

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO ACADÊMICO DE ENGENHARIA DE PRODUÇÃO
CURSO DE ESPECIALIZAÇÃO EM BIOINFORMÁTICA

BÁRBARA MORIEL

**BANCO DE DADOS DOS GENES PERTENCENTES AO SISTEMA DE
SECRETÃO TIPO VI PRESENTE NAS ESPÉCIES DO GÊNERO *AEROMONAS***

MONOGRAFIA DE ESPECIALIZAÇÃO

LONDRINA
2016

BÁRBARA MORIEL

**BANCO DE DADOS DOS GENES PERTENCENTES AO SISTEMA DE
SECREÇÃO TIPO VI PRESENTE NAS ESPÉCIES DO GÊNERO *AEROMONAS***

Monografia apresentada ao Curso de Especialização em Bioinformática da Universidade Tecnológica Federal do Paraná, como requisito parcial para a obtenção do título de Especialista em Bioinformática.

Orientador: Prof. Dr. Alessandro Botelho Bovo.

LONDRINA
2016

**BANCO DE DADOS DOS GENES PERTENCENTES AO SISTEMA DE
SECREÇÃO TIPO VI PRESENTE NAS ESPÉCIES DO GÊNERO *AEROMONAS***

Monografia apresentada ao Curso de Especialização em Bioinformática da Universidade Tecnológica Federal do Paraná, como requisito parcial para a obtenção do título de Especialista em Bioinformática.

Prof. Dr. Alessandro Botelho Bovo
Universidade Tecnológica Federal do Paraná

Prof. Dr. Fabrício Martins Lopes
Universidade Tecnológica Federal do Paraná

Prof. Dr. Alexandre Rossi Paschoal
Universidade Tecnológica Federal do Paraná

Londrina, 22 de agosto de 2016

Banco de Dados dos Genes Pertencentes ao Sistema de Secreção Tipo VI Presente nas Espécies do Gênero *Aeromonas*

Resumo

As espécies do gênero *Aeromonas* são bactérias em forma de bacilos gram-negativos e muitas são patógenos humanos. Com o advento das novas técnicas de sequenciamento, este gênero tem sido melhor pesquisado. Dados genéticos são introduzidos diariamente em uma rede de banco de dados, porém nem sempre são verificadas a veracidade e a qualidade desses dados. Assim, informações duvidosas podem ser propagadas, levando a resultados e conclusões erradas. Dadas essas dificuldades, foi criado um banco de dados para o sistema de secreção tipo VI em *Aeromonas*, a fim de facilitar a busca, análise e aquisição das sequências proteicas pertencentes a esse sistema que está relacionado diretamente com virulência. Para a construção do banco de dados foram obtidos os genomas das espécies de *Aeromonas* depositadas no NCBI. Esses genomas foram anotados pelo software RAST e as sequências comparadas com a espécie referência *Aeromonas hydrophila* SSU. A anotação automática realizada no RAST mostrou que das 115 espécies de *Aeromonas*, 72 apresentavam o sistema de secreção tipo IV. O banco de dados está disponível publicamente em [http:// bio.seita.co](http://bio.seita.co).

Palavras-chaves: Sistema de secreção tipo VI; *Aeromonas*; Virulência.

A database of genes belonging to the type VI secretion system in species of genus *Aeromonas*

Abstract

Species of the genus *Aeromonas* are gram-negative bacteria and many are human pathogens. With the advent of new sequencing techniques, this genus has been better investigated. Genetic data have been introduced into a database network on a daily basis. However, but the truth and quality of this data is not always verified. Thus, doubtful information can be propagated, leading to misleading results and conclusions. Given these difficulties, a database has been created for the type VI

secretion system in *Aeromonas*, in order to facilitate the search, analysis and acquisition of protein sequences belonging to this system, which is closely related to virulence. In order to do that, we obtained the genomes of the *Aeromonas* species deposited in the NCBI. These genomes have been annotated by the RAST software and the sequences compared to the reference species *Aeromonas hydrophila* SSU. The automatic annotation performed by the RAST showed that of the 115 *Aeromonas* species, 72 had the type IV secretion system. The database is publicly available at: <http://http://bio.seita.co>.

Keywords: Type VI Secretion System; *Aeromonas*; Virulence.

1. INTRODUÇÃO

As espécies do gênero *Aeromonas* são bactérias em forma de bacilos gram-negativos. Muitas delas são patógenos humanos capazes de causar uma variedade de infecções que vão de gastroenterite à septicemia, sendo frequentemente associadas com diarreia (JANDA; ABBOTT, 2010; PARKER; SHAW, 2011).

No país ainda existem poucos estudos (HOFER et al., 2006, GUERRA et al., 2007; PEREIRA et al., 2008; SUREK et al., 2010; ASSIS, 2012) disponíveis sobre essas bactérias. Porém com o advento das novas técnicas de sequenciamento juntamente com sua redução de custo, proporcionou um aumento nas investigações sobre este gênero e, com isso, os dados genômicos disponíveis sobre este grupo têm crescido abundantemente.

Essas bactérias apresentam diversos fatores associados com virulência, ou seja, a capacidade de causar doença. Dentre eles, está o sistema de secreção tipo VI (T6SS), o qual é de grande importância na interação de bactérias patogênicas com células hospedeiras (TOMÁS, 2012).

O sistema é amplamente encontrado em grande parte das bactérias gram-negativas e parece ser muito versátil. Pode ser utilizado para atingir as células eucarióticas ou células bacterianas concorrentes e provavelmente desempenha um papel importante nas interações inter-bacterianas, estrutura das comunidades em todos os tipos de nichos ambientais, na capacidade de muitas bactérias patogênicas para competir

com rivais e colonizar locais de infecção altamente competitivos, pode atuar também na interação direta de certos patógenos e simbioses com seus hospedeiros eucarióticos (SUAREZ et al., 2008; COULTHURST, 2013).

As sequências de DNA e RNA obtidas em pesquisas podem ser encontrados em uma rede de banco de dados chamada *International Nucleotide Sequence Database Collaboration* (INSDC – <http://www.insdc.org>), da qual fazem parte o *National Center for Biotechnology Information* (NCBI – <http://www.ncbi.nlm.nih.gov>), o *European Molecular Biology Laboratory* (EMBL – <https://www.ebi.ac.uk>) e DNA Data Bank of Japan (DDJB – <http://www.ddbj.nig.ac.jp>). Nestes bancos de dados a introdução de várias sequências são feitas diariamente, porém nem sempre a veracidade e a qualidade desses dados são confirmados ou verificados. Assim, informações errôneas compartilhadas por esses bancos de dados podem ser altamente propagadas, levando a uma conclusão ou resultados fora da realidade biológica.

Dessa maneira, é importante realizar uma apuração dos resultados e consultas obtidos nesses bancos de dados, a fim evitar a disseminação de erros. É também importante reportar e corrigir as falhas já existentes, melhorando assim, a qualidade das informações disponíveis nos bancos de dados para consulta pública.

A partir da dificuldade em se analisar a veracidade de uma sequência e do crescente interesse pelas espécies do gênero *Aeromonas*, foi criado um banco de dados para o sistema de secreção tipo VI dessas bactérias, a fim de facilitar a consulta, análise e aquisição das sequências proteicas dos genes pertencentes a esse sistema intimamente relacionado com virulência.

2. MATERIAIS E MÉTODOS

2.1. Amostragem

Para a montagem do banco de dados do sistema de secreção tipo IV foram obtidos os genomas completos e incompletos, em formato FASTA, das espécies de *Aeromonas* depositadas no banco de dados do NCBI até o mês de outubro de 2015.

