

**UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE GENÉTICA**

**PROGRAMA DE PÓS-GRADUAÇÃO
EM GENÉTICA E BIOLOGIA MOLECULAR
TESE DE DOUTORADO**

**EVOLUÇÃO GENÔMICA E DA ILHA DE
ALTA PATOGENICIDADE DE *Yersinia***

NARA SUZY AGUIAR DE FREITAS

VIRTUS IMPAVIDA

**RECIFE
2010**

**UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE GENÉTICA**

**PROGRAMA DE PÓS-GRADUAÇÃO EM
GENÉTICA E BIOLOGIA MOLECULAR**

**EVOLUÇÃO GENÔMICA E DA ILHA DE ALTA
PATOGENICIDADE DE *Yersinia***

NARA SUZY AGUIAR DE FREITAS

Tese apresentada ao
Programa de Pós-graduação
em Genética e Biologia
Molecular da Universidade
Federal de Pernambuco como
requisito para obtenção do
grau de Doutor em Genética
pela UFPE

Orientador: Prof^a Dr^a Alzira Maria Paiva de Almeida

Co-orientador: Prof^o Dr. Valdir de Queiroz Balbino

**RECIFE
2010**

Freitas, Nara Suzy Aguiar de

Evolução genômica e da ilha de alta patogenicidade de Yersinia / Nara Susy Aguiar de Freitas. – Recife: O Autor, 2010.

104 folhas : il., fig., tab.

Tese (doutorado) – Universidade Federal de Pernambuco. CCB.
Departamento de Genética. Doutorado em Genética e Biologia
Molecular.

Inclui bibliografia e anexos.

1. Genética molecular 2. Genoma 3. Genômica I. Título.

572.8 CDD (22.ed.)

UFPE/ CCB – 2010- 069

**UNIVERSIDADE FEDERAL DE PERNAMBUCO
PROGRAMA DE PÓS-GRADUAÇÃO EM
GENÉTICA E BIOLOGIA MOLECULAR**

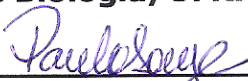
**Parecer da comissão examinadora da tese de:
NARA SUZY AGUIAR DE FREITAS**

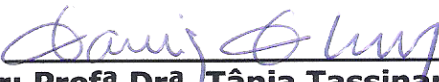
**EVOLUÇÃO GENÔMICA E DA ILHA DE ALTA PATOGENICIDADE DE
*Yersinia***

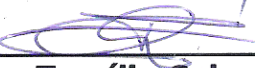
**A comissão examinadora considera o presente trabalho
Aprovado**

**Portanto, cumpridas todas as exigências regimentais, NARA SUZY
AGUIAR DE FREITAS faz jus ao grau de Doutor em Genética e Biologia
Molecular pela UFPE. Recife, 12 de fevereiro de 2010**


1º Examinador: Profº Dr. Martin Alejandro Montes
Deptº de Biologia/UFRPE


2º Examinador: Profº Dr. Paulo Roberto Eleutério
Deptº de Biologia/UFRPE

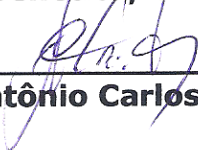

3º Examinador: Profª Drª. Tânia Tassinari Rieger
Deptº de Genética, CCB/UFPE


4º Examinador: Profº Dr. Tercílio Calsa Júnior
Deptº de Genética, CCB/UFPE


5º Examinador: Profº. Dr. Marcos Antonio de Moraes Junior
Deptº de Genética, CCB/UFPE


Orientador: Profª. Drª. Alzira Maria Paiva de Almeida
Deptº. de Microbiologia/CPqAM/Fiocruz


Co-orientador: Profº Dr. Valdir de Queiroz Albino
Deptº de Genética, CCB/UFPE


Coordenador: Profº Dr. Antônio Carlos Freitas

*"Se aprende com as
diferenças e não
com as igualdades"*

Paulo Freire

*Para todos que vão
citar o presente
trabalho de uma
forma diferente.*

AGRADECIMENTOS

Aos meus orientadores, Prof^a Alzira Almeida e Prof^o Valdir Balbino, pelas ideias e tempo dedicado a esta pesquisa.

Aos colaboradores, César Costa (UFPE), Cristina Leal-Balbino (CPqAM/Fiocruz), Manoel Sena-Júnior (UFPE) e Tiago Ferreira (UFRPE), pela confiança e apoio que recebi.

A todos que fazem o Laboratório de Genética Molecular Humana da UFPE, em especial a Prof^a Rosilda Santos e ao Prof^o Luiz Mauricio-Silva, por me fazerem sentir em casa mesmo trabalhando.

Aos colegas da Bioinformática e Biologia Evolutiva da UFPE, em especial a Carlos Santiago, Klaudia Emanuela, Lidiane Gomes, Marcos Batista, Nadia Aragão, Paula Almeida, Simone Medeiros, Tiago Levi e Wellington Cunha pelo incentivo e sugestões.

À Universidade Federal de Pernambuco e aos professores e colegas do Programa de Pós-Graduação em Genética pela aprendizagem, apoio e amizade.

À Universidade Federal Rural de Pernambuco pelo suporte durante o período de realização deste trabalho.

Aos Professores e amigos Antônio Carlos, José Ferreira e Tânia Rieger pelo carinho nos momentos difíceis.

Meus sinceros agradecimentos àqueles profissionais que em mim acreditaram e contribuíram para a realização deste trabalho.

À minha família, minha razão de viver.

Sumário

	Página
<i>Lista de Figuras</i>	i
<i>Lista de Tabelas</i>	v
<i>Lista de abreviações</i>	vii
Resumo	viii
1. Introdução	1
2. Revisão Bibliográfica:	3
2.1. Evolução da <i>Yersinia pestis</i>	3
2.2. Ilha de patogenicidade	5
2.3. Mecanismos de transmissão das Ilhas de Patogenicidade (PAIs)	7
2.3.1 Elementos transponíveis	8
2.4. Fenótipo pigmentação (Pgm)	9
2.4.1. Aquisição horizontal dos fatores de virulência das cepas de <i>Yersinia enterocolitica</i>	11
2.4.1.1 Microevolução de <i>Y. pestis</i>	13
2.5. O sistema <i>Quorum Sensing</i> e os polissacarídeos	15
2.5.1 Os polissacarídeos	17
2.6. <i>Yersínias</i> nos bancos de dados da web	18
2.7. Evolução das sequências HGT, genômicas e a reconstrução filogenética	21
3. Objetivos	26
3.1. Objetivo Geral	26

3.2. Objetivos Específicos	26
4. Referências Bibliográficas	27
5. Artigo 1 - Evolution of codon usage in the High-Pathogenicity Island of <i>Yersinia</i> and identification of association patterns of adjacent molecule groups	35
6. Artigo 2 - Ancestrais da ilha de alta patogenicidade de <i>Yersinia pestis</i>	52
7. Informações complementares	86
7.1. Resultados dos alinhamentos dos LCBs nas regiões up_HPI_down das cepas de <i>E. coli</i>	86
7.2. Resultados dos alinhamentos dos LCBs entre o plasmídeo pPBPR1 de <i>P. profundum</i> e o plasmídeo pYV de <i>Y. enterocolitica</i> 8081	88
8. Conclusões Gerais	90
9. Abstract	91
10. Anexos	92
10.1. Instruções para Autores - <i>Genetics and Molecular Research</i>	93
10.2. Instruções para Autores - <i>Genetics</i>	97
10.3. Relação das cepas avaliadas e suas principais características.	104

Lista de Figuras

Página

Revisão Bibliográfica

Figura 1 - Ilha de alta patogenicidade das *Yersínias* flanqueada por sítios *attB* e formada por uma região de seqüências variadas e pelas seqüências dos genes *ybt* adjacente do gene *int* e inserida no *asn**tRNA* (Baseada em Lesic e Carniel, 2004). 10

Figura 2 – Monofilia das *Yersínias* e aquisição de elementos genéticos móveis: YAPI (Ilha de Patogenicidade de Aderência das *Yersínias*), pYV (plasmídeo de virulência das *Yersínias*) e HPI (Ilha de Alta Patogenicidade das *Yersínias*). As setas idicam que os elementos genéticos foram adquiridos por transferência horizontal (Baseada em Carniel, 2008). 13

Figura 3 - Microevolução da *Yersinia pestis*: Antes da divergência da *Yersinia pseudotuberculosis*, uma cepa da *Yersinia pestis* ancestral adquiriu o plasmídeo pFra e posteriormente o plasmídeo pPla e o fago filamentoso YpfΦ dando origem aos diferentes biovars. As setas indicam que os elementos genéticos foram adquiridos por transferência horizontal (Baseada em Carniel, 2008). 15

Figura 4 – Mecanismo de atuação do sistema *Quorum Sensing*: molécula sinalizadora denominada de auto-indutora, como a N-acil homoserina lactona (AHL), que permite que genes não sejam ativados prematuramente. A ativação ocorre apenas com o aumento da densidade populacional das bactérias. 17

Artigo 1

Figure 1 - Phylogenetic reconstruction based on 11 concatenated *ybt* gene sequences of *Yersinia pestis*, *Yersinia pseudotuberculosis*, *Yersinia enterocolitica* and *Escherichia coli* strains using PAUP* version 4.0b10 (Swofford, 2002) by exhaustive search (bootstrap test of 10.000 pseudo-replications). The heuristic Maximum Parsimony (MP) and Maximum Likelihood (ML) criteria in phylogenetic tree reconstruction were 46

explored. The per-character consistency homoplasy index (HI) and rescaled consistency index (RC) were used to evaluate the phylogram.

Figura 2 - Phylogenetic reconstruction using the high-pathogenicity island 47 of the complete host genomes of the strains: *Yersinia pestis*, *Yersinia pseudotuberculosis*, *Yersinia enterocolitica* and *Escherichia coli*, based on codon usage frequency (classification|hierarchical cluster analysis|Euclidian distance), INCA2.1 and SPSS.13 softwares.

Figura 3 - Estimated distance between genomes the strains of *Yersinia* 48 *pestis* (91001, Angola, Antiqua, CO92, KIM and Pestoides F) or “Yp”, *Yersinia pseudotuberculosis* IP32953 or “Ypseud”, *Yersinia enterocolitica* 8081 or “Yent” and *Escherichia coli* (536, APEC01, CFT073 and UTI89) or “Ec” as related to the GC3s frequency classes. A) containing all mRNAs; B) containing the mRNAs of the high-pathogenicity island (HPI) and those located 30kb upstream and downstream the HPI, obtained by Correspondence Analysis, software SPSS.13 (Euclidian distance). The proximity of points on the x-axis indicates that classes of GC3 frequency of the strains are more homogeneous. The strains in x-axis inferior right quadrant tend to exhibit higher contrast to that in the y-axis superior left quadrant.

Figure 4 - Locally Collinear Blocks (LCBs) formed by aligning ~90kb of 49 sequences of *Yersinia pestis* and *Escherichia coli* strains, over 30kb upstream and downstream the high-pathogenicity island: A) 42 LCBs formed by aligning *Y. pestis* CO92 with 91001, Angola, Antiqua, KIM and Pestoides F; B) LCBs formed by aligning *Y. enterocolitica* with *Y. pseudotuberculosis* (software MAUVE 2.1). The LCBs shown in the horizontal line are the units in synteny of each strain and in the vertical lines coming from the LCBs of each strain represent the possible homologous genes in the other strains.

Artigo 2

Figura 1 - Reconsruição filogenética baseada no método Neighbor-Joining 79 e teste Bootstrap (4000 réplicas), software MEGA4.1: A) gene 16S

(*rimM*); e B) genes *tufA* e 16S (*rimM*) concatenados.

Figura 2 - Reconstruções filogenéticas baseadas nos genes concatenados: A) 11 genes *ybt* (*fyuA-ybtE-ybtT-ybtU-irp1-irp2-ybtA-ybtP-ybtQ-ybtX-ybtS*), 16S, *tufA* *cdtMPQ*, *luxR* e *ybtD*; B) 11 genes *ybt*; e C) *fyuA-ybtE-ybtT-ybtU-irp1-irp2-ybtA-ybtX-ybtS*; usando o método de Máxima Parcimônia (software PAUP* v. 4.0b10), busca exaustiva (bootstrap com 10.000 pseudo-replicações) e índice de homoplasia (HI) e consistência de reescalonamento (RC).

Figura 3 – Filograma construído a partir das 18 classes de frequências de GCs em cada posição do códon dos mRNAs de 15 genomas hospedeiros dos genes *ybt*, usando os softwares MatLab/ SPSS.13.

Figura 4 – Rearranjos dos Blocos Colineares Local (LCBs), formados pelas sequências cromossômicas das cepas de *Yersinia pestis* (91001, Angola, Antiqua, CO92, Nepal516, KIM e Pestoides F), produzidos pelos alinhamentos par a par com a cepa CO92 (software MAUVE 2.3). Os LCBs nas linhas horizontais são homólogos a *Y. pestis* CO92 e nas linhas verticais indicam as posições dos LCBs no cromossomo alinhado. A baixo de cada par de cepa, em vermelho, encontram-se os números totais de LCBs e os apresentados nos alinhamentos.

Figura 5 – Rearranjos entre os Blocos Colineares Local (LCBs) de 15 cepas, formados pelos alinhamentos dos genes *tufA* *cdtMPQ*, *ybtD*, *luxR*, onze *ybt* e 16S (*rimM*), homólogos da cepa de *Y. pestis* CO92 (software MAUVE 2.3). Maior intensidade de cor nos LCBs indica maior similaridade de DNA.

Informações complementares

Figura 7.1.1. - 15 Blocos Colineares Local (LCBs) formados pelos alinhamentos das sequências 30kb antes e depois da HPI das cepas de *Escherichia coli* 536, APEC01, CFT073 e UTI89 (software MAUVE 2.3).

Figura 7.2.1. - 7 Blocos Colineares Local (LCBs) formados pelos alinhamentos entre os plasmídeos pPBPR1 de *Photobacterium profundum* SS9 e o plasmídeo pYV de *Yersinia enterocolitica* 8081 (software MAUVE

2.3).

Figura 7.2.2. – Variações das frequências de aminoácidos entre os 89 plasmídeos pPBPR1 de *Photobacterium profundum* SS9 e o plasmídeo pYV de *Yersinia enterocolitica* 8081, baseadas nos mRNAs dos arquivos *ffn/NCBI (software INCA).

Lista de Tabelas

Página

Revisão Bibliográfica

Tabela 1 – Relação dos projetos genômicos de cepas de *Yersinia* patogênicas disponíveis nos bancos de dados da web (Fonte: NCBI, outubro de 2009). 20

Tabela 2 – Relação dos projetos genômicos de cepas de *Yersinia* não patogênicas disponíveis nos bancos de dados da web (Fonte: NCBI, até outubro de 2009). 21

Artigo 1

Tabela 1 - Estimate of the ratio (R) between the non-synonymous (dN) and synonymous (dS) substitutions per site of the loci *ybt* and *int* found in the strains of *Escherichia coli* (536, APEC01, CFT073 and UTI89), *Yersinia pestis* (91001, Angola, Antiqua, CO92, KIM and Pestoides F), *Yersinia pseudotuberculosis* IP32953 and *Yersinia enterocolitica* 8081, using the *neighbor-joining* method and modified Nei-Gojobori (p-distance), model MEGA4 Software. 50

Tabela 2 - Numbers of mRNAs (genes) of the high-pathogenicity island (HPI), of the regions 30kb upstream and downstream the HPI (up_HPI_down) and of the host genomes (HG), sizes in kb, GC-contents and GC3 frequencies in strains of *Escherichia coli*, *Yersinia pestis*, *Yersinia pseudotuberculosis* IP32953 and *Yersinia enterocolitica* 8081 (INCA2.1 software). 51

Artigo 2

Tabela 1 - Matriz de similaridade das correlações das taxas da 84
razão entre as substituições não sinônimas e sinônimas das
sequências dos genes *ybt* das cepas de *Escherichia coli* (536,
APEC01 e UTI89), *Yersinia pestis* (91001, Angola, Antiqua, CO92,
KIM e Pestoides F), *Yersinia pseudotuberculosis* (IP32953 e
PB1/+), *Yersinia enterocolitica* 8081, *Citrobacter koseri* ATCC BAA-
895, *Photobacterium luminescens* TTO1 e *Photobacterium profundum*
SS9, baseada no método de Pamilo-Biachi-Li (1993), softwares
SPSS.13/MatLab.

Tabela 2 – Resultados do blast de nucleotídeos das sequências 85
completas da Ilha de Aderência de Yersínia (YAPI) da cepa de
Yersinia pseudotuberculosis 32777 (NCBI).

Lista de abreviações

kb	Kilobase
pb	Pares de base
ICE	Elementos integrativos e conjugativos
Ybt	Sideróforo yersiniabactina
PAI	Ilha de patogenicidade
HPI	Ilha de alta patogenicidade das yersínias
up_HPI_down	Regiões 30kb antes e depois da ilha de alta patogenicidade
GI	Ilha Genômica
HG	Genoma Hospedeiro
HGT	Transferência horizontal de genes
GC	Guanina ou Citosina
AT	Adenina ou Timina
GC1	Guanina ou Citosina na primeira base do códon
GC2	Guanina ou Citosina na segunda base do códon
GC3	Guanina ou Citosina na terceira base do códon
IS	Sequências de Inserção
DNA	Ácido Desoxirribonucléico
YAPI	Ilha de patogenicidade de aderência de <i>Yersinia</i>
MP	Máxima parcimônia
ML	Máxima verossimilhança
HI	Índice de consistência de Homoplasia
RC	Índice de consistência Rescalonado
LCB	Bloco colinear Local
dN	Substituição não sinônima
dS	Substituições sinônimas
QS	Quorum Sensing
AHL	N-acil homoserina lactona

RESUMO

A evolução dos mecanismos de virulência bacteriana está diretamente relacionada com a aquisição dos elementos transponíveis. A compreensão desses fenômenos teve grande avanço com a produção de sequências completas de genomas. Atualmente, sete sequências completas de cepas de *Yersinia pestis* são conhecidas. Nossas análises dessas sequências revelaram novos aspectos sobre a evolução da ilha de alta patogenicidade (HPI), onde estão as sequências que codificam o sideróforo yersiniabactina (Ybt). As evidências de ciclos adaptativos, mesmo na ausência de genomas intermediários, proporcionaram informações de que a HPI das yersínias foi herdada monofileticamente, podendo a Vibrionaceae ser sua família ancestral. A metodologia foi baseada nas análises das sequências vizinhas às HPIs, assinaturas seletivas dos genes do operon *ybt* e seus homólogos e frequências do uso de códons dos genomas hospedeiros, que revelaram novos aspectos da evolução desta ilha genômica. Os padrões e as diferenças de conteúdos de GC e das substituições sinônimas e não sinônimas indicam que os genes *ybt* e seus genomas hospedeiros passaram por diferentes pressões seletivas. Assim, grupos ancestrais diferentes encontraram formas distintas de preservar suas HPIs. Observamos que as sequências vizinhas das HPIs são reguladas por *quorum sensing*, indicando uma rede geral de regulação que envolve os genes *ybt*. A análise da organização cromossômica das sete cepas representativas de diferentes regiões do mundo revelou também uma dinâmica de rearranjos que poderia justificar o reconhecimento de novas variantes de *Y. pestis* até recentemente considerada uma espécie muito homogênea.

