificação taxonômica incluem *gyrB*, que codifica a subunidade B da DNA girase (Yanez et al., 2003); (Sun et al., 2021), e *rpoD/rpoB*, que codificam respectivamente a subunidade sigma 70 e a subunidade beta da RNA polimerase (Soler et al., 2004); (Navarro; Martínez-Murcia, 2018). Estes genes foram utilizados para identificar várias espécies (Martínez-Murcia et al., 2016); (Figueras et al., 2017). Porém, o uso de apenas um gene “housekeeping” muitas vezes não é conclusivo. O uso de vários genes concatenados pode ajudar na resolução da classificação taxonômica das bactérias do gênero *Aeromonas* (Martínez-Murcia; Lamy, 2015); (Fernández-Bravo; Figueras, 2020).

2.4 Métodos baseados em genotipagem

A partir de métodos moleculares, é possível rastrear se dois isolados de *Aeromonas* são pertencentes ou não ao mesmo clone, sendo assim, partilham de uma mesma relação epidemiológica (Fernández-Bravo; Figueras, 2020). Os principais métodos são: DNA-PCR polimórfico amplificado aleatoriamente (RAPD-PCR); eletroforese em gel de campo pulsado (PFGE); tipagem de sequência multilocus (MLST); polimorfismo de comprimento de fragmento amplificado (AFLP); PCR de consenso intergênico repetitivo enterobacteriano (ERIC-PCR) (Martínez-Murcia; Lamy, 2015).

O MLST baseia-se na análise das sequências de DNA, sendo estes fragmentos dos genes *housekeeping*, para caracterizar isolados de espécies microbianas. Uma coleção de bancos de dados (PubMLST), com curadoria, que integram dados de sequência populacional de espécies e gêneros microbianos (Jolley; Bray; Maiden, 2018), ajuda os pesquisadores a compararem os seus resultados. Os dados de *Aeromonas* foram implementados ao banco em 2010 com base nos dados obtidos por Martino e colaboradores, usando seis genes (*gyrB*, *groL*, *gltA*, *metG*, *ppsA* e *recA*) (Martino et al., 2006). Uma das possíveis limitações do uso deste método, é o número de cepas disponíveis no banco de dados, mas que vem crescendo nos últimos anos. Últimos dados

do banco (Sequências Alélicas: 4,925 – última atualização 16/03/22); (Isolados: 1,163 - última atualização 12/01/22); (Genomas: 94 – última atualização 11/10/21). Dados somente para *Aeromonas spp.* Acessado em 16/03/2022.

2.5 Métodos baseados em genômica

Entre as ferramentas baseadas na comparação do genoma para classificação taxonômica cita-se: isDDH (hibridização DNA-DNA in silico); MLPA (Amplificação de sonda dependente de ligação multiplex); ANI (Identidade média de nucleotídeos) (Meier-Kolthoff et al., 2013); (Beaz-Hidalgo et al., 2015).

O método baseado em isDDH foi descrito por Kolthoff e colaboradores. Utiliza uma calculadora de distância de genoma-a-genoma (GGDC) desenvolvida pelo DSMZ (Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany) (Meier-Kolthoff et al., 2013). Este método mostrou ser uma ferramenta útil para identificar e determinar a similaridade genética entre dois genomas bacterianos. Richter e Rosselló-Mora utilizaram ANI para analisar dois genomas e identificar sua espécie. O valor de corte estabelecido para ANI foi de 95-96%, e os resultados concordaram com os dados do isDDH (Richter; Rosselló-Mora, 2009); (Kim et al., 2014).

Beaz-Hidalgo e colaboradores, identificaram novamente 44 genomas que se encontravam depositados no banco de dados NCBI, utilizando análise de sequência multilocus de genes *housekeeping* (*gyrB*, *rpoD*, *recA*, *dnaJ*, *gyrA*, *dnaX*) e ANI. Os resultados demonstraram que 14 genomas acabaram sendo rotulados de forma errônea. Estes dados demonstram a importância de verificar e analisar a classificação taxonômica dos genomas antes de submeter para o NCBI ou outras bases de dados (Beaz-Hidalgo et al., 2015).

A identidade média de nucleotídeos (ANI), demonstra índices de similaridade entre um determinado par de genomas que pode ser aplicável a organismos procarióticos, independentemente de seu conteúdo de G+C. Se o ANI atingir uma pontuação de corte de >95%, sugerem que os genomas analisados pertencem à mesma espécie. Diferentes estudos utilizaram estes métodos para realizar a identificação e classificação taxonômica das *Aeromonas* de uma forma mais robusta, evidenciando assim, a necessidade de se fazer estudos mais aprofundados sobre o tema (Talagrand et al., 2017); (Awan et al., 2018).

2.6 Métodos baseados em genoma completo

O BLAST, um algoritmo de alinhamento, é muito utilizado para realizar comparações de sequências biológicas (Altschul et al., 1990). Alinhar grandes conjuntos de dados requer grande quantidade de tempo, e acaba se tornando inviável quando genomas completos precisam ser estudados (Vinga; Almeida. 2003). Uma das limitações desse método é quando a identidade de sequência é baixa devido as matrizes de substituição, que consequentemente, podem interferir de forma negativa nos resultados do alinhamento (Zielezinski et al., 2017).

Devido a essas restrições, se percebeu a necessidade do desenvolvimento de alternativas para aumentar a velocidade para as comparações de dados estruturados (Chen et al., 2016); (Li et al., 2017). Vários métodos livres de alinhamento foram construídos com objetivo de analisar grandes conjuntos de dados de sequência biológica, voltados para análises comparativas de genomas completos (Zielezinski et al., 2017); (Zhang et al., 2017); (Li et al., 2017); (Leimeister et al., 2019). Entretanto, estudos mais aprofundados destas técnicas ainda se tornam necessários para corroborar a sua eficácia. Sendo assim, abordagens fundamentadas no mapeamento de frequências relativas de palavras (*k-mers*) em espaços vetoriais estão sendo discutidas em vários estudos (Boden et al., 2013); (Vinga, 2014); (Nóe; Martin, 2014); (Zielezinski et al., 2017); (Leimeister et al., 2019). As representações vetoriais de proteínas ajudam durante o processamento das análises, facilitando o uso de mineração de dados para reconhecer propriedades importantes que podem estar escondidas em sequências biológicas (Asgari; Mofrad, 2015); (Leimeister et al., 2019).

3 MATERIAIS E MÉTODOS

3.1 Genomas

Ao todo, 177 genomas completos representando 15 espécies definidas foram utilizados neste estudo. São eles: *A. hydrophila* (n = 42); *A. caviae* (n = 29); *A. veronii* (n = 25); *A. salmonicida* (n = 19); *A. media* (n = 12); *A. jandaei* (n = 7); *A. dhakensis* (n = 4); *A. enteropelogenes* (n = 4); *A. schubertii* (n = 2); *A. allosaccharophila* (n = 1); *A. encheleia* (n = 1); *A. rivipollensis* (n = 1); *A. rivuli* (n = 1); *A. sanarellii* (n = 1); *A. simiae* (n = 1). Adicionalmente, genomas completos de 27 cepas identificadas apenas como *Aeromonas* sp foram incluídos neste estudo. As sequências codificantes (proteomas) foram baixadas em formato aminoácidos e nucleotídeos a partir do banco de dados NCBI, até a data de 13/01/2022. As informações

completas sobre os genomas podem ser observadas na (tabela suplementar S1).

3.2 Ferramentas para análise e visualização

As análises filogenéticas foram conduzidas por meio da ferramenta alignment-free SWeeP (DE PIERRI *et al.*, 2020). O SWeeP é um modelo computacional que pode ser aplicado para análises de machine learning de maneira geral. O modelo baseia-se na representação de sequências biológicas em vetores, o que permite a manipulação de grandes quantidades de genomas completos de maneira ágil e eficiente. Por meio de uma janela deslizante (máscara), é realizada a varredura das sequências biológicas, onde as combinações entre os pares de aminoácidos são representadas em uma matriz de tamanho 400x400. As colunas dessas matrizes são concatenadas em vetores, cada um contendo 160.000 atributos, onde cada vetor representa um proteoma.

Uma das vantagens da utilização dos vetores SWeeP para as análises filogenéticas é a possibilidade de redução de dimensionalidade, dado o tamanho da projeção definida pelo usuário. Nesta tese, utilizamos o tamanho de projeção de 1369, e a máscara '11011', conforme sugerido por Raittz *et al.*, (2021) e De Pierri *et al.*, (2022, no prelo) como sendo o melhor parâmetro para análises de proteomas bacterianos. A inferência filogenética foi realizada o modelo Neighbor-joining (SAITOU; NEI, 1987) e para a visualização das árvores foi utilizado o Dendroscope 3 (HUSON; SCORNAVACCA, 2012).

FastANI: Para o cálculo do ANI utilizamos a ferramenta chamada FastANI (Jain *et al.*, 2018). A definição deste método é calcular a identidade nucleotídica média de pares de genes ortólogos, que são compartilhados entre os genomas de estudo. Suporta análises para *draft* ou genoma completo. Para realizar suas análises, utiliza Mashmap (Jain *et al.*, 2017), responsável pelo mapeamento das sequências, sendo seu algoritmo livre de alinhamento. Fornece os valores de ANI na faixa de identidade de nucleotídeos entre 80 a 100%, sendo estipulado valores >95% pertencentes a mesma espécie. Os valores de identidade gerados foram coloridos da seguinte forma na tabela: abaixo de 95% em amarelo; acima de 95% em verde e valores igual ou superior a 99% em azul.

Dendroscope3, versão 3.8.1 (Huson; Scornavacca, 2012) foi utilizado para a visualização dos resultados das árvores filogenéticas geradas pelo SWeeP. Para edição das árvores, o programa Adobe Illustrator foi usado.

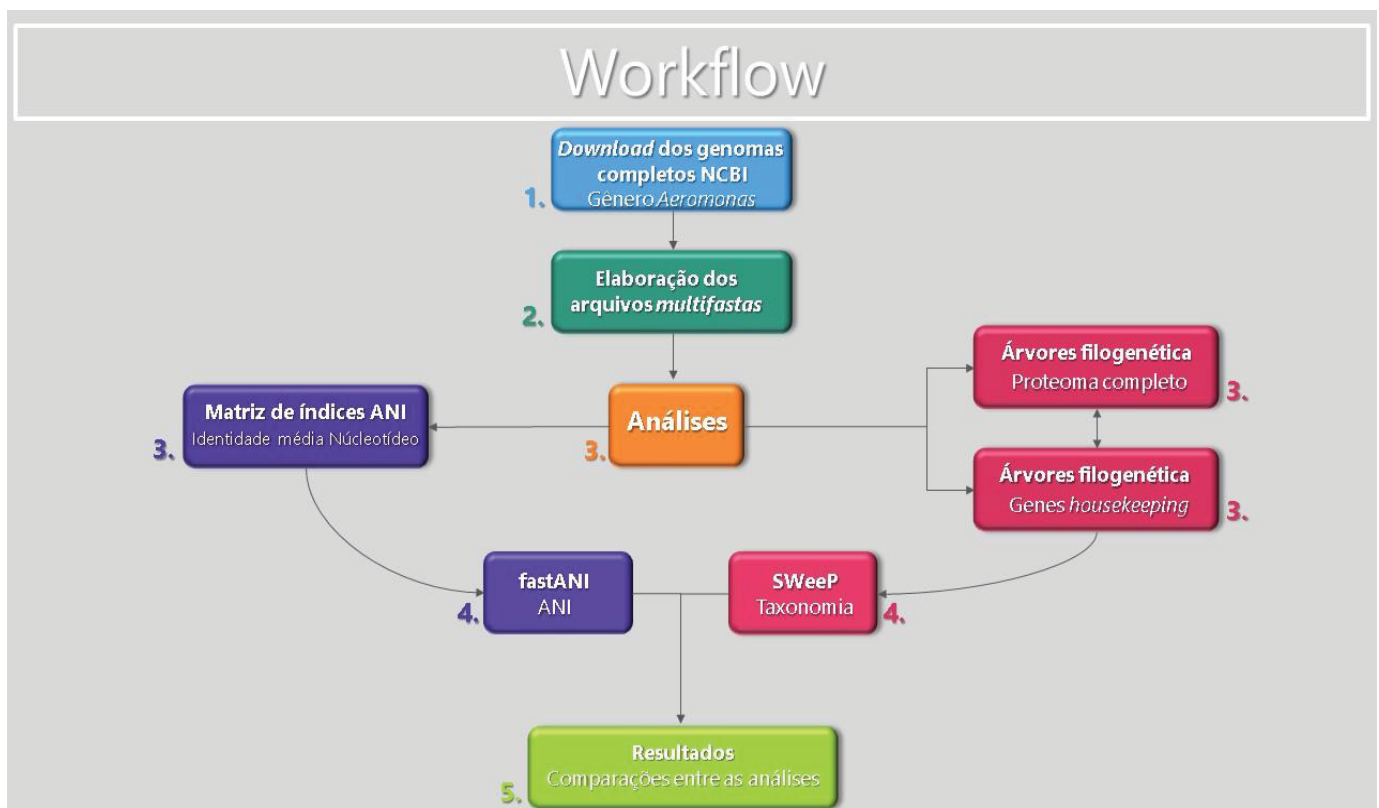
3.3 Tratamento das sequências para análise

SWeeP: Após baixadas as sequências dos proteomas, o cabeçalho (*headers*) de cada proteína foi retirado. Com isso, concatenamos todas as sequências em formato de amino ácido, montando o proteoma completo para cada organismo. Depois construímos um único arquivo, chamado de “multifasta” que possui todos os proteomas dos genomas estudados, sendo que cada um deles é separado e distinguido por “>” seguido do nome do organismo.

Genes *housekeeping*: A partir das sequências baixadas dos proteomas, localizamos os três genes de interesse *gyrB*, *rpoD*, *rpoB*. Copiamos a sua sequência em formato de aminoácido e montamos um novo “multifasta”, somente com estes genes, para análise. Os genes de cada organismo foram concatenados e separados por “>” seguido do nome do organismo.

ANI: Para análise deste método utilizamos o proteoma completo em formato de nucleotídeos. A partir das sequências baixadas, montamos um “multifasta”, retiramos novamente o cabeçalho (*headers*) das sequências e concatenamos. Cada organismo está separado por “>” seguido do nome do organismo. *Workflow* dos processos esta representado na figura 1, abaixo.

FIGURA 1. Descrição das etapas do trabalho.



FONTE: O autor (2022).