2.2. Anotação dos genomas

Anotação automática

Os genomas foram anotados automaticamente por meio do *web service* RAST (*Rapid Annotation using Subsystem Technology* – <http://rast.nmpdr.org>) (AZIZ et al., 2008; OVERBEEK et al., 2014; BRETTIN et al., 2015) com o objetivo de selecionar as bactérias que apresentam o T6SS.

Anotação manual

Como a anotação automática pode gerar erros, foi necessária a verificação manual dos genes e sequências anotados para atestar a fidelidade e compatibilidade com os dados biológicos.

Primeiramente, no *web service* RAST, foram selecionadas as bactérias que apresentavam o T6SS. Para verificação das sequências que seriam introduzidas no banco de dados, foi necessário a escolha de uma espécie, priorizando *Aeromonas*, a qual apresentava um sistema de secreção funcional e de preferência comprovado experimentalmente, para ser utilizada como referência.

A verificação manual das sequências dos genes do T6SS foi realizada através da comparação com a espécie de referência utilizando o algoritmo de alinhamento local BLASTp (*Protein Basic Local Alignment Search Tool*), que se encontra disponível no próprio RAST.

2.3. Critérios para a seleção das sequências

Foram utilizados os seguintes critérios para a estruturação do banco de dados e seleção das bactérias a serem introduzidas: (a) o sistema de secreção deveria contemplar todos ou a maioria dos genes; (b) exclusão dos genes que apresentavam sequência incompleta, interrompidas por *stop codons* ou com mudanças de *frame (frameshift)* em relação a sequência de referência; (c) foram obtidas somente as sequências de aminoácidos, pois a sequência de nucleotídeos de alguns genes não estava disponível no banco de dados do NCBI.

2.4. Banco de dados

Inicialmente foi necessário especificar um modelo de dados para representar as informações do domínio. Para isso, criou-se um modelo entidade-relacionamento por meio da ferramenta MySQL Workbench. O Sistema Gerenciador de Banco de Dados utilizado foi o MySQL.

As sequências de aminoácidos dos genes do T6SS foram salvas em formato FASTA e introduzidas no banco de dados por meio de um *script* personalizado em linguagem Python.

2.5. Aplicação WEB

Para o desenvolvimento da aplicação WEB foi adotada a linguagem Python, por meio do *framework* Flask, e as linguagens HTML e CSS.

2. RESULTADOS E DISCUSSÃO

2.6. Amostragem

No banco de dados do NCBI até o mês de outubro de 2015 foram sequenciadas 115 espécies de *Aeromonas*, dessas 16 apresentavam o genoma completo e 134 genomas incompleto (*draft*). A listagem com as espécies bacterianas utilizadas estão disponíveis no material suplementar 1 (Apêndice A).

2.7. Anotação dos genomas

Anotação automática

A anotação automática realizada no RAST revelou que dentre as 115 espécies de *Aeromonas* analisadas 72 apresentavam o sistema de secreção tipo VI, sendo 9 genomas completos e 63 genomas incompletos. No material suplementar 2 (Apêndice B) estão listadas, em ordem alfabética, todas as espécies que

apresentam esse sistema e a quantidade de sequências para cada gene anotado. As espécies destacadas em negrito possuem o genoma completo.

Anotação manual

Como a anotação automática é passível de erros, foi necessário fazer uma verificação manual dos genes e sequências para assegurar a veracidade dos dados e apurar se eles são condizentes com as informações biológicas.

A partir disso, foi selecionada a espécie *Aeromonas hydrophila* SSU como referência para a anotação manual, pois já existiam dados da literatura (SUAREZ et al., 2008, 2010a, 2010b; SHA et al., 2013; GRIM et al., 2014) mostrando experimentalmente a presença e funcionalidade do T6SS. As sequências de proteínas do sistema de secreção tipo VI em *Aeromonas hydrophila* SSU estão disponíveis no banco de dados do NCBI sob o número de acesso DQ667172.

No próprio anotador RAST as sequências de aminoácidos do T6SS foram alinhadas por meio do algoritmo BLASTp com as sequências da espécie de referência, listadas no material suplementar 3 (Apêndice C).

Para a análise das sequências e montagem do banco de dados foi adotada a nomenclatura descrita por Suarez (2008) e Coulthurst (2013) para as espécies de *Aeromonas*, por isso os nomes estão diferenciados dos mostrados pelo anotador RAST. Abaixo segue uma tabela com o nome de cada gene e seus respectivos sinônimos.

Tabela 01: Nomenclatura adotada para a montagem do banco de dados e seus respectivos sinônimos.

Utilizada	RAST
Hcp-2	Hcp
VgrG-2	VgrG
ImpB	
ImpC	
Pvc109	
VasA	ImpG
VasB	ImpH

VasC	Impl
VasD	
VasE	ImpJ
VasF	MotB
ClpB	
VasH	
VasI	
ImpA	
VasK	IcmF
hipotética	
uropatogênica	
VgrG-3	VgrG
Hcp-1	Hcp
VgrG-1	VgrG

Alguns genes do T6SS não foram anotados automaticamente pelo RAST, sendo o principal deles o Hcp. Também não foram anotados os genes MotB, uma proteína hipotética conservada e uma proteína uropatogênica. Além disso, as proteínas VgrG-1, VgrG-2 e VgrG-3, presentes na espécie de referência não possuem essa classificação no RAST, o que dificultou bastante a análise, pois elas são muito similares entre si, não sendo possível indicar com precisão a qual gene estava relacionada, levando, assim, à exclusão de várias espécies.

2.8. Critérios para a seleção das sequências

A partir dos critérios de seleção foram excluídas 31 bactérias das 72 anotadas com o T6SS. A Tabela 2 apresenta de forma resumida a quantidade de espécies utilizadas na análise, a distribuição do T6SS nas espécies de *Aeromonas* e a quantidade de espécies introduzidas no banco de dados de acordo com os critérios de seleção.

Tabela 02: Distribuição da quantidade de bactéria por espécie encontradas no banco de dados de NCBI, anotadas com a presença do T6SS pelo RAST e selecionadas segundo os critérios pré-estabelecidos.

Espécies	NCBI	T6SS	Selecionadas
<i>Aeromonas allosaccharophila</i>	2	0	0
<i>Aeromonas aquatica</i>	1	1	0
<i>Aeromonas australiensis</i>	1	0	0
<i>Aeromonas bestiarum</i>	1	1	1
<i>Aeromonas bivalvium</i>	1	1	1
<i>Aeromonas caviae</i>	11	6	3
<i>Aeromonas dhakensis</i>	4	4	3
<i>Aeromonas diversa</i>	2	0	0
<i>Aeromonas encheleia</i>	1	1	1
<i>Aeromonas enteropelogenes</i>	4	4	1
<i>Aeromonas eucrenophila</i>	1	1	1
<i>Aeromonas finlandiensis</i>	1	1	0
<i>Aeromonas fluvialis</i>	1	0	0
<i>Aeromonas hydrophila</i>	38	26	16
<i>Aeromonas jandaei</i>	3	3	0
<i>Aeromonas lacus</i>	1	1	1
<i>Aeromonas media</i>	4	0	0
<i>Aeromonas molluscorum</i>	1	0	0
<i>Aeromonas piscicola</i>	1	1	1
<i>Aeromonas popoffii</i>	1	1	0
<i>Aeromonas rivuli</i>	1	1	0
<i>Aeromonas salmonicida</i>	10	7	6
<i>Aeromonas sanarellii</i>	1	0	0
<i>Aeromonas schubertii</i>	1	1	0
<i>Aeromonas simiae</i>	1	1	1
<i>Aeromonas sobria</i>	1	1	1
<i>Aeromonas sp.</i>	6	3	0
<i>Aeromonas taiwanensis</i>	1	0	0
<i>Aeromonas tecta</i>	1	1	1
<i>Aeromonas veronii</i>	12	5	2
TOTAL	115	72	40

2.9. Banco de dados

Para a estruturação dos dados foram criadas duas tabelas, uma contendo as espécies bacterianas e outra contendo os genes. Essas duas tabelas foram

relacionadas através de uma relação muitos para muitos (n:m), pois o T6SS apresenta vários genes, os quais estão presentes em várias bactérias, bem como as várias bactérias apresentam o mesmo gene para o T6SS. Na tabela relacional foram introduzidas as sequências de aminoácidos dos genes.