Palavras-chave: *Yersinia pestis*; uso de códon, evolução, HPI, genômica comparativa.

1. INTRODUÇÃO

A *Yersinia pestis*, bactéria Gram-negativa da família Enterobacteriaceae, é o agente causador da peste, doença que pelas suas características de alta letalidade e capacidade de disseminação marcou a história da humanidade. Atualmente a peste permanece em diversos focos silvestres em vários países e a *Y. pestis* está classificada como um dos agentes biológicos de Bioterrorismo.

Os trabalhos com filogenia molecular apontam para uma monofilia do gênero *Yersinia*, que apresenta três espécies patogênicas: *Y. pestis*, *Y. pseudotuberculosis* e *Y. enterocolitica*. Esses patógenos compartilham uma ilha de alta patogenicidade (HPI) que é essencial para a virulência e responsável pela biossíntese do sideróforo yersiniabactina (Ybt). Os genes *ybt* apresentam conteúdos de GC conservados entre as cepas patogênicas das yersínias e em outras espécies com diferentes conteúdos de GC genômico. Provavelmente, essa alta similaridade seja um dos motivos que levaram esses genes a não serem avaliados quanto aos tipos de seleção natural ou taxas de substituições sinônimas e não sinônimas.

A seleção natural é tida na biologia evolutiva como o principal fator responsável pela evolução das espécies. Os processos básicos da evolução, as discussões interdisciplinares e a genômica comparativa, com seus novos métodos de identificação de biotipos geográficos, permitem que a dinâmica das populações seja visualizada de forma mais clara. Porém, os processos evolutivos nem sempre são considerados nas discussões sobre os agrupamentos hierárquicos filogenéticos.

Outra dificuldade em abordar a evolução das espécies que persiste nos dias atuais é devido a falta de modelos matemáticos nas avaliações dos fenótipos e dos genótipos. Algumas abordagens descrevem as relações ancestrais-descendentes baseadas na presença ou ausência de um caractere, sem uma avaliação que possa correlacionar os padrões e variações dos fenótipos e genótipos dos grupos de indivíduos durante a evolução.

Os programas de biologia computacional podem efetuar avaliações quantitativas e qualitativas de cromossomos inteiros, mas geralmente esses dados são pouco discutidos em conjunto. As frequências de uso de GC1, GC2 e

GC3 (GC em cada posição do códon) em cada gene dos cromossomos podem ser obtidas de forma relativamente simples, mas o volume de dados não é facilmente organizado em uma única apresentação gráfica.

Outro fator importante, relativo às classes de frequências de conteúdos de GC dos genomas, é que não há método descrito para uma avaliação do número de genes envolvidos em análises com vários genomas. Esses dados apresentados simultaneamente podem facilitar a discussão sobre o processo de homogeneização do genoma e as avaliações dos genes ricos ou pobres em GC durante a evolução das espécies. No caso do conteúdo de GC1 e GC2, os dados são menos explorados, mas podem ser significativos nos genes das ilhas genômicas e enriquecer uma abordagem evolutiva das sequências dos cromossomos de uma espécie ou grupos de genes transferidos horizontalmente.

Os modelos de avaliação das classes de frequências de GC em cada posição do códon e as maneiras de visualização dos dados ainda não foram discutidas, mas os conteúdos GC e GC3 são fornecidos por ferramentas computacionais em diversos tipos de frequências, assim a separação da frequência de GC1, GC2 e GC3 nos códons de um gene ou de milhares de genes por classes podem ser obtidos a partir de alguns modelos de biologia que empregam algoritmos e que estejam no cerne da biologia de sistema computacional. Em seguida, essas frequências podem ser programadas para ficarem em uma determinada classe. Desta forma, é possível analisar antigas questões em aberto ao definir vários “níveis” do cenário adaptativo.

No presente trabalho a HPI da *Y. pestis* foi avaliada *in silico* para identificação dos tipos de seleção natural que atuam sobre os genes *ybt* e responder algumas questões sobre a evolução da HPI dentro e fora do gênero *Yersinia*. Ferramentas da bioinformática para análises filogenéticas foram discutidas e modelos de biologia de sistemas foram propostos para análise filogenética dos genomas. O trabalho é apresentado em duas partes, a primeira aborda a evolução do uso de códons da HPI e a identificação de padrões das sequências adjacentes e os grupos de moléculas associadas. A segunda parte discute sobre os ancestrais da HPI de *Y. pestis*.

2. REVISÃO BIBLIOGRÁFICA

2.1. Evolução da *Yersinia pestis*

O gênero *Yersinia* da Família Enterobacteriaceae, contém atualmente 14 espécies que apresentam diferentes níveis de patogenicidade. Três espécies: *Y. pestis*, *Y. pseudotuberculosis* e *Y. enterocolitica* são patogênicas. A *Y. pestis* é o agente causador da peste, doença primária dos roedores, transmitida geralmente por pulgas, que pode afetar o homem. A *Y. pseudotuberculosis* e a *Y. enterocolitica* são bactérias enteropatogênicas, responsáveis por um quadro clínico denominado Yersiniose, e são veiculadas principalmente pela água contaminada e alimentos (Perry e Fetherston, 1997; Carniel, 2008).

Onze espécies: *Y. intermedia*, *Y. kristensenii*, *Y. frederiksenii*, *Y. aldovae*, *Y. rohdei*, *Y. ruckeri*, *Y. bercovieri*, *Y. mollaretii*, *Y. aleksiciae*, *Y. massiliensis* e *Y. similis* são ambientais e geralmente não são patogênicas aos humanos (Sulakvelidze, 2000; Sprague e Neubauer, 2005; Sprague et al. 2008; Merhej et al. 2008).

A *Y. pestis* foi responsável por três pandemias que afetaram parcelas significativas das populações humanas durante a era Cristã: A primeira pandemia, denominada "Peste de Justiniano" (542 a 605 d.C), teve início na África Central e se estendeu para o Egito. A segunda pandemia, ou "Peste Negra", teve início na Ásia e estendeu-se por toda Europa e parte do norte da África, persistindo do século XIV ao século XVIII. A terceira pandemia, ou "Pandemia Contemporânea", teve início na China, estendendo-se para Hong Kong em 1894, de onde se disseminou pelo mundo através do transporte marítimo, chegando aos Estados Unidos, América do Sul, África do Sul e Madagascar (Pollitzer, 1954).

Classicamente a partir de diferenças nas reações de fermentação do glicerol (G) e redução do nitrato (N) a *Y. pestis* foi subdividida em três biovars ou variedades geográficas: Antiqua (G^+N^+), Medievalis (G^+N^-) e Orientalis (G^-N^+) (Devignat, 1951). As observações epidemiológicas e os registros históricos conduziram à hipótese de que o biovar Antiqua foi responsável pela primeira

pandemia, o Medievalis pela segunda e o biovar Orientalis pela terceira e atualmente está difundido em todos os continentes, exceto a Oceania (Perry e Fetherston, 1997).

Zhou e colaboradores (2004) propuseram um quarto biovar “Microtus” para outro grupo de cepas, avirulentas para o homem, porém patogênicas para o roedor *Microtus bandtii*, presente em focos da China, e que pelas características bioquímicas (G^+N^-) seriam do biovar Medievalis. Contudo, diferentemente dos outros o biovar Microtus não fermenta arabinose. Cepas de *Y. pestis* atípicas, como as chamadas pestóides, que foram isoladas de roedores nos focos da antiga União Soviética, fermentam a ramnose e a melibiose, são G^-N^- e não se enquadram nestes biovars (Anisimov *et al.* 2004).

A análise dos genomas das cepas de *Y. pestis* Orientalis (cepa CO92), Medievalis (cepa KIM) e Microtus (91001) revelou a ausência de grandes segmentos no cromossomo da cepa 91001 e a presença de alguns pseudogenes que podem ter contribuído para sua avirulência aos humanos (Deng *et al.* 2002; Parkhill *et al.* 2001; Song *et al.* 2004).

Segundo Achtman e colaboradores (1999), a *Y. pestis* derivou de *Y. pseudotuberculosis* de 1.500 a 20.000 anos atrás, provavelmente pouco antes da primeira pandemia. A *Y. enterocolitica* é uma espécie heterogênea e bioquimicamente foi caracterizada em seis biotipos (1A, 1B, 2, 3, 4, e 5). O biotipo 1A inclui a maioria das cepas não patogênicas, enquanto as cepas dos biotipos de 2 a 5 são pouco patogênicas e o biotipo 1B é altamente patogênico para camundongos. A distribuição geográfica dos diversos biotipos é distinta. O biotipo 1B é encontrado mais frequentemente na América do Norte e é referido como biotipo do Novo-Mundo e os biotipos 2 a 5 são predominantes na Europa e Japão e referidos como biotipos do Velho-Mundo (Thomson *et al.* 2006). A análise de sequências multiloco sugere que *Y. enterocolitica* e *Y. pseudotuberculosis* divergiram há cerca de 200 milhões de anos (Thomson *et al.* 2006; Carniel, 2008).

2.2. Ilha de patogenicidade

O termo ilha de patogenicidade (Pathogenicity island = PAI) foi primeiramente empregado por Hacker e colaboradores (1997) para descrever duas regiões genômicas em cepas de *Escherichia coli* uropatogênicas relacionadas a fatores de virulência. As PAIs são segmentos de 10-200 kb e são um tipo de ilha genômica (Genomic island = GI) de bactérias patogênicas que contém fatores de virulência, tais como adesinas, toxinas, invasinas e proteínas dos sistemas de secreção (Hacker e Kaper, 2000). Após a descoberta da primeira PAI outras GIs foram caracterizadas e classificadas de acordo com as atividades desenvolvidas pelos seus conjuntos gênicos. As mais conhecidas são as ilhas de simbiose, as ilhas de resistência à antibióticos e as ilhas metabólicas (Dobrindt *et al.* 2004).

Os genes relacionados aos fatores de virulência e de resistência à antibióticos contidos em GIs estão presentes tanto em bactérias patogênicas como em não patogênicas (Hacker e Kaper, 2000; Dobrindt *et al.* 2004). Algumas GIs são específicas de algum sorotipo ou de uma espécie patogênica e podem ser usadas como marcadores de diagnóstico.

As GIs desconhecidas podem ser identificadas *in silico*, caso a sequência do genoma seja conhecida (Oelschlaeger e Hacker, 2004). Um dos métodos mais simples de identificação de GIs é a análise de sequências homólogas com conteúdos de GC (Guanina e Citosina) diferentes dos cromossomos (Yoon *et al.* 2005).

As GIs promovem alterações genéticas importantes para a evolução dos microrganismos e são adquiridas por transferência horizontal de genes (Horizontal gene transfer = HGT). As análises de HGT podem considerar o uso de códon, a abundância relativa de dinucleotídeos, a detecção de *repeats*, as sequências de inserção (*IS*) e os transposons (Dobrindt *et al.* 2004).

As HGT permitem que os genomas bacterianos estejam em um estado constante de fluxo gênico, o que em princípio indica que qualquer segmento de DNA em uma população bacteriana pode ser horizontalmente transferido. Porém, apenas uma pequena porcentagem do DNA adquirido será mantida na célula

receptora e, então, transferida para gerações subsequentes, pois muitos mecanismos limitam as aquisições horizontais de genes. Dentre esses mecanismos, encontram-se modulações fisiológicas, ambientais e genéticas (Solomon e Grossman, 1996; Arber, 2000; Oelschlaeger e Hacker, 2004). Calcula-se que cerca de 20 a 50 proteínas estejam relacionadas à competência bacteriana, uma das primeiras etapas da transformação (Thomas e Nielsen, 2005).

A transferência horizontal é uma importante ferramenta na adaptação dos procariotos à um nicho específico, pois a aquisição de um conjunto gênico fortemente selecionado aumenta a adaptabilidade dos genomas hospedeiros (Lawrence, 2002). Portanto, pode-se considerar as GIs como uma das mais fortes evidências da HGT quando os mesmos genes estão presentes em distintos organismos e apresentam as mesmas funções em diferentes contextos ecológicos (Hacker e Kaper, 2000).

Os impactos das HGT apenas podem ser completamente entendidos quando são combinados os conhecimentos dos mecanismos genéticos envolvidos, a dinâmica populacional dos procariotos e as pressões seletivas sobre as unidades avaliadas. Esse conjunto de informações permite determinar a história dos eventos que ocorreram com as HGT durante a evolução desses seres vivos. Fato esse que afeta toda a biodiversidade da terra, uma vez que muitos procariotos estão diretamente relacionados à manutenção da vida dos demais seres (Arber, 2000; Inge-vechtomov, 2002).

O cenário evolutivo das HGT entre as bactérias de relações mais distantes ainda é pouco conhecido, embora homólogos de GIs presentes em bactérias filogeneticamente distantes tenham sido descritos (Collyn *et al.* 2004; Collyn *et al.* 2006). Os caminhos evolutivos desses genes e seus genomas hospedeiros são melhor investigados com a obtenção das sequências completas de DNA e a associação de ferramentas da bioinformática (Collyn *et al.* 2006).

2.3. Mecanismos de transmissão das Ilhas de Patogenicidade (PAIs)

Os mecanismos de transmissão das PAIs ainda não foram elucidados, mas o fato de estarem inseridas em tRNA e usarem fago da integrase como maquinaria de integração/excisão sugere que foram originadas a partir dos bacteriófagos (Carniel 2008). As SaPI de *Staphylococcus aureus*, por exemplo, são pequenas ilhas móveis de 14-17kb que carregam genes de fatores de virulência (Ubeda *et al.* 2007). Um modelo semelhante de transferência para a ilha de alta patogenicidade das yersínias (High pathogenicity island = HPI) foi demonstrado experimentalmente por Antonenka e colaboradores (2005). Neste modelo, a ilha interage dentro de um sítio de *repeats attB* (17pb) carregado por um plasmídeo conjugativo, que serve como uma “nave” de transferência da HPI a um novo hospedeiro bacteriano.

Outro modelo de transferência da HPI foi observado na cepa de *E. coli* ECOR31. Essa cepa difere das demais que produzem o sideróforo yersiniabactina (Ybt) por abrigar um fragmento adicional de 35 kb na região 3' da HPI. A sequência é semelhante ao DNA dos *elementos integrativos e conjugativos* (Integrative and conjugative elements = ICE) e contém genes envolvidos na transferência conjugativa do DNA. Segundo Schubert e colaboradores (2004), essa cepa poderia ser o progenitor da HPI, que foi transferida horizontalmente para *Y. pestis*. A proposta se baseia também no fato de que o ICE putativo de *E. coli* ECOR31 (35 kb) está em um loco diferente da HPI em outras cepas ou ausente (Schubert *et al.* 2004; Carniel, 2008).

Entre as yersínias a transferência horizontal da HPI foi observada entre cepas de *Y. pseudotuberculosis* hospedeira e não hospedeira. O processo foi baseado apenas na recombinação do DNA e a transferência da HPI ocorreu sem a ajuda de uma maquinaria transportadora de excisão/integração (Lesic e Carniel, 2005). Nesta bactéria a transferência horizontal ocorreu em baixas frequências (10^{-8}) e sob circunstâncias específicas (baixa temperatura e ausência de ferro em meio líquido). A recombinação ocorreu entre as regiões

flanqueadoras da HPI e a troca de fragmentos homólogos algumas vezes não foi restrita apenas a ilha.

2.3.1 Elementos transponíveis

Na década de 1950, quando os elementos transponíveis foram descobertos por Barbara McClintock, em estudos com milho, acreditava-se que a frequência dos elementos genéticos móveis nos genomas era baixa, mas agora se sabe que são provavelmente os componentes mais abundantes dos genomas e compreendem a maior parte do DNA medianamente repetitivo (Wessler, 2006).

Os elementos transponíveis ou transposons são segmentos de DNA que têm a capacidade de se moverem e replicarem dentro do genoma em que estão associados, aumentando a variabilidade genética quando são inseridos/deletados em diferentes pontos do cromossomo bacteriano (Mascaretti, 2003; Hacker *et al.* 1997).

Os transposons podem mover-se do cromossomo para um plasmídeo e vice-versa. Os elementos *IS* são transposons simples que podem inativar genes de virulência e quando combinados podem também promover a mobilização de grandes porções de DNA (Hacker *et al.* 1997; Hacker e Kaper, 2000). Ao mesmo tempo eles podem contribuir para o valor adaptativo (*fitness*) das células bacterianas ativando *operons* do sistema de catabolismo celular (Hall, 1999). O aporte de um transposon em uma posição alvo do DNA é mediado por uma integrase, que nas GIs geralmente estão integradas dentro ou próximas a um tRNA.

Os elementos *IS* são caracterizados por serem curtas sequências de DNA, que codificam enzimas necessárias para a sua transposição, flanqueadas em ambas as extremidades por sequências idênticas ou muito semelhantes de nucleotídeos de orientação inversa, conhecidas como repetições invertidas (10 a 25pb) ou *repeats* (Mascaretti, 2003; Lesic e Carniel, 2005; Ubeda *et al.* 2007).

As cepas de *Y. pestis* dos diversos biovars completamente sequenciadas exibem diferenças quanto ao número de *IS* (Achtman *et al.* 2004; Chain *et al.* 2006), semelhante ao que é observado entre *Y. pseudotuberculosis* IP32953 e *Y.*

enterocolitica 8081 (Thomson *et al.* 2006). As modificações nas estruturas genéticas por HGT permitem que as populações bacterianas tenham novas combinações de genes e consequentemente novos agentes patogênicos e caráter ecológico (Hacker e Kaper, 2000), com trocas de *IS* ocorrendo não apenas nas regiões que flanqueiam as *GIs*, mas geralmente nas unidades menores que compreendem a PAI (Hacker e Carniel, 2001; Thomson *et al.* 2006).

2.4. Fenótipo pigmentação (Pgm)

Fetherston e colaboradores (1992) denominaram de “loco *pgm*” um segmento de DNA instável do cromossomo de *Y. pestis*, de cerca de 102 kb, limitado em ambas as extremidades por um elemento *IS100*, que pode deletar espontaneamente. As culturas Pgm⁻ são incapazes de utilizar a hemina ou o pigmento vermelho Congo do meio de cultura e são brancas (ou não pigmentadas) e avirulentas. As culturas Pgm⁺ são pigmentadas (vermelho) e virulentas.

Posteriormente foram identificadas no loco *pgm* duas regiões física e funcionalmente independentes. Um segmento de cerca de 35kb denominado “segmento de aquisição de ferro” ou HPI e um segmento de cerca de 68kb denominado “segmento de pigmentação” ou *hms* (Iteman *et al.* 1993; Fetherston e Perry, 1994; Buchrieser *et al.* 1998; 1999).