LEGENDA: (1) *Download* dos genomas completos do gênero *Aeromonas* no NCBI. (2) Elaboração dos arquivos *multifasta*. Após baixadas as sequências dos proteomas, o cabeçalho (*headers*) de cada proteína foi retirado. Com

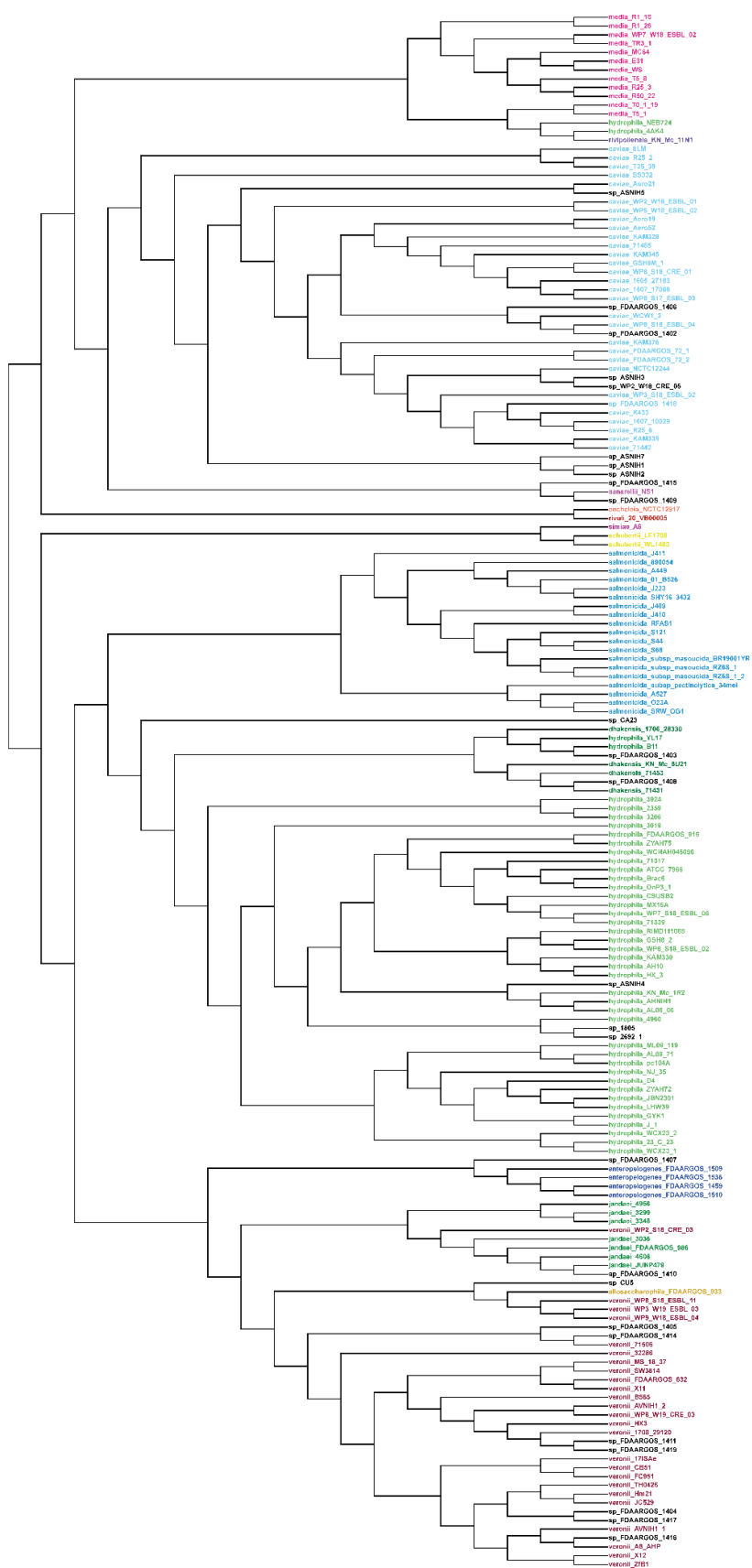
isso, concatenamos todas as sequências em formato de amino ácido, montando o proteoma completo para cada organismo. Depois construímos um único arquivo, chamado de *multifasta*, que possui todos os proteomas dos genomas estudados, sendo que cada um deles é separado e distinguido por “>” seguido do nome do organismo. Isto foi feito através de ferramenta de editor de texto. (3) Análises. Matriz de índices de ANI (identidade média dos nucleotídeos) com sequências formato de nucleotídeo (*multifasta*); SWeeP com sequências do proteoma completo em formato de aminoácido (*multifasta*) para construção de árvores filogenéticas. (4) fastANI para gerar a matriz de índices do ANI. SWeeP para gerar as árvores filogenéticas (árvore de proteomas e genes housekeeping *rpoD*, *rpoB*, *gyrB*). (5) Comparação dos resultados e visualização das árvores filogenéticas geradas, utilizando a ferramenta Dendroscope, e matriz dos índices do ANI (Planilha Excel).

4 RESULTADOS E DISCUSSÃO

Os resultados de cada análise, duas árvores filogenéticas baseadas respetivamente no proteoma (figura 2); genes housekeeping *gyrB*, *rpoD*, *rpoB* (figura 3) de *Aeromonas* geradas com o modelo SWeeP, e a matriz construída com os dados do ANI foram considerados isoladamente e então comparados entre si. A tabela dos índices de ANI, obtidos da comparação entre as *Aeromonas* e gerada pela ferramenta fastANI se encontra no material suplementar (tabela S2).

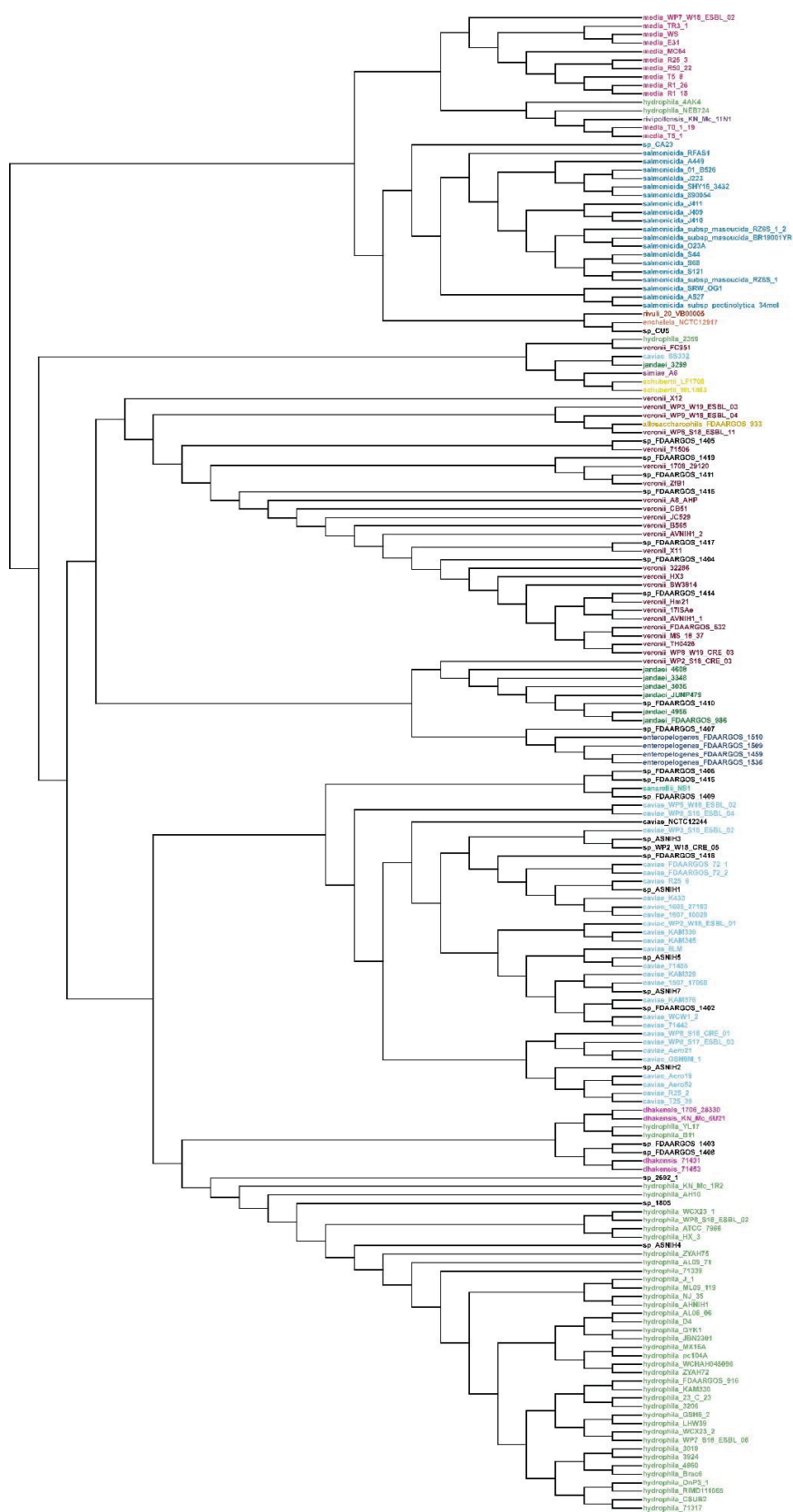
Na figura S1 abaixo, da árvore de proteoma, classifica as *Aeromonas* em dois grandes ramos filogenéticos. Um deles incluem as espécies *A. encheleia*, *A. rivuli*, *A. rivipollensis*, *A. media*, *A. caviae*, *A. sanarelli*. Outro ramo, incluem, *A. simiae*, *A. schubertii*, *A. salmonicida*, *A. dhakensis*, *A. hydrophila*, *A. enteropelogenes*, *A. jandaei*, *A. allosaccharophila*, *A. veronii*. *Aeromonas* sp. foram incluídas em várias espécies. Em alguns casos, observa-se ramos contendo espécies diferentes, o que sugere possível identificação errônea.

FIGURA 2. Árvore filogenética (proteoma)



Fonte: O Autor, 2022.

Na árvore filogenética baseada nos genes *housekeeping* *gyrB*, *rpoD*, *rpoB* (figura 3), foi separado também em dois grandes ramos, no entanto, a distribuição das espécies foi distinta em relação aos observados na análise realizada com proteoma (figura 2). Na segunda árvore, um dos grupos incluem *A. media*, *A. rivipollensis*, *A. rivuli*, *A. encheleia*, como observados na análise do proteoma. No entanto, *A. salmonicida* foi incluída neste grupo, enquanto *A. caviae* e *A. sanarelli* não fazem parte do mesmo. Este segundo grupo contém dois ramos principais, um deles agrupa *A. simiae* e *A. schubertii*, juntamente com um representante de *A. hydrophila*, *A. caviae*, *A. jandaei*, *A. veronii*. E outro contém as espécies *A. hydrophila*, *A. dhakensis*, *A. caviae*, *A. sanarelli*, *A. enteropelogenes*, *A. jandaei*, *A. veronii*.



Fonte: O Autor, 2022.

Na matriz dos índices do ANI, verifica-se um grande bloco contendo *A. hydrophila* com valores de ANI superiores a 95%, exceto para as cepas YL17, 4AK4, B11, NEB724. Outros blocos contêm respectivamente todas as cepas de *A. caviae*, *A. salmonicida*, *A. jandaei*, *A. dhakensis*, *A. enteropelogenes*, *A. schubertii*. Entre *A. veronii*, 20 cepas apresentam ANI superior a 95%, ao passo que as cepas (*A. veronii* 71506, *A. veronii* WP2 S18 CRE 03, *A. veronii* WP3 W19 ESBL 03, *A. veronii* WP8 S18 ESBL 11, *A. veronii* WP9 W18 ESBL 04), tem ANI inferior, sugerindo que não pertencem a esta espécie. Para *A. media*, apenas duas cepas (*A. media* T0 1 19 e *A. media* T5 1), apresentam ANI inferior aqueles observados para as demais cepas. Outras 6 espécies têm apenas 1 representante. Entre as 27 cepas de *Aeromonas* identificados apenas a nível de gênero, apenas 3 (*Aeromonas* sp CA23, *Aeromonas* sp CU5, *Aeromonas* sp FDAARGOS 1415) não foram atribuídas entre algumas das espécies analisadas. A classificação das espécies de *Aeromonas* com base nos resultados da matriz do índice ANI é similar a distribuição observada na árvore filogenética baseada em proteoma.

Após essa primeira análise, os resultados indicados nas duas árvores (figura 2 árvore proteoma, (figura 3, árvore genes *housekeeping*) e na (tabela suplementar S2) foram comparados entre si considerando cada espécie individualmente. Os resultados estão indicados a seguir.

4.1 *Aeromonas media* (12 cepas)

Para esta espécie as duas árvores (proteoma e genes *housekeeping*) apresentaram algumas diferenças em relação as espécies mais próximas (como discutido acima). Embora, todas as cepas estejam no mesmo ramo, existem dois subgrupos. Um deles contém apenas as cepas *A. media* T0 1 19, e T5 1; estas estão agrupadas com *A. hydrophila* 4AK4, *A. hydrophila* NEB724, *A. rivipollensis* KN Mc 11N1. Os valores de ANI são superiores a 95% entre as cepas *A. media* T0 1 19, e T5 1; mas são inferiores a este valor quando estas são comparadas as demais *A. media*. Isto sugere, que estas duas cepas pertencem a uma espécie diferente. Como mostrado na (tabela S2), estas cepas apresentam ANI superiores a 95% com *A. hydrophila* 4AK4, *A. hydrophila* NEB724, *A. rivipollensis* KN Mc 11N1, com as quais possivelmente constituem uma única espécie.

TABELA 1. Comparação dos valores de ANI entre espécies, *A. hydrophila*, *A. media*, *A. rivipollensis*.

Bactérias	A.h 4AK4	A.h NEB724	A.m T0 1 19	A.m T5 1	A.r KN Mc 11N1
A.h 4AK4		96.03%	95.79%	95.72%	95.28%
A.h NEB 724	96.03%		96.60	96.60%	96.69%
A.m T0 1 19	95.79%	96.60%		99.95%	96.27%
A. m T5 1	95.72%	96.60%	99.95%		96.25%
A. r KN Mc 11N1	95.28%	96.69%	96.27%	96.25%	
A. m E31	93.36%	93.30%	94.05%	94.03%	93.32%

FONTE: O autor (2022).

LEGENDA: A.h (*Aeromonas hydrophila*); A.m (*Aeromonas media*); A.r (*Aeromonas rivipollensis*); os valores de ANI entre cada cepa estão indicados em porcentagem.

4.2 *Aeromonas hydrophila* (42 cepas)

Nestas espécies também houve diferenças entre a classificação taxonômica em relação as espécies mais próximas. A principal diferença foi em relação à *A. salmonicida* incluída no mesmo ramo na análise pelo proteoma, e excluída na análise pelos genes housekeeping. Por sua vez, nesta última análise *A. caviae* foi incluída no mesmo ramo que *A. hydrophila*.

Outra diferença entre as duas árvores filogenéticas, foi a classificação da cepa *A. hydrophila* 2359 que foi incluída na análise pelos genes housekeeping juntamente com outras espécies distintas (*A. veronii* FC951, *A. caviae* SS332, *A. jandaei* 3299, *A. simiae* A6, *A. schubertti* LF1708, *A. schubertti* WL1483). Os valores de ANI para *A. hydrophila* 2359 não corroboram com esta classificação, e indicam que árvore filogenética baseada na análise do proteoma apresenta maior confiabilidade.

Na (tabela S2) observa-se que os valores de ANI são superiores a 95% frente a grande maioria das *A. hydrophila*, exceto pelas cepas previamente identificadas como *A. hydrophila* 4AK4, *A. hydrophila* NEB724, que foram agrupadas com *A. media*; e *A. hydrophila* YL17 e *A. hydrophila* B11, que foram agrupadas com *A. dhakensis* (figuras 2 e 3, tabela S2). Estas cepas apresentaram valores de ANI inferiores a 95% em relação as demais *A. hydrophila* (tabela S2).