Na tabela *bacteria* foram introduzidos os nomes das espécie, já na tabela *gene* foram adicionados os nomes dos genes do T6SS. Na tabela relacional *sequencia_bacteria_gene* foram inseridas as sequências de aminoácidos dos genes de cada bactéria. Por meio dessa tabela foi possível fazer uma relação entre a sequência de aminoácidos com o gene e a bactéria, e também foi possível relacionar as tabelas *bacteria* e *gene*. A Figura 1 exemplifica o modelo utilizado no banco de dados.

Dessa maneira, é possível fazer uma consulta no banco de dados de quais genes estão presentes em determinada bactéria, quais são suas sequências de aminoácidos, bem como, quais bactérias apresentam determinado gene e quais as sequências relacionadas à elas.

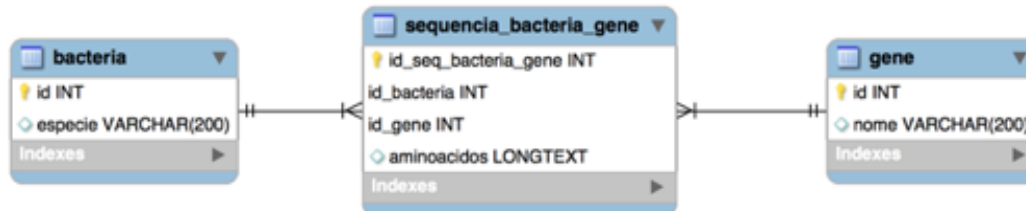


Figura 01: Modelo do banco de dados do sistema de secreção tipo VI feito no MySQL Workbench, mostrando a relação n:m entre as tabelas *bacteria* e *gene*.

2.10. Aplicação WEB

A primeira versão do banco de dados está disponível para consulta pública em <http://bio.seita.co>. Nesta aplicação é possível fazer uma busca dos genes do T6SS presentes em uma determinada espécie, bem como quais espécies apresentam um determinado gene. Também é possível extrair a sequência de aminoácidos de cada gene em cada bactéria. Outras funcionalidades como o alinhamento de sequências através do algoritmo BLAST serão adicionadas posteriormente.

4. REFERÊNCIAS

ASSIS, F. E. A. **Prevalência e perfil de susceptibilidade aos antimicrobianos de bactérias enteropatogênicas isoladas no Paraná.** 2012. Dissertação (Mestrado em Ciências Farmacêuticas) - Programa de Pós-Graduação em Ciências Farmacêuticas, Universidade Federal do Paraná, 2012.

AZIZ, R. K.; BARTELS, D.; BEST, A. A.; DEJONGH, M.; DISZ, T.; EDWARDS, R. A.; FORMSMA, K.; GERDES, S.; GLASS, E. M.; KUBAL, M.; MEYER, F.; OLSEN, G. J.; OLSON, R.; OSTERMAN, A. L.; OVERBEEK, R. A.; MCNEIL, L. K.; PAARMANN, D.; PACZIAN, T.; PARRELLO, B.; PUSCH, G. D.; REICH, C.; STEVENS, R.; VASSIEVA, O.; VONSTEIN, V.; WILKE, A.; ZAGNITKO, O. The RAST Server: rapid annotations using subsystems technology. **BMC Genomics**, 2008.

BRETTIN T., DAVIS, J. J.; DIZ, T.; EDWARDS, R. Aa; GERDES, S.; OLSEN, G. J.; OLSON, R.; OVERBEEK, R.; PARRELLO, B.; PUSCH, G. D.; SHUKLA, M.; THOMASON, J. A.; STEVENS, R.; VONSTEIN, V; WATTAM, A.R.; XIA, F. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. **Sci Rep.** 5:8365, 2015.

COULTHURST, S. J. The Type VI secretion system e a widespread and versatile cell targeting system. **Research in Microbiology**, v. 164, p. 640-654, 2013.

GRIM, C. J; KOZLOVA, E. V.; PONNUSAMY, D.; FITTS, E. C.; SHA, J.; KIRTLEY, M. L.; VAN LIER, C. J.; TINER, B. L.; EROVA, T. E.; JOSEPH, S. J.; READ, T. D.; SHAK, J. R.; JOSEPH, S. W.; SINGLETARY, E.; FELLAND, T.; BAZE, W. B.; HORNEMAN, A. J., CHOPRA, A. K. Functional Genomic Characterization of Virulence Factors from Necrotizing Fasciitis-Causing Strains of *Aeromonas hydrophila*. **Appl Environ Microbiol.** 80(14):4162-83, 2014.

GUERRA, I. M. F.; FADANELLI, R.; FIGUEIRÓ, M.; SCHREINER, F.; DELAMARE, A. P. L.; WOLLHEIM, C.; COSTA, S. O. P.; ECHEVERRIGARAY, S. *Aeromonas* associated diarrhoeal disease in South Brazil: prevalence, virulence factors and

antimicrobial resistance. **Brazilian Journal of Microbiology.** v. 38, p. 638-643, 2007.

HOFER, E.; dos REIS, C. M. F.; THEOPHILO, G. N. D.; CAVALCANTI, V. O.; de LIMA, N. V.; HENRIQUES, M. F. C. M. Envolvimento de *Aeromonas* em surto de doença diarreica aguda em São Bento do Una, Pernambuco. **Revista da Sociedade Brasileira de Medicina Tropical.** v. 39, n. 2, p. 217-220, 2006.

JANDA, J. M.; ABBOTT, S. L. The Genus *Aeromonas*: Taxonomy, Pathogenicity, and Infection. **Clinical Microbiology Reviews.** v. 23, n. 1, p. 35-73, 2010.

OVERBEEK, R.; OLSON, R.; PUSH, G. D.; OLSEN, G. J.; DAVIS, J. J.; DISZ, T.; EDWARDS, R. A.; GERDES, S.; PARRELLO, B.; SHUKLA, M.; VONSTEIN, V.; WATTAM, A. R.; XIA, F.; STEVENS, R. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). **Nucleic Acids Research.** 2014.

PARKER, J. L.; SHAW J. G. *Aeromonas* spp. clinical Microbiology and Disease. **Journal of Infection.** p1-10, 2010.

PEREIRA, C. S.; AMORIM, S. D.; SANTOS, A. F. M.; dos REIS, C. M. F.; THEOPHILO, G. N. D.; RODRIGUES, D. P. Caracterização de *Aeromonas* spp isoladas de neonatos hospitalizados. **Revista da Sociedade Brasileira de Medicina Tropical.** v. 41, n. 2, p. 179-182, 2008.

SHA, J.; ROSENZWEIG, J. A.; KOZLOVA, E. V.; WANG, S.; EROVA, T. E.; KIRTLEY, M. L.; VAN LIER, C. J.; CHOPRA, A. K. Evaluation of the roles played by Hcp and VgrG type 6 secretion system effectors in *Aeromonas hydrophila* SSU pathogenesis. **Microbiology.** 159 (Pt 6):1120-35.

SUAREZ, G.; SIERRA, J. C.; SHA, J.; WANG, S.; EROVA, T. E.; FADL, A. A.; FOLTZ, S. M.; HORNEMAN, A. J.; CHOPRA, A. K. Molecular characterization of a functional type VI secretion system from a clinical isolate of *Aeromonas hydrophila*. **Microb Pathog.** 44(4):344-61, 2008.