A região de aquisição de ferro ou HPI (Figura 1) é a região relacionada com a alta patogenicidade das yersínias e possui as sequências para os genes que participam da síntese do sideróforo yersiniabactina (Ybt) (Carniel *et al.* 1996; Buchrieser *et al.* 1998; Hacker e Carniel, 2001).

Os genes que codificam a Ybt estão distribuídos extensamente em outros gêneros da família Enterobacteriaceae, sobretudo em cepas de *E. coli* responsáveis por infecções extra-intestinais (Schubert *et al.* 2002). Até o momento não se sabe a origem precisa da HPI, mas o fato dela manter sua mobilidade e ser altamente conservada em vários gêneros bacterianos sugere

que foi adquirida recentemente e pode ter retido sua habilidade de ser transmitida horizontalmente a novos hospedeiros bacterianos (Carniel, 2008).

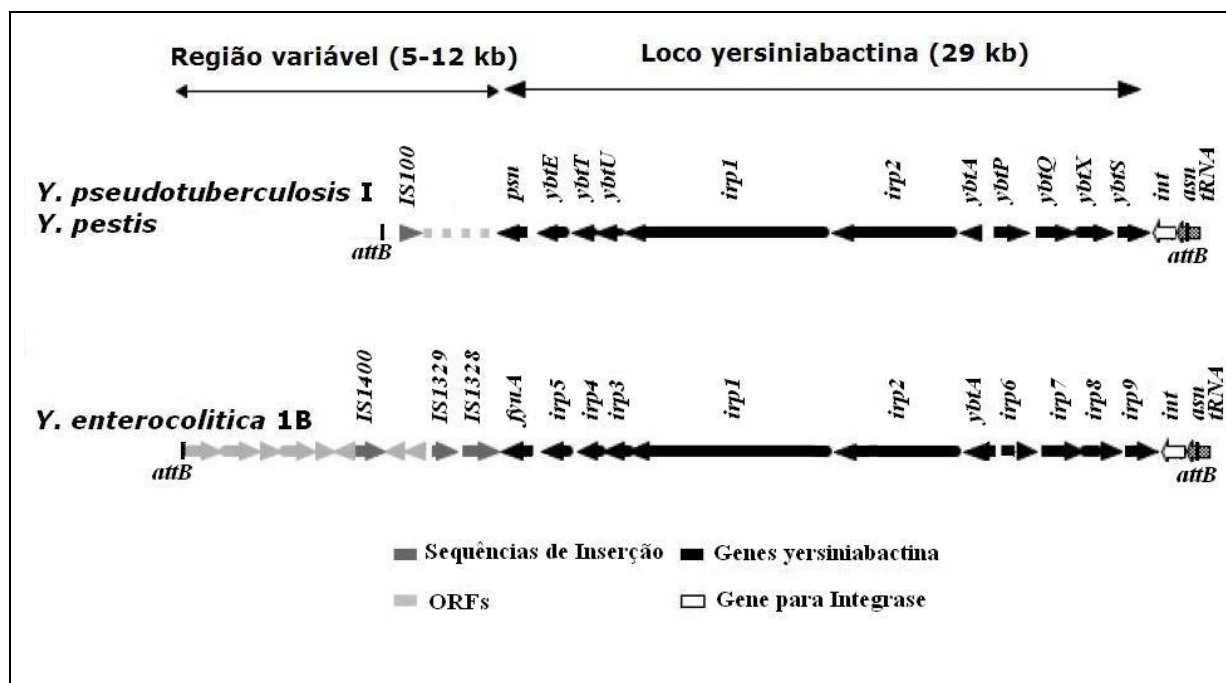


Figura 1 - Ilha de alta patogenicidade das Yersínias flanqueada por sítios *attB* e formada por uma região de seqüências variadas e pelas seqüências dos genes *ybt* adjacente do gene *int* e inserida no *asn*tRNA (Baseada em Lesic e Carniel, 2004).

O segmento da HPI da cepa *Y. pestis* KIM ou loco *ybt*, que codifica o sideróforo Ybt, possui os genes *psn*, *ybtE*, *ybtT*, *ybtU*, *irp1*, *irp2*, *ybtA*, *ybtP*, *ybtQ*, *ybtX*, *ybtS* e *int*. O gene *int* é homólogo ao bacteriófago-P4 e codifica a integrase, uma recombinase que media rearranjos de DNA durante a transferência horizontal dos genes (Buchrieser *et al.* 1998, 1999).

Os genes *ybt* codificam as seguintes proteínas: HMWP1 e HMWP2, codificadas pelos genes *irp1* e *irp2* respectivamente; YbtT/Irp4 uma proteína desconhecida; YbtE/Irp5 uma salicyl-AMP ligase; YbtS/Irp9, uma antranilato sintase que converte diretamente o corismato em salicilato (Fetherston *et al.* 1999; Pelludat *et al.* 2003).

Os produtos dos genes *ybtP/irp6*, *ybtQ/irp7* são proteínas de transporte ABC e juntamente com o produto do gene *psn/fyuA*, uma proteína da membrana externa (receptor da yersiniabactina e da pesticina), estão envolvidos com a

internalização do ferro (Brem *et al.* 2001; Perry *et al.* 2003). O gene *ybtA* codifica uma proteína AraC, que regula a expressão da biossíntese do sistema Ybt e está localizado junto do *operon ybtPQXS* (Fetherston *et al.* 1999). Os genes *ybtX/irp8* e *ybtU/irp3* codificam as proteínas transporte da superfamília MFS de sinal de transdutor e S-HMWP1 - tiazolinil redutase, respectivamente.

O fenótipo Hms⁺ em *Y. pestis* é essencial para a transmissão da bactéria por pulgas para os mamíferos (Lillard *et al.* 1997; Bobrov *et al.* 2008). A produção de Hms envolve um complexo enzimático que necessita de ao menos seis proteínas (Hms) codificadas por genes organizados em três operons: o *hmsHFRS* e os *hmsT* e *hmsP*. As proteínas HmsH e HmsF estão localizadas na parte externa da membrana enquanto que as HmsR, HmsS HmsT e HmsP se encontram no interior da membrana (Lillard *et al.* 1997; Buchrieser *et al.* 1999; Bobrov *et al.* 2008). A matriz extracelular dependente de *hmsHFRS* apresenta um polímero de β -1,6-N-acetil-D-glucosamina e esse exopolissacarídeo é controlado positivamente pela proteína HmsT, diguanilato ciclase, e negativamente por HmsP, fosfodiesterase (Bobrov *et al.* 2008).

Atualmente, não há uma definição clara sobre a regulação em nível molecular dos diferentes grupos de moléculas presentes no loco Pgm, mas ao contrário das proteínas Hms, dependentes do sistema *quorum sensing*, as do loco Ybt são fatores de virulência, adquiridos horizontalmente que ainda não foram associados a esse sistema, apresentado no item 2.5.

2.4.1. Aquisição horizontal dos fatores de virulência das cepas de *Yersinia* enteropatogênicas

Yersinia e *Escherichia* são grupos de Enterobacteriaceae bem distintos, porém os genes *ybt* de *Yersinia* e *E. coli* apresentam alta similaridade. Esse fato foi utilizado por Schubert e colaboradores (2004) como base para a proposta de que a aquisição horizontal da HPI para os dois gêneros tenha ocorrido recentemente, muito após suas divergências de um ancestral comum.

Mesmo sem a identidade da espécie membro das Enterobacteriaceae doadora da HPI para os gêneros *Escherichia* e *Yersinia*, Carniel (2008) formulou

uma proposta para a origem e aquisição da HPI entre os membros patogênicos do gênero *Yersinia*.

Pelo fato da HPI ser sistematicamente ausente em todas as espécies não patogênicas foi proposto que ela tenha sido especificamente adquirida pelo ramo patogênico do gênero *Yersinia*. Desta forma, a HPI de *Y. enterocolitica* e a HPI de *Y. pseudotuberculosis* foram adquiridas independentemente após a divergência da espécie ancestral comum.

Segundo Carniel (2008), existem pelo menos três evidências a favor desta proposta: (i) entre as espécies enteropatogênicas a HPI se encontra apenas nos subgrupos de cepas patogênicas (*Y. enterocolitica* 1B e *Y. pseudotuberculosis* I e III), (ii) a HPI de *Y. enterocolitica* tem uma organização genética diferente da HPI de *Y. pseudotuberculosis*, e (iii) as identidades entre os genes *ybt* de *Y. enterocolitica* e de *Y. pseudotuberculosis* são muito maiores que entre os seus genes cromossômicos.

O gênero *Yersinia* carrega outra PAI (98kb), que apresenta conteúdos de GC heterogênicos, e foi denominada de Ilha de patogenicidade de aderência de *Yersinia* (*Yersinia* Adhesion Pathogenicity Island = YAPI). Inicialmente a YAPI foi identificada em *Y. pseudotuberculosis*, mas se encontra presente também em cepas de *Y. enterocolitica* 1B (Collyn *et al.* 2004; Thomson *et al.* 2006; Carniel, 2008) e ausente em todas as cepas de *Y. pestis*.

A alta identidade das sequências de YAPI da *Y. enterocolitica* e *Y. pseudotuberculosis* sugere que a ilha foi adquirida do mesmo ancestral comum (Collyn *et al.* 2006). Se assim for, a YAPI deve ter sido secundariamente perdida por alguns subconjuntos de cepas de *Y. pseudotuberculosis* (Carniel, 2008), Figura 2.

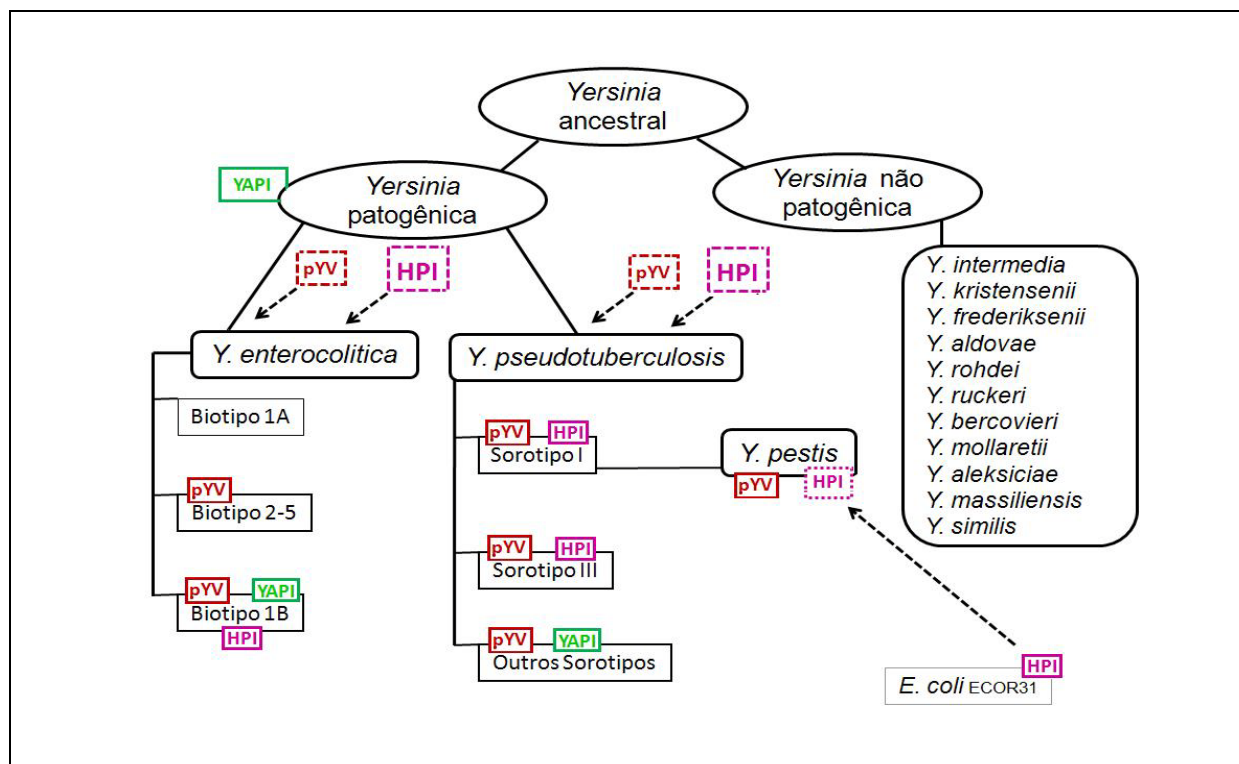


Figura 2 – Monofilia das Yersínias e aquisição de elementos genéticos móveis: YAPI (Ilha de Patogenicidade de Aderência das Yersínias), pYV (plasmídeo de virulência das Yersínias) e HPI (ilha de alta patogenicidade das Yersínias). As setas indicam que os elementos genéticos foram adquiridos por transferência horizontal (Baseada em Carniel, 2008).

2.4.1.1 Microevolução de *Y. pestis*

A aquisição horizontal de dois plasmídios e de um fago filamentoso, durante ou após a divergência de *Y. pseudotuberculosis*, contribuiu para o aumento da virulência da *Y. pestis* e sua adaptação a um novo estilo de vida nos hospedeiros e dos mecanismos de transmissão via vetor-artrópode (Carniel, 2008).

As três yersínias patogênicas possuem um plasmídeo pYV/pCD1 (70kb) indispensável para a virulência. As sequências do pYV codificam um sistema de secreção do tipo III (TTSS), que permite às bactérias secretar e translocar proteínas efetoras (Yops) para dentro das células do hospedeiro. O mesmo ocorre com o sistema de secreção tipo III de várias outras bactérias, sugerindo

uma transmissão horizontal desse segmento do pYV (Cornelis, 2002; Carniel, 2008). O pYV apresenta ainda sequências para uma adesina (YadA) e o antígeno V ou LcrV que tem dupla função: (1) protege o patógeno da resposta imune do hospedeiro e (2) controle da expressão das Yops (Cornelis, 2002; Sing *et al.* 2005).

Dois plasmídeos específicos da *Y. pestis* pPla (ou pPst ou pPCP) e pFra (ou pMT) foram adquiridos após a divergência da *Y. pseudotuberculosis* e codificam fatores de virulência e de transmissão. O plasmídeo pPla está presente em todas as cepas de *Y. pestis*, exceto na variedade Pestoides F. O plasmídeo pPla codifica uma protease (Pla), ou ativador do plasminogênio, que parece atuar na transmissão da bactéria pelas pulgas (Worsham e Roy, 2003; Carniel, 2008). Já o plasmídeo pFra apresenta sequências que codificam uma proteína de envoltório altamente imunogênica para o homem e outros animais, denominada de fração antigênica F1. Ele também é responsável por codificar a proteína Ymt (anteriormente denominada toxina murina), importante na transmissão de *Y. pestis* pelas pulgas.

O fago filamentoso da *Y. pestis* YpfPhi (ou YpfΦ) possui alta identidade com um profago filamentoso (CUS1), presente nas cepas altamente virulentas de *E. coli* O18:K1:H7. Esse fato sugere que estes organismos foram infectados recentemente por um virion similar a partir de um doador desconhecido e por um mecanismo não esclarecido (Figura 3). YpfPhi está integrado de maneira estável nas cepas de *Y. pestis* biovar Orientalis, mas nos biovars Antiqua e Medievalis está presente como um replicon extracromossômico instável. A função de YpfPhi ainda não está esclarecida, no entanto, não parece necessário para a transmissão da bactéria pelo vetor, mas parece participar do processo infeccioso por um mecanismo ainda ignorado. A estabilização de YpfPhi no biovar Orientalis pode ter contribuído para o aumento do seu potencial pandêmico (Derbise *et al.* 2007; Carniel, 2008).

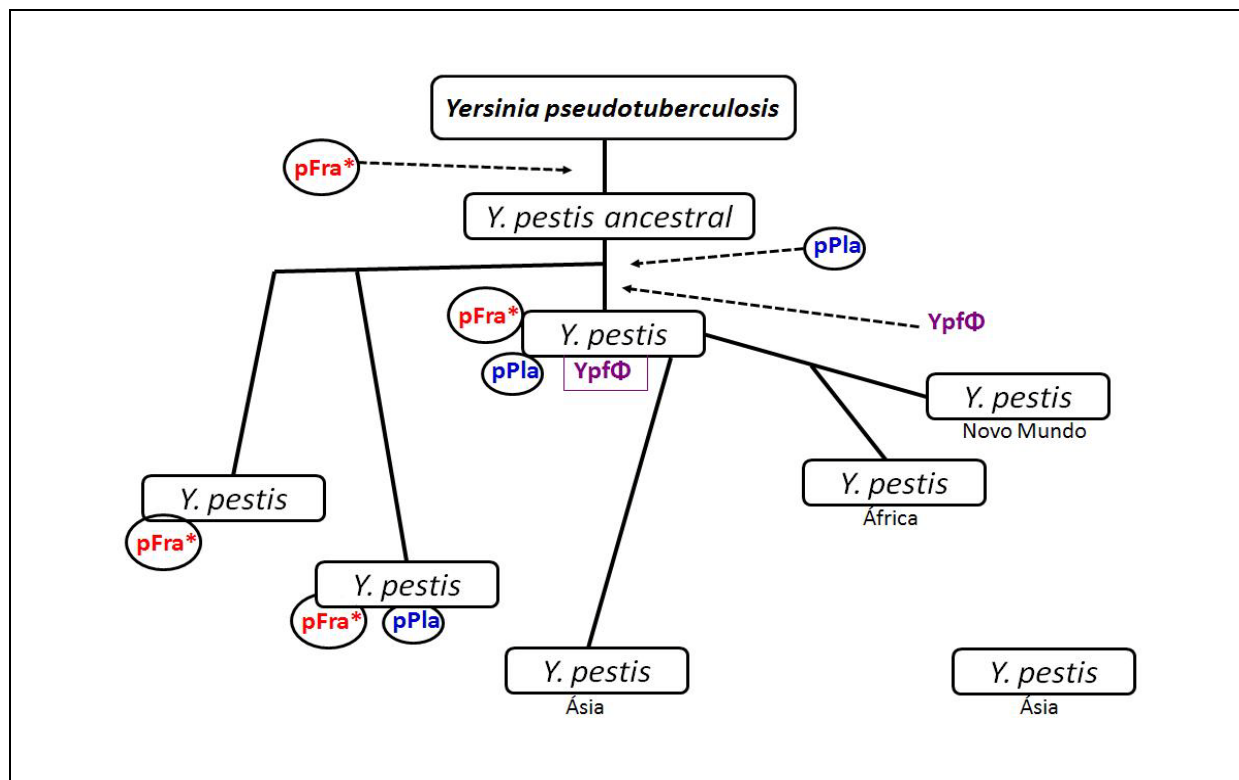


Figura 3 - Microevolução da *Yersinia pestis*: Antes da divergência da *Yersinia pseudotuberculosis*, uma cepa da *Yersinia pestis* ancestral adquiriu o plasmídeo *pFra* e posteriormente o plasmídeo *pPla* e o fago filamentosso *YpfΦ* dando origem aos diferentes biovars. As setas indicam que os elementos genéticos foram adquiridos por transferência horizontal (Baseada em Carniel, 2008).