Nas duas árvores, um total de três *Aeromonas* sp 1805; 2692 1 e ASNIH4 estão agrupadas com a espécie *A. hydrophila*. Todas apresentam ANI acima de 95% quando comparados com as bactérias da espécie *A. hydrophila*. Os resultados de ANI dessas cepas comparadas a *A. hydrophila* ATCC 7966 estão descritos na tabela abaixo.

TABELA 2. Valores de ANI obtidos para *A. hydrophila* ATCC 7966 e *Aeromonas* sp.

Bactérias	<i>A. hydrophila</i> ATCC 7966
<i>A. hydrophila</i> 2359	96.64%
<i>Aeromonas</i> sp 1805	96.88%
<i>Aeromonas</i> sp 2692 1	96.90%
<i>Aeromonas</i> sp ASNIH4	96.20%

FONTE: O Autor (2022)

LEGENDA: Valores de ANI da espécie *A. hydrophila* contra *Aeromonas* sp.

Portanto, entre as bactérias depositadas no NCBI como *A. hydrophila* existem cepas erroneamente identificadas; entre as *Aeromonas* sp., três foram incluídas nesta espécie.

4.3 *Aeromonas caviae* (29 cepas)

A principal diferença entre as árvores é *A. caviae* SS332. Na árvore gerada com dados do proteoma ela se manteve com as outras cepas da mesma espécie, com as quais compartilha valores de ANI acima de 95% (Tabela 3, Tabela S2). Entretanto, na árvore dos genes *housekeeping* está agrupada com (*A. veronii* FC951, *A. caviae* SS332, *A. jandaei* 3299, *A. simiae* A6, *A. schubertii* LF1708, *A. schubertii* WL1483), em um ramo distinto.

Nas duas árvores, nove *Aeromonas* sp, foram agrupadas com *A. caviae*. Os valores de ANI dessas cepas são superiores a 95% quando comparadas com *A. caviae*, indicando que pertencem a essa espécie. A tabela 3, mostra os valores de ANI dessas bactérias em comparação com *A. caviae* 8LM.

TABELA 3. Comparação de *A. caviae* 8LM com *A. caviae* SS332 e *Aeromonas* sp. usando ANI

Bactéria	<i>A. caviae</i> 8LM
<i>A. caviae</i> SS332	96.88%
<i>Aeromonas</i> sp ASNIH3	97.78%
<i>Aeromonas</i> sp ASNIH5	97.05%
<i>Aeromonas</i> sp FDAARGOS_1402	97.79%
<i>Aeromonas</i> sp FDAARGOS_1406	96.82%
<i>Aeromonas</i> sp FDAARGOS_1418	96.96%
<i>Aeromonas</i> sp WP2_W18_CRE_05	97.04%
<i>Aeromonas</i> sp ASNIH1	97.75%
<i>Aeromonas</i> sp ASNIH2	97.34%
<i>Aeromonas</i> sp ASNIH7	97.05%

FONTE: O Autor (2022)

LEGENDA: Porcentagem dos valores de ANI entre espécies *A. caviae* e *Aeromonas* sp.

4.4 *Aeromonas veronii* (25 cepas)

Entre as 25 cepas previamente identificadas como *A. veronii* no NCBI, 20 apresentaram

valores de ANI entre si superiores a 95%. Entre as outras cinco cepas, *A. veronii* WP2 S18 CRE 03 foi agrupada como *A. jandaei* em ambas as árvores apresentando ANI de 96.12% com *A. jandaei* 3036 e acima de 95% com as outras cepas dessa espécie, indicando que foi erroneamente identifica como *A. veronii*.

As cepas de *A. veronii* WP3 W19 ESBL 03, WP9 W18 ESBL 04 e WP8 S18 ESBL 11 não se encontram no ramo principal com as demais cepas da espécie nas duas árvores analisadas. Além disso, os seus valores de ANI são inferiores a 95% quando comparadas com as demais *A. veronii* (Tabela S2). Os resultados das três análises, indicam que essas três cepas estão relacionadas a espécie *A. allosaccharophila*.

Finalmente, *A. veronii* 71506 foi agrupada com *Aeromonas* sp FDAARGOS 1405, *Aeromonas* sp FDAARGOS 1414 na árvore baseada proteoma e apenas com a *Aeromonas* sp FDAARGOS 1405 na árvore baseada em genes *housekeeping*. Em relação a análise do ANI, *A. veronii* 71506 não atingi valores suficientes para ser considerada como membro dessa espécie, e valores superiores a 95% de ANI são observados apenas com *Aeromonas* sp FDAARGOS 1405, *Aeromonas* sp FDAARGOS 1414.

Na árvore gerada com os genes *housekeeping*, *A. veronii* FC951 se encontra junto com (*A. caviae* SS332, *A. jandaei* 3299, *A. simiae* A6, *A. schubertti* LF1708, *A. schubertti* WL1483), em um ramo distinto. Porém, na árvore gerada com o proteoma (figura 2) a cepa *A. veronii* FC951 se encontra com as demais da mesma espécie, com as quais apresenta valor de ANI maiores que 95%, usando a *A. veronii* AVNIH1 1 de exemplo para comparação na (tabela 4).

Cinco *Aeromonas* sp estão agrupadas com *A. veronii* nas duas árvores e contém valor de ANI, superior a 95%. Na tabela abaixo mostra os valores de ANI dessas *Aeromonas* sp usando a *A. veronii* AVNIH1 1 para comparação.

TABELA 4. Comparação dos valores de ANI de *A. veronii*; *Aeromonas* sp; *A. jandaei*; *A. allosaccharophila*

Bactérias	A. j 3036	A. all FDAARGOS 933	A. v AVNIH1 1
A. v WP2 S18 CRE 03	96.12%	90.09%	90.67%
A. v WP3 W19 ESBL 03	90.12%	96.05%	93.53%
A. v WP8 S18 ESBL 11	90.12%	95.81%	93.20%
A. v WP9 W18 ESBL 04	90.02%	96.00%	93.55%
A. sp FDAARGOS 1404	90.59%	86.00%	96.02%
A. sp FDAARGOS 1411	90.46%	93.11%	95.41%
A. sp FDAARGOS 1416	90.80%	93.67%	96.13%
A. sp FDAARGOS 1417	90.42%	93.73%	96.12%
A. sp FDAARGOS 1419	90.48%	93.09%	95.31%
A. v FC951	90.08%	93.06%	95.26%

FONTE: O Autor (2022).

LEGENDA: A. v (*Aeromonas veronii*); A. sp (*Aeromonas* sp); A. j (*Aeromonas jandaei*); A. all (*Aeromonas allosaccharophila*); valores de ANI estão mostrados em porcentagem.

4.5 *Aeromonas salmonicida* (19 cepas)

De acordo (Abbott et al., 2003), com base em diversos ensaios fenotípicos, *A. salmonicida* é parte do complexo *A. hydrophila*, juntamente com *A. bestiarum*. Essas espécies estão em ramos distinto de acordo com análises filogenéticas utilizando *gyrb* e *rpoD* (Soler et al., 2004). No presente estudo, foram observadas diferenças na classificação taxonômica de *A. salmonicida* entre as duas árvores filogenéticas (figura 2 e 3). Na árvore baseada no proteoma, as cepas de *A. salmonicida* estão próximas de *A. hydrophila*, o que está de acordo com os dados de Abbott e colaboradores. Já na árvore baseado nos genes *housekeeping*, *A. salmonicida* se encontra no grupo com *A. media* e distante de *A. hydrophila*. De acordo com Abbott e colaboradores, *A. media*, *A. eucrenophila* e *A. caviae* constituem o complexo *A. caviae*. Portanto, observa-se um conflito entre as duas metodologias, proteoma e genes *housekeeping*. No entanto, todas as cepas da espécie estão agrupadas no mesmo ramo, independente da metodologia e apresentam valores de ANI superiores a 95%.

4.6 *Aeromonas jandaei* (7 cepas)

As cepas de *A. jandaei* se encontram agrupadas pela análise do proteoma. Já na árvore baseada em genes *housekeeping*, a cepa *A. jandaei* 3299 se encontra distante das demais *A. jandaei* e agrupadas com *A. veronii* FC951, *A. caviae* SS332, *A. simiae* A6, *A. schubertti* LF1708, *A. schubertti* WL1483. Além disso, *A. veronii* WP2 S18 CRE 03, esta agrupada com a espécie *A. jandaei* em ambas as árvores, e os valores de ANI estão indicados abaixo. Os resultados indicam que esta cepa foi erroneamente identificada. A cepa identificada apenas a nível de gênero,

Aeromonas sp FDAARGOS 1410, também pode ser classificada dentro da espécie *A. jandaei*, de acordo com os dados filogenéticos (figura 2 e 3) e de ANI (tabela 5 e S2).

TABELA 5. *A. jandaei* em comparação com espécies *A. veronii* e *A. sp.*

Bactérias	<i>A. jandaei</i> 3036
<i>A. jandaei</i> 3299	95.85%
<i>A. veronii</i> WP2 S18 CRE 03	96.12%
<i>Aeromonas</i> sp FDAARGOS 1410	96.36%

FONTE: O autor (2022).

LEGENDA: Comparação dos valores de ANI entre espécies.

4.7 *Aeromonas dhakensis* (4 cepas)

A distribuição de *A. dhakensis* foi similar nas duas árvores, e os seus valores de ANI entre as 4 cepas foram superiores a 95%. No entanto, observa-se nas duas árvores filogenéticas que as cepas *A. hydrophila* B11, *A. hydrophila* YL17 estão agrupadas com *A. dhakensis*, com as quais apresentam ANI superior a 95%. Também, duas cepas identificadas apenas a nível de gênero (*Aeromonas* FDAARGOS 1403 e *Aeromonas* sp FDAARGOS 1408) podem ser incluídas dentro desta espécie, de acordo com as análises filogenéticas e resultados de ANI. Os valores de ANI entre essas bactérias e *A. dhakensis* estão descritos na tabela abaixo.

TABELA 6. Comparação ANI entre espécie *A. hydrophila* e *A. dhakensis*

Bactérias	A.h B11	A.h YL17	A. dha 71431	A. dha 71453
A.h B11		97.29%	96.90%	97.03%
A.h YL17	97.29%		96.94%	96.92%
A. sp FDAARGOS 1403	97.15%	97.24%	96.87%	96.83%
A. sp FDAARGOS 1408	96.96%	96.94%	98.23%	97.83%

FONTE: O Autor (2022).

LEGENDA: A.h (*Aeromonas hydrophila*); A. dha (*Aeromonas dhakensis*). A. sp (*Aeromonas* sp); valores em porcentagem indicam o resultado do ANI.

4.8 *Aeromonas enteropelogenes* (4 cepas)

Em ambas as árvores filogenéticas, os representantes desta espécie ficaram agrupados entre si, juntamente com *Aeromonas* sp FDAARGOS 1407. Todas estas bactérias apresentam, quando comparadas entre si, valores de ANI acima de 95%, o que suporta a inclusão da cepa *Aeromonas* sp FDAARGOS 1407 dentro da espécie *A. enteropelogenes*.

TABELA 7. Valores de ANI de *Aeromonas* sp FDAARGOS 1407 comparados com *A. enteropelogenes*

Bactérias	<i>Aeromonas</i> sp FDAARGOS 1407
<i>A. enteropelogenes</i> FDAARGOS_1459	96.54%
<i>A. enteropelogenes</i> FDAARGOS_1509	96.87%
<i>A. enteropelogenes</i> FDAARGOS_1510	96.45%
<i>A. enteropelogenes</i> FDAARGOS_1536	96.47%

FONTE: O autor (2022).

LEGENDA: Valores de ANI para *Aeromonas* sp contra *A. enteropelogenes*; valores em porcentagem indicam o resultado do ANI.

4.9 *Aeromonas schubertii* (2 cepas)

As duas árvores apresentaram os mesmos resultados, com as duas cepas agrupadas e valores de ANI superiores a 95%. A estirpe *A. simiae* A6 foi agrupada próxima de *A. schubertii*, no entanto elas representam espécies distintas em relação aos valores de ANI.

4.10 *Aeromonas simiae* (1 cepa)

A única cepa dessa bactéria se encontra próxima da espécie *A. schubertii*, nas duas árvores filogenéticas. Porém, elas representam espécies distintas, e valores de ANI inferiores 95%.

4.11 *Aeromonas allosaccharophila* (1 cepa)

Resultados apresentados juntamente com *A. veronii*.

4.12 *Aeromonas encheleia* (1 cepa)

Na árvore dos genes *housekeeping* esta bactéria está agrupada com *A. rivuli* 20 VB00005 e *Aeromonas* sp CU5. Enquanto na árvore do proteoma, esta agrupada com *A. rivuli* 20 VB00005. No entanto, com os resultados de ANI, indicam valores inferiores a 95% entre elas. Portanto, pertencem a espécies distintas.

4.13 *Aeromonas rivipollensis* (1 cepa)

Resultados apresentados junto com *A. media*.

4.14 *Aeromonas sanarellii* (1 cepa)

Existe apenas um representante desta espécie, que em ambas as árvores está agrupada com *Aeromonas* sp FDAARGOS 1409. Os valores de ANI entre estas duas bactérias são de

96.98%, indicando que pertencem a mesma espécie.

4.15 *Aeromonas rivuli* (1 cepa)

Resultados apresentados juntos com *A. encheleia*.

5 CONCLUSÕES

No presente estudo, três métodos (proteoma, genes *housekeeping*, ANI) distintos foram empregados para verificar a classificação taxonômica de 177 cepas do gênero *Aeromonas*, depositadas no NCBI. Embora esses métodos tenham apresentado resultados parecidos, a maior similaridade foi observada entre informações da árvore filogenética baseada em proteoma e o ANI. Foi possível determinar que 11 bactérias estão erroneamente identificadas, descritas na tabela abaixo.

TABELA 8. Bactérias classificadas erroneamente a nível de espécie no NCBI.

NCBI	Espécie sugerida	Método de análise
<i>A. media</i> T0 1 19	<i>A. rivipollensis</i> #	Proteoma; <i>housekeeping</i> ; ANI
<i>A. media</i> T5 1	<i>A. rivipollensis</i> #	Proteoma; <i>housekeeping</i> ; ANI
<i>A. hydrophila</i> 4AK4	<i>A. rivipollensis</i> #	Proteoma; <i>housekeeping</i> ; ANI
<i>A. hydrophila</i> NEB724	<i>A. rivipollensis</i> #	Proteoma; <i>housekeeping</i> ; ANI
<i>A. veronii</i> 71506	ND*	Proteoma; ANI
<i>A. hydrophila</i> YL17	<i>A. dhakensis</i>	Proteoma; <i>housekeeping</i> ; ANI
<i>A. hydrophila</i> B11	<i>A. dhakensis</i>	Proteoma; <i>housekeeping</i> ; ANI
<i>A. veronii</i> WP2 S18 CRE 03	<i>A. jandaei</i>	Proteoma; <i>housekeeping</i> ; ANI
<i>A. veronii</i> WP3 W19ESBL 03	<i>A. allosaccharophila</i>	Proteoma; <i>housekeeping</i> ; ANI
<i>A. veronii</i> WP9 W18ESBL 04	<i>A. allosaccharophila</i>	Proteoma; <i>housekeeping</i> ; ANI
<i>A. veronii</i> WP8 S18 ESBL 11	<i>A. allosaccharophila</i>	Proteoma; <i>housekeeping</i> ; ANI

FONTE: O Autor (2022).