SUAREZ, G.; SIERRA, J. C.; EROVA, T. E; SHA, J.; HORNEMAN, A. J.; CHOPRA, A. K. A type VI secretion system effector protein, VgrG1, from *Aeromonas hydrophila* that induces host cell toxicity by ADP ribosylation of actin. **J Bacteriol.** 192(1):155-68, 2010a.

SUAREZ, G.; SIERRA, J. C.; KIRTLEY, M. L.; CHOPRA, A. K. Role of Hcp, a type 6 secretion system effector, of *Aeromonas hydrophila* in modulating activation of host immune cells. **Microbiology.** 156(Pt 12):3678-88, 2010b.

SUREK, M.; VIZZOTTO, B. S.; SOUZA, E. M.; PEDROSA, F. O.; DALLAGASSA, C. B.; FARAH, S. M. S. S.; FADEL-PICHETH, C. M. T. Identification and antimicrobial susceptibility of *Aeromonas* spp. isolated from stool samples of Brazilian subjects with diarrhoea and healthy controls. **Journal of Medical Microbiology.** v. 59, p. 373-374, 2010.

APÊNDICE A – MATERIAL SUPLEMENTAR 1

Abaixo estão elencadas todas as bactérias, em ordem alfabética, utilizadas para a busca do sistema de secreção tipo IV e posterior montagem do banco de dados. As espécies destacadas em negrito possuem genoma completo.

Aeromonas allosaccharophila BVH88
Aeromonas allosaccharophila CECT 4199
Aeromonas aquatica AE235
Aeromonas australiensis CECT 8023
Aeromonas bestiarum CECT 4227
Aeromonas bivalvium CECT 7113
Aeromonas caviae 429865
***Aeromonas caviae* 8LM**
Aeromonas caviae A23
Aeromonas caviae Ae398
Aeromonas caviae CECT 838
Aeromonas caviae CECT 4221
***Aeromonas caviae* FDAARGOS_72**
***Aeromonas caviae* FDAARGOS_75**
Aeromonas caviae FDAARGOS_76
Aeromonas caviae L12
Aeromonas caviae YL12
Aeromonas dhakensis AAK1
Aeromonas dhakensis CECT 7289
Aeromonas dhakensis CIP 107500
Aeromonas dhakensis SSU
Aeromonas diversa 2478-85
Aeromonas diversa CECT 4254
Aeromonas encheleia CECT 4342
Aeromonas enteropelogenes 1999lcr
Aeromonas enteropelogenes CECT 4255T
Aeromonas enteropelogenes CECT 4487
Aeromonas enteropelogenes LK14
Aeromonas eucrenophila CECT 4224
Aeromonas finlandiensis 4287D
Aeromonas fluvialis LMG 24681
Aeromonas hydrophila 14
Aeromonas hydrophila 116
Aeromonas hydrophila 145
Aeromonas hydrophila 173
Aeromonas hydrophila 187
Aeromonas hydrophila 226
Aeromonas hydrophila 259
Aeromonas hydrophila 277
Aeromonas hydrophila 48_AHYD
***Aeromonas hydrophila* 4AK4**

Aeromonas hydrophila 50_AHYD
Aeromonas hydrophila 52_AHYD
Aeromonas hydrophila 53_AHYD
Aeromonas hydrophila 56_AHYD
Aeromonas hydrophila AD9
Aeromonas hydrophila Ae34
***Aeromonas hydrophila* AH10**
***Aeromonas hydrophila* AL06-06**
***Aeromonas hydrophila* AL09-71**
Aeromonas hydrophila BWH65
Aeromonas hydrophila FDAARGOS_78
***Aeromonas hydrophila* J-1**
Aeromonas hydrophila KOR1
Aeromonas hydrophila L14f
Aeromonas hydrophila M013
Aeromonas hydrophila M023
Aeromonas hydrophila M062
***Aeromonas hydrophila* ML09-119**
Aeromonas hydrophila NF1
Aeromonas hydrophila NF2
***Aeromonas hydrophila* NJ-35**
***Aeromonas hydrophila* pc104A**
Aeromonas hydrophila RB-AH
Aeromonas hydrophila SNUFPC-A8
Aeromonas hydrophila SSU
***Aeromonas hydrophila* YL17**
***Aeromonas hydrophila* subsp. *hydrophila* ATCC7966**
Aeromonas hydrophila subsp. *ranae* CIP 107985
Aeromonas jandaei CECT 4228
Aeromonas jandaei L14h
Aeromonas jandaei Riv2
Aeromonas lacus AE122
Aeromonas media ARB13
Aeromonas media ARB20
Aeromonas media CECT 4232
***Aeromonas media* WS**
Aeromonas molluscorum 848
Aeromonas piscicola LMG 24783
Aeromonas popoffii CIP 105493
Aeromonas rivuli DSM 22539
Aeromonas salmonicida CBA 100
Aeromonas salmonicida subsp. *achromogenes* AS03
Aeromonas salmonicida subsp. *masoucida* NBRC 13784
Aeromonas salmonicida subsp. *pectinolytica* 34mel
Aeromonas salmonicida subsp. *salmonicida* 01-B526
Aeromonas salmonicida subsp. *salmonicida* 2004-05MF26
Aeromonas salmonicida subsp. *salmonicida* 2009-144K3
***Aeromonas salmonicida* subsp. *salmonicida* A449**
Aeromonas salmonicida subsp. *salmonicida* CIP 103209

Aeromonas salmonicida subsp. *salmonicida* JF3224
Aeromonas sanarellii LMG 24682
Aeromonas schubertii CECT 4240
Aeromonas simiae CIP 107798
Aeromonas sobria CECT 4245
Aeromonas sp. 159
Aeromonas sp. HZM
Aeromonas sp. L_1B5_3
Aeromonas sp. MDS8
Aeromonas sp. ZOR0001
Aeromonas sp. ZOR0002
Aeromonas taiwanensis LMG 24683
Aeromonas tecta CECT 7082
Aeromonas veronii AER39
Aeromonas veronii AER397
Aeromonas veronii AMC34
Aeromonas veronii AMC35
Aeromonas veronii ARB3
***Aeromonas veronii* B565**
Aeromonas veronii CECT 4486
Aeromonas veronii CIP 107763
Aeromonas veronii Hm21
Aeromonas veronii Phln2
Aeromonas veronii bv. *sobria* LMG 13067
Aeromonas veronii bv. *veronii* CECT 4257

APÊNDICE B – MATERIAL SUPLEMENTAR 2

Relação das espécies de *Aeromonas* e quantidade de sequências dos genes pertencentes ao sistema de secreção tipo VI anotados pelo RAST.