2.5. O sistema *Quorum Sensing* e os polissacarídeos

O termo *Quorum Sensing* (QS) foi utilizado pela primeira vez por Fuqua e colaboradores (1994) para descrever padrões de comportamento cooperativo em bactérias. Essa cooperação depende de uma molécula sinalizadora, de baixo peso molecular denominada auto-indutora e sua concentração extracelular está relacionada com a densidade populacional das bactérias (Boettcher e Ruby, 1995; Qazi *et al.* 2006).

Existem duas categorias principais de moléculas que desempenham o papel de sinalizadores de QS: uma formada por aminoácidos de curtas cadeias peptídicas, que são comuns em bactérias Gram-positivas (hepta ou

octapeptídeos) e se ligam aos receptores localizados na superfície das células bacterianas (Qazi *et al.* 2006). A outra é formada por moléculas derivadas de ácidos graxos, comuns em bactérias Gram-negativas. A N-acil homoserina lactona – AHL é a molécula sinalizadora mais abundante nesse grupo de bactérias (Qazi *et al.* 2006; Heermann e Fuchs, 2008).

O sistema QS está associado às moléculas que controlam a expressão de diversas funções fisiológicas da bactéria e seu principal objetivo é evitar a expressão prematura de fatores de virulência nas unidades bacterianas isoladas (Figura 4). A sinalização por QS se dá em função das flutuações ambientais, tais como diferenças de pH, temperatura, disponibilidade de nutrientes e outras, refletindo os múltiplos sistemas que permitem a adaptação das bactérias ao meio ambiente (Brito *et al.* 2004; Higgins *et al.* 2007; Heermann e Fuchs, 2008).

A bactéria *Photorhabdus luminescens*, assim como a *Y. enterocolitica*, que tem como vetor os flebotomíneos, é entérica e altamente patogênica para insetos. Heermann e Fuchs (2008) ao verificarem que nos genomas de ambas as bactérias há uma sobreposição de sequências reguladas por QS as denominaram de uma nova classe de LuxR-reguladores, responsável pela virulência nos insetos. Em *Vibrio cholerae*, o agente da cólera em humanos, o gene *luxR* está associado à patogenicidade e a formação de biofilmes ou polissacarídeos extracelulares.

O comportamento evolutivo dos genes homólogos de *luxR* e *luxI* pode estar relacionado com a própria história do filo *Proteobacteria* (dividido em cinco classes: alfa, beta, gama, delta e ípsilon). A família das Enterobacteriaceae (Beta-Proteobactéria), assim como as demais, compartilha ambos os reguladores de QS. Porém não se sabe como esses genes surgiram e toda a distância evolutiva entre eles precisa ser explorada.

Segundo Gray e Garey (2001), a análise filogenética de *luxR* e *luxI* sugere que o sistema QS divergiu dentro de cada classe de Proteobactéria, embora casos de HGT também tenham ocorrido. Os homólogos de *luxR* parecem ter sido adquiridos de formas distintas e os de *luxI* tendem a apresentar múltiplas cópias.

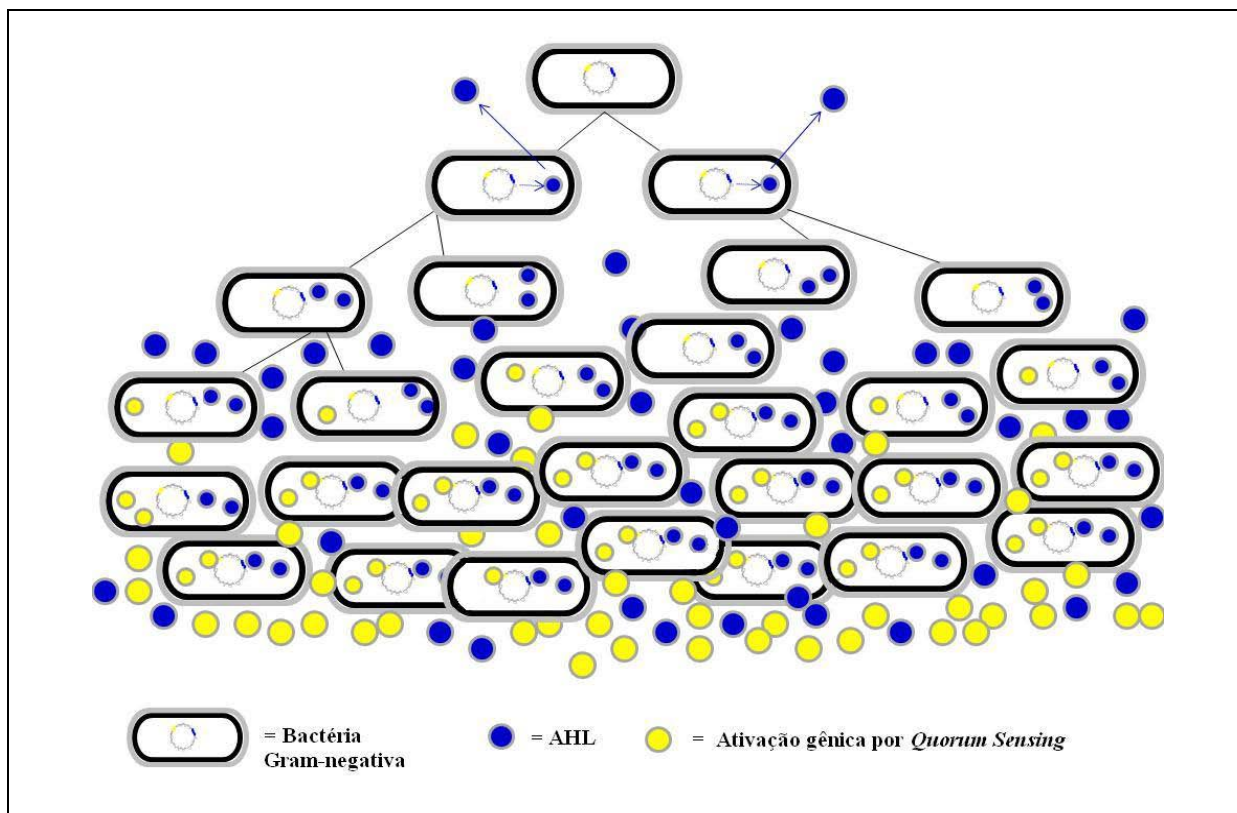


Figura 4 – Mecanismo de atuação do sistema *Quorum Sensing*: molécula sinalizadora denominada de auto-indutora, como a N-acil homoserina lactona (AHL), que permite que genes não sejam ativados prematuramente. A ativação ocorre apenas com o aumento da densidade populacional das bactérias.

No gênero *Yersinia* os genes *lux* são altamente conservados nas espécies patogênicas e não patogênicas. A *Y. pseudotuberculosis*, por exemplo, produz pelo menos oito AHLs diferentes e possui dois homólogos de *luxI* (família AHL sintase) e dois *luxR* (família de AHL-reguladores dependente de resposta). Em *Y. pestis*, os genes *luxR* apresentam perda ou rearranjo de segmentos, que aparentemente podem refletir sua recente divergência de *Y. pseudotuberculosis* (Atkinson *et al.* 2006).

2.5.1 Os polissacarídeos

Todas as Proteobactéria são Gram-negativas com uma parede celular formada principalmente de lipopolissacarídeos. Nas Beta-Proteobactérias a

análise filogenética de genes associados à produção de biofilmes indica que a microbiodiversidade das populações dessa classe apresenta perfis microbianos que correspondem aos padrões de adaptação de seus membros relacionados ecologicamente (Brummer *et al.* 2006; Waters *et al.* 2008).

A produção de biofilme ou associações de microrganismos e de seus produtos extracelulares que se encontram aderidos a superfícies (bióticas ou abióticas) permitem a agregação de células, proporcionando vantagens as bactérias patogênicas (Watnick e Kolter, 2000; Higgins *et al.* 2007). A produção de biofilmes é regulada pelo sistema QS, dependente da regulação da concentração intracelular de c-di-GMP, assim como a motilidade e fatores de virulência (Tamayo *et al.* 2007).

A produção de biofilme em *Y. pestis* é essencial para a transmissão da peste das pulgas para os mamíferos. Essa produção é dependente dos genes contidos no “segmento de pigmentação” ou *hms*, vizinho a HPI. A utilização de anticorpos específicos permitiu a Bobrov e colaboradores (2008) identificar o polímero β -1,6-N-acetil-D-glucosamina como parte da matriz extracelular de *hmsHFRS* dependente para a formação de biofilme. Na cepa *E. coli* K-12 os genes *pgaABCD* estão associados à produção de biofilmes e codificam proteínas similares as Hms de yersínias (variando de 34% a 83% de identidade), que também são responsáveis pela biossíntese de um EPS que contém o polímero β -1,6-N-acetil-D-glucosamina (Wang *et al.* 2004).

2.6. Yersínias nos bancos de dados da web

O número de sequências gênicas das diversas cepas de yersínia depositadas nos bancos de dados na web é bastante elevado e vem crescendo continuamente. Atualmente, existem doze cepas com representantes das três espécies patogênicas aos humanos completamente sequenciadas e depositadas no GenBank (<ftp.ncbi.nih.gov>). As doze cepas correspondem a um representante de *Y. enterocolitica*, quatro de *Y. pseudotuberculosis* e sete de *Y. pestis*.

Por outro lado, entre as doze espécies do gênero *Yersinia* mais estudadas (Sulakvelidze 2000; Carniel 2008) onze já tiveram cepas como foco de projetos genômicos. Na Tabela 1 encontram-se os projetos dos representantes das três espécies patogênicas ao homem e na Tabela 2 os demais representantes do gênero com sequenciamento genômico em andamento, faltando apenas um projeto para a *Y. aleksiciae*.

Os genomas das doze cepas de *Yersinia* com status de completo (Tabela 1) foram produzidos por bibliotecas que geraram sequências para uma cobertura completa dos genomas. Todos apresentam dados relativos à *TaxMap* (distribuições taxonômicas das proteínas homólogas), *CDD* (domínios conservados nos bancos de dados), *COG* (clusters dos grupos de ortólogos de genes e proteínas) e FTP (protocolo de transferência de arquivos). Nos FTP estão armazenados diversos diretórios de cada acesso dos genomas completos, como por exemplo, o diretório *.fnn, onde estão armazenadas todas as sequências dos mRNA (GenBank/NCBI). O FTP se baseia em um método de recuperar/receber arquivos numa rede (interna ou externa) de computadores diretamente da fonte para o computador do usuário (<ftp.ncbi.nih.gov>).

Os projetos genômicos de *Y. pestis* envolvem no momento 28 cepas, em um total de 51 cepas de *Yersinia* isoladas de diversas regiões biogeográficas e de diferentes hospedeiros. A maioria dos projetos encontra-se ainda em fase de montagem, em andamento ou em fase inicial de produção de bibliotecas (www.ncbi.nih.gov). Todas as bibliotecas foram projetadas com base no método WGS (*Whole Genome Shotgun*), que em sua estratégia de sequenciamento, o DNA de alto peso molecular é picotado/clivado em fragmentos aleatórios. Posteriormente, os fragmentos são selecionados por tamanho (geralmente em fragmentos de 2, 10, 50, e 150 kb) e clonados em um vetor apropriado. Os clones são então sequenciados em ambas as terminações (3' e 5'). A distância entre os pares pode ser inferida se o tamanho da biblioteca for conhecido e tiver uma faixa limitada de desvio (Fleischmann *et al.* 1995).

Tabela 1 – Relação dos projetos genômicos de cepas de *Yersinia* patogênicas disponíveis nos bancos de dados da web (Fonte: NCBI, outubro de 2009).

Genomas	ID*	Status	isolado	Centro Responsável
<i>Y. enterocolitica</i> 8081	190(1)	Completo	septicemia humana	Welcome Trust Sanger Institute
<i>Y. pestis</i> Angola	16067	Completo	–	TIGR
<i>Y. pestis</i> Antiqua	16645(2)	Completo	solo/Congo	US DOE Joint Genome Institute
<i>Y. pestis</i> CA88-4125	18241	Montagem	Monterey,USA	ERIC/Joint Genome Institute
<i>Y. pestis</i> CO92	34(3)	Completo	EUA	Welcome Trust Sanger Institute
<i>Y. pestis</i> KIM	288(4)	Completo	–	University of Wisconsin-Madison, USA
<i>Y. pestis</i> Nepal516	16646(5)	Completo	solo/Nepal	US DOE Joint Genome Institute
<i>Y. pestis</i> Pestoides F	16700(6)	Completo	–	US DOE Joint Genome Institute
<i>Y. pestis</i> Microtus 91001	10638(7)	Completo	<i>Microtus</i> /China	AMMS/IME, China
<i>Y. pestis</i> D106004	36507	Andamento	–	State Key Lab. IDP e col.
<i>Y. pestis</i> D182038	36545	Andamento	–	State Key Lab. IDP e col.
<i>Y. pestis</i> EV76	39793	Andamento	–	Army Edgewood Chemical Biological Center
<i>Y. pestis</i> G1670	38409	Andamento	–	University of Florida
<i>Y. pestis</i> G3768	38305	Andamento	–	Army Edgewood Chemical Biological Center
<i>Y. pestis</i> Pestoides G	38307	Andamento	–	Army Edgewood Chemical Biological Center
<i>Y. pestis</i> Z176003	36547	Andamento	–	State Key Lab. IDP e col.
<i>Y. pestis</i> Medievalis Harbin 35	30505	Andamento	–	ERIC/Los Alamos National Laboratory
<i>Y. pestis</i> FV-1	1642(8)	Montagem	–	TGen
<i>Y. pestis</i> Nepal516	30503	Montagem	solo/Nepal	ERIC/Joint Genome Institute
<i>Y. pestis</i> Pestoides A	30511	Montagem	EUA	ERIC/Joint Genome Institute
<i>Y. pestis</i> Antiqua B42003004	17685	Montagem	–	ERIC/Joint Genome Institute
<i>Y. pestis</i> Antiqua E1979001	17683	Montagem	<i>Eothenomys</i> /China	TIGR
<i>Y. pestis</i> Antiqua UG05-0454	17691	Montagem	–	TIGR
<i>Y. pestis</i> Mediaevalis K1973002	17687	Montagem	Hetian/china	TIGR
<i>Y. pestis</i> Orientalis F1991016	17681	Montagem	Cangyuan/China	TIGR
<i>Y. pestis</i> Orientalis IP275	16376	Montagem	Madagascar	TIGR
<i>Y. pestis</i> Orientalis India 195	30507	Montagem	India	ERIC/Joint Genome Institute
<i>Y. pestis</i> Orientalis MG05-1020	17689	Montagem	Madagascar	TIGR
<i>Y. pestis</i> Orientalis PEXU2	30509	Montagem	–	ERIC/Joint Genome Institute
<i>Y. pseudotuberculosis</i> IP 31758	16070(9)	Completo	1B/Rússia	TIGR
<i>Y. pseudotuberculosis</i> IP 32953	12950(10)	Completo	–	Lawrence Livermore National Lab.
<i>Y. pseudotuberculosis</i> PB1/+	28745	Completo	1B	DOE Joint Genome Institute
<i>Y. pseudotuberculosis</i> YPIII	28743	Completo	O3	US DOE Joint Genome Institute

*Projeto no NCBI e (Nº) de Referências: 1)Thomson *et al.* (2006); 2 e 5) Chain *et al.* (2006); 3) Parkhill *et al.* (2001); 4) Deng *et al.* 2002; 6) Worsham *et al.* (2003); 7) Song *et al.* (2004); 8) Touchman *et al.* (2007); 9) Eppinger *et al.* (2007) e 10) Zhou *et al.* 2004.

Tabela 2 – Relação dos projetos genômicos de cepas de *Yersinia* não patogênicas disponíveis nos bancos de dados da web (Fonte: NCBI, até outubro de 2009).

Genomas*	Status	Isolado*	Mbp	GC%	Proteínas	RNAs
<i>Y. aldovae</i> ATCC 35236	inicial	Água/Checoslováquia	4.3	47.0	3967	66
<i>Y. bercovieri</i> ATCC 43970	inicial	Fezes humana/França	4.3	48.0	3880	67
<i>Y. frederiksenii</i> ATCC 3641	inicial	Esgoto/Dinamarca	4.8	46.0	4358	69
<i>Y. frederiksenii</i> 7175	inicial	Porco	–	–	–	–
<i>Y. frederiksenii</i> 867	inicial	Humano/Bélgica	–	–	–	–
<i>Y. frederiksenii</i> CNY 867	inicial	–	–	–	–	–
<i>Y. frederiksenii</i> WA 933	inicial	Alemanha	–	–	–	–
<i>Y. frederiksenii</i> WA 935	inicial	Alemanha	–	–	–	–
<i>Y. frederiksenii</i> WE 83/02	inicial	Bélgica	–	–	–	–
<i>Y. intermedia</i> ATCC 29909	inicial	urina humana	4.7	47.0	4122	73
<i>Y. kristensenii</i> ATCC 33638	inicial	urina humana	5.0	47.0	4542	69
<i>Y. kristensenii</i> 105	inicial	Dinamarca	–	–	–	–
<i>Y. kristensenii</i> 490	inicial	França	–	–	–	–
<i>Y. kristensenii</i> WA 948	inicial	Filândia	–	–	–	–
<i>Y. kristensenii</i> WE 180/98	inicial	Belgica	–	–	–	–
<i>Y. mollaretii</i> ATCC 43969	inicial	solo/EUA	4.5	49.0	4002	62
<i>Y. rohdei</i> ATCC 43380	inicial	–	4.3	46	3853	59
<i>Y. ruckeri</i> ATCC 29473	inicial	Truta	3.7	47	3365	53

*Projeto sem número de ID; Naval Medical Research Center responsável por todos os projetos

2.7. Evolução das sequências HGT, genômicas e a reconstrução filogenética

A transferência horizontal de genes ou a troca de material genético entre células ou genomas de espécies não relacionadas ocorre nos procariotos mediante os processos de conjugação, transdução e, principalmente, transformação (Hacker e Kaper, 2000) conhecidos desde 1928. Porém os impactos da HGT na evolução dos procariotos foram de fato estabelecidos nas últimas décadas, após o sequenciamento completo dos genomas e através de experimentos em laboratório com organismos modelo (Achtman *et al.* 2004; Collyn *et al.* 2004; Chain *et al.* 2006; Thomson *et al.* 2006; Derbise *et al.* 2007).

A composição de aminoácidos das proteínas varia substancialmente entre grupos taxonômicos, que podem assim evoluir. Segundo Jordan e colaboradores (2005), há uma tendência universal nas mudanças das frequências dos aminoácidos, que foi verificada a partir da análise comparativa de 15 grupos

envolvendo bactérias (archaeas e eucariotos). Essas análises sugeriram que as substituições de aminoácidos foram representadas pelo ganho de Cys, Met, His, Ser e Phe e perda de Pro, Ala, Glu e Gly em 14 grupos de espécies e permitem supor que, de uma forma geral, os aminoácidos em declínio de frequência devem ser os mais antigos no código genético e, inversamente, todos os aminoácidos com frequências crescentes (com exceção de Ser), provavelmente são mais recentes (Jordan *et al.* 2005).