LEGENDA: #Grupo distinto de todas as outras, mas agrupadas com *A. rivipollensis* KN Mc 11N1, única cepa dessa espécie com genoma completo, apresentando um ANI superior a 95%. *Não foi possível definir espécie, uma vez que agrupou com *Aeromonas* sp e não tem valor de ANI compatível com outras espécies; agrupadas somente com *Aeromonas* sp FDAARGOS 1405, *Aeromonas* sp FDAARGOS 1414.

Entre as 27 cepas identificadas como *Aeromonas* sp., 22 foram identificadas e atribuídas para alguma espécie; 3 não tiveram nenhum resultado significativo entre as análises dos três métodos; e 2 se agruparam com *A. veronii* 71506 que apresenta ANI inferior a 95% com as demais *A. veronii* (resultado descrito na tabela 8, acima).

TABELA 9. Resultado da nova classificação de *Aeromonas* sp.

Aeromonas sp NCBI	Nova classificação	Método de análise
A. sp 1805	<i>A. hydrophila</i>	Proteoma; <i>housekeeping</i> ; ANI
A. sp 2692 1	<i>A. hydrophila</i>	Proteoma; <i>housekeeping</i> ; ANI
A. sp ASNIH4	<i>A. hydrophila</i>	Proteoma; <i>housekeeping</i> ; ANI
A. sp ASNIH3	<i>A. caviae</i>	Proteoma; <i>housekeeping</i> ; ANI
A. sp ASNIH5	<i>A. caviae</i>	Proteoma; <i>housekeeping</i> ; ANI
A. sp FDAARGOS 1402	<i>A. caviae</i>	Proteoma; <i>housekeeping</i> ; ANI
A. sp FDAARGOS 1406	<i>A. caviae</i>	Proteoma; <i>housekeeping</i> ; ANI
A. sp FDAARGOS 1418	<i>A. caviae</i>	Proteoma; <i>housekeeping</i> ; ANI
A. sp WP2 W18 CRE 05	<i>A. caviae</i>	Proteoma; <i>housekeeping</i> ; ANI
A. sp ASNIH1	<i>A. caviae</i>	Proteoma; <i>housekeeping</i> ; ANI
A. sp ASNIH2	<i>A. caviae</i>	Proteoma; <i>housekeeping</i> ; ANI
A. sp ASNIH7	<i>A. caviae</i>	Proteoma; <i>housekeeping</i> ; ANI
A. sp FDAARGOS 1404	<i>A. veronii</i>	Proteoma; <i>housekeeping</i> ; ANI
A. sp FDAARGOS 1411	<i>A. veronii</i>	Proteoma; <i>housekeeping</i> ; ANI
A. sp FDAARGOS 1416	<i>A. veronii</i>	Proteoma; <i>housekeeping</i> ; ANI
A. sp FDAARGOS 1417	<i>A. veronii</i>	Proteoma; <i>housekeeping</i> ; ANI
A. sp FDAARGOS 1419	<i>A. veronii</i>	Proteoma; <i>housekeeping</i> ; ANI
A. sp FDAARGOS 1410	<i>A. jandaei</i>	Proteoma; <i>housekeeping</i> ; ANI
A. sp FDAARGOS 1403	<i>A. dhakensis</i>	Proteoma; <i>housekeeping</i> ; ANI
A. sp FDAARGOS 1408	<i>A. dhakensis</i>	Proteoma; <i>housekeeping</i> ; ANI
A. sp FDAARGOS 1407	<i>A. enteropelogenes</i>	Proteoma; <i>housekeeping</i> ; ANI
A. sp FDAARGOS 1409	<i>A. sanarellii</i>	Proteoma; <i>housekeeping</i> ; ANI

FONTE: O autor (2022).

LEGENDA: A tabela mostra a classificação atual das bactérias no NCBI e a classificação proposta com base nas análises realizadas nesse trabalho. Das 27 *Aeromonas* sp. apenas três não foram identificadas a nível de espécie (*Aeromonas* sp CU51, *Aeromonas* sp CA231, *Aeromonas* sp FDAARGOS 14151).

Estes dados mostram que por mais que os bancos de dados de referência (NCBI) tenham algumas etapas para conferência dos dados, ainda acabam ocorrendo alguns equívocos na hora do depósito dos genomas. Isso pode ocorrer por dois motivos:

1. Em decorrência de utilização de metodologias inadequadas para a identificação de *Aeromonas*, bem como a falta de revisão dos genomas anteriormente ao depósito nos bancos de dados públicos.
2. Ineficácia (ou inexistência) de métodos validação no *pipeline* no banco de dados para a confirmação de gênero e espécie. A conferência automatizada dos dados depositados auxiliaria a comunidade científica de maneira geral, além de impulsionar o desenvolvimento de novos métodos para ajudar na resolução desta problemática.

O estudo de Beaz-Hidalgo et al., (2015) ap resenta um exemplo clássico, onde foram analisadas as espécies do gênero *Aeromonas* por meio de ANI e MLST. Foram reclassificadas diversas cepas, entre elas 4AK4 e YL17, entretanto, o banco de dados NCBI as mantém identificadas como *A. hydrophila*, erroneamente até o momento (fato que também evidenciamos no presente trabalho).

É importante ressaltar que várias dúvidas e questionamentos sobre os dados depositados

e publicados ainda ficam em aberto, sendo ainda necessário comparar resultados entre diferentes abordagens para chegar a melhor conclusão possível. Mesmo diante de trabalhos futuros, neste estudo foi possível constatar que a utilização de um único método para a identificação do gênero *Aeromonas* pode levar a conclusões precipitadas.

REFERÊNCIAS

- ABBOTT, S. L., CHEUNG, W. K., JANDA, J. M. The genus *Aeromonas*: biochemical characteristics, atypical reactions, and phenotypic identification schemes. **Journal of clinical microbiology**, 41(6), 2348-2357. 2003.
- ABBOTT, S.L. *Aeromonas*. In: Baron EJ, Pfaller MA, Jorgensen JH, Tenover FC, editors. **Manual of Clinical Microbiology**. 8th ed. Washington DC: ASM Press; 2010. pp. 701–5.
- AL-BENWAN, K.; ABBOTT, S.; JANDA, J. M.; HUYS, G.; ALBERT, M. J. Cystitis caused by *Aeromonas caviae*. **Journal of Clinical Microbiology**, v. 45, n. 7, p. 2348–2350, 2007.
- AL-JABRI, Z., ZAMUDIO, R., HORVATH-PAPP, E., RALPH, J. D., AL-MUHARRAMI, Z., RAJAKUMAR, K., & OGGIONI, M. R. Integrase-controlled excision of metal-resistance genomic islands in *Acinetobacter baumannii*. **Genes**, 9(7), 366. 2018.
- AL-JASSIM, N.; ANSARI, M. I.; HARB, M.; HONG, P. Y. Removal of bacterial contaminants and antibiotic resistance genes by conventional wastewater treatment processes in Saudi Arabia: Is the treated wastewater safe to reuse for agricultural irrigation? **Water Research**, v. 73, p. 277–290, 2015. Elsevier Ltd.
- ALPERI, A.; FIGUERAS, M. J.; INZA, I.; MARTÍNEZ-MURCIA, A. J. Analysis of 16S rRNA gene mutations in a subset of *Aeromonas* strains and their impact in species delineation. **International Microbiology**, v. 11, n. 3, p. 185–194. 2008
- ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W., LIPMAN, D. J. Basic local alignment search tool. **Journal of molecular biology**, 215(3), 403-410. 1990.
- ALTSCHUL, S. F.; MADDEN, T. L.; SCHÄFFER, A. A.; et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Oxford University Press**, 1997.
- ANDERSSON, D. I.; HUGHES, D. Microbiological effects of sublethal levels of antibiotics. **Nature Reviews Microbiology**, 2014.
- ARAUJO, R. M., PARES, R., LUCENA, F. The effect of terrestrial effluents on the incidence of *Aeromonas* spp. in coastal waters. **Journal of Applied Bacteriology**, 69(3), 439-444. 1990.
- ARAVENA-ROMÁN, M.; BEAZ-HIDALGO, R.; INGLIS, T. J. J.; et al. *Aeromonas australiensis* sp. nov., isolated from irrigation water. **International Journal of Systematic and Evolutionary Microbiology**, v. 63, n. PART6, p. 2270–2276, 2013.
- ARAVENA-ROMÁN, M.; INGLIS, T. J. J.; HENDERSON, B.; RILEY, T. V.; CHANG, B. J. Antimicrobial susceptibilities of *Aeromonas* strains isolated from clinical and environmental sources to 26 antimicrobial agents. **Antimicrobial Agents and Chemotherapy**, v. 56, n. 2, p. 1110–1112, 2012.
- ARVEY, A. J.; AZAD, R. K.; RAVAL, A.; LAWRENCE, J. G. Detection of genomic islands via segmental genome heterogeneity. **Nucleic Acids Research**, v. 37, n. 16, p. 5255–5266, 2009.
- ASGARI, E., MOFRAD, M. R. Continuous distributed representation of biological sequences for deep proteomics and genomics. **PloS one**, 10(11), e0141287. 2015.
- ASSIS, F. E., WOLF, S., SUREK, M., DE TONI, F., SOUZA, E. M., PEDROSA, F. O., ... & FADEL-

- PICHETH, C. M. Impact of *Aeromonas* and diarrheagenic *Escherichia coli* screening in patients with diarrhea in Paraná, southern Brazil. **The Journal of Infection in Developing Countries**, 8(12), 1609-1614. 2014.
- AWAN, F., DONG, Y., LIU, J., WANG, N., MUSHTAQ, M. H., LU, C., LIU, Y. Comparative genome analysis provides deep insights into *Aeromonas hydrophila* taxonomy and virulence-related factors. **BMC genomics**, 19(1), 1-18. 2018.
- AZAD, R. K.; LAWRENCE, J. G. Towards more robust methods of alien gene detection. **Nucleic Acids Research**, v. 39, n. 9, 2011.
- BALALI, G. I., YAR, D. D., AFUA DELA, V. G., ADJEI-KUSI, P. Microbial contamination, an increasing threat to the consumption of fresh fruits and vegetables in today's world. **International Journal of Microbiology**, 2020. 2020.
- BANERJEE, P. Discovering genomic islands using DNA sequence embedding (Doctoral dissertation, Iowa State University). 2021.
- BARCELLOS, F. G.; MENNA, P.; BATISTA, J. S. D. S.; HUNGRIA, M. Evidence of horizontal transfer of symbiotic genes from a *Bradyrhizobium japonicum* inoculant strain to indigenous diazotrophs *Sinorhizobium* (Ensifer) *fredii* and *Bradyrhizobium elkanii* in a Brazilian savannah soil. **Applied and Environmental Microbiology**, v. 73, n. 8, p. 2635–2643, 2007.
- BATRA, P.; MATHUR, P.; MISRA, M. C. *Aeromonas* spp.: An Emerging Nosocomial Pathogen. **Journal of Laboratory Physicians**, v. 8, n. 01, p. 001–004, 2016. Georg Thieme Verlag KG.
- BEATSON, S. A., DAS GRAÇAS DE LUNA, M., BACHMANN, N. L., ALIKHAN, N. F., HANKS, K. R., SULLIVAN, M. J. HENDERSON, I. R. Genome sequence of the emerging pathogen *Aeromonas caviae*. **Journal of Bacteriology**, 193(5), 1286-1287. 2011.
- BEAZ-HIDALGO, R., HOSSAIN, M. J., LILES, M. R., & FIGUERAS, M. J. Strategies to avoid wrongly labelled genomes using as example the detected wrong taxonomic affiliation for *Aeromonas* genomes in the GenBank database. **PloS one**, 10(1), e0115813. 2015.
- BEAZ-HIDALGO, R.; ALPERI, A.; BUJÁN, N.; ROMALDE, J. L.; FIGUERAS, M. J. Comparison of phenotypical and genetic identification of *Aeromonas* strains isolated from diseased fish. **Systematic and Applied Microbiology**, v. 33, n. 3, p. 149–153, 2010.
- BEAZ-HIDALGO, R.; ALPERI, A.; FIGUERAS, M. J.; ROMALDE, J. L. *Aeromonas piscicola* sp. nov., isolated from diseased fish. **Systematic and Applied Microbiology**, v. 32, n. 7, p. 471–479, 2009.
- BEAZ-HIDALGO, R.; FIGUERAS, M. J. *Aeromonas* spp. whole genomes and virulence factors implicated in fish disease. **Journal of Fish Diseases**, Apr. 2013.
- BELETE, T. M. Novel targets to develop new antibacterial agents and novel alternatives to antibacterial agents. **Human Microbiome Journal**, 11, 100052. 2019.
- BELLO-LÓPEZ, J. M., CABRERO-MARTÍNEZ, O. A., IBÁÑEZ-CERVANTES, G., HERNÁNDEZ-CORTEZ, C., PELCASTRE-RODRÍGUEZ, L. I., GONZALEZ-AVILA, L. U., CASTRO-ESCARPULLI, G. Horizontal gene transfer and its association with antibiotic resistance in the genus *Aeromonas* spp. **Microorganisms**, 7(9), 363. 2019.
- BENTLEY, S. D.; PARKHILL, J. Comparative genomic structure of prokaryotes. **Annual Review of Genetics**, 2004.