Espécies	ImpB	ImpC	Pvc109	ImpG	ImpH	Impl	VasD	ImpJ	CipB	VasH	VasI	ImpA	IcmF	VgrG3	VgrG	Hcp
<i>Aeromonas</i> AE235 <i>aquatica</i>	1	1	1	1	1	1	1	1	2	2	1	0	2	1	2	0
<i>Aeromonas</i> CECT 4227 <i>bestiarum</i>	1	1	1	1	1	1	1	1	2	2	1	1	1	4	0	0
<i>Aeromonas</i> CECT 7113 <i>bivalvium</i>	1	1	1	1	1	1	1	1	2	2	1	1	1	3	0	0
<i>Aeromonas caviae</i> 8LM	1	1	1	1	1	1	1	1	2	2	1	1	1	3	0	0
<i>Aeromonas caviae</i> CECT 838	1	1	1	1	1	1	1	1	2	2	1	1	1	3	0	0
<i>Aeromonas</i> FDAARGOS_72 <i>caviae</i>	1	1	1	1	1	2	1	1	3	3	1	1	1	4	0	0
<i>Aeromonas</i> FDAARGOS_75 <i>caviae</i>	1	1	1	1	1	1	1	1	2	3	1	1	1	3	0	0
<i>Aeromonas caviae</i> L12	1	1	1	1	1	1	1	1	2	2	1	1	1	3	0	0
<i>Aeromonas caviae</i> YL12	1	1	1	1	1	1	1	1	2	2	1	1	1	3	0	0
<i>Aeromonas</i> AAK1 <i>dhakensis</i>	1	1	1	1	1	1	1	1	2	2	1	1	1	4	0	0
<i>Aeromonas</i> CECT 7289 <i>dhakensis</i>	1	1	1	1	1	1	1	1	2	2	1	2	1	5	0	0
<i>Aeromonas</i> CIP 107500 <i>dhakensis</i>	1	1	1	1	1	1	1	1	2	2	1	1	1	4	0	0
<i>Aeromonas</i> SSU <i>dhakensis</i>	1	1	1	1	1	1	1	2	2	2	1	1	1	4	1	0
<i>Aeromonas</i> CECT 4342 <i>encheleia</i>	1	1	1	1	1	1	1	1	2	2	1	0	2	4	0	0
<i>Aeromonas enteropelogenes</i> 1999lcr	1	1	1	1	1	1	1	1	2	3	1	1	1	4	0	0
<i>Aeromonas enteropelogenes</i> CECT 4255T	1	1	1	1	1	1	1	1	2	3	1	1	1	3	0	0
<i>Aeromonas enteropelogenes</i> CECT 4487	1	1	1	1	1	1	1	1	2	3	1	1	1	1	2	0
<i>Aeromonas enteropelogenes</i> LK14	1	1	1	1	1	1	1	1	2	2	1	1	1	5	0	0
<i>Aeromonas</i> CECT 4224 <i>eucrenophila</i>	1	1	1	1	1	1	1	1	2	2	1	1	1	3	0	0
<i>Aeromonas</i> <i>finlandiensis</i>	1	1	1	1	1	1	1	1	1	2	1	1	1	3	1	0

4287D

<i>Aeromonas hydrophila</i> 14	1	1	1	1	1	1	1	1	2	2	1	1	1	0	5	0
<i>Aeromonas hydrophila</i> 116	1	1	1	1	1	1	1	1	2	2	1	1	1	0	5	0
<i>Aeromonas hydrophila</i> 187	1	1	1	1	1	1	1	1	2	3	1	1	1	0	5	0
<i>Aeromonas hydrophila</i> 259	1	1	1	1	1	1	1	1	2	2	1	1	1	0	5	0
<i>Aeromonas hydrophila</i> 277	1	1	1	1	1	1	1	1	2	2	1	1	1	0	4	0
<i>Aeromonas hydrophila</i> 48_AHYD	1	1	1	1	1	1	1	1	2	2	1	1	1	4	0	0
<i>Aeromonas hydrophila</i> 4AK4	1	1	1	1	1	1	1	1	2	3	1	1	1	0	3	0
<i>Aeromonas hydrophila</i> 50_AHYD	1	1	1	1	1	1	1	1	2	2	1	1	1	4	0	0
<i>Aeromonas hydrophila</i> 52_AHYD	1	1	1	1	1	1	1	1	2	2	2	1	1	2	1	0
<i>Aeromonas hydrophila</i> 53_AHYD	1	1	1	1	1	1	1	1	2	2	1	1	1	4	0	0
<i>Aeromonas hydrophila</i> 56_AHYD	1	1	1	1	1	1	1	1	2	2	1	1	1	4	0	0
<i>Aeromonas hydrophila</i> AD9	1	1	1	1	1	1	1	1	2	2	1	1	1	4	0	0
<i>Aeromonas hydrophila</i> AH10	1	1	1	1	1	1	1	1	2	2	1	1	1	5	0	0
<i>Aeromonas hydrophila</i> BWH65	1	1	1	1	1	1	1	1	2	2	1	1	1	3	0	0
<i>Aeromonas hydrophila</i> FDAARGOS_78	1	1	1	1	1	2	1	1	2	2	1	1	1	5	0	0
<i>Aeromonas hydrophila</i> J-1	1	1	1	1	1	1	1	1	2	2	1	3	1	4	0	0
<i>Aeromonas hydrophila</i> KOR1	1	1	1	1	1	1	1	1	2	2	1	1	1	4	0	0
<i>Aeromonas hydrophila</i> L14f	1	1	1	1	1	1	1	1	2	2	1	1	1	3	0	0
<i>Aeromonas hydrophila</i> M013	1	1	1	1	1	1	1	1	2	2	1	1	1	3	2	0
<i>Aeromonas hydrophila</i> M062	1	1	1	1	1	1	1	1	2	2	1	1	1	2	1	0
<i>Aeromonas hydrophila</i> NF1	1	1	1	1	1	1	1	1	2	2	1	2	1	5	0	0
<i>Aeromonas hydrophila</i> NF2	1	1	1	1	1	1	1	1	2	3	1	1	1	3	2	0
<i>Aeromonas hydrophila</i>	1	1	1	1	1	1	1	1	2	2	1	3	1	5	0	0

NJ-35																	
<i>Aeromonas</i> SNUFPC-A8	<i>hydrophila</i>	1	1	1	1	1	1	1	1	2	2	1	1	1	1	3	0
<i>Aeromonas</i> SSU	<i>hydrophila</i>	1	1	1	1	1	1	1	1	2	2	1	1	1	0	4	0
<i>Aeromonas</i> subsp. ATCC7966	<i>hydrophila</i> <i>hydrophila</i>	1	1	1	1	1	1	1	1	2	2	1	1	1	1	4	0
<i>Aeromonas</i> CECT 4228	<i>jandaei</i>	1	1	1	1	1	1	1	1	2	1	1	1	1	4	0	0
<i>Aeromonas jandaei</i> L14h		1	1	1	1	1	1	1	1	2	2	1	1	1	3	0	0
<i>Aeromonas jandaei</i> Riv2		1	1	1	1	1	1	1	1	2	1	1	1	1	1	2	0
<i>Aeromonas lacus</i> AE122		1	1	1	1	1	1	1	1	2	1	1	1	1	4	0	0
<i>Aeromonas</i> LMG 24783	<i>piscicola</i>	1	1	1	1	1	1	1	1	2	2	1	0	2	7	0	0
<i>Aeromonas popoffii</i> CIP 105493		1	1	1	1	1	1	1	1	2	2	1	1	1	11	0	0
<i>Aeromonas rivuli</i> DSM 22539		1	1	1	1	1	1	1	1	2	2	1	1	1	4	0	0
<i>Aeromonas salmonicida</i> CBA 100		1	1	1	1	1	1	1	1	2	2	1	0	2	5	0	0
<i>Aeromonas salmonicida</i> subsp. salmonicida 01- B526		1	2	1	1	1	1	1	1	2	2	1	0	2	0	1	0
<i>Aeromonas salmonicida</i> subsp. salmonicida 2004- 05MF26		1	2	1	1	1	1	1	1	2	2	1	0	2	0	1	0
<i>Aeromonas salmonicida</i> subsp. salmonicida 2009- 144K3		1	2	1	1	1	1	1	1	2	2	1	0	2	0	1	0
<i>Aeromonas</i> <i>salmonicida</i> subsp. salmonicida A449	subsp.	1	2	1	1	1	1	1	1	2	2	1	0	2	0	1	0
<i>Aeromonas salmonicida</i> subsp. salmonicida CIP 103209		1	2	1	1	1	1	1	1	2	2	1	0	2	0	1	0
<i>Aeromonas salmonicida</i> subsp. salmonicida JF3224		1	2	1	1	1	1	1	1	2	2	1	0	2	0	1	0
<i>Aeromonas</i> CECT 4240	<i>schubertii</i>	1	1	1	1	1	1	1	1	3	2	1	1	1	3	1	0
<i>Aeromonas simiae</i> CIP 107798		1	1	1	1	1	1	1	1	2	1	1	1	1	4	0	0
<i>Aeromonas sobria</i> CECT 4245		1	1	1	1	1	1	1	1	2	1	1	1	1	6	1	1
<i>Aeromonas sp.</i> 159		1	1	1	1	1	1	1	1	2	1	1	1	1	2	2	0