As tendências aparentes no ganho e na perda de aminoácidos durante a evolução de uma proteína devem-se à variação quase neutra. Por exemplo, se um aminoácido *A* é favorecido pela seleção natural e sofre mutação para um aminoácido *B* e a seleção contra *B* for bastante fraca (suavemente deletéria), *B* pode permanecer por um tempo antes de ser substituído novamente por *A* (McDonald, 2006). Geralmente se aceita que as proteínas estão em equilíbrio ou em evolução neutra, admitindo-se a existência de uma tendência dos ancestrais de perder aminoácidos (Nei e Gojobori, 1986).

A compreensão da história evolutiva das HGT requer a análise das populações, seus ambientes e a compreensão prévia da pressão seletiva a que estão submetidas as unidades gênicas avaliadas. A amplitude dos valores da razão entre as taxas de substituições sinônimas e não sinônimas diminui o sinal filogenético e prejudica a formação dos clusters baseados em alinhamentos (Nei e Gojobori, 1986; Nei e Kumar, 2000; Arber, 2000; Budd *et al.* 2004; McDonald, 2006).

A presença de substituições silenciosas pode obscurecer os sinais totais da assimetria das diferentes trincas de códons de um aminoácido, que deriva na maior parte das substituições de recolocação, criando problemas nos modelos filogenéticos baseados em aminoácidos. Os genes corretamente determinados no ortólogo *p* são necessários para calibrar estimativas da divergência. Especificamente, se no alinhamento das bases de duas duplicatas de sequências de aminoácidos divergentes assimetricamente, uma das duplicatas se tornar mais divergente do que a sequência de DNA verdadeira do ortólogo *p* (Conant e Wagner, 2003).

Duas estratégias diferentes têm sido usadas na genômica comparativa e ambas apresentam três aspectos. A primeira estratégia é baseada na

comparação entre sequências e permite: verificar os alinhamentos múltiplos entre sequências de DNA homólogas ou unidades colineares (os genes semelhantes são concatenados); quantificar as frequências dos ortólogos, e verificar a heterogeneidade do uso de códons; avaliar os padrões dentro do genoma a partir das diferenças de conteúdos nucleotídeos entre os genes (Darling *et al.* 2004).

A segunda estratégia é baseada em assinaturas de nucleotídeos ou dos genomas e reflete o uso de códons, permitindo as seguintes avaliações: comparar sequências de DNA não-homólogas, medindo vícios de dinucleotídeos e especificando as relações evolucionárias entre cepas de vírus (Karlin *et al.* 1994); medir distâncias taxonômicas entre grupos de espécies, plasmídeos e organelas de eucariotos (Wang e Hickey, 2007; Jorgensen *et al.* 2007); detectar PAI em bactérias (Collyn *et al.* 2006).

Os métodos mais amplamente utilizados para calcular as distâncias são baseados em frequências de oligonucleotídeos, como a distância Euclidiana e a de Pearson, e permitem verificar os vícios no uso de códons sinônimos, que por sua vez estão associados a diversos fatores biológicos, tais como: o nível de expressão gênica, comprimento do gene, início do sinal de tradução do gene, composição de aminoácido das proteínas, estrutura protéica, abundância de tRNA, frequências e padrões das mutações e conteúdos de GC. A quantificação do viés do uso de códons ajuda a entender a história evolutiva dos seres vivos a partir da análise das sequências homólogas (Karlin *et al.* 1994; Supek e Vlahovicek, 2004; Tamura *et al.* 2007).

A análise de sequências homólogas *in silico* requer três conceitos prévios: ortologia, quando as sequências têm um mesmo ancestral comum; paralogia, quando se originam de uma duplicação gênica; xenologia, quando se originam por incursão lateral ou horizontal. Apenas sequências ortólogas poderão fornecer informações filogenéticas da história dos organismos. Paradoxalmente, a topologia de uma árvore construída com genes xenólogos poderá evidenciar a propagação dos genes pela HGT (Budd *et al.* 2004).

Assim, o paradigma é que a topologia das árvores construídas com genes transferidos lateralmente pode evidenciar a história evolutiva dos grupos de bactérias que já receberam esses genes de um ancestral comum. Para detectar a

HGT de regiões de PAI a identidade no BLAST/NCBI (www.ncbi.nlm.nih.gov) do homólogo deve ser de 80% para a sequência de DNA ou 25% para a sequência da proteína e a região alinhada de ambos com cerca 70% dos comprimentos das sequências de entrada no BLAST ou *query* (Todd *et al.* 2001; Yoon *et al.* 2005).

As sequências que corresponderem a cada loco de uma PAI podem ser identificadas a partir das regiões que contiverem quatro ou mais homólogos dos genes da mesma ascensão de PAI e unidas as regiões vizinhas (Yoon *et al.* 2005). Um loco de uma PAI deve conter ao menos um gene de virulência com alta identidade entre as sequências (>80%), para inferir que as suas funções são idênticas. A identificação de baixa identidade (<30%) sugere diferenças de funções mesmo havendo proteínas claramente relacionadas pela evolução (Todd *et al.* 2001).

Para um alinhamento coerente é necessário observar os diferentes tipos de substituições de nucleotídeos. Segundo Nei e Kumar (2000), metodologias filogenéticas são essencialmente estatísticas e podem ser classificadas em dois grupos principais, de acordo com seus critérios quantitativos (métodos de distância) e qualitativos (métodos de parcimônia e verossimilhança). A partir da matriz gerada podem ser construídas as árvores com um dado algoritmo, sendo os mais utilizados o UPGMA (Unweighted pair group method with arithmetic mean) e *neighbor-joining* (Tamura *et al.* 2007).

Para construir múltiplos alinhamentos de genomas completos e analisar grandes eventos em escala evolutiva, como rearranjo e inversão, Darling e colaboradores (2004) desenvolveram um programa de alinhamentos múltiplos de genomas que fornece uma base para as investigações em matéria de genômica comparativa e o estudo da dinâmica evolutiva. Alinhar múltiplos genomas completos é um problema fundamentalmente diferente do alinhamento de curtas sequências. O processo pode ser realizado em um tempo relativamente curto pelo programa MAUVE (Multiple genome alignment).

O programa MAUVE foi desenvolvido com a ideia de que alinhamentos múltiplos de genomas devem exigir apenas modestos recursos computacionais, empregando técnicas algorítmicas progressivas quando novas sequências são alinhadas. Por exemplo, um par de genomas de cepas de *Y. pestis* pode ser alinhado em menos de um minuto, enquanto um grupo de nove divergentes

genomas de Enterobacteriaceae pode ser alinhado em poucas horas (Darling *et al.* 2004). Existe também o pressuposto de que os genomas são colineares e permitem a construção de blocos colineares local (LCBs).

Um programa muito versátil que oferece uma variedade de recursos úteis na análise do uso de códons de todos os genes dos genomas bacterianos é o *Interactive codon usage analysis* ou INCA (Supek e Vlahovicek, 2004). Esse programa também fornece vários índices relacionados aos códons, ele produz as frequências de GC3 e utiliza gráficos interativos. Por outro lado, o programa *Statistical package for the social* ou SPSS (Norusis, 1998) oferece uma ampla gama de capacidades de análise estatística, tais como os testes de correlação, multicolinearidade e de hipóteses e também permite contagens de frequências, ordenar dados, reorganizar a informação além de possuir saídas gráficas.

Na necessidade de acrescentar-se um novo modelo de análise para associar múltiplos dados das sequências não fornecidos pelo INCA pode-se desenvolver seleção de parâmetros para programação e criar um *script* utilizando ambiente de modelagem computacional, como por exemplo, no *MATrix LABoratory* ou MatLab (www.mathworks.com).

O MatLab é um programa interativo de alto desempenho, voltado para o cálculo numérico. A versatilidade desse programa permite integrar análise numérica, cálculo com matrizes, processamento de sinais e construção de gráficos em ambientes onde problemas e soluções são expressos somente como eles são escritos matematicamente, ao contrário da programação tradicional (www.mathworks.com).

A associação de ferramentas computacionais pode não apenas efetuar avaliações quantitativas e qualitativas de sequências de cromossomos inteiros como pode permitir, por exemplo, reavaliar a história da Ilha de alta patogenicidade das Yersínias, a partir de seus padrões e variações genéticos e evolutivos.

3. Objetivos

3.1. Objetivo Geral

Avaliar genomas bacterianos e as regiões relacionadas ao sistema yersiniabactina para determinar associações de moléculas, padrões e variações evolutivas da HPI dentro e fora do gênero *Yersinia*.

3.2. Objetivos Específicos

Construir bancos de dados local com sequências homólogas dos genes *ybt*, genes reguladores de *quorum sensing* e genes responsáveis pela produção de polissacarídeos e com as sequências de seus respectivos genomas hospedeiros;

Fazer uma análise comparativa das reconstruções filogenéticas dos genes *ybt* dos genomas hospedeiros e verificar a existência de correlação entre as frequências de aminoácidos das HPIs e seus hospedeiros;

Avaliar os padrões evolutivos das frequências do uso de códons, aminoácidos, conteúdos de GC nas HPIs, suas regiões vizinhas e dos genomas hospedeiros;

Determinar a razão entre substituições sinônimas e não sinônimas dos sítios das sequências *ybt* para avaliar seus valores adaptativos;

Comparar as HPIs e suas regiões vizinhas para detectar padrões e divergências entre sequências homólogas com similaridades variadas e avaliar a presença de sequências associadas ao sistema *quorum sensing* e produção de polissacarídeos;

Verificar eventuais eventos de rearranjos cromossômicos entre *ybt* e possíveis padrões de associações com as sequências adjacentes, assim como das sequências cromossômicas dos genomas hospedeiros das cepas de *Y. pestis*;

Desenvolver parâmetros para programação biológica de uma avaliação simultânea de conteúdos de GC1, GC2 e GC3 nos genomas hospedeiros da HPI para estudos filogenômicos.

4. Referências Bibliográficas

- Achtman M, Morelli G, Zhu P, Wirth T, Diehl I, *et al.* (2004) Microevolution and history of the plague bacillus, *Yersinia pestis*. *Proc Natl Acad Sci USA*. 101:17837-42.
- Achtman M, Zurth K, Morelli G, Torrea G, Guiyoule A, *et al.* (1999) *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci U S A*. 96:14043-8.
- Anisimov AP, Lindler LE, Pier GB (2004) Intraspecific diversity of *Yersinia pestis*. *Clin Microbiol Rev*. 17:434-64.
- Antonenka U, Nölting C, Heesemann J, Rakin A (2005) Horizontal transfer of *Yersinia* high-pathogenicity island by the conjugative RP4 *attB* target-presenting shuttle plasmid. *Mol Microbiol*. 57:727-34.
- Arber W (2000) Genetic variation: molecular mechanisms and impact on microbial evolution. *FEMS Microbiol Rev*. 24:1-7.
- Atkinson S, Sockett RE, Cámara M, Williams P (2006) Quorum sensing and the lifestyle of *Yersinia*. *Curr Issues Mol Biol*. 8:1-10.
- Bobrov AG, Kirillina O, Forman S, Mack D, Perry RD (2008) Insights into *Yersinia pestis* biofilm development: topology and co-interaction of Hms inner membrane proteins involved in exopolysaccharide production. *Environ Microbiol*. 10:1419-32.
- Boettcher KJ, Ruby EG (1995) Detection and quantification of *Vibrio fischeri* autoinducer from symbiotic squid light organs. *J Bacteriol*. 177:1053-8.
- Brem D, Pelludat C, Rakin A, Jacobi CA, Heesemann J (2001) Functional analysis of yersiniabactin transport genes of *Yersinia enterocolitica*. *Microbiology*. 147:1115-27.
- Brito CFA, Carvalho CMB, Santos FR, Gazzinelli RT, Oliveira SC, *et al.* (2004) *Chromobacterium violaceum* genome: molecular mechanisms associated with pathogenicity. *Genet Mol Res*. 3:148-161.

- Brummer IH, Felske A, Wagner-Döbler I (2006) Diversity and seasonal variability of beta-Proteobacteria in biofilms of polluted rivers: analysis by temperature gradient gel electrophoresis and cloning. *Appl Environ Microbiol.* 69:4463-73.
- Buchrieser C, Prentice M, 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.* 180:2321-29.
- Buchrieser C, Rusniok C, Frangeul L, Couve E, Billault A, *et al.* (1999) The 102-kilobase *pgm* locus of *Yersinia pestis*: sequence analysis and comparison of selected regions among different *Yersinia pestis* and *Yersinia pseudotuberculosis* strains. *Infect. Immun.* 67:4851-61.
- Budd A, Blandin S, Levashina EA, Gibson TJ (2004) Bacterial α_2 -macroglobulins: colonization factors acquired by horizontal gene transfer from the metazoan genome?. *Genome Biol.* 5: R38.
- Carniel E (2008) Pathogenic *Yersinia*: Stepwise Gain of Virulence due to Sequential Acquisition of Mobile Genetic Elements. *In* Horizontal gene transfer in the evolution of pathogenesis (Eds. H Hensel and M Schmidt). Cambridge University Press, Cambridge, 194-216.
- Carniel E, Guilvout I, Prentice M (1996) Characterization of a large chromosomal "high-pathogenicity island" in biotype 1B *Yersinia enterocolitica*. *J Bacteriol.* 178:6743-51.
- Chain PS, Hu P, Malfatti SA, Radnedge L, Larimer F, Vergez LM, *et al.* (2006) Complete genome sequence of *Yersinia pestis* strains Antiqua and Nepal516: evidence of gene reduction in an emerging pathogen. *J Bacteriol.* 188:4453-63.
- Collyn F, Billault A, Mullet C, Simonet M, Marceau M (2004) YAPI, a new *Yersinia pseudotuberculosis* pathogenicity island. *Infect Immun.* 72:4784-90.
- Collyn F, Guy L, Marceau M, Simonet M, Roten CA (2006) Describing ancient horizontal gene transfers at the nucleotide and gene levels by comparative pathogenicity island genomics. *Bioinformatics.* 22:1072-1079.
- Conant GC, Wagner A (2003) Asymmetric sequence divergence of duplicate genes. *Genome Res.* 13:2052-8.

- Cornelis GR (2002) *Yersinia* type III secretion: send in the effectors. J Cell Biol. 158:401-8.
- Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14:1394-403.
- Deng W, Burland V, Plunkett G, Boutin A, Mayhew GF, et al. (2002) Genome sequence of *Yersinia pestis* KIM. J Bacteriol. 184:4601-11.
- Derbise A, Chenal-Francisque V, Pouillot F, Fayolle C, Prévost MC, et al (2007) A horizontally acquired filamentous phage contributes to the pathogenicity of the plague bacillus. Mol Microbiol. 63:1145-57.
- Devignat R (1951) Varieties of *Pasteurella pestis*: new hypothesis. Bull World Health Organ. 4:247-63.
- Dobrindt U, Hochhut B, Hentschel U, Hacker J (2004) Genomic islands in pathogenic and environmental microorganisms. Nat Rev Microbiol. 2:414–424.
- Eppinger M, Rosovitz MJ, Fricke WF, Rasko DA, Kokorina G, et al. (2007) The complete genome sequence of *Yersinia pseudotuberculosis* IP31758, the causative agent of Far East scarlet-like fever. PLoS Genet. 8:e142.
- Fetherston JD, Bertolino VJ, Perry RD (1999) YbtP and YbtQ: two ABC transporters required for iron uptake in *Yersinia pestis*. Mol Microbiol. 32:289-99.
- Fetherston JD, Perry RD (1994) The pigmentation locus of *Yersinia pestis* KIM6⁺ is flanked by an insertion sequence and includes the structural genes for pesticin sensitivity and HMWP2. Mol Microbiol. 13:697-708.
- Fetherston JD, Schuetze P, Perry RD (1992) Loss of the pigmentation phenotype in *Yersinia pestis* is due to the spontaneous deletion of 102 kb of chromosomal DNA which is flanked by a repetitive element. Mol Microbiol. 6:2693-704.
- Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science. 269:496-512.

- Fuqua WC, Winans SC, Greenberg EP (1994) Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *J Bacteriol.* 176:269-75.
- Gray KM and Garey JR (2001) The evolution of bacterial LuxI and LuxR quorum sensing regulators. *Microbiology.* 147:2379–2387.
- Hacker J, Blum-Oehler G, Muhldorfer I, Tschape H (1997) Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol.* 23:1365-2958.
- Hacker J, Carniel E (2001) Ecological fitness, genomic islands and bacterial pathogenicity. *EMBO Rep.* 2:376–381.
- Hacker, J, Kaper JB (2000) Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.* 54:641-679.
- Hall BG (1999) Transposable elements as activators of cryptic genes in *E. coli*. *Genetica.* 107:181-7.
- Heermann R, Fuchs TM (2008) Comparative analysis of the *Photorhabdus luminescens* and the *Yersinia enterocolitica* genomes: uncovering candidate genes involved in insect pathogenicity. *BMC Genomics.* 9:40.
- Higgins DA, Pomianek ME, Kraml CM, Taylor RK, Semmelhack MF, *et al.* (2007) The major *Vibrio cholerae* autoinducer and its role in virulence factor production. *Nature.* 450:883-886.
- Iteman I, Guiyoule A, de Almeida AM, Guilvout I, Baranton G, *et al.* (1993) Relationship between loss of pigmentation and deletion of the chromosomal iron-regulated *irp2* gene in *Yersinia pestis*: evidence for separate but related events. *Infect Immun.* 61: 2717-2722.
- Jordan IK, Kondrashov FA, Adzhubei IA, Wolf YI, Koonin EV, *et al.* (2005) Universal trend of amino acid gain and loss in protein evolution, *Nature.* 433:633-638.
- Jorgensen FG, Schierup MH, Clark AG (2007) Heterogeneity in regional GC content and differential usage of codons and amino acids in GC-poor and GC-rich regions of the genome of *Apis mellifera*. *Mol Biol Evol.* 24:611-9.
- Karlin S, Mocarski ES, Schachtel GA (1994) Molecular evolution of herpesviruses: genomic and protein sequence comparisons. *J Virol.* 68:1886-902.