- BERG, K. A.; LYRA, C.; NIEMI, R. M.; et al. Virulence genes of *Aeromonas* isolates, bacterial endotoxins and cyanobacterial toxins from recreational water samples associated with human health symptoms. **Journal of Water and Health**, v. 9, n. 4, p. 670–679, 2011.
- BERG, K. A.; LYRA, C.; SIVONEN, K.; et al. High diversity of cultivable heterotrophic bacteria in association with cyanobacterial water blooms. **ISME Journal**, v. 3, n. 3, p. 314–325, 2009.
- BERTELLI, C., TILLEY, K. E., & BRINKMAN, F. S. Microbial genomic island discovery, visualization and analysis. **Briefings in bioinformatics**, 20(5), 1685–1698. 2019.
- BERTELLI, C.; LAIRD, M. R.; WILLIAMS, K. P.; et al. IslandViewer 4: Expanded prediction of genomic islands for larger-scale datasets. **Nucleic Acids Research**, v. 45, n. W1, p. W30–W35, 2017. Oxford University Press.
- BHOWMICK, U. D., BHATTACHARJEE, S. Bacteriological, clinical and virulence aspects of *Aeromonas*-associated diseases in humans. **Polish Journal of Microbiology**, 67(2), 137. 2018.
- BIDDANDA, B., DILA, D., WEINKE, A., MANCUSO, J., VILLAR-ARGAIZ, M., MEDINA-SÁNCHEZ, J. M., CARRILLO, P. Housekeeping in the hydrosphere: Microbial cooking, cleaning, and control under stress. **Life**, 11(2), 152. 2021.
- BODEN, M., SCHÖNEICH, M., HORWEGE, S., LINDNER, S., LEIMEISTER, C., & MORGENSTERN, B. Alignment-free sequence comparison with spaced k-mers. In **German Conference on Bioinformatics 2013**. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
- BORRELL, N.; FIGUERAS, M. J.; GUARRO, J. **Phenotypic identification of *Aeromonas* genomospecies from clinical and environmental sources**. 1998.
- BÖTTGER, E. C. **Rapid determination of bacterial ribosomal RNA sequences by direct sequencing of enzymatically amplified DNA**. 1989.
- BROWN-JAQUE, M., CALERO-CÁCERES, W., MUNIESA, M. Transfer of antibiotic-resistance genes via phage-related mobile elements. **Plasmid**, 79, 1–7. 2015.
- BUKIN, Y. S., GALACHYANTS, Y. P., MOROZOV, I. V., BUKIN, S. V., ZAKHARENKO, A. S., & ZEMSKAYA, T. I. The effect of 16S rRNA region choice on bacterial community metabarcoding results. **Scientific Data**, 6(1), 1–14. 2019.
- BURRUS, V.; PAVLOVIC, G.; DECARIS, B.; GUÉDON, G. **Conjugative transposons: the tip of the iceberg**. 2002.
- BURRUS, V.; WALDOR, M. K. Shaping bacterial genomes with integrative and conjugative elements. **Research in Microbiology**, 2004. Elsevier Masson SAS.
- CAO, H., WANG, M., WANG, Q., XU, T., DU, Y., LI, H., ... & LIU, B. Identifying genetic diversity of O antigens in *Aeromonas hydrophila* for molecular serotype detection. **PloS one**. 2018.
- CARNAHAN, A. M.; CHAKRABORTY, T.; FANNING, G. R.; et al. ***Aeromonas trota* sp. nov., an Ampicillin-Susceptible Species Isolated from Clinical Specimens**. 1991
- CARRARO, N., RIVARD, N., BURRUS, V., & CECCARELLI, D. Mobilizable genomic islands, different strategies for the dissemination of multidrug resistance and other adaptive traits. **Mobile genetic elements**, 7(2), 1–6. 2017.
- CASTRO-ESCARPULLI, G.; FIGUERAS, M. J.; AGUILERA-ARREOLA, G.; et al. Characterization of *Aeromonas* spp. isolated from frozen fish intended for human consumption in Mexico. **International Journal of Food Microbiology**, v. 84, n. 1, p. 41–49, 2003. Elsevier.

CHANG, Y.-C.; WANG, J.-Y.; SELVAM, A.; et al. **Multiplex PCR Detection of Enterotoxin Genes in *Aeromonas* spp. from Suspect Food Samples in Northern Taiwan**. 2008.

CHARETTE, S. J., BROCHU, F., BOYLE, B., FILION, G., TANAKA, K. H., DEROME, N. Draft genome sequence of the virulent strain 01-B526 of the fish pathogen *Aeromonas salmonicida*. 2012.

CHATTERJEE, R.; CHAUDHURI, K.; CHAUDHURI, P. On detection and assessment of statistical significance of Genomic Islands. **BMC Genomics**, v. 9, 2008.

CHE, D.; HASAN, M. S.; WANG, H.; et al. Prediction Model EGID: an ensemble algorithm for improved genomic island detection in genomic sequences. **print) Bioinformatics**, v. 7, n. 6, p. 311, 2011.

CHE, D.; HOCKENBURY, C.; MARMELESTEIN, R.; RASHEED, K. Classification of genomic islands using decision trees and their ensemble algorithms. **BMC Genomics**, v. 11, n. SUPPL. 2, 2010.

CHE, D.; WANG, H.; FAZEKAS, J.; CHEN, B. An Accurate Genomic Island Prediction Method for Sequenced Bacterial and Archaeal Genomes. **J Proteomics Bioinform Citation**, v. 7, n. 8, p. 214–221, 2014.

CHEN, C. Y.; WANG, G. H.; TSENG, I. H.; CHUNG, Y. C. Analysis of bacterial diversity and efficiency of continuous removal of Victoria Blue R from wastewater by using packed-bed bioreactor. **Chemosphere**, v. 145, p. 17–24, 2016. Elsevier Ltd.

CHEN, F., SUN, J., HAN, Z., YANG, X., XIAN, J. A., LV, A., SHI, H. Isolation, identification and characteristics of *Aeromonas veronii* from diseased crucian carp (*Carassius auratus gibelio*). **Frontiers in microbiology**, 2742. 2019.

CHEN, I.; CHRISTIE, P. J.; DUBNAU, D. The ins and outs of DNA transfer in bacteria. **Science**, 2. Dec. 2005.

CHEN, L., QIU, Q., JIANG, Y. U., WANG, K., LIN, Z., LI, Z., WANG, W. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. **Science**, 364(6446), eaav6202. 2019.

CHEN, S., DENG, L. Y., BOWMAN, D., SHIAU, J. J. H., WONG, T. Y., MADAHIAN, B., & LU, H. H. S. Phylogenetic tree construction using trinucleotide usage profile (TUP). In **BMC bioinformatics** (Vol. 17, No. 13, pp. 117-130). BioMed Central. 2016.

CHRISTENSEN, H., OLSEN, J. E. Sequence-based classification and identification of prokaryotes. In **Introduction to bioinformatics in microbiology** (pp. 121-134). Springer, Cham. 2018.

CHURCH, D. L., CERUTTI, L., GÜRTLER, A., GRIENER, T., ZELAZNY, A., EMLER, S. Performance and application of 16S rRNA gene cycle sequencing for routine identification of bacteria in the clinical microbiology laboratory. **Clinical Microbiology Reviews**, 33(4), e00053-19. 2020.

CODJOE, F.; DONKOR, E. Carbapenem Resistance: A Review. **Medical Sciences**, v. 6, n. 1, p. 1, 2017. MDPI AG.

COLEMAN, M.L; SULLIVAN, M.B; MARTINY, A.C; STEGLICH, C; BARRY, K; DELONG E,F; CHISHOLM,S.W. **Science**. 2006 Mar 24; 311(5768):1768-70.

COLSTON, S. M., FULLMER, M. S., BEKA, L., LAMY, B., GOGARTEN, J. P., & GRAF, J. Bioinformatic genome comparisons for taxonomic and phylogenetic assignments using *Aeromonas* as a test case. **MBio**, 5(6), e02136-14. 2014.

COLWELL, R. R.; MACDONELL, M. T.; DE, J.; et al. **Proposal to Recognize the Family Aeromonadaceae fam. nov.** 1986.

CUI, B., LIANG, S. Monitoring opportunistic pathogens in domestic wastewater from a pilot-scale anaerobic biofilm reactor to reuse in agricultural irrigation. **Water**, 11(6), 1283. 2019.

DARLING, A. C. E.; MAU, B.; BLATTNER, F. R.; PERNA, N. T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. **Genome Research**, v. 14, n. 7, p. 1394–1403, 2004.

DE PIERRI, C.R. Representação vetorial de sequências e textos para aprimorar a consistência da taxonomia bacteriana. Tese de doutorado. Programa de Pós-graduação em Ciências - Bioquímica, Universidade Federal do Paraná. Curitiba, 2022. (no prelo).

DIAS, C.; SERRA, C. R.; SIMÕES, L. C.; et al. Extended-spectrum β -lactamase and carbapenemase-producing *Aeromonas* species in wild animals from Portugal. **Veterinary Record**, v. 174, n. 21, p. 532, 2014. British Veterinary Association.

DOBRINDT, U.; HOCHHUT, B.; HENTSCHEL, U.; HACKER, J. Genomic islands in pathogenic and environmental microorganisms. **Nature Reviews Microbiology**, May 2004.

EL-BAHAR, H. M., ALI, N. G., ABOYADAK, I. M., KHALIL, S. A. E. S., IBRAHIM, M. S. Virulence genes contributing to *Aeromonas hydrophila* pathogenicity in *Oreochromis niloticus*. **International Microbiology**, 22(4), 479-490. 2019.

ELBEHIRY, A., MARZOUK, E., ABDEEN, E., AL-DUBAIB, M., ALSAYEQH, A., IBRAHEM, M., HEMEG, H. A. Proteomic characterization and discrimination of *Aeromonas* species recovered from meat and water samples with a spotlight on the antimicrobial resistance of *Aeromonas hydrophila*. **Microbiologyopen**, 8(11), e782. 2019.

ELHAI, J.; LIU, H.; TATON, A. Detection of horizontal transfer of individual genes by anomalous oligomer frequencies. **BMC Genomics**, v. 13, n. 1, 2012.

ERDEM, B., KARIPTAŞ, E., CIL, E., IŞIK, K. Biochemical identification and numerical taxonomy of *Aeromonas* spp. isolated from food samples in Turkey. **Turkish Journal of Biology**, 35(4), 463-472. 2011.

EUZEBY, J. P. List of Bacterial Names with Standing in Nomenclature: A Folder Available on the Internet. **International Union of Microbiological Societies**. 1997.

EUZEBY, J. P., TINDALL B. J. Valid publication of new names or new combinations: making use of the validation lists. **ASM News** .70:258-259.2004.

FARMER J.J., ARDUINO M.J., HICKMAN-BRENNER F.W. **The Prokaryotes**. Springer Science and Business Media; Berlin/Heidelberg, Germany: 2006. The genera *Aeromonas* and *Plesiomonas*; pp. 564–596.

FERNÁNDEZ-BRAVO, A., FIGUERAS, M. J. An update on the genus *Aeromonas*: Taxonomy, epidemiology, and pathogenicity. **Microorganisms** 8(1), 129. 2020.

FERNANDEZ-CASSI, X.; SILVERA, C.; CERVERO-ARAGÓ, S.; et al. Evaluation of the microbiological quality of reclaimed water produced from a lagooning system. **Environmental**

Science and Pollution Research, v. 23, n. 16, p. 16816–16833, 2016. Springer Verlag.

FIGUERAS M.J., BEAZ-HIDALGO R. *Aeromonas*. Academic Press; Norfolk, UK: 2015. **Aeromonas infections in humans**. pp. 65–108.

FIGUERAS M.J., BEAZ-HIDALGO R. **Encyclopedia of Food Microbiology**. Academic Press; London, UK: 2014. *Aeromonas*: Introduction, detection by cultural and modern techniques; pp. 25–30.

FIGUERAS, M. J., BEAZ-HIDALGO, R., COLLADO, L., & MARTÍNEZ-MURCIA, A. J. Recommendations for a new bacterial species description based on analyses of the unrelated genera *Aeromonas* and *Arcobacter*. **Bull. BISMIS**, 2, 1-16. 2011.

FIGUERAS, M. J., LATIF-EUGENÍN, F., BALLESTER, F., PUJOL, I., TENA, D., BERG, K., LILES, M. R. 'Aeromonas intestinalis' and 'Aeromonas enterica' isolated from human faeces, 'Aeromonas crassostreae' from oyster and 'Aeromonas aquatilis' isolated from lake water represent novel species. **New microbes and new infections**, 15, 74-76. 2017.

FIGUERAS, M. J.; SUAREZ-FRANQUET, A.; CHACÓN, M. R.; et al. First record of the rare species *Aeromonas culicicola* from a drinking water supply. **Applied and Environmental Microbiology**, v. 71, n. 1, p. 538–541, 2005.

FIGUERAS, M.J; BORREGO, J. J. New perspectives in monitoring drinking water microbial quality. **International Journal of Environmental Research and Public Health**, 2010. MDPI.

FINLAY, B. B.; FALKOW, S. **Common Themes in Microbial Pathogenicity Revisited**. 1997.

FONTES, M. C.; SAAVEDRA, M. J.; MONERA, A.; MARTINS, C.; MARTÍNEZ-MURCIA, A. Phylogenetic identification of *Aeromonas simiae* from a pig, first isolate since species description. **Veterinary Microbiology**, v. 142, n. 3–4, p. 313–316, 2010.

FOSSE, T.; GIRAUD-MORIN, C.; MADINIER, I. Phénotypes de résistance aux β -lactamines dans le genre *Aeromonas*. *Pathologie Biologie*. **Anais**. v. 51, p.290–296, 2003. Elsevier Masson SAS.

FROST, L. S.; LEPLAE, R.; SUMMERS, A. O.; TOUSSAINT, A. Mobile genetic elements: The agents of open-source evolution. **Nature Reviews Microbiology**, Sep. 2005.

FU, S., WANG, Q., WANG, R., ZHANG, Y., LAN, R., HE, F., & YANG, Q. Horizontal transfer of antibiotic resistance genes within the bacterial communities in aquacultural environment. **Science of The Total Environment**, 153286. 2022.

GAL-MOR, O.; FINLAY, B. B. Pathogenicity islands: A molecular toolbox for bacterial virulence. **Cellular Microbiology**, Nov. 2006.

GALUSHKO, A., KUEVER, J. *Bergey's Manual of Systematics of Archaea and Bacteria*. In **Bergey's Manual of Systematics of Archaea and Bacteria** (pp. 1-4). 2020.

GHENGHESH, K. S.; RAHOUMA, A.; ZORGANI, A.; et al. *Aeromonas* in Arab countries: 1995-2014. **Comparative Immunology, Microbiology and Infectious Diseases**, 1. Oct. 2015. Elsevier Ltd.

GOMES, S., FERNANDES, C., MONTEIRO, S., CABECINHA, E., TEIXEIRA, A., VARANDAS, S., SAAVEDRA, M. J. The role of aquatic ecosystems (River Tua, Portugal) as reservoirs of multidrug-resistant *Aeromonas* spp. **Water**, 13(5), 698. 2021.

GONÇALVES, O. S., DE QUEIROZ, M. V., & SANTANA, M. F. Potential evolutionary impact of integrative and conjugative elements (ICEs) and genomic islands in the *Ralstonia solanacearum*

species complex. **Scientific reports**, 10(1), 1-9. 2020.

GUZ, L., NOWAKIEWICZ, A., PUK, K., ZIĘBA, P., GNAT, S., MATUSZEWSKI, Ł. Virulence and Antimicrobial Resistance Pattern of *Aeromonas* spp. Colonizing European Pond Turtles *Emys orbicularis* and Their Natural Environment. First Study from Poland. **Animals**, 11(10), 2772. 2021.

HACKER, J.; CARNIEL, E. **Ecological fitness, genomic islands and bacterial pathogenicity A Darwinian view of the evolution of microbes**. 2001.

HACKER, J.; KAPER, J. B. **PATHOGENICITY ISLANDS AND THE EVOLUTION OF MICROBES**. 2000.

HACKER, JORG; BENDER, LARISA; OTT, MANFRED; et al. **Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extraintestinal *Escherichia coli* isolates**. 1990.

HAMMAD, A. M.; MOUSTAFA, A. E. H.; MANSOUR, M. M.; et al. Molecular and phenotypic analysis of hemolytic *aeromonas* strains isolated from food in Egypt revealed clinically important multidrug resistance and virulence profiles. **Journal of Food Protection**, v. 81, n. 6, p. 1015–1021, 2018. International Association for Food Protection.

HARAYAMA, S., KASAI, H. Bacterial phylogeny reconstruction from molecular sequences. **Molecular identification, systematics, and population structure of prokaryotes** (pp. 105-139). Springer, Berlin, Heidelberg. 2006.

HARMSSEN, D., KARCH, H. 16S rDNA for diagnosing pathogens: a living tree. **ASM News**. 70:19-24. 2004.

HASAN, M. S.; LIU, Q.; WANG, H.; et al. Software Volume 8(4) GIST: Genomic Island suite of tools for predicting genomic islands in genomic sequences. **Bioinformatics**, v. 8, n. 4, p. 203–205, 2012.