<i>Aeromonas</i> sp. L_1B5_3	1	1	1	1	1	1	1	1	2	1	1	1	1	5	0	0
<i>Aeromonas</i> sp. MDS8	1	1	1	1	1	1	1	2	2	2	1	1	1	1	7	0
<i>Aeromonas tecta</i> CECT 7082	1	1	1	1	1	1	1	1	2	2	1	0	2	3	0	0
<i>Aeromonas veronii</i> AMC34	1	1	1	1	1	1	1	1	2	2	1	0	2	1	4	0
<i>Aeromonas veronii</i> ARB3	1	1	1	1	1	1	1	1	2	2	1	1	1	1	2	0
<i>Aeromonas veronii</i> CECT 4486	1	1	1	1	1	1	1	1	2	1	1	1	1	1	2	0
<i>Aeromonas veronii</i> Hm21	1	1	1	1	1	1	1	1	2	2	1	1	1	0	4	0
<i>Aeromonas veronii</i> PhIn2	1	1	1	1	1	1	1	1	1	1	1	1	1	0	3	0

APÊNDICE C – MATERIAL SUPLEMENTAR 3

>Hcp-2

MPTPCYISIEGKTQGNITAGFTSDSVGNIFVQGHEDMLVQEFQHVVTVPDTPQSGQPAGQRVHKP
FKFTVALNKAVPLMYNSLASGEMLPKVTWKWYRTSVEGKQEHFFSTVLTDAIVDIDCQMPHCQDPAK
SDFTQLIQVSLAYRKIDWEHTVAGTSGADDWRAPIEA

>VgrG-2

MADSTGLQFTVKV GALPENTFVVAEFALDEALNRPFNRLRELASAQPDIIDFGAVLDQPCELLWVYNGE
LQRRVCGVVSDFAQGDSGFRTRYQLMVQPALWRLSLRQNSRIFQAQKPEILSILLQEHGITDYAFA
LKNEHAKREYCVQYRETGLDFVNRLAAEEGMFYFHEFEVVGKHRIVFADDSAALTAGPELFFNLGNRS
LEQGPYVRQFHRYEAVRPSDELKDYSFKTPAYGLSHKVGAEALTHQRDTYQHFDFFPGRYKEDPSG
KFAAQHRLDALRNDVAVAGSGKSNSSALQPGQTFSLTEHPNDSLNTDWQIVRIQHTGLQSQALEEEGG
SGPTVYHNEFGVVKASTTWRARIGSPEAPHKPMVDGPQIAVVVGPEGEEIYCDEHGRVKLQFPWDR
YGSSNDQSSCWVRSQGWAGGQYGMMAIPRIGHEVIVSFLEGDPDQPIVTGRTYHATNRPPYELPA
NKTRTVLRTETHQGEFNELEFEDQAGKEEYIHGQKDLNVLINDAAWHIKHDQHTDIDNERVTRIKA
NDHLTVEGEKRDQIKADYSLTVDASLHQKLGQSLLEAGSEVHHKAGMKIVMEAGAEELTKVGGSFV
KIDPSGVTLSGGSIKMNSGGSPGSGS

>ImpB

MGKIGDGGTVAPKERINIKYVPATGGQQAIEPLTMMVGNMKGRTEDTPIEERQTVSIDKNNFSSV
MKESALELKFVSNRLEENSQDELPSIKQSLEDFTPDSSVAQQVPELRKLELREALVALKGPLGNIP
AFRNRLQDLLSSDDAREQLLKELDLIKPAE

>ImpC

MSLVEEQVQAGASAASSLLDEIMAQARITPVDEGYSAKQGVAAALIANILDSGTTSEPVNKALVDSMI
VELDKKLSKQMDVILHAKELQEMESSWRSLKLLVDRTDFRENIKIQLVHATKEELLEDFFESPEITQSG
LYKHVYSTGYGQFGGQPVGAVIGDYAFTHSSPDIKLMQYVSAVGAMAHAPFISSVAPAFFGVDSFTDL
PSIKDLKSVFEGPAYTKWRSLESEDARYLGLTAPRFLARLPYDPTENPIKGFNYQEDISSDHDHYLW
GNTAYLMGTSLTDSFAKYRWCPIIGPQSGGAIHDLPVHVEAMGQLQAKIPTEVLITDRREYELSEE
GFITLTMRKDSNAAFFSANSVQKPKVFPNTKEGKEAETNYKLTQLPYMFIINRLAHYIKVLQREQIG
SWKERQDLERELNGWIKQYVADQENPPADVRSRRPLRAAQIKVLDVEGEPGWYQVAMAVRPHFKY
MGASFELSLVGRDLKE

>Pvc109

MPHLSSWDRGNAASLFDRIERGEGLSSPRSEVDELVDVSKRQLDQILNTRPGNCRSAPDLGVIDLND
ATQGSADIKGRIREAIRQCIRRYEPRIIHVDVRSPTYQASPLEMSFQVTAHVRLEHIEQVTSFNVHMDS
HRHYRMV

>VasA

MSLEHYFRDELAFRLQGREFADAYPELTRFLSEQNTDPDVERLLEGFAFLTGNLRAKIEDEFPELTH
GLLNMLWPNYLRPVPSMTIMQFSVIPGAIQPALVRQGCQLDSLPMDEVTCHFQTCCHDTWVYPADIR
HIAAQSGNDLSTISLDIALHAPLPLSELQDKLRFLLGGDSYTAYELYFWLSNQLSHIELEIDGKCFRQE
AKALKPVGFERGDALLPYPNNVYSGYRILQEYFCFPESFLFFELSGGDWPKQPLPVSEFKVHFCFDR
PLPAELKIRPDSFMLNCVPAINLFQHDSEPVNLNQRQAEPKASYRHADGFEIFSVQVEGWVEGNL
GRSRGTPRIYQPFESFQHIERAKQRLALYYRVRVKEAVSGDGFHSLSFVRGDETTTVELDESISVT
LTCTNRSRAARLKVGSVCVPTGSSPSFATFRNLIRPTRPLRPALDGSLSLHWTLISNLSLNYVSLLRDAL
VQVLRTYDFPALHDKQAEQASRKRLAGIEEIEETKPVDRLVRGMPVRGLKSVLSIRQSAFGSEGELYLF
STVLAHFFSLYASVNAFHLLLEVNLNKNERYQWPVQIQGHSLM

>VasB

MTSGKDAPRFNFFQLVELLNRLDGTQERGLDYLPDENIRFKATASLGFPTSDVLQIGRDGQGRHE
LEVAFLGLHGSQSPMPGYLLDSLAWEYAQGEQKLGFLDFFHHRLLTLLHRIWRKYRYHVRFDQNGE
DGFSRMLFALVGLGNEAVCQSLPVNRAKMLSYAGMLASPSRSPEVVAGLVAHCFDLADVEVSAWQ
WRKVPIHQDQQNRLGGACATLGDDFVIGDKVNDCAKFKLLKINNLSFGQFLGFLPNGEHFHALVTFV
SFILRDQLAWDLRLGFGQEQARGRLGEEQSARLGWSTFLGQPPSDPFVTICVQE