- Lawrence JG (2002) Gene transfer in bacteria: speciation without species? *Theor Popul Biol.* 61:449-60.
- Lesic B and Carniel E. (2004) The High-Pathogenicity Island: a broad-host- range pathogenicity island. In *Yersinia* molecular and cellular biology (Eds. E. Carniel and BJ Hinnebusch). Horizon Biociensie, 285-306.
- Lesic B, Carniel E (2005) Horizontal Transfer of the High-Pathogenicity Island of *Yersinia pseudotuberculosis*. *J Bacteriol.* 187: 3352–3358.
- Lillard JW Jr, Fetherston JD, Pedersen L, Pendrak ML, Perry RD (1997) Sequence and genetic analysis of the hemin storage (*hms*) system of *Yersinia pestis*. *Gene.* 193:13-21.
- Mascaretti AO (2003) Molecular genetics of bacteria. In *Bacteria versus antibacterial agents: an integrated approach*. ASM Press, Washington, DC. 69-86.
- McDonald JH (2006) Apparent trends of amino acid gain and loss in protein evolution due to nearly neutral variation. *Mol Biol Evol.* 23:240-4.
- Merhej V, Adékambi T, Pagnier I, Raoult D, Drancourt M. (2008) *Yersinia massiliensis* sp. nov., isolated from fresh water. *Int J Syst Evol Microbiol.* 58:779-84.
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol.* 5:418-26.
- Nei M, Kumar S (2000) *Molecular evolution and phylogenetics*. Oxford University Press: Oxford, pp 333.
- Norusis MJ (1998) *Statistical Package for the Social Sciences (SPSS)*. Chicago (IL), software version 9.0. Published on the internet at <http://www.spss.com/>.
- Oelschlaeger TA, Hacker J (2004) Impact of pathogenicity islands in bacterial diagnostics. *APMIS.* 112:930-936.
- Parkhill J, Wren BW, Thomson NR, Titball RW, Holden MT, *et al.* (2001) Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature.* 413:523-527.
- Pelludat C, Brem D, Heesemann J (2003) *Irp9*, encoded by the high-pathogenicity island of *Yersinia enterocolitica*, is able to convert chorismate

- into salicylate, the precursor of the siderophore yersiniabactin. *J Bacteriol.* 185:5648-5653.
- Perry RD, Fetherston JD (1997) *Yersinia pestis*-etiologic agent of plague. *Clin Microbiol Rev.* 10:35-66.
- Perry RD, Shah J, Bearden SW, Thompson JM, Fetherston JD (2003) *Yersinia pestis* TonB: role in iron, heme, and hemoprotein utilization. *Infect Immun.* 71:4159-62.
- Pollitzer R (1954) Plague. *W. H. O. Monogr Ser.* 22:1–698.
- Qazi S, Middleton B, Muharram SH, Cockayne A, Hill P, *et al.* (2006) N-acyl homoserine lactones antagonize virulence gene expression and quorum sensing in *Staphylococcus aureus*. *Infect Immun.* 74:910-9.
- Schubert S, Dufke S, Sorsa J, Heesemann J (2004) A novel integrative and conjugative element (ICE) of *Escherichia coli*: the putative progenitor of the *Yersinia* high-pathogenicity island. *Mol Microbiol.* 51:837-48.
- Schubert S, Picard B, Gouriou S, Heesemann J, Denamur E (2002) *Yersinia* high-pathogenicity island contributes to virulence in *Escherichia coli* causing extraintestinal Infections. *Infect Immun.* 70: 5335–5337.
- Sing A, Reithmeier-Rost D, Granfors K, Hill J, Roggenkamp A, Heesemann J (2005) A hypervariable N-terminal region of *Yersinia* LcrV determines Toll-like receptor 2-mediated IL-10 induction and mouse virulence. *Proc Natl Acad Sci U S A.* 102:16049-54.
- Solomon JM, Grossman AD (1996) Who's competent and when: regulation of natural genetic competence in bacteria. *Trends Genet.* 12:150-5.
- Song Y, Tong Z, Wang J, Wang L, Guo Z, *et al.* (2004) Complete genome sequence of *Yersinia pestis* strain 91001, an isolate avirulent to humans. *DNA Res.* 11:179-97.
- Sprague LD, Neubauer H. (2005) *Yersinia aleksiciae* sp. nov. *Int J Syst Evol Microbiol.* 55:831-5.
- Sprague LD, Scholz HC, Amann S, Busse HJ, Neubauer H. (2008) *Yersinia similis* sp. nov. *Int J Syst Evol Microbiol.* 58:952-8.

- Sulakvelidze A (2000) *Yersiniae* other than *Y. enterocolitica*, *Y. pseudotuberculosis*, and *Y. pestis*: the ignored species. *Microbes Infect.* 2:497-513.
- Supek F, Vlahovicek K (2004) INCA: synonymous codon usage analysis and clustering by means of self-organizing map. *Bioinformatics.* 20:2329-2330.
- Tamayo R, Pratt JT, Camilli A (2007) Roles of cyclic diguanylate in the regulation of bacterial pathogenesis. *Annu Rev Microbiol.* 61:131-48.
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol.* 24:1596-99.
- Thomas CM, Nielsen KM (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol.* 3:711-21.
- Thomson NR, Howard S, Wren BW, Holden MT, Crossman L, *et al.* (2006) The complete genome sequence and comparative genome analysis of the high pathogenicity *Yersinia enterocolitica* strain 8081. *PLoS Genet.* 2:2039-2051.
- Todd AE, Orengo CA, Thornton JM (2001) Evolution of function in protein superfamilies, from a structural perspective. *J Mol Biol.* 307:1113-1143.
- Touchman JW, Wagner DM, Hao J, Mastrian SD, Shah MK, *et al.* (2007) A North American *Yersinia pestis* draft genome sequence: SNPs and phylogenetic analysis. *PLoS One.* 2:e220.
- Ubeda C, Barry P, Penadés JR, Novick RP (2007) A pathogenicity island replicon in *Staphylococcus aureus* replicates as an unstable plasmid. *Proc Natl Acad Sci U S A.* 104:14179-81.
- Wang HC, Hickey DA (2007) Rapid divergence of codon usage patterns within the rice genome. *BMC Evol Biol.* 7 Suppl 1:S6.
- Wang X, Preston JF 3rd, Romeo T (2004) The *pgaABCD* locus of *Escherichia coli* promotes the synthesis of a polysaccharide adhesin required for biofilm formation. *J Bacteriol.* 186:2724-34.
- Waters CM, Lu W, Rabinowitz JD, Bassler BL (2008) Quorum sensing controls biofilm formation in *Vibrio cholerae* through modulation of cyclic di-GMP levels and repression of *vpsT*. *J Bacteriol.* Apr. 190:2527-2536.
- Watnick P, Kolter R (2000) Biofilm, city of microbes. *J Bacteriol.* 182:2675-9.

- Wessler SR (2006) Eukaryotic transposable elements and genome evolution special feature: Transposable elements and the evolution of eukaryotic genomes. *Proc Natl Acad Sci U S A*. 103: 17600–17601.
- Whitfield C (1988) Bacterial extracellular polysaccharides. *Can J Microbiol*. 34:415-20.
- Worsham PL, Roy C (2003) Pestoides F, a *Yersinia pestis* strain lacking plasminogen activator, is virulent by the aerosol route. *Adv Exp Med Biol*. 529:129-31.
- Yoon SH, Hur CG, Kang HY, Kim YH, Oh TK, *et al.* (2005) A computational approach for identifying pathogenicity islands in prokaryotic genomes. *BMC Bioinformatics*. 184:1-11.
- Zhou D, Tong Z, Song Y, Han Y, Pei D, *et al.* (2004) Genetics of metabolic variations between *Yersinia pestis* biovars and the proposal of a new biovar, Microtus. *J Bacteriol*. 186:5147-52.

5. Artigo Científico

(Submetido a revista: ***Genetics and Molecular Research***)

Artigo 1 - Evolution of codon usage in the High-Pathogenicity Island of Yersinia and identification of association patterns of adjacent molecule groups

Evolution of codon usage in the High-Pathogenicity Island of *Yersinia* and identification of association patterns of adjacent molecule groups

N.S.A. Freitas^{1,2}, T.A.E. Ferreira³, M.R. da Sena-Júnior⁴, C.R.L. Costa-Junior², T.C. Leal-Balbino⁵, A.M.P. de Almeida⁵, V.Q. Balbino²

¹Departamento de Biologia, Universidade Federal Rural de Pernambuco, Recife, PE, Brasil

²Departamento de Genética, Universidade Federal de Pernambuco, Recife, PE, Brasil

³Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco, Recife, PE, Brasil

⁴Departamento de Estatística, Universidade Federal de Pernambuco, Recife, PE, Brasil

⁵Departamento de Microbiologia, Centro de Pesquisas Aggeu Magalhães, Recife, PE, Brasil

Corresponding author: Valdir de Queiroz Balbino

E-mail: vqbalbino@pq.cnpq.br

Departamento de Genética, Universidade Federal de Pernambuco, Av. Prof. Moraes Rego
S/N Cidade Universitária, 50732-970 - Recife, PE, Brazil

Phone/Fax: (5581) 21268512

ABSTRACT: Our assessment of the neighboring sequences of *Yersinia* high-pathogenicity island (HPI), *ybt* genes selective signatures and codon usage frequency of host genomes revealed new aspects of the HPI evolution. The analysis of the *ybt* sequences of the *Y. pestis* strain CO92 and their homologous in other species suggests that contrary to previous understanding the three pathogenic *Yersinia* had a single HPI donor. This is supported by a null selective signature found in *Y. pestis* and its ancestor *Y. pseudotuberculosis ybt* genes. In *Y. enterocolitica* the rate of GC3-rich genes is lower than in the other two species and the selective signatures of the *ybt* genes are negative. A significant correlation was found between the amino acids frequency of the HPI and host genomes. Probably a higher selection pressure on the *Y. enterocolitica* HPI produced a higher frequency of non-synonymous codons. Furthermore, *Y. enterocolitica* and *Y. pseudotuberculosis*, but *Y. pestis*, share unique sequences in the neighboring HPI regions what strengthen the likelihood of a single HPI donor and a common ancestor. Different gene sequences from all neighboring HPI regions were regulated by *quorum-sensing*, indicating a general regulatory network involving Ybt.

Key words: *Yersinia pestis*, codon usage, evolution, comparative genomics.

INTRODUCTION

The genus *Yersinia* from the Enterobacteriaceae family has 14 species, three of which are pathogenic: *Y. pestis*, the etiologic agent of plague, and the enteropathogens *Y. enterocolitica* and *Y. pseudotuberculosis* (Perry and Fetherston, 1997; Sulakvelidze 2000; Sprague et al., 2008; Merhej et al., 2008). It is accepted that *Y. enterocolitica* and *Y. pseudotuberculosis* diverged about 200 million years ago and *Y. pestis* originated from *Y. pseudotuberculosis* between 1.5 and 20 thousand years ago (Achtman et al., 2004; Carniel, 2008).

The evolution of bacterial virulence is related to horizontal acquisition of genetic elements, in particular the genomic islands (GIs), which encode pathogenicity factors and are also known as pathogenicity islands (PAIs) (Hacker and Kaper, 2000; Dobrindt et al., 2004; Collyn et al., 2006). Upon transfer to the host genome the GI undergo codon usage selection, which may be faster or slower, depending on the GC/AT content of both sequences and on the

characteristics of neighboring sequences in the island insertion site. The changes in codon usage can be faster when the number of GI is high. Distortions of the GC content and the genome homogenization process indicate high levels of rearrangements (Oda et al., 2008). This information allied to the data of GC3 (GC content of the third codon position) evolution show the genome as a selection unit. Grantham et al. (1980) proposed that each gene in a genome tends to have a GC3 value according to its species what can explain the molecular selection mechanisms (Li, 1987).

The selective value of a gene in different species can vary during evolution, suggesting that identification of beneficial or harmful selection signatures might allow understanding the evolution of groups of genes of interest, especially by simultaneous use of qualitative and quantitative methods and detection of a basal taxonomic unit.

A determinant PAI of the *Yersinia* virulence evolution, the High-Pathogenicity Island (HPI), is a mobile sequence flanked by *attB* sites adjacent to the integrase (*int*) mobility gene insertion site, homologous to that of the bacteriophage P4. The *Yersinia* HPI harbors the *ybt* locus coding for a set of proteins responsible for biosynthesis, transport and regulation of the yersiniabactin (Ybt) siderophore. The *ybt* genes appear to be organized into four groups: the first comprises genes (*irp2*, *irp1*, *ybtU*, *ybtT* and *ybtE*) involved in the Ybt biosynthesis; the second includes genes (*ybtP*, *ybtQ*, *ybtX* and *ybtS*) responsible for the Ybt uptake; the third is the outer membrane receptor (*fyuA*); and the fourth is a regulator (*ybtA*) (Buchrieser et al., 1999; Fetherston et al., 1999; Geoffroy et al., 2000).

Previous studies on the HPI evolution were limited to multiple alignment evaluation which did not include the analysis of codon usage. The purpose of our study was to analyze the evolutionary pattern of the *Yersinia* HPI based on the codon usage bias and synteny of neighboring sequences. Moreover, an analysis based on the selection signatures of the *ybt* genes, the correlation among the HPI and their host genomes and a phylogenetic reconstruction based on complete genomes, were performed for the first time in the present study.

MATERIAL AND METHODS

Selection of strains with entire HPI

The HPI sequences of the *Y. pestis* strain CO92 (accession number AL590842), representative of the last plague pandemic (Perry and Fetherston, 1997; Parkhill et al., 2001) were obtained from the GenBank at NCBI (<http://www.ncbi.nlm.nih.gov>) and used for selection of species with entire HPI. The sequences of 11 genes involved in the biosynthesis, transport and regulation of the yersiniabactin (Ybt) siderophore and the *int* adjacent sequence were identified in the NCBI and compiled in a local database. The concatenated *ybt* sequences were then used as query for the localization of homologous of other bacterial genomes by Blastn/NCBI.

Determination of selection signatures and phylogenetic reconstruction of the HPI

The homologous sequences of *int* and 11 *ybt* genes of the selected genomes were organized in two local databases: one with homologous sequences of *int* and *ybt* genes, and the other containing the concatenated sequences of the *ybt* genes. Subsequently, these sequences were aligned using ClustalW v.1.6 software, implemented in MEGA v.4.1 (Tamura et al., 2007).

After alignment the ratio (*R*) of synonymous and non-synonymous substitution rates (dN/dS) was calculated for each group of sequences by the modified Nei-Gojobori method (Nei and Gojobori, 1986) using MEGA. With the aligned sequences, an evolutionary model

was evaluated for the dataset using the Akaike information criterion (AIC), by the Modeltest 3.7 (Posada et al., 2004).

The phylogenetic analysis of the *ybt* genes was performed using PAUP* version 4.0b10 (Swofford, 2002) by exhaustive search. The branches consistence was evaluated by 10.000 bootstrap pseudo-replications (Felsenstein, 1985). The TreeView program version 4.0 (Page, 2002) was used to visualize the dendrograms. The heuristic Maximum Parsimony (MP) and Maximum Likelihood (ML) criteria in phylogenetic tree reconstruction were explored. The per-character consistency homoplasy index (HI) and rescaled consistency index (RC) were used to evaluate the phylogram.

Phylogenetic reconstruction of the host genome and evaluation of neighboring HPI regions

The sequences of all mRNAs of the selected host genomes (HG) were localized in the GenBank FTP/NCBI directory and compiled in a local database. To study the patterns of association among adjacent molecule groups, the sequences of the *ybt* locus and all genes located 30kb upstream (up_HPI) and downstream (HPI_down) were identified and compiled in local databases and named up_HPI_down. The files with the sequences of HPI, up_HPI_down and HG were submitted to INCA v. 2.1 software (Supek and Vlahovicek, 2004) to obtain the matrices of codon usage, amino acid and GC3 frequencies of each ORF (Open Reading Frame) with more than 10 codons.

The matrices of codon usage and amino acid frequencies were used for phylogenetic analysis of the HG and the HPI-HG correlation respectively. A dendrogram was constructed based on the hierarchical cluster analysis (between-group linkage) and the similarity correlation was based on Pearson's correlation coefficient. To investigate the GC3 frequencies of the *ybt* and *int* plus up_HPI_down regions and the HG, the values were divided into nine classes of frequency (0.11_0.20 to 0.91_1.00) for Correspondence Analysis. These evaluations were performed using software SPSS (SPSS for Windows, Rel. 13, Chicago, IL, USA), with the metric Euclidean distance.

To identify patterns of association among groups of adjacent molecules, the *ybt* sequences and their organization and the files containing the sequences of the up_HPI_down regions were submitted to Mauve.2.1 software (Darling et al., 2004) followed by analysis of the locally collinear blocks (LCBS). In case of doubt on the identities of ORFs (hypothetical protein), the sequence pairs were evaluated using the Blast 2 sequences tool (Tatusova and Madden, 1999) or Blast X (Altschul et al., 1990).

RESULTS

Eleven *ybt* genes homologous to those of the *Y. pestis* strain CO92 were analyzed in the genomes of five strains of *Y. pestis* (KIM, Microtus or 91001, Angola, Antiqua and Pestoides F), one *Y. pseudotuberculosis* IP32953, one *Y. enterocolitica* 8081 and four *E. coli* (strains 536, APEC01, CFT073 and UTI89). We observed that as *ybtS* in *Y. pestis* Angola, *int* is a pseudogene in *Y. pseudotuberculosis* and *Y. enterocolitica*. In *E. coli* CFT073 we detected an *IS200* (c2425) within the ORFs of *irp2* and three ORFs (c2427-28-29) homologous to different fragments of *irp1*. To include this strain in the analysis the *IS200* was manually removed and the homologous fragments were linked.

The phylogenetic reconstruction was based on the concatenated homologous *ybt* sequences. The TMV+I+G evolutionary model and the topology were obtained using the same algorithm for the ML and MP. All clades presented were supported by >60% bootstrap. The trees were not rooted although the *E. coli* strains behaved as outgroup. Thus, the topology was automatically rooted, highlighting the monophyly of the HPI which reflects the

monophyly of the Yersiniaceae group with 100% bootstrap for MP and ML (Figure 1), supported with HI = 0.033 and RC = 0.773.

The types of selection or selective signatures determined by *R* on the *ybt* sequences are shown in Table 1. The selection signatures of *Y. pestis* and *Y. pseudotuberculosis* strains were null ($dN/dS = 0$) in all *ybt* units. In *E. coli* the selection was positive for *ybtE*, *ybtU* and *irp2* ($dN/dS > 1$), neutral for *ybtP* ($dN/dS = 1$), null for *ybtT*, and negative for the other genes analyzed ($dN/dS < 1$) revealing different selection pressures over *ybt*. The number of conserved nucleotides and codons were, respectively, 27772 (97.49%) and 9021 (95.01%).

The phylogenetic reconstruction was also based on the codon usage frequency of the host genomes (HG). This allowed clustering the *E. coli* strains into a single group, placing *Y. enterocolitica* as the oldest species of the genus *Yersinia* but do not differentiate *Y. pseudotuberculosis* from *Y. pestis* (Figure 2).

Information on the number of mRNAs, average sizes in kb, GC% and GC3% in the regions up_HPI_down, HPI and HG are shown on Table 2. Although different in the four species studied, no intraspecific variation was observed for the average GC3 frequency of the HG. The GC-content and GC3 of the HPI were uniform. In spite of variations in number and gene size the up_HPI_down regions, tended to have average GC-contents and GC3 frequencies similar to those of the HG and HPI.