HAYATGHEIB, N., MOREAU, E., CALVEZ, S., LEPELLETIER, D., POULIQUEN, H. A review of functional feeds and the control of *Aeromonas* infections in freshwater fish. **Aquaculture International**, 28(3), 1083-1123. 2020.

HE, T., SHEN, J., SCHWARZ, S., WU, C., WANG, Y. Characterization of a genomic island in *Stenotrophomonas maltophilia* that carries a novel *flaR* gene variant. **Journal of Antimicrobial Chemotherapy**, 70(4), 1031-1036. 2015.

HEISS, C.; WANG, Z.; THURLOW, C. M.; et al. Structure of the capsule and lipopolysaccharide O-antigen from the channel catfish pathogen, *Aeromonas hydrophila*. **Carbohydrate Research**, v. 486, 2019.

HOEL, S.; VADSTEIN, O.; JAKOBSEN, A. N. Species distribution and prevalence of putative virulence factors in Mesophilic *Aeromonas* spp. isolated from fresh retail sushi. **Frontiers in Microbiology**, v. 8, n. MAY, 2017.

HOEL, S.; VADSTEIN, O.; JAKOBSEN, A. N. The significance of mesophilic *aeromonas* spp. In minimally processed ready-to-eat seafood. **Microorganisms**, v. 7, n. 3, 2019.

HOSSAIN, M. J., WALDBIESER, G. C., SUN, D., CAPPS, N. K., HEMSTREET, W. B., CARLISLE, K., LILES, M. R. Implication of lateral genetic transfer in the emergence of *Aeromonas hydrophila* isolates of epidemic outbreaks in channel catfish. **PLoS One**. 2013.

HSIAO, W. W. L.; UNG, K.; AESCHLIMAN, D.; et al. Evidence of a Large Novel Gene Pool

Associated with Prokaryotic Genomic Islands. (C. Fraser, Ed.) **PLoS Genetics**, v. 1, n. 5, p. e62, 2005.

HSIAO, W.; WAN, I.; JONES, S. J.; BRINKMAN, F. S. L. IslandPath: Aiding detection of genomic islands in prokaryotes. **Bioinformatics**, v. 19, n. 3, p. 418–420, 2003.

HUA, H. T.; BOLLET, C.; TERCIAN, S.; DRANCOURT, M.; RAOULT, D. *Aeromonas popoffii* urinary tract infection. **Journal of Clinical Microbiology**, v. 42, n. 11, p. 5427–5428, 2004.

HUSON, D. H., SCORNAVACCA, C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. **Systematic biology**, 61(6), 1061-1067. 2012.

IGBINOSA, I. H.; IGUMBOR, E. U.; AGHDASI, F.; TOM, M.; OKOH, A. I. Emerging *Aeromonas* species infections and their significance in public health. **The Scientific World Journal**, 2012.

JAIN, C., DILTNEY, A., KOREN, S., ALURU, S., & PHILLIPPY, A. M. A fast approximate algorithm for mapping long reads to large reference databases. In **International Conference on Research in Computational Molecular Biology** (pp. 66-81). Springer, Cham. 2017.

JAIN, C., RODRIGUEZ-R, L. M., PHILLIPPY, A. M., KONSTANTINIDIS, K. T., & ALURU, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. **Nature communications**, 9(1), 1-8. 2018.

JANDA, J. M.; ABBOTT, S. L. The genus *Aeromonas*: Taxonomy, pathogenicity, and infection. **Clinical Microbiology Reviews**, Jan. 2010.

JANDA, J. MICHAEL; ABBOTT, SHARON L. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. **Journal of clinical microbiology**, v. 45, n. 9, p. 2761-2764, 2007.

JJEMBA, P. K.; WEINRICH, L. A.; CHENG, W.; GIRALDO, E.; LECHEVALLIER, M. W. Regrowth of potential opportunistic pathogens and algae in reclaimed-water distribution systems. **Applied and Environmental Microbiology**, v. 76, n. 13, p. 4169–4178, 2010.

JOHN, N.; VIDYALAKSHMI, V. B.; HATHA, A. A. M. Effect of pH and Salinity on the Production of Extracellular Virulence Factors by *Aeromonas* from Food Sources. **Journal of Food Science**, v. 84, n. 8, p. 2250–2255, 2019. Blackwell Publishing Inc.

JOLLEY, K. A., BRAY, J. E., MAIDEN, M. C. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. **Wellcome open research**, 3. 2018.

JUHAS, M. Genomic Islands and the Evolution of Multidrug-Resistant Bacteria. **Horizontal Gene Transfer**, 143-153. 2019.

KAMPFER, R.; ERHART, R.; BEIMFOHR, C.; et al. **MICROBIAL ECOLOGY Characterization of Bacterial Communities from Activated Sludge: Culture-Dependent Numerical Identification Versus In Situ Identification Using Group-and Genus-Specific rRNA-Targeted Oligonucleotide Probes**. 1996.

KHAJANCHI, B. K.; FADL, A. A.; BORCHARDT, M. A.; et al. Distribution of virulence factors and molecular fingerprinting of *Aeromonas* species isolates from water and clinical samples: Suggestive evidence of water-to-human transmission. **Applied and Environmental Microbiology**, v. 76, n. 7, p. 2313–2325, 2010.

KIM, F. J. P., SILVA, A. E. M., SILVA, R. V. S., KIM, P. C. P., ACOSTA, A. C., SILVA, S. M. B. C., MOTA, R. A. High frequency of *Aeromonas* spp. and virulence genes in Nile Tilapia farms

(*Oreochromis niloticus*) in semi-arid regions of Pernambuco, Brazil. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, 71(5), 1609-1615. 2019.

KIM, M., OH, H. S., PARK, S. C., & CHUN, J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. **International journal of systematic and evolutionary microbiology**, 64(Pt_2), 346-351. 2014.

KING, G. E.; WERNER, S. B.; KIZER, K. W. **Epidemiology of Aeromonas Infections in California**. 1992.

KLENK H.-P., ZILLIG W. DNA-dependent RNA polymerase subunit B as a tool for phylogenetic reconstructions: branching topology of the archeal domain. **J Mol Evol**. 1994.

KOONIN, E. V. Horizontal gene transfer: Essentiality and evolvability in prokaryotes, and roles in evolutionary transitions. **F1000Research**, 2016.

KORCZAK B., CHRISTENSEN H., EMLER S., FREY J., KUHNERT P. Phylogeny of the family Pasteurellaceae based on rpoB sequences. **Int J Syst Evol Microbiol**. 2004.

KRIEG, N. R. HOLT, J. G. Bergey's manual of systematic bacteriology. The Williams & Wilkins Co., Baltimore, Md 1984.

KÜMMERER, K. Antibiotics in the aquatic environment - A review - Part II. **Chemosphere**, Apr. 2009.

Lamy B., Kodjo A., Laurent F. Identification of *Aeromonas* isolates by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. **Diagn. Microbiol. Infect. Dis.** 71 1–5. 10.1016/j.diagmicrobio.2011.04.014. 2011.

Lamy, B., Baron, S., Barraud, O. *Aeromonas*: the multifaceted middleman in the One Health world. **Current opinion in microbiology**, 65, 24-32. 2022.

LAMY, B.; KODJO, A.; LAURENT, F. Prospective nationwide study of *Aeromonas* infections in France. **Journal of Clinical Microbiology**, v. 47, n. 4, p. 1234–1237, 2009. American Society for Microbiology.

LANGILLE, M. G. I.; HSIAO, W. W. L.; BRINKMAN, F. S. L. Detecting genomic islands using bioinformatics approaches. **Nature Reviews Microbiology**, Mar. 2010.

LANGILLE, M. G. I.; HSIAO, W. W. L.; BRINKMAN, F. S. L. Evaluation of genomic island predictors using a comparative genomics approach. **BMC Bioinformatics**, v. 9, 2008.

LARBIG, K. D.; CHRISTMANN, A.; JOHANN, A.; et al. Gene islands integrated into tRNAGly genes confer genome diversity on a *Pseudomonas aeruginosa* clone. **Journal of Bacteriology**, v. 184, n. 23, p. 6665–6680, 2002.

LATIF EUGENÍN, F. L. *Aeromonas, un microorganismo ambiental de importancia en salud humana y animal* (Doctoral dissertation, Universitat Rovira i Virgili). 2015.

LATIF-EUGENÍN, F.; BEAZ-HIDALGO, R.; FIGUERAS, M. J. A culture independent method for the detection of *Aeromonas* sp. From water samples. **Italian Journal of Food Safety**, v. 5, n. 1, p. 11–14, 2016b. Page Press Publications.

LATIF-EUGENÍN, F.; BEAZ-HIDALGO, R.; FIGUERAS, M. J. Evaluation of different conditions and culture media for the recovery of *Aeromonas* spp. from water and shellfish samples. **Journal of Applied Microbiology**, v. 121, n. 3, p. 883–891, 2016a. Blackwell Publishing Ltd.

- LATIF-EUGENÍN, F.; BEAZ-HIDALGO, R.; SILVERA-SIMÓN, C.; FERNANDEZ-CASSI, X.; FIGUERAS, M. J. Chlorinated and ultraviolet radiation -treated reclaimed irrigation water is the source of *Aeromonas* found in vegetables used for human consumption. **Environmental Research**, v. 154, p. 190–195, 2017. Academic Press Inc.
- LAVRINIENKO, A., JERNFORS, T., KOSKIMÄKI, J. J., PIRTILÄ, A. M., WATTS, P. C. Does intraspecific variation in rDNA copy number affect analysis of microbial communities? **Trends in microbiology**, 29(1), 19-27. 2021.
- LAWRENCE, J. G. Common themes in the genome strategies of pathogens. **Current Opinion in Genetics and Development**, Dec. 2005.
- LEE, C. C.; CHEN, Y. P. P.; YAO, T. J.; et al. GI-POP: A combinational annotation and genomic island prediction pipeline for ongoing microbial genome projects. **Gene**, v. 518, n. 1, p. 114–123, 2013.
- LEIMEISTER, C. A., SCHELLHORN, J., DÖRRER, S., GERTH, M., BLEIDORN, C., & MORGENSTERN, B. Prot-SpaM: Fast alignment-free phylogeny reconstruction based on whole-proteome sequences. **GigaScience**, 8(3), giy148. 2019.
- LEPLAE, R.; LIMA-MENDEZ, G.; TOUSSAINT, A. A first global analysis of plasmid encoded proteins in the ACLAME database. **FEMS Microbiology Reviews**, Nov. 2006.
- LESIC, B.; BACH, S.; GHIGO, J. M.; et al. Excision of the high-pathogenicity of *Yersinia pseudotuberculosis* requires the combined of actions of its cognate integrase and Hef, a new recombination directionality factor. **Molecular Microbiology**, v. 52, n. 5, p. 1337–1348, 2004.
- LI, F.; WANG, W.; ZHU, Z.; et al. Distribution, virulence-associated genes and antimicrobial resistance of *Aeromonas* isolates from diarrheal patients and water, China. **Journal of Infection**, v. 70, n. 6, p. 600–608, 2015. W.B. Saunders Ltd.
- LI, Y., HE, L., LUCY HE, R., & YAU, S. S. T. A novel fast vector method for genetic sequence comparison. **Scientific reports**, 7(1), 1-11. 2017.
- LI, Y., LIU, Y., ZHOU, Z., HUANG, H., REN, Y., ZHANG, Y. WANG, L. Complete genome sequence of *Aeromonas veronii* strain B565. 2011.
- LIU, F., YUWONO, C., TAY, A. C. Y., WEHRHAHN, M. C., RIORDAN, S. M., ZHANG, L. Analysis of global *Aeromonas veronii* genomes provides novel information on source of infection and virulence in human gastrointestinal diseases. **BMC genomics**, 23(1), 1-15. 2022.
- LIU, J., DONG, Y., WANG, N., LI, S., YANG, Y., WANG, Y., LIU, Y. TETRAHYMENA thermophila predation enhances environmental adaptation of the carp pathogenic strain *Aeromonas hydrophila* NJ-35. **Frontiers in cellular and infection microbiology**, 8, 76. 2018.
- LIU, X., SUN, W., ZHANG, Y., ZHOU, Y., XU, J., GAO, X., ZHANG, X. Impact of *Aeromonas hydrophila* and infectious spleen and kidney necrosis virus infections on susceptibility and host immune response in Chinese perch (*Siniperca chuatsi*). **Fish & Shellfish Immunology**, 105, 117-125. 2020.
- LOWRY, R.; BALBOA, S.; PARKER, J. L.; SHAW, J. G. *Aeromonas* Flagella and Colonisation Mechanisms. **Advances in Microbial Physiology**. v. 65, p.203–256, 2014. Academic Press.
- LU, B., LEONG, H. W. GI-Cluster: detecting genomic islands via consensus clustering on multiple features. **Journal of bioinformatics and computational biology**, 16(03), 1840010. 2018.

LU, B.; LEONG, H. W. Computational methods for predicting genomic islands in microbial genomes. **Computational and Structural Biotechnology Journal**, 2016a. Elsevier B.V.

LU, B.; LEONG, H. W. GI-SVM: A sensitive method for predicting genomic islands based on unannotated sequence of a single genome. *Journal of Bioinformatics and Computational Biology*. **Bioinform Comput Biol**. 2016b.

MAIQUES, E.; ÚBEDA, C.; TORMO, M. Á.; et al. Role of staphylococcal phage and SaPI integrase in intra- and interspecies SaPI transfer. **Journal of Bacteriology**, v. 189, n. 15, p. 5608–5616, 2007.

MANSON, J. M.; GILMORE, M. S. Pathogenicity Island integrase cross-talk: A potential new tool for virulence modulation. **Molecular Microbiology**, Aug. 2006.

MANTRI, Y.; WILLIAMS, K. P. Islander: A database of integrative islands in prokaryotic genomes, the associated integrases and their DNA site specificities. **Nucleic Acids Research**, v. 32, n. DATABASE ISS., 2004.

MARCOUX, P. É., VINCENT, A. T., MASSICOTTE, M. A., PAQUET, V. E., DOUCET, É. J., HOSSEINI, N., CHARETTE, S. J. Systematic analysis of the stress-induced genomic instability of Type Three Secretion System in *Aeromonas salmonicida* subsp. *salmonicida*. **Microorganisms**, 9(1), 85. 2020.

MARTIN-CARNAHAN A., JOSEPH S.W. Order XII. Aeromonadales ord. nov. In: Brenner D.J., Krieg N.R., Staley J.T., Garrity G.M., editors. **Bergey's Manual of Systematic Bacteriology**. Williams & Wilkins; Philadelphia, PA, USA: 2005. pp. 556–578.

MARTÍNEZ-MURCIA A., LAMY B. *Aeromonas*. Academic Press; Norfolk, UK. **Molecular diagnostics by genetic methods**; pp. 155–200. 2015.

MARTÍNEZ-MURCIA A. J. Phylogenetic positions of *Aeromonas encheleia*, *Aeromonas popoffii*, *Aeromonas* DNA hybridization group 11 and *Aeromonas* group 501. **Int J Syst Bacteriol**. 49:1403–1408. 1999.

MARTÍNEZ-MURCIA, A. J., SOLER, L., SAAVEDRA, M. J., CHACÓN, M. R., GUARRO, J., STACKEBRANDT, E., et al. Phenotypic, genotypic, and phylogenetic discrepancies to differentiate *Aeromonas salmonicida* from *Aeromonas bestiarum*. **Int. Microbiol**. 8, 259–269. 2005.