>VasC

MDDFNQQSLVVLNSEQLDGSSQQVQYRFDEMGGTLGASEQDDWQLRDRGLGAVLPAHARIELNDGR
FCLCDLSGQTYINGATSPIGRARKVHLEQGDDELIVGPFRLRTYLGAITPEQSLQQVLGNRATEQLDEW
LTGDEPTARTDDPRTLADPLLALQQERRTPNSLMDGLPTAGVPTDGPLADSRLOASPRSQPDVDVS
PLAPFSAVENTMNQEFLDMPESIENHPDYQLSLDGVHDVALTPLMRGLGQPLQLQDTQQAHDMLEEM
GKTVRAMVEGLLQLQTEQAALADKHLRPIEDNPLRLGLDYDETLAVLFAEQKSPVHLSAPAAVAESLH
NVRIHHVANQQAIGAALDSILQAFSPEALIGRFEQYRRSGASGMADEGWAWNMYQHYYRELT SARQ
QGFDKLFHQVYAQAYDQAVRQQQGLI

>VasD

MIRALMLGAALLALAGCTTMGKMADVAMNPDIQVGSNDSQPSTLGLSLLAEPDVNPNESGEAAPIEF
QVLLAEDSKLLATDYDQVTADVEKALGKNYLHDQDYTLPLGQFKYLPPVKLDEKTRYIGVIARYADP
DSEAERKVIKIKSKGAAYQILVHLRLDEVELQKEEE

>VasE

MSSRNRVIWREGLFIKQHFQQQQRHSDYALHARLSALSDYFYGLQSLAINEDYLGFGRIALVGATGI
LPDGTVFNIPNDMLPTPLEITDASVANQKVYLALPLSVSGVNEVGQGGQVATRLQHRHDVRLDHS
EGGDVVSLEVGRVSLRMLEREDRSAYASLAIARILDKRPDGGVLDPNFMPCSSISVSAIPTLKRFLGE
SAGLVAERARLSQRIAAPGQQGVADVAEFMMLQLLNRAQPQLSHLARLGLTLPERLHEALVQLCGE
LMTFTDESRLPPEFPAYRHDDQVVSFEPVMLALRQALSTVLSPRAVSIQRKHQYGIMVAMVNESEL
MKSADFLAVRARMPQEQLRKQLLQQTKVASSDKIRELISLQLPGIPLPLPVAPRQLPYHAGYSYFQL
DRQSPAWQMLAVSNTLAFHIAGDFPELDMQLWAIRSQ

>VasF

MTTDIIKNEQLSDLLFDHAEQLDMDSYWFRLRGQSINPMIDAVTPLLGLVQRVRLLSRYDQVPELYQ
RVVTEIQAIIEQELMAQGYENGVVLSFRYILCTFIDEAVMGRDWGSQSEWSQHSLLARFHNETWGGEK
VFVLLARLQEDPVRYRDILEFIYLCCLGFEGRYKVMSSQGREEFERIVRQLHKQLAHEAGGEAPSVFH
LDLGQQASRYQLRKQVSLRSLFMGGALILALIFGLYHHQLNNTQDQVLRQLGELLK

>ClpB

MIRIELPVLVERLNPICRHMLEEAAALCVNHQGAIEIRIEHLLLKMLETPLSDVRQILKVAEVDVEELKALL
QPSPADNGYGGQYPSFSPMLVEWLQDSWLLASAELQHAQLRSGVMMLVLLMTPQRYLPGSVTRLLA
KVNRELLRQQFDEWVKESAETQVTTTTGGKTAQAALPADASLLARFTVNVTEQARQGS�DPVLCRD
HEIDLMDILSRRRKNPIVGEAGVGKSALIEGLALRMVAGQVPEKLRGVELMTLDLGMQAGASVK
GEFEKRFKGVMEVKDAVQPVILFIDEAHTLIGAGNQAGGLDVSNLIKPALARGELRTIAATTWGEYKK
YVEKDAALSRRFQLVKVGEPADEATVILRGLRSIYKAHGVLIDEEALQASAQLSARYISGRQLPKAI
DVLDTACARVAINLTPPRAVSHLQNALRQRELEIRQLERQSLIGLDNAERLAELQTAQQACREELA
EQEARWQQQGLVHQVELRAALLADQQDEMLAREALDLADAPLDPQAAAEQLATLERELAEQQG
EVLVSAHVDKTQIAAVIAEWTGVPLNRISQGELDVVTRLPEYLGELIKGQDVAVAHLHKHLLTARADLR
RPGRPLGAFLLVGPSPGVGKTETVLQIAELMFGGRQYLTTINMSEYQEKHTVSRIGSPPGYVGFGE
GVLTEAIRQKPYSVLLDEVEKAHPDVNLNFYQAFDKGELADGEGRIIDCKNVVFFLTSNLGFQTVL
AEQPDVLLDALYPELAFFKPALLARMEVIPYLPGLHDTLVQIVGGKLNRLVKLLKERFGAEVLLDDE
AEEILLRANRSENGARMLESVIDGALLPPVSLQLLQRLSAGEPIKRVHFSVADHQFVAEVGA

>VasH

MEQALAFALALTAQRDEPHLCHWWSSTLHASFPKGLLLGMLDVSGRQLECKGWVRGKDVALLGLA
VDDFHSPLAYVLHKAQSRTWDSLYGGARIEHAGFRALLADLGQQCGLHAYPLLDNSNGKPFVALLMD
EGEVLRAWADGPELAQLSQVFCNQLTLIRDLGRSRRDQGVLRDLSLRQMKGEGERLRQHEKLGIDQL
VGQSAVIRGLREQINQAGQHKLTVLIQGETGSGKEVVARLVHQCSDRANKPFVAINCAAIPENLIESEL
FGYQKGAFGALANKTGLVAQANGGTLFLDEVGDMPAAMQAKLLRVLETRSYRPLGAEQEYHSDFR
LIAATHQPLTRHVEEGQFRADLYHRLCQCCLLIAPLREHIEDVPLLCQHFMAQFAAQDGKTLAGLQRKF
LKQLQTYDFPGNVRELRLNLEVACAHTRSGEVGLEALPPELRERVCVELPGYMDDYNHIRDLRRAM
QQYEASVIEARLRHFHGNRMLVAESLNIPKRTLHDKCKQKLEVN

>VasI

MSVIALPLLLATSAAEAPLDMARWQACRQEPSPLVRLACYDAIGNGAASATEGNPKSAAWQAIWA
QEARTPASAPFLLQSDEARGSETLTRPALRGATLSIGCVDSITHIRLRLDQPWSGEEKVQVALDGQPS
AGSQSWFIRDQGLLLEYGRGLPAIEELKRWLGHRELQVRANNGALLRVDLSGLKEALAPLRQQCRW

>ImpA

MSYQHPWCARLLTSLPDEQIRGAVLADEPRWDYVETELVKLGLSLAHSQVDLNAVAEACLGLLESRTK

DMRVLAQLLRCLQHPAKATPLGAAISLLEAWIQAYWLLAWPGNASQKQRLMVQIVKRFEGALPRICES
ASAAELAQLLAQAEQLEQVWLAQCPCDKGELLDPVLMGLKRAQRQQVAQAQADAAGQPQSSSAAAT
GSPASVASTASGAGAMVLSGGGGSAGIDVDSSNDRWRQTQLKVAELLIERQPEAAVGYRLRRHAV
WAGITAVPMMSGAGNKTPLAPMSADMVDEYRAAMNAPDQGLWQRIEQSLTAPYWFEGHRLSAEVA
QKLGFGAVAQAI AEELGTFLQRLPALRELA FSDGSPFLSPECSRWLQPAKGG SAGSGETGLAE EVAQ
RHGEQGVAALALLDERIAQLKEPRDRFHALLVQAEELLAQEGMEALARQHYQHLWQEASRLGLSHW
EPGLVNRLES LAAPLSK

>Vask

MFKTIFTFLRQQLPKLKPSWPLLGAVLWVLA LILVWWLGPRL ELRGAKPFEPLWGRVVFTLLWLWLLL
GVVSWRVWRKMQQLKAERQHEVVLEQDPVKGLIDRQALFLDRWLQALNTHLGK GALYAMPWYLVL
GLPGSGKSSLIHRANPANKLNPRLDTEL RDVAQDQLVDCWLGEQAVMLDPAGVLLSQSEAE LDPQA
RKHERLWLHLLGWLNEHRRRQPLNGLVLTVDLAWLSHASVAERKAYAQLMRSRLQEV SATMNTRLP
LYVTFTKLDLLRGFDVIYQQLDKEAREAVLGVTFKPGADWQQDLALFWDQWVDNLNQNLPELMLSRL
DAAQRNALFSFVRQLAGLKDYVTSLLAETLAI EESKPLLVRGVYVSSVYQQGVPFDAFAQAASRRYNL
PEPIHSALRGESNTYFVRQLFSSII FPEAHLAGENRLHTLYRRRRMAIGLSCLSLFSAALIGGWHYFYRV
NEEAGRNVLTKAQAFMETNEVADAHAFGVSQ LPRNLNIREATLSFGNYRERMPLVADLGLYQGD EIG
PYVEGSYLQLLSLRFLPAQMQLLEDLNQAPAGS EEKLA ILRVMRMLDDASGRNKELVAQYMASRW
QKAFPGQGAVQEQLMGHLDYALDHTN WYGARAERDQAAITAFVPFKEPVYGAQRELGKLP MYQRV
YQNLVVKANDVLPDDLNRDEVGPTFDTVFAL RSDNAGQVPRLLTWPGFNDFFLKQDKALIDLTAMD
AWVLGQRKLSQLSEADRKEITRQVNDRYVTDYVNQWQKLLTNLDVQTLESPEQALDVLAAITGNDQP
FQVRLASLDDNTRIRKISDVEGDPAQAISARIGRPFMATNGVLAGRGEQGPLIQEVNQKLV ELQHYLEL
IVNATEPGQSALKAVQLRMTNKYADPVFALQQYARSLPAPLDRWVGQLSEQSSRLVIDLAMSSLNQE
WQDKVLAPFNSQLAGRYPFDPSSNKDVPLSEMERFFAPGGTLD SFYQVNLKPMVESGLMEGEFSSP
IQAE LVKQLDRAARIRQIFFSQQGNLEVQFALEPIELTANKRRSVLNLDGQLLEYAHGRRTKIPLVWP N
TMRDGAESKITLVPAARERSPRSEGFVGPWAMFRLMDKGELTQVNDATFDVRFVVDQGAMTYRVYT
DSAQNPF TGGLFSQFRLPESLY

>hipotética

MSQNNQQGQALKVGRDPRMLPEYEALRAEINKLSHASRPEVDWQRIHQ LASLIFEKHGVDLQTAIYF
TLARSRLQGLSGFTEGCEFLANLIVTQWESFWPPVHQERARIEMLDWFIARISDVIRQYQISHEDKR LI
YRCERALQLISEKLHNADLSRIPRVENLVHFIEGYTHLFD ETEIVVSDEPGLKEDLQIPPMVFFKSDME
SDHGGTVSASTPHLPQGSILVGREKGQVKPTVLKIEQHRRQRP AWFWFGCGLLSCALPVMGWL GW
QQHQQEKTVAARQLVQPAELPRALSYDDIRQARIVLGEQTLQNMESDLVARYQNQLTRLEQTSPLY
WYRYGEGLRNSLQMLYPDSLAVKALDKQWQTQLG SQQGDVISVPTYLDARAGVDALLDQ LLELRQ
RKTVTISYLSQLYE VQKNLMQNIPFSLRLDELEARKASQEPITAAELKSLENDLKALNIRLYQLQQGAS
GS

>uropatogênica

MCPSIALLGDIGTDHEGFHPSPVIAASPDVFLDGK PVARQGDPLAPHDKPNNPPHPRSISGGVGSVLV
NGKPIAVTGTAVGCGGVVIGSGSGQAG

>VgrG-3

MADSTGLQFTVKVGALPESTFVVAEFALDEGLNRPFNLRLELASRQPDIDFGAVLDQPCELLVWYNG
ELQRRVCGVSDMAQGD SGFRTRYQLLVLPALWRLSLRQNSRIFQAQKPDEILSILLQEHGITDYAF
ALKNEHAKREYCVQYRETDLDFVNRLAAEEGMFYFHEFEAGKHRIVFADDAAL TQGPELFFNLGNR
SLEQGPYVRQFH YREAVRPSDVELKDYSFKTPAYGLSHKKQGADLDHQRDTYQHFD FPGRYKQGG
SGKAFAQHRLDALRND AVAGSGKSNCAALQPGQTFSLTEHPNGGLNTDWQVVRIQHTGLQPQALEE
EGSGSPTVYHNEFGVVKASSTWRARIGSPEALHKPMVDGPQIAMVVGPEEGEEIY CDEHGRVKLQFP
WDRYGSSNDQSSCWVRVSQGWAGGQYGMMAIPRIGHEVIVSFLEGD PDQPIVTGRTYHATNRPPY
ELPANKARTVLR TETHQGE GFNELRFEDQAGQEEIYIHGQKDLNVL IENNAVWHIKHDEHRDIDNERV
TRIKANDHLTVEGEKR DQIKADYSLTVDASLHQQLGQSLLVEAGQEVHIKVGDKLVLEAGSEITLKG GG
SFVKVDPSGIKLMGPAIKLNAGGSAGSGSGWGGKAPVVPK GVEVVKAPELVELVKAIPTEKAMEALLK
EQGPAQTFVPSR

>Hcp-1

MPTPCYISIEGKTQGNITAGAF TSDSVGNIFVQGHED EMLVQEFQHVVTVPTDPQSGQPAGQRVHKP
FKFTVALNKAVPLMYNALASGEMLPKVTLK WYRTSVEGKQEHFFSTVLT DATIVDIDCQMPHCQDPAK
SDFTQLIQVSMAYRKIDWEHTVAGTSGADDWRAP IEA

>VgrG-1

MADSTGLQFTVKVGALPENTFVVAEFALDEALNRPFNLRLELASAQPDIIDFGAVLDQPCELLWYNGE
LQRRVCGVVSDFAQGDSGFRTRYQLRVLPALWRLSLRQNSRIFQAQKPDEILSILLQEHGITDYAFA
LKNEHAKREYCVQYRESLDLDFVNRLAAEEGMFYFHEFEAGKHRIVFADDDAAALTQGPELFFNLGNRS
LEQGPYVRQFHRYREAVRPSDVELKDYSFKTPAYGLSHKKVGAELTHQRDTYQHFDFFPGRYKEDPSG
KAFQAHRDLALRNDVAVAGQAKSNCAALLPGQSFSLTEHPNGSLNTDWQIVRIQHTGLQPQALEEEGG
SGPTVYHNEFGVVKASTTWRARIGSPEAPHKPMVDGPQIAIVVGPDGEEIYCDEHGRVKLQFPWDY
GSSNDQSSCWVRVSQGWAGGQYGMMAPRIGHEVIVSFLEGDPDQPIVTGRTYHATNRPPYELPAN
KTRTVLRTETHQGEFNELELRFEDQVGQEEIYIHGQKDLNVLIENTAAWHIKHDEHTDVDNERVTRIKA
NDHLTVEGEKRDQIKADYSLTVDTSMHQKLGDSWLTQAGQEVHVKAGAKVVLEAGSEITVKVGGCFI
KVDGGGVTLVGPTIKMNSGGSPSSGSGWGGKSPVDPLGVSVPKPKVPLTPAQLATMKSAAPFCEE
CEKCKEGGCEI