The GC3 frequencies of the HG varied in classes from 0.11_0.20 to 0.91_1.00 and that of the up_HPI_down regions from 0.11_0.20 to 0.71_0.80. The Correspondence Analysis of both distances is shown in Figures 3A and 3B. The GC3 frequency of the HG (Figure 3A) shows that the proportion of GC3-rich genes among *E. coli*, *Y. pestis* and *Y. pseudotuberculosis* is similar, unlike those of *Y. enterocolitica*, which has a smaller number of GC3-rich genes. Nonetheless, all strains have a higher GC3 distribution in the classes of central frequencies. The only difference is that *E. coli* strains have more genes with GC3 in the higher classes of 0.71_0.80 and *Y. pseudotuberculosis* and *Y. pestis* in the classes of 0.41_0.50. The variation of the classes of GC3 frequency of the up_HPI_down regions (Figure 3B) indicates the occurrence of more genes in GC3 classes with frequencies above average, each strain presenting different proportions of this variable on both sides of their HPI. The two highest GC3-rich classes were found among the *ybt* genes. The distance among the strains in the graph of the Correspondence Analysis is due to the heterogeneity of the HPI neighboring regions (Figure 3B). The HPI-HG correlation of the amino acid frequencies was significant at 0.01.

The sequences alignment of up_HPI_down regions of the 12 strains analyzed revealed LCBs corresponding to the genes *ybt* and *int* only. The synteny among the preserved LCBs of the same species allowed a good view of the ORFs and homologous and non-homologous rearrangements in the up_HPI_down regions. The multiple sequences alignments of *E. coli* 536 with APEC01, CFT073 and UTI89 revealed 14 LCBs (data not shown) and the alignments of *Y. pestis* CO92 with 91001, Angola, Antiqua, KIM and Pestoides F strains revealed 42 LCBs (Figure 4A).

The up_HPI_down regions of the six *Y. pestis* strains showed more variations and higher number of inversions than the four *E. coli* strains. All *Y. pestis* KIM sequences were inverted in relation to CO92 strain. With the up_HPI region of CO92 as reference the *Y. pestis* LCB1 has homologous sequences coding for one hypothetical protein, UvrY and UvrC (Figure 4A), the LCB1 of *Y. pestis* 91001 corresponds to the genes *mutT*, *sdaA*, *hpaC* and *hpaB* without significant similarities with other strains over 5kb. The 91001 and Angola strains displayed a preserved LCB order compared to CO92, except the position of LCB4 of 91001, which is the LCB2 of CO92.

Y. pestis CO92 LCB2 starts with a sequence homologous to PgsA phosphatidyl glycerophosphate synthetase (YPO1867), however the subsequent sequences are not homologous. LCB3 is represented by a putative transposase TnpA. LCB4 by a series of non-homologous, hypothetical proteins and LCB5 contains sequences homologous to UgpB, which participates in carbohydrate formation and is neighbor to other series of non-homologous sequences. LCB6 harbors homologous *IS100* (YPO1901) and LCB7 and LCB8 codes for hypothetical proteins homologous to those of *Y. pestis* and *Y. pseudotuberculosis* strains. In *Y. pestis* Angola (YpAngola_A2092c) however the LCB8 sequence is *yp42* and is 30% identical to a structural protein of the white spot bacilliform virus (wsv242). In LCB9-19 the *ybt* genes are in synteny with *int* (LCB20) and tRNA_{asn} (LCB21). In the region HPI_down the LCBs22-25 are represented by ORFs coding for fimbrial protein and putative two-component response regulator homologous to that of *Y. pseudotuberculosis*.

The LCB26 (Figure 4A, colorless area) contains a pseudogene for histidine kinase, a virulence sensor protein and LCB27-30 harbor sequences for a two-component response regulator, 4-hydroxybutyrate coenzyme A transferase, transcriptional regulator and citrate lyase beta chain. LCB31 contains a sequence homologous to *luxR* (quorum-sensing-regulated). LCB32-35 contains sequences coding for a lipoprotein and sugar transport and LCB36 codes for a LysR-family of transcriptional regulators. One ORF of the nitrogen fixation family (LCB37) is in synteny, followed by three additional sequences for transcriptional regulators (LCB38-40). The last two *Y. pestis* strains LCBs contain sequences for a glutaminase and one membrane protein (MFS superfamily Multidrug-efflux transporter). From LCB27 to LCB40 the *Y. pestis* strains analyzed share full homology.

The alignments of the up_HPI_down regions of *Y. pestis* CO92 with 91001, Angola, Antiqua, KIM, Pestoides F and *Y. pseudotuberculosis* revealed 16 LCBs. The three LCBs of the up_HPI region correspond to the homologous sequences of an *IS100*, one hypothetical protein and *yp42* of the Angola strain, respectively. After the 12 LCBs for *ybt* and *int* genes, the LCB16, HPI_down region, of *Y. pestis* CO92 represents a sequence coding for a putative exported protein (YPO1919), and an alpha-galactosidase (YPTB1608). In *Y. pseudotuberculosis*, no significant similarity was observed in the alignment of both sequences (BLAST (bl2seq)/NCBI).

The LCB sequences alignment among the *Y. pseudotuberculosis* and *Y. enterocolitica* strains showed homology among four sets of ORFs coding for units of the TRAP-T family C4-dicarboxylate transporter: large DctM subunit (YPTB1604 and YE2646); small DctQ subunit (YPTB1605 and YE2645); DctP subunit (YPTB1606 and YE2644) and one sequence for phosphopantetheinyl transferase (YPTB1607 and YE2642). These units of carrier proteins are located in the regions HPI_down of *Y. pseudotuberculosis* and up_HPI of *Y. enterocolitica* (Figure 4B).

In the up_HPI_down region of *Y. pseudotuberculosis*, there are sequences for histidine kinase, cold shock protein and a hypothetical protein (YPTB1628), which has a GGDEF/PAS domain. In *Y. enterocolitica*, sequences for putative fimbrial chaperone (YE2602) and *citA* sensor kinase protein (YE2654) were found plus sequences for products similar to those of the *Y. pestis* strains evaluated. In total, the *Y. pseudotuberculosis* and *Y. enterocolitica* strains displayed 32 LCBs. However, there was high similarity only among the TRAP-T units, the *ybt* locus and the pseudogene *int*. Synteny and homology in them were probably affected by the presence of four *IS* and transposases after *fyuA* of *Y. enterocolitica*.

DISCUSSION

The topology of the concatenated sequences of the *ybt* genes (Figure 1) reconstructed reliably and with high bootstrap values the monophyly of the *Yersinia* suggesting vertical

acquisition of the HPI. The clade of *Y. enterocolitica* appears as the sister group of *Y. pseudotuberculosis* and *Y. pestis*. The tree topology (Figure 1) also places *Y. pestis* Pestoides F as the oldest representative of the species as previously shown (Achtman et al. 2004; Thomson et al. 2006).

The variation of the ratio of synonymous and non-synonymous substitutions (Table 1) between the HPI of *Y. enterocolitica* and other species of the genus allowed a better understanding of the evolution of the *Yersinia* *ybt* genes. On the other hand, since only one genome of *Y. enterocolitica* was available, the lack of previous information on GC3 between and within host genomes and saturation of conserved sites in the alignments may have hampered the understanding of the HPI monophyly. Our approach to assessing the number of GC-rich genes using a simple statistical method showed more clearly the diversity of the HG studied and revealed new evolutionary facts.

High sequence similarities usually indicate a common evolutionary history, but do not necessarily correspond to homology between species or to a common ancestry. The selective signatures indicate that the transmission of the *Y. enterocolitica* HPI to other species is unlikely. Moreover, the functional loss of the integrase gene, YE2611, (pseudogene) impaired its ability as HPI donor. Activity of an integrase adjacent to the locus *ybt* is required for the HPI donation process by recombination (Lesica and Carniel 2005).

In molecular terms the selective signatures reflect Darwinian selection and are based on powerful statistical methods. However, although the theory of evolution by natural selection is accepted, the models for molecular evolution have been under debate (Yang and Bielawski, 2000). In statistical terms, *Error* can represent the measurement uncertainty, but does not necessarily imply that it is wrong. In this case, two errors are considered: one affects the precision (random) and the other the accuracy (systematic).

The sequences of amino acids and especially the sequence quality can produce inconsistent clades. The difficulties of accepting the Nei-Gojobori (1986) and ML methods are that they do not consider the transition/transversion bias and the bias of codon usage in protein analysis. Usually the results of both methods are similar, based on similar assumptions (Li et al., 1985). The heuristic tree reconstruction by ML has the advantage of not depending on ancestral sequences and can easily incorporate characteristics of the evolution of mRNA sequence, with both the transition/transversion bias and the bias of codon usage.

The phylogenetic reconstruction (Figure 1) based on the MP and ML methods and their reliability values showed a low homoplasy index, indicating that the tree is formed by clades resulting from homologous *ybt* sequences. This is corroborated by the high index level of rescheduled consistency and indicates that the *ybt* genes are synapomorphic characters shared by the *Yersinia* during evolution.

The differences on the concept of the HPI independent evolution and our findings are because studies to uncover the evolution of mRNA sequences based on qualitative and quantitative methods are still rare.

The positive correlation observed between HG-HPI indicates similarity between the groupings derived from the frequencies of amino acids of HPI and HG. By the amino acids standards use of the HG or HPI, *Yersinia* species share a common ancestor-descendant history. This is also reflected on the frequency of codon usage (Figure 2). The details of the synteny of the neighboring regions of *ybt* genes of *Y. enterocolitica* and *Y. pseudotuberculosis* are shown in Figure 4B.

The presence of sequences similar to *Y. pestis* PMT plasmid in *Photobacterium* may indicate that insecticide genes loss causes insect pathogenicity loss (Wilkinson et al., 2009). Both *Y. enterocolitica* and *P. luminescens* have a common evolutionary history with insect

pathogenicity. Furthermore, *P. luminescens* also produces a siderophore, the photobactin (Heermann and Fuchs, 2008).

The presence of the HPI only in the pathogenic Yersinia, suggests its independent acquisition by *Y. enterocolitica* and *Y. pseudotuberculosis* (Carniel, 2008). Our results may admit the loss of pathogenicity genes by non-pathogenic species of Yersinia. The HPI donor could be an ancestor of *P. luminescens* and *Y. enterocolitica*. This ancestor, sharing common environmental characteristics, would be older than *P. luminescens*, which has been appointed as *Y. enterocolitica* ancestor (Isabel et al. 2008; Heermann and Fuchs, 2008).

The evaluation of complete genomes allowed observing the distortions in the GC contents of *ybt* genes (GC-rich), their neighbors (GC-poor) and the genomes homogeneity (Table 2 and Figure 3). These data explain the maintenance of protein functions in the HPI neighborhood regions, without similarities of the corresponding nucleotides. A possible evaluation based on the frequency classes of GC3 of *ybt* and *int* genes of the four species would place all them close to a single central point in the graph of the Correspondence Analysis (Figure 3) due to their high GC3 similarities, reflecting the values described for HPI.

The diversity of *Y. enterocolitica* biotypes and its higher number of GIs (Thomson et al., 2006) may be related to the ecological conditions and selection pressures observed by the different GC3 patterns (Figure 3A) and their respective up_HPI_down regions (Figure 3B). The genetic fitness of *Y. enterocolitica* and its GIs allowed faster changes in codon usage.

Similarly as the LCBs revealed intraspecific variation among *Y. pestis* strains, the different HG-GC3 frequency classes indicate that the strains Pestoides F and Angola, which have a larger number of GC-poor genes, are less related than the other *Y. pestis* strains. Concerning *Y. enterocolitica*, its distance with the other Yersinia can be because the number of genes with GC3 frequencies above the mean (0.61-0.70 to 0.91-1.0) is 883 units while in the other genomes the average number of genes, in the same frequency classes is 1649 units. This significantly lower number of GC3-rich genes must have caused a higher selective pressure on the *Y. enterocolitica* genome.

The up_HPI_down regions of *Y. pestis*, *Y. pseudotuberculosis*, *Y. enterocolitica*, and *E. coli* strains have low interspecific similarity sequences coding for several products of yet unknown functions. The data in the functional annotations in NCBI, however, revealed functional patterns of the molecules encoded from these sequences. The functional similarities between the products of the up_HPI_down regions of HG (e.g. genes of the nitrogen fixation group, histidine kinase, polysaccharide and multidrug efflux system) suggests the probable existence of a regulatory network involving Ybt and quorum-sensing systems.

The *Y. pestis* sequences located in up_HPI_down region are homologous to *pgsA* and *ugpB* (Figure 4A) involved in capsular polysaccharide biosynthesis (Ashbaugh et al., 1998; O'May et al., 2009). On the other hand, the sequence *luxR* and other homologous encoding histidine kinase located in up_HPI_down regions of *Y. pseudotuberculosis*, are classically known as quorum-sensing-regulated.

Some gene products observed in *Y. pestis* and *E. coli* up_HPI_down regions may also be related to sequences of *Y. pseudotuberculosis* and *Y. enterocolitica* corresponding regions, e.g., the coding sequences of histidine kinase (YPTB1612-13) and cold shock proteins (YPTB1624) in *Y. pseudotuberculosis* and putative fimbrial chaperone (YE2602) and putative sugar kinase (YE2605) in *Y. enterocolitica*. These four proteins have homologous sequences considered as members of the quorum-sensing system (Munavar et al., 2005, Jung et al., 2007), although without similarities to form LCBs. On the other hand, the genes coding for the TRAP-T family (transporter of C4-dicarboxylate) and 4'-phosphopantetheinyl transferase formed LCBs in the alignment of the up_HPI_down regions of the two species and correspond to portions of the chromosomes and not of HPI.

Sequences for DctM, DctQ, DctP and 4'-phosphopantetheinyl transferase were detected in neighboring chromosomal segments outside the *Y. enterocolitica* and *Y. pseudotuberculosis* HPI, suggesting that their common ancestor had this characteristic. Another striking characteristic of these sequences was that even when subjected to a higher number of chromosomal rearrangements, their *ybt* system continue nearly identical and are involved in sequences for *quorum-sensing* system, due to the high number of *IS* and transposons.

Concluding, contrary to previous understanding our results suggests that the three pathogenic *Yersiniae* had a single HPI donor. The clades based on concatenated *ybt* indicate that these nucleotide sets represent a strong molecular signature of the *Yersiniae* phylogeny, since they were transferred vertically and reflect the genus phylogeny. The negative selection signature of the *Y. enterocolitica* 8081 *ybt* and its lower number of GC3-rich genes in comparison with the other host genomes indicate that this strain maintained its HPI and survived the process of genome homogenization under a higher selection pressure. Based on the different selective values of the *ybt* genes, it may be suggested that the same gene in different species may vary along the evolutionary history. The codon usage and GC3 evaluation models used in this study are readily applicable and the data matrices can be evaluated by other statistical software. The first model revealed hierarchical clustering of genomes along the lines of the classical design model, which supports only a limited number of sequences. The model GC3 on the other hand showed details of the distance dynamics among the genomes and also of groups of HPI-neighboring genes.

ACKNOWLEDGEMENTS: The authors wish to thank the following Brazilian institutions: Universidade Federal Rural de Pernambuco; Universidade Federal de Pernambuco; Fundação Oswaldo Cruz/Centro de Pesquisas Aggeu Magalhães; Fundação de Amparo a Ciência e Tecnologia do Estado de Pernambuco and Conselho Nacional de Desenvolvimento Científico e Tecnológico.

REFERENCES

- Achtman M, Morelli G, Zhu P, Wirth T, Diehl I, Kusecek B, Vogler AJ, Wagner DM, Allender CJ, Easterday WR, Chenal-Francisque V, Worsham P, Thomson NR, Parkhill J, Lindler LE, Carniel E, Keim P (2004). Microevolution and history of the plague bacillus, *Yersinia pestis*. *Proc Natl Acad Sci USA*. 101:17837-42.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol*. 5:403-10.
- Ashbaugh CD, Albertí S, Wessels MR (1998). Molecular analysis of the capsule gene region of group A *Streptococcus*: the *hasAB* genes are sufficient for capsule expression. *J Bacteriol*. 180:4955-9.
- Bach S, de Almeida AMP, Carniel E (2000). The *Yersinia* high-pathogenicity island is present in different members of the family Enterobacteriaceae. *FEMS Microbiol Lett*. 183:289-94.
- Buchrieser C, Rusniok C, Frangeul L, Couve E, Billault A, Kunst F, Carniel E, Glaser P (1999). The 102-kilobase *pgm* locus of *Yersinia pestis*: sequence analysis and comparison of selected regions among different *Yersinia pestis* and *Yersinia pseudotuberculosis* strains. *Infect Immun*. 67:4851-61.
- Carniel E (2008). Pathogenic *Yersinia*: Stepwise Gain of Virulence due to Sequential Acquisition of Mobile Genetic Elements. *In* Horizontal Gene Transfer in the Evolution of Pathogenesis (eds. H Hensel and M Schmidt). Cambridge University Press, Cambridge, 194-216.

- Collyn F, Guy L, Marceau M, Simonet M, Roten CA (2006). Describing ancient horizontal gene transfers at the nucleotide and gene levels by comparative pathogenicity island genomics. *Bioinformatics*. 22:1072-79.
- Darling AC, Mau B, Blattner FR, Perna NT (2004). Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*. 14:1394-1403.
- Dobrindt U, Hochhut B, Hentschel U, Hacker J (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol*. 2:414-24.
- Felsenstein J (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 39: 783-791.
- Fetherston JD, Bertolino VJ, Perry RD (1999). YbtP and YbtQ: two ABC transporters required for iron uptake in *Yersinia pestis*. *Mol Microbiol*. 32:289-99.
- Geoffroy VA, Fetherston JD, Perry, RD (2000). *Yersinia pestis* YbtU and YbtT are involved in synthesis of the siderophore yersiniabactin but have different effects on regulation. *Infect. Immun*. 68:4452-4461.
- Grantham R, Gautier C, Gouy M, Mercier R, Pavé A (1980). Codon catalog usage and the genome hypothesis. *Nucleic Acids Res*. 8:r49-r62.
- Hacker J, Kaper JB (2000). Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol*. 54:641-79.
- Heermann R, Fuchs TM (2008). Comparative analysis of the *Photothabdus luminescens* and the *Yersinia enterocolitica* genomes: uncovering candidate genes involved in insect pathogenicity. *BMC Genomics*. 9: 40.
- Isabel S, Leblanc E, Boissinot M, Boudreau DK, Grondin M, Picard FJ, Martel EA, Parham NJ, Chain PS, Bader DE, Mulvey MR, Bryden L, Roy PH, Ouellette M, Bergeron MG (2008). Divergence among genes encoding the elongation factor Tu of *Yersinia* Species. *J Bacteriol*. 190:7548-58.
- Jung K, Odenbach T, Timmen M (2007). The quorum-sensing hybrid histidine kinase LuxN of *Vibrio harveyi* contains a periplasmically located N terminus. *J Bacteriol*. 189:2945-2948.
- Lesic B, Carniel E (2005). Horizontal Transfer of the High-Pathogenicity Island of *Yersinia pseudotuberculosis*. *J Bacteriol*. 187: 3352-3358.
- Li WH (1987). Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. *J Mol Evol*. 24:337-45.
- Li WH, Wu CI, Luo CC (1985). A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol Biol Evol*. 2:150-74.
- Merhej V, Adékambi T, Pagnier I, Raoult D, Drancourt M (2008). *Yersinia massiliensis* sp. nov., isolated from fresh water. *Int J Syst Evol Microbiol*. 58:779-84.
- Munavar H, Zhou Y, Gottesman S (2005). Analysis of the *Escherichia coli* Alp phenotype: heat shock induction in *ssrA* mutants. *J Bacteriol*. 187:4739-51.
- Nei M, Gojobori T (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol*. 3:418-26.
- Oda Y, Larimer FW, Chain PS, Malfatti S, Shin MV, Vergez LM, Hauser L, Land ML, Braatsch S, Beatty JT, Pelletier DA, Schaefer AL, Harwood CS (2008). Multiple genome sequences reveal adaptations of a phototrophic bacterium to sediment microenvironments. *Proc Natl Acad Sci USA*. 105: 18543-18548.
- O'May GA, Jacobsen SM, Longwell M, Stoodley P, Mobley HL, Shirtliff ME (2009). The high-affinity phosphate transporter Pst in *Proteus mirabilis* HI4320 and its importance in biofilm formation. *Microbiology*. 155:1523-35.

- Page RD (2002). Visualizing phylogenetic trees using TreeView. *Curr Protoc Bioinformatics*. Chapter 6:Unit 6.2.
- Parkhill J, Wren BW, Thomson NR, Titball RW, Holden MT, Prentice MB, Sebaihia M, James KD, Churcher C, Mungall KL, Baker S, Basham D, Bentley SD, Brooks K, Cerdeño-Tárraga AM, Chillingworth T, Cronin A, Davies RM, Davis P, Dougan G, Feltwell T, Hamlin N, Holroyd S, Jagels K, Karlyshev AV, Leather S, Moule S, Oyston PC, Quail M, Rutherford K, Simmonds M, Skelton J, Stevens K, Whitehead S, Barrell BG (2001). Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature*. 413:523-27.
- Perry RD, Fetherston JD (1997). *Yersinia pestis* - etiologic agent of plague. *Clin Microbiol Rev*. 10:35-66.
- Posada D, Buckley TR (2004). Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. *Syst Biol*. 53:793-808.
- Sprague LD, Scholz HC, Amann S, Busse HJ, Neubauer H (2008). *Yersinia similis* sp. nov. *Int J Syst Evol Microbiol*. 58:952-8.
- Sulakvelidze A (2000). *Yersiniae* other than *Y. enterocolitica*, *Y. pseudotuberculosis*, and *Y. pestis*: the ignored species. *Microbes Infect*. 2:497-513.
- Supek F, Vlahovicek K (2004). INCA: synonymous codon usage analysis and clustering by means of self-organizing map. *Bioinformatics*. 20:2329-30.
- Swofford, DL (2002). PAUP*: Phylogenetic Analysis Using Parsimony (* and Other Methods) (Sinauer, Sunderland, MA/USA), Version 4.0b10.
- Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol*. 24:1596-99.
- Tatusova TA, Madden TL (1999). BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett*. 15:247-50.
- Thomson NR, Howard S, Wren BW, Holden MT, Crossman L, Challis GL, Churcher C, Mungall K, Brooks K, Chillingworth T, Feltwell T, Abdellah Z, Hauser H, Jagels K, Maddison M, Moule S, Sanders M, Whitehead S, Quail MA, Dougan G, Parkhill J, Prentice MB (2006). The complete genome sequence and comparative genome analysis of the high pathogenicity *Yersinia enterocolitica* strain 8081. *PLoS Genet*. 2:2039-51.
- Wilkinson P, Waterfield NR, Crossman L, Corton C, Sanchez-Contreras M, Vlisidou I, Barron A, Bignell A, Clark L, Ormond D, Mayho M, Bason N, Smith F, Simmonds M, Churcher C, Harris D, Thompson NR, Quail M, Parkhill J, Ffrench-Constant RH (2009). Comparative genomics of the emerging human pathogen *Phototribdus asymbiotica* with the insect pathogen *Phototribdus luminescens*. *BMC Genomics*. 10:302.
- Yang Z, Bielawski JP (2000). Statistical methods for detecting molecular adaptation. *Trends Ecol Evol*. 15:496-503.

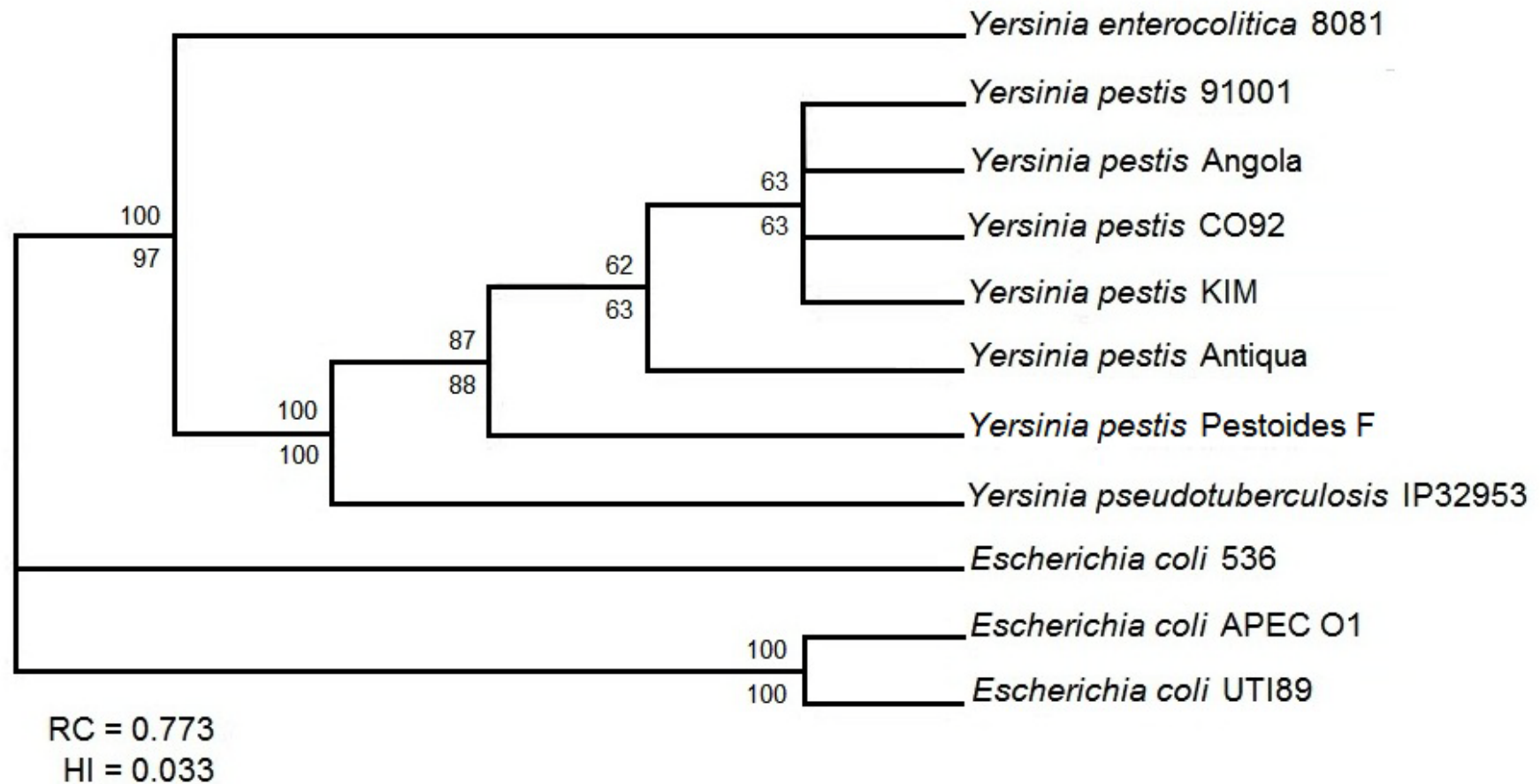


Figure 1 - Phylogenetic reconstruction based on 11 concatenated *ybt* gene sequences of *Yersinia pestis*, *Yersinia pseudotuberculosis*, *Yersinia enterocolitica* and *Escherichia coli* strains using PAUP* version 4.0b10 (Swofford, 2002) by exhaustive search (bootstrap test of 10.000 pseudo-replications). The heuristic Maximum Parsimony (MP) and Maximum Likelihood (ML) criteria in phylogenetic tree reconstruction were explored. The per-character consistency homoplasy index (HI) and rescaled consistency index (RC) were used to evaluate the phylogram.

*****HIERARCHICAL CLUSTER ANALYSIS*****

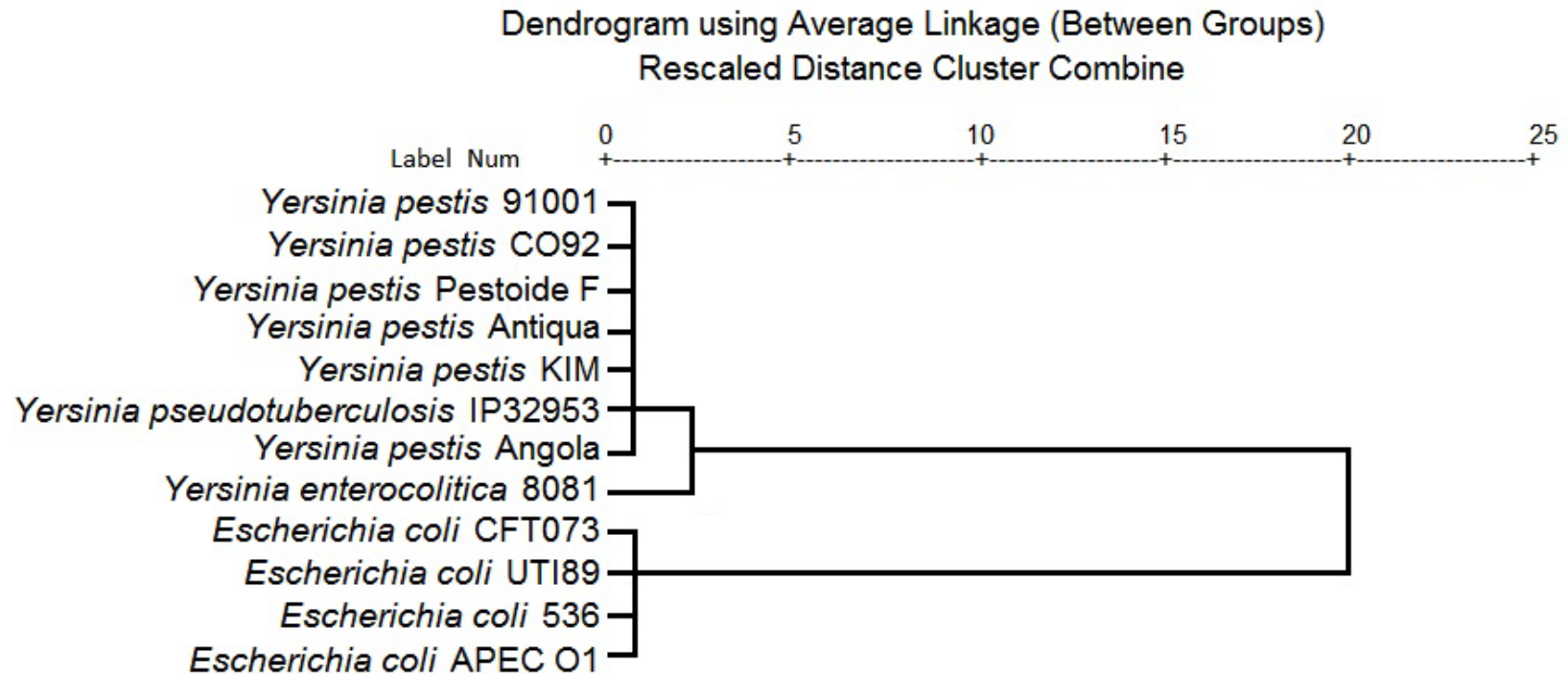


Figure 2 - Phylogenetic reconstruction using the high-pathogenicity island of the complete host genomes of the strains: *Yersinia pesti*, *Yersinia pseudotuberculosis*, *Yersinia enterocolitica* and *Escherichia coli*, based on codon usage frequency (classification|hierarchical cluster analysis|Euclidian distance), INCA2.1 and SPSS.13 softwares.

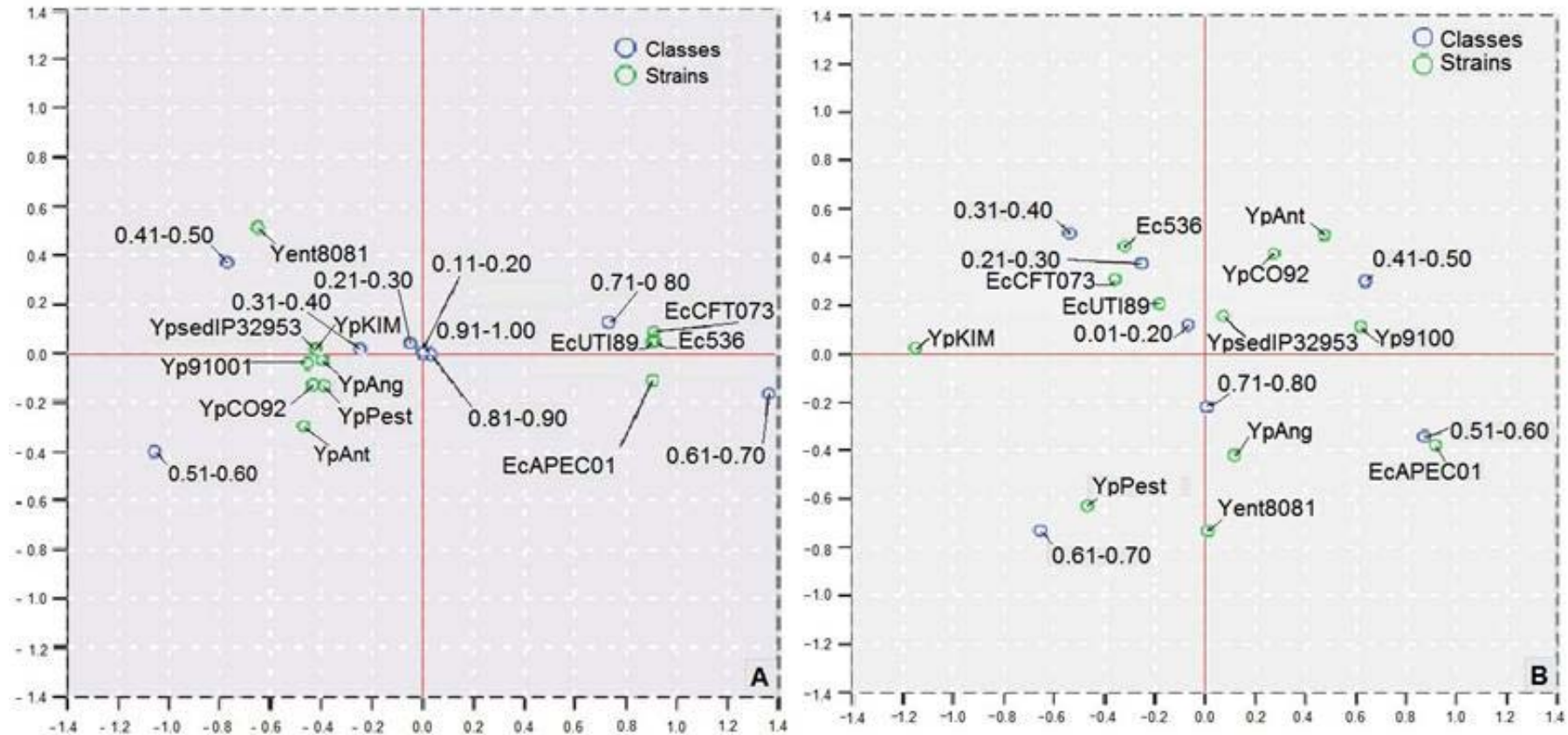


Figure 3 - Estimated distance between genomes the strains of *Yersinia pestis* (91001, Angola, Antiqua, CO92, KIM and Pestoides F) or “Yp”, *Yersinia pseudotuberculosis* IP32953 or “Ypseud”, *Yersinia enterocolitica* 8081 or “Yent” and *Escherichia coli* (536, APEC01, CFT073 and UTI89) or “Ec” as related to the GC3s frequency classes. A) containing all mRNAs; B) containing the mRNAs of the high-pathogenicity island (HPI) and those located 30kb upstream and downstream the HPI, obtained by Correspondence Analysis, SPSS.13 software (Euclidian distance). The proximity of points on the x-axis indicates that classes of GC3 frequency of the strains are more homogeneous. The strains in x-axis inferior right quadrant tend to exhibit higher contrast to that in the y-axis superior left quadrant.

Table 1 - Estimate of the ratio (R) between the non-synonymous (dN) and synonymous (dS) substitutions per site of the loci *ybt* and *int* found in the strains of *Escherichia coli* (536, APEC01, CFT073 and UTI89), *Yersinia pestis* (91001, Angola, Antiqua, CO92, KIM and Pestoides F), *Yersinia pseudotuberculosis* IP32953 and *Yersinia enterocolitica* 8081, using the *neighbor-joining* method and modified Nei-Gojobori (p-distance), model MEGA4 Software.

Groups	All <i>Yersinia</i>	<i>Y. pestis</i> and <i>Y. pseudotuberculosis</i>	<i>Y. pestis</i>	<i>Y. pestis</i> and <i>E. coli</i>	<i>E. coli</i>	All strains
Genes	$R = dN/dS$					
<i>fyuA</i>	0.167	0.000	0.000	0.500	0.250	0.167
<i>ybtE</i>	0.000	0.000	0.000	0.500	1.500	0.375
<i>ybtT</i>	0.222	0.000	0.000	0.500	0.000	0.429
<i>ybtU</i>	0.273	0.000	0.000	1.000	1.500	0.375
<i>irp1</i>	0.200	0.000	0.000	0.286	0.250	0.250
<i>irp2</i>	0.429	0.000	0.000	1.200	1.375	0.778
<i>ybtA</i>	0.077	0.000	0.000	0.286	0.143	0.133
<i>ybtP</i>	0.077	0.000	0.000	0.000	1.000	0.083
<i>ybtQ</i>	0.222	0.000	0.000	0.222	0.500	0.222
<i>ybtX</i>	0.222	0.000	0.000	0.200	0.167	0.200
<i>ybtS</i>	0.100	0.000	0.000	0.333	0.333	0.111
All <i>ybt</i>	0.250	0.000	0.000	0.600	0.571	0.400
<i>int</i>	0.286	1.000	0.000	1.000	1.200	0.714

Table 2 - Numbers of mRNAs (genes) of the high-pathogenicity island (HPI), of the regions 30kb upstream and downstream the HPI (up_HPI_down) and of the host genomes (HG), sizes in kb, GC-contents and GC3 frequencies in strains of *Escherichia coli*, *Yersinia pestis*, *Yersinia pseudotuberculosis* IP32953 and *Yersinia enterocolitica* 8081 (INCA2.1 software).

Strains	mRNAs			kb			GC%			GC3%		
	HPI	up_HPI_down	HG	HPI	up_HPI_down	HG	