MARTINEZ-MURCIA, A. J.; BENLLOCH, S.; COLLINS, M. D. Phylogenetic interrelationships of members of the genera *Aeromonas* and *Plesiomonas* as determined by 16S ribosomal DNA sequencing: lack of congruence with results of DNA-DNA hybridizations. **International Journal of systematic and evolutionary microbiology**, v. 42, n. 3, p. 412–421, 1992.

MARTÍNEZ-MURCIA, A., BEAZ-HIDALGO, R., NAVARRO, A., CARVALHO, M., ARAVENA-ROMÁN, M., CORREIA, A., SAAVEDRA, M. *Aeromonas lusitana* sp. nov., isolated from untreated water and vegetables. **Current microbiology**, 72(6), 795–803. 2016.

MARTINO, M. E., FASOLATO, L., MONTEMURRO, F., ROSTEGHIN, M., MANFRIN, A., PATARNELLO, T., CARDAZZO, B. Determination of microbial diversity of *Aeromonas* strains on the basis of multilocus sequence typing, phenotype, and presence of putative virulence genes. **Applied and environmental microbiology**, 77(14), 4986–5000. 2011.

MCLELLAN, S. L., ROGUET, A. The unexpected habitat in sewer pipes for the propagation of microbial communities and their imprint on urban waters. **Current opinion in biotechnology**, 57,

34-41. 2019.

MCLELLAN, S. L.; HUSE, S. M.; MUELLER-SPITZ, S. R.; ANDREISHCHEVA, E. N.; SOGIN, M. L. Diversity and population structure of sewage-derived microorganisms in wastewater treatment plant influent. **Environmental Microbiology**, v. 12, n. 2, p. 378–392, 2010.

MEIER-KOLTHOFF, J. P., GÖKER, M., SPRÖER, C., & KLENK, H. P. When should a DDH experiment be mandatory in microbial taxonomy? **Archives of Microbiology**, 195(6), 413-418. 2012.

MENDOZA-BARBERÁ, E., MERINO, S., TOMÁS, J. Surface Glucan Structures in *Aeromonas* spp. **Marine Drugs**, 19(11), 649. 2021.

MERINO S., TOMÁS J. Bacterial capsules and evasion of immune responses. **eLS**. 2010 doi: 10.1002/9780470015902.a0000957.pub4.

MERKL R. SIGI: score-based identification of genomic islands. **BMC Bioinf.** 2004.

MIDDENDORF, B.; HOCHHUT, B.; LEIPOLD, K.; et al. Instability of Pathogenicity Islands in Uropathogenic *Escherichia coli* 536. **Journal of Bacteriology**, v. 186, n. 10, p. 3086–3096, 2004.

MOLLET C., DRANCOURT M., RAOULT D. rpoB sequence analysis as a novel basis for bacterial identification. **Mol Microbiol**. 1997.

MORANDI A., ZHAXYBAYEVA O., GOGARTEN J. P., GRAL. Evolutionary and diagnostic implications of intragenomic heterogeneity in the 16S rRNA gene in *Aeromonas* strains. **J Bacteriol**. 2005.

MORANDI, A.; ZHAXYBAYEVA, O.; GOGARTEN, J. P.; GRAF, J. Evolutionary and diagnostic implications of intragenomic heterogeneity in the 16S rRNA gene in *Aeromonas* strains. **Journal of Bacteriology**, v. 187, n. 18, p. 6561–6564, 2005.

MORANDI, Alessia et al. Evolutionary and diagnostic implications of intragenomic heterogeneity in the 16S rRNA gene in *Aeromonas* strains. **Journal of Bacteriology**, v. 187, n. 18, p. 6561-6564, 2005.

MORIEL, B., DE CAMPOS PREDIGER, K., DE SOUZA, E. M., PEDROSA, F. O., FADEL-PICHETH, C. M., CRUZ, L. M. In silico comparative analysis of *Aeromonas* type VI secretion system. **Brazilian Journal of Microbiology**, 52(1), 229-243. 2021.

MURALI, A., BHARGAVA, A., WRIGHT, E. S. IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. **Microbiome**, 6(1), 1-14. 2018.

MURRAY, P. R., E. J. BARON, M. A. PFALLER, F. C. TENOVER AND R. H. YOLKEN (ED.). **Manual of clinical microbiology**, 7th ed. ASM Press, Washington, D.C. 1999.

NAGAR, V.; SHASHIDHAR, R.; BANDEKAR, J. R. Characterization of *Aeromonas* strains isolated from Indian foods using rpoD gene sequencing and whole cell protein analysis. **World Journal of Microbiology and Biotechnology**, v. 29, n. 4, p. 745–752, 2013. Kluwer Academic Publishers.

NAGAR, V.; SHASHIDHAR, R.; BANDEKAR, J. R. Prevalence, Characterization, and Antimicrobial Resistance of *Aeromonas* Strains from Various Retail Food Products in Mumbai, India. **Journal of Food Science**, v. 76, n. 7, 2011.

NAKAMURA, Y.; ITOH, T.; MATSUDA, H.; GOJOBORI, T. Biased biological functions of horizontally-transferred genes in prokaryotic genomes. **Nature Genetics**, v. 36, n. 7, p. 760–766, 2004.

- NASH, J. H. E.; FINDLAY, W. A.; LUEBBERT, C. C.; et al. Comparative genomics profiling of clinical isolates of *Aeromonas salmonicida* using DNA microarrays. **BMC Genomics**, v. 7, 2006.
- NAVARRO, A., MARTÍNEZ-MURCIA, A. Phylogenetic analyses of the genus *Aeromonas* based on housekeeping gene sequencing and its influence on systematics. **Journal of applied microbiology**, 125(3), 622-631. 2018.
- NDI, O. L.; BARTON, M. D. Incidence of class 1 integron and other antibiotic resistance determinants in *Aeromonas* spp. from rainbow trout farms in Australia. **Journal of Fish Diseases**, v. 34, n. 8, p. 589–599, 2011.
- NIEHUS, R.; MITRI, S.; FLETCHER, A. G.; FOSTER, K. R. Migration and horizontal gene transfer divide microbial genomes into multiple niches. **Nature Communications**, v. 6, 2015. Nature Publishing Group.
- NOÉ, L., MARTIN, D. E. A coverage criterion for spaced seeds and its applications to support vector machine string kernels and k-mer distances. **Journal of Computational Biology**, 21(12), 947-963. 2014.
- NOVICK, R. P., RAM, G. The floating (pathogenicity) island: a genomic dessert. **Trends in Genetics**, 32(2), 114-126. 2016.
- OU, H. Y.; CHEN, L. L.; LONNEN, J.; et al. A novel strategy for the identification of genomic islands by comparative analysis of the contents and contexts of tRNA sites in closely related bacteria. **Nucleic Acids Research**, v. 34, n. 1, 2006.
- OU, H. Y.; HE, X.; HARRISON, E. M.; et al. MobilomeFINDER: Web-based tools for in silico and experimental discovery of bacterial genomic islands. **Nucleic Acids Research**, v. 35, n. SUPPL.2, 2007.
- PABLOS, M.; HUYS, G.; CNOCKAERT, M.; et al. Identification and epidemiological relationships of *Aeromonas* isolates from patients with diarrhea, drinking water and foods. **International Journal of Food Microbiology**, v. 147, n. 3, p. 203–210, 2011.
- PABLOS, M.; REMACHA, M. A.; RODRÍGUEZ-CALLEJA, J. M.; et al. Identity, virulence genes, and clonal relatedness of *Aeromonas* isolates from patients with diarrhea and drinking water. **European Journal of Clinical Microbiology and Infectious Diseases**, v. 29, n. 9, p. 1163–1172, 2010.
- PALLEN, M. J., TELATIN, A., OREN, A. The next million names for Archaea and Bacteria. **Trends in Microbiology**, 29(4), 289-298. 2021.
- PANG, M., JIANG, J., XIE, X., WU, Y., DONG, Y., KWOK, A. H., LIU, Y. Novel insights into the pathogenicity of epidemic *Aeromonas hydrophila* ST251 clones from comparative genomics. **Scientific reports**. 2015.
- PARK, S. Y., HAN, J. E., KWON, H., PARK, S. C., KIM, J. H. RECENT insights into *Aeromonas salmonicida* and its bacteriophages in aquaculture: A comprehensive review. **J Microbiol Biotechnol**. 28;30(10):1443-1457.2020.
- PARKER, J. L., SHAW, J. G. *Aeromonas* spp. clinical microbiology and disease. **Journal of Infection**, 62(2), 109-118. 2011.
- PERCIVAL, S. L., WILLIAMS, D. W. Chapter Three-*Aeromonas*. In *Microbiology of Waterborne Diseases-Microbiological Aspects and Risks (Second Edition)* (pp. 49-64). Academic Press. 2014.

PESSOA, R. B. G., DE OLIVEIRA, W. F., MARQUES, D. S. C., DOS SANTOS CORREIA, M. T., DE CARVALHO, E. V. M. M., & COELHO, L. C. B. B. The genus *Aeromonas*: A general approach. **Microbial pathogenesis**, 130, 81-94. 2019.

PFISTER, P.; RISCH, M.; BRODERSEN, D. E.; BÖTTGER, E. C. Role of 16S rRNA helix 44 in ribosomal resistance to hygromycin B. **Antimicrobial Agents and Chemotherapy**, v. 47, n. 5, p. 1496–1502, 2003.

PIANETTI, A.; BRUSCOLINI, F.; SABATINI, L.; COLANTONI, P. Microbial characteristics of marine sediments in bathing area along Pesaro-Gabicce coast (Italy): A preliminary study. **Journal of Applied Microbiology**, v. 97, n. 4, p. 682–689, 2004.

PIDIYAR V. J., JANGID K., DAYANANDA K. M., KAZNOWSKI A., GONZALEZ J. M., PATOLE M. S., SHOUCHE Y. S. 2003. Phylogenetic affiliation of *Aeromonas culicicola* MTCC 3249T based on *gyrB* gene sequence and PCR-amplicon sequence analysis of cytolytic enterotoxin gene. **Syst Appl Microbiol**.2003.

PIOTROWSKA, M.; POPOWSKA, M. Insight into the mobilome of *Aeromonas* strains. **Frontiers in Microbiology**, 2015. Frontiers Media S.A.

PIOTROWSKA, M.; PRZYGODZINSKA, D.; MATYJEWICZ, K.; POPOWSKA, M. Occurrence and variety of β -lactamase genes among *Aeromonas* spp. isolated from Urban Wastewater treatment plant. **Frontiers in Microbiology**, v. 8, n. MAY, 2017. Frontiers Media S.A.

PREDIGER, K. D. C., DALLAGASSA, C. B., MORIEL, B., VIZZOTTO, B. S., VOLANSKI, W., SOUZA, E. M., FADEL-PICHETH, C. M. (2020). Virulence characteristics and antimicrobial resistance of *Aeromonas veronii* biovar *sobria* 312M, a clinical isolate. **Brazilian Journal of Microbiology**, 51(2), 511-518. 2020.

PREDIGER, K. D. C., PEREIRA, R. D. S., WINCKLER NETO, C. H. D. P., SANTOS, R. C. V., FADEL-PICHETH, C. M. T., VIZZOTTO, B. S. A prospective study on *Aeromonas* in outpatients with diarrhea in the central region of Rio Grande do Sul State. **Brazilian Journal of Microbiology**, 43, 966-968. 2012.

RAJAN, I.; ARAVAMUTHAN, S.; MANDE, S. S. Identification of compositionally distinct regions in genomes using the centroid method. **Bioinformatics**, v. 23, n. 20, p. 2672–2677, 2007.

RAMSAY, J. P.; SULLIVAN, J. T.; STUART, G. S.; LAMONT, I. L.; RONSON, C. W. Excision and transfer of the *Mesorhizobium loti* R7A symbiosis island requires an integrase *IntS*, a novel recombination directionality factor *RdfS*, and a putative relaxase *RlxS*. **Molecular Microbiology**, v. 62, n. 3, p. 723–734, 2006.

RASMUSSEN-IVEY, C. R.; HOSSAIN, M. J.; ODOM, S. E.; et al. Classification of a hypervirulent *aeromonas hydrophila* pathotype responsible for epidemic outbreaks in warm-water fishes. **Frontiers in Microbiology**, v. 7, n. OCT, 2016. Frontiers Media S.A.

RASMUSSEN-IVEY, Cody. Comparative Genomic Analysis and Evaluation of Control Strategies for Hypervirulent *Aeromonas hydrophila* in Channel Catfish (*Ictalurus punctatus*). 2019.

RAVATN, R.; STUDER, S.; SPRINGAEL, D.; ZEHNDER, A. J. B.; ROELOF VAN DER MEER, J. **Chromosomal Integration, Tandem Amplification, and Deamplification in *Pseudomonas putida* F1 of a 105-Kilobase Genetic Element Containing the Chlorocatechol Degradative Genes from *Pseudomonas* sp. Strain B13**. 1998.

RAVENHALL, M.; ŠKUNCA, N.; LASSALLE, F.; DESSIMOZ, C. Inferring Horizontal Gene

Transfer. **PLoS Computational Biology**, v. 11, n. 5, 2015. Public Library of Science.

REITH, M. E., SINGH, R. K., CURTIS, B., BOYD, J. M., BOUEVITCH, A., KIMBALL, J., ... & BROWN, L. L. The genome of *Aeromonas salmonicida* subsp. *salmonicida* A449: insights into the evolution of a fish pathogen. *BMC genomics*, 9(1), 1-15. 2008.

REYES-RODRIGUEZ, N. E., SALGADO-MIRANDA, C., FLORES-VALLE, I. T., GONZALEZ-GOMEZ, M., SORIANO-VARGAS, E., PELAEZ-ACERO, A., VEGA-SANCHEZ, V. Molecular identification and virulence potential of the genus *Aeromonas* isolated from wild rainbow trout (*Oncorhynchus mykiss*) in Mexico. **Journal of Food Protection**, 82(10), 1706-1713. 2019.

RICHTER, M., ROSSELLÓ-MÓRA, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. **Proceedings of the National Academy of Sciences**, 106(45), 19126-19131. 2009.

ROGER, F., MARCHANDIN, H., JUMAS-BILAK, E., KODJO, A., & LAMY, B. Multilocus genetics to reconstruct aeromonad evolution. **BMC microbiology**, 12(1), 1-23. 2012.

ROMERO, A.; SARACENI, P. R.; MERINO, S.; et al. The animal model determines the results of *Aeromonas* virulence factors. **Frontiers in Microbiology**, v. 7, n. OCT, 2016. Frontiers Media S.A.

ROSSATO, A. M., PREDIGER, K. D. C., GOMES, C. N., NETO, C. H. D. P. W., PEREIRA, R. D. S., DALLAGASSA, C. D. B., VIZZOTTO, B. S. Molecular identification of aeromonas spp isolated from patients with diarrhea at santa maria-rs, Brazil. **Medicina Tropical**, 41, 179-82. 2008.

ROSSELLÓ-MÓRA, R., WHITMAN, W. B. Dialogue on the nomenclature and classification of prokaryotes. **Systematic and applied microbiology**, 42(1), 5-14. 2019.

SALVAT, M.J.F., ASHBOLT, N. **Global Water Pathogen Project**. University of Alberta; Edmonton, AB, Canada: 2019.

SÁNCHEZ-CÉSPEDES, J.; FIGUERAS, M. J.; ASPIROZ, C.; et al. Development of imipenem resistance in an *Aeromonas veronii* biovar *sobria* clinical isolate recovered from a patient with cholangitis. **Journal of Medical Microbiology**, v. 58, n. 4, p. 451–455, 2009.

SANFORD, R. A., LLOYD, K. G., KONSTANTINIDIS, K. T., LÖFFLER, F. E. Microbial taxonomy run amok. **Trends in Microbiology**, 29(5), 394-404. 2021.

SCHMIDT, H.; HENSEL, M. Pathogenicity Islands in Bacterial Pathogenesis. **Clinical Microbiology Reviews**, Jan. 2004.

SCHOUTEN, J. P., MCELGUNN, C. J., WAAIJER, R., ZWIJNENBURG, D., DIEPVENS, F., & PALS, G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. **Nucleic acids research**, 30(12), e57-e57. 2002.

SCHUBERT, R. H. The taxonomy and nomenclature of the genus *Aeromonas* Kluyver and van Niel 1936: Part I. Suggestions on the taxonomy and nomenclature of the aerogenic *Aeromonas* species. **International Journal of Systematic and Evolutionary Microbiology**, 17(1), 23-37. 1967.

SELVARAJ, G. K., WANG, H., ZHANG, Y., TIAN, Z., CHAI, W., & LU, H. Class 1 In-Tn5393c array contributed to antibiotic resistance of non-pathogenic *Pseudoxanthomonas mexicana* isolated from a wastewater bioreactor treating streptomycin. **Science of The Total Environment**, 821, 153537. 2022.

SESHADRI, R.; JOSEPH, S. W.; CHOPRA, A. K.; et al. Genome sequence of *Aeromonas hydrophila* ATCC 7966T: Jack of all trades. **Journal of Bacteriology**, v. 188, n. 23, p. 8272–8282, 2006.

SHRIVASTAVA, S.; SIVA KUMAR REDDY, C. V.; MANDE, S. S. Indegenius, a new method for high-throughput identification of specialized functional islands in completely sequenced organisms. **Journal of Biosciences**, v. 35, n. 3, p. 351–364, 2010.

SIMON, H. Y., SIDDLE, K. J., PARK, D. J., SABETI, P. C. Benchmarking metagenomics tools for taxonomic classification. **Cell**, 178(4), 779-794. 2019.

SLEVIN, N. J.; OPPENHEIM, B. A.; DEAKIN, D. P. **Aeromonas hydrophila septicaemia and muscle abscesses associated with immunosuppression**. 1988.

SNEATH, P. H. A. Evidence from *Aeromonas* for genetic crossing-over in ribosomal sequences. **Int J Syst Bacteriol**. 43:626–629. 1993.

SNEATH, P. H. A. International code of nomenclature of bacteria. **American Society for Microbiology**. Washington, DC. 1992.

SOARES, S. C.; ABREUVINÍ, V. A. C.; RAMOS, R. T. J.; et al. PIPS: Pathogenicity island prediction software. **PLoS ONE**, v. 7, n. 2, 2012.

SOARES, S. C.; GEYIK, H.; RAMOS, R. T. J.; et al. GIPSy: Genomic island prediction software. **Journal of Biotechnology**, v. 232, p. 2–11, 2016. Elsevier B.V.

SOLER, L., YANEZ, M. A., CHACON, M. R., AGUILERA-ARREOLA, M. G., CATALÁN, V., FIGUERAS, M. J., MARTÍNEZ-MURCIA, A. J. Phylogenetic analysis of the genus *Aeromonas* based on two housekeeping genes. **International journal of systematic and evolutionary microbiology**, 54(5), 1511-1519. 2004.

SORIANO VARGAS, EDGARDO et al. Caracterización fenotípica y genotípica de aislamientos de *aeromonas* spp. obtenidos de trucha arcoíris (*oncorhynchus mykiss*). 2014.

SOUCY, S. M.; HUANG, J.; GOGARTEN, J. P. Horizontal gene transfer: Building the web of life. **Nature Reviews Genetics**, 21. Aug. 2015. Nature Publishing Group.

STACKEBRANDT E., FREDERIKSEN W., GARRITY G. M. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. **Int J Syst Evol Microbiol**. 2002.

STACKEBRANDT, E., FREDERIKSEN, W., GARRITY, G. M., GRIMONT, P. A., KÄMPFER, P., MAIDEN, M. C., WHITMAN, W. B. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. **International journal of systematic and evolutionary microbiology**, 52(3), 1043-1047. 2002.

SUN, Y., ZHAO, Y., XU, W., FANG, R., WU, Q., HE, H., ZHOU, T. Taxonomy, virulence determinants and antimicrobial susceptibility of *Aeromonas* spp. isolated from bacteremia in southeastern China. **Antimicrobial Resistance & Infection Control**, 10(1), 1-9. 2021.

SUREK, M., VIZZOTTO, B. S., SOUZA, E. M., PEDROSA, F. D. O., DALLAGASSA, C. B., FARAH, S. M., & FADEL-PICHETH, C. M. Identification and antimicrobial susceptibility of *Aeromonas* spp. isolated from stool samples of Brazilian subjects with diarrhoea and healthy controls. **Journal of medical microbiology**, 59(3). 2010.

SUSSMAN, J. I.; BARON, E. J.; TENENBAUM, M. J.; et al. **Viridans Streptococcal Endocarditis: Clinical, Microbiological, and Echocardiographic Correlations**. 1986.

- TALAGRAND-REBOUL, E., ROGER, F., KIMPER, J. L., COLSTON, S. M., GRAF, J., LATIF-EUGENÍN, F., LAMY, B. Delineation of taxonomic species within complex of species: *Aeromonas media* and related species as a test case. **Frontiers in microbiology**, 8, 621. 2017.
- TANAKA, K. H.; DALLAIRE-DUFRESNE, S.; DAHER, R. K.; FRENETTE, M.; CHARETTE, S. J. An insertion sequence-dependent plasmid rearrangement in *aeromonas salmonicida* causes the loss of the type three secretion system. **PLoS ONE**, v. 7, n. 3, 2012.
- TANAKA, K. H.; VINCENT, A. T.; EMOND-RHEAULT, J. G.; et al. Plasmid composition in *Aeromonas salmonicida* subsp. *salmonicida* 01-B526 unravels unsuspected type three secretion system loss patterns. **BMC Genomics**, v. 18, n. 1, 2017. BioMed Central Ltd.
- TEKEDAR, H. C., KUMRU, S., BLOM, J., PERKINS, A. D., GRIFFIN, M. J., ABDELHAMED, H., LAWRENCE, M. L. Comparative genomics of *Aeromonas veronii*: Identification of a pathotype impacting aquaculture globally. **PloS one**, 14(8), e0221018. 2019.
- TENA, D.; GONZÁLEZ-PRAETORIUS, A.; GIMENO, C.; PÉREZ-POMATA, M. T.; BISQUERT, J. Extraintestinal infection due to *Aeromonas* spp.: Review of 38 cases. **Enfermedades Infecciosas y Microbiología Clínica**, v. 25, n. 4, p. 235–241, 2007. Ediciones Doyma, S.L.
- THOMAS, S. G., GLOVER, M. A., PARTHASARATHY, A., WONG, N. H., SHIPMAN, P. A., HUDSON, A. O. Expression of a shiga-like toxin during plastic colonization by two multidrug-resistant bacteria, *Aeromonas hydrophila* RIT668 and *Citrobacter freundii* RIT669, isolated from endangered turtles (*Clemmys guttata*). **Microorganisms**, 8(8), 1172. 2020.
- TOMÁS, J. M. The Main *Aeromonas* Pathogenic Factors. **ISRN Microbiology**, v. 2012, p. 1–22, 2012. Hindawi Limited.
- TOUSSAINT, A.; MERLIN, C. Mobile elements as a combination of functional modules. **Plasmid**, v. 47, n. 1, p. 26–35, 2002. Academic Press Inc.
- TREANGEN, T. J.; ROCHA, E. P. C. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. **PLoS Genetics**, v. 7, n. 1, 2011.
- TSAI, Y. H.; SHEN, S. H.; YANG, T. Y.; et al. Monomicrobial Necrotizing Fasciitis Caused by *Aeromonas hydrophila* and *Klebsiella pneumoniae*. **Medical Principles and Practice**, v. 24, n. 5, p. 416–423, 2015. S. Karger AG.
- TSENG, T. T.; TYLER, B. M.; SETUBAL, J. C. Protein secretion systems in bacterial-host associations, and their description in the Gene Ontology. **BMC Microbiology**, 2009.
- TU, Q.; DING, D. Detecting pathogenicity islands and anomalous gene clusters by iterative discriminant analysis. **FEMS Microbiology Letters**, v. 221, n. 2, p. 269–275, 2003.
- TYRRELL, C.; BURGESS, C. M.; BRENNAN, F. P.; WALSH, F. Antibiotic resistance in grass and soil. **Biochemical Society Transactions**, 2019. Portland Press Ltd.
- UEDA, K.; SEKI, † TATSUJI; KUDO, T.; YOSHIDA, T.; KATAOKA, M. **Two Distinct Mechanisms Cause Heterogeneity of 16S rRNA**. 1999.
- VARELA, A. R.; NUNES, O. C.; MANAIA, C. M. Quinolone resistant *Aeromonas* spp. as carriers and potential tracers of acquired antibiotic resistance in hospital and municipal wastewater. **Science of the Total Environment**, v. 542, p. 665–671, 2016. Elsevier.
- VÁVROVÁ, A., BALÁŽOVÁ, T., SEDLÁČEK, I., TVRZOVÁ, L., & ŠEDO, O. Evaluation of the MALDI-TOF MS profiling for identification of newly described *Aeromonas* spp. **Folia**

microbiologica, 60(5), 375-383. 2015.

VEGA-SÁNCHEZ V. Caracterización fenotípica y genotípica de aislamientos de *Aeromonas* spp. obtenidos de trucha arcoíris (*Oncorhynchus mykiss*) **Universidad Autónoma del Estado de México**; Toluca, Mexico: 2015.

VEGA-SÁNCHEZ, V.; LATIF-EUGENÍN, F.; SORIANO-VARGAS, E.; et al. Re-identification of *Aeromonas* isolates from rainbow trout and incidence of class 1 integron and β -lactamase genes. **Veterinary Microbiology**, v. 172, n. 3–4, p. 528–533, 2014. Elsevier B.V.

VERNIKOS G.S. 2008. In Silico Prediction of Genomic Islands in Microbial Genomes. **(Ph.D. thesis)**.

VERNIKOS, G. S.; PARKHILL, J. Interpolated variable order motifs for identification of horizontally acquired DNA: Revisiting the *Salmonella* pathogenicity islands. **Bioinformatics**, v. 22, n. 18, p. 2196–2203, 2006.

VERNIKOS, G. S.; PARKHILL, J. Resolving the structural features of genomic islands: A machine learning approach. **Genome Research**, v. 18, n. 2, p. 331–342, 2008.

VINCENT, A. T.; FERNÁNDEZ-BRAVO, A.; SANCHIS, M.; et al. Investigation of the virulence and genomics of *Aeromonas salmonicida* strains isolated from human patients. **Infection, Genetics and Evolution**, v. 68, p. 1–9, 2019. Elsevier B.V.

VINGA, S. Alignment-free methods in computational biology. **Briefings in Bioinformatics**, 15(3), 341-342. 2014.

VINGA, S., ALMEIDA, J. Alignment-free sequence comparison—a review. **Bioinformatics**, 19(4), 513-523. 2003.

WAACK, S.; KELLER, O.; ASPER, R.; et al. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. **BMC Bioinformatics**, v. 7, 2006.

WILLIAMS, K. P. **Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies**. 2002.

WOODRING, J.; SRIJAN, A.; PURIPUNYAKOM, P.; et al. Prevalence and Antimicrobial susceptibilities of *Vibrio*, *Salmonella*, and *Aeromonas* isolates from various uncooked seafoods in Thailand. **Journal of Food Protection**, v. 75, n. 1, p. 41–47, 2012. International Association for Food Protection.

WU, C. J., WANG, H. C., CHEN, C. S., SHU, H. Y., KAO, A. W., CHEN, P. L., & KO, W. C. Genome sequence of a novel human pathogen, *Aeromonas aquariorum*. **Journal of Bacteriology**. 2012.

WU, C. J.; CHEN, P. L.; TANG, H. J.; et al. Incidence of *Aeromonas* bacteremia in southern Taiwan: *Vibrio* and *Salmonella* bacteremia as comparators. **Journal of Microbiology, Immunology and Infection**, v. 47, n. 2, p. 145–148, 2014.

YAMAMOTO S., BOUVET P. J. M., HARAYAMA S. Phylogenetic structures of the genus *Acinetobacter* based on *gyrB* sequences: comparison with the grouping by DNA–DNA hybridization. **Int J Syst Bacteriol**. 1999.

YAMAMOTO S., HARAYAMA S. Phylogenetic analysis of *Acinetobacter* strains based on the nucleotide sequences of *gyrB* genes and on the amino acid sequences of their products. **Int J Syst Bacteriol**. 1996.

- YAMASAKI, O., TAKAHASHI, E., NODA, K., KANAYA, N., TANABE, S., SHIRAKAWA, Y., MORIZANE, S. Surgical site infection caused by *Aeromonas hydrophila* presenting as necrotizing soft tissue infection after esophagectomy. **The Journal of Dermatology**, 47(6), 673-676. 2020.
- YANEZ, M. A., CATALÁN, V., APRÁIZ, D., FIGUERAS, M. J., & MARTÍNEZ-MURCIA, A. J. Phylogenetic analysis of members of the genus *Aeromonas* based on *gyrB* gene sequences. **International journal of systematic and evolutionary microbiology**, 53(3), 875-883. 2003.
- YANO, Y.; HAMANO, K.; TSUTSUI, I.; et al. Occurrence, molecular characterization, and antimicrobial susceptibility of *Aeromonas* spp. in marine species of shrimps cultured at inland low salinity ponds. **Food Microbiology**, v. 47, p. 21–27, 2015. Academic Press.
- ZHANG R., OU H.Y., GAO F., LUO H. Identification of horizontally-transferred Genomic Islands and genome segmentation points by using the GC profile method. **Curr Genomics**. 2014.
- ZHANG, Q., JUN, S. R., LEUZE, M., USSERY, D., & NOOKAEW, I. Viral phylogenomics using an alignment-free method: A three-step approach to determine optimal length of k-mer. **Scientific reports**, 7(1), 1-13. 2017.
- ZHOU, Y., YU, L., NAN, Z., ZHANG, P., KAN, B., YAN, D., & SU, J. Taxonomy, virulence genes and antimicrobial resistance of *Aeromonas* isolated from extra-intestinal and intestinal infections. **BMC infectious diseases**, 19(1), 1-9. 2019.
- ZHU, Q., MAI, U., PFEIFFER, W., JANSSEN, S., ASNICAR, F., SANDERS, J. G., KNIGHT, R. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. **Nature communications**, 10(1), 1-14. 2019.
- ZIELEZINSKI, A., VINGA, S., ALMEIDA, J., KARLOWSKI, W. M. Alignment-free sequence comparison: benefits, applications, and tools. **Genome biology**, 18(1), 1-17. 2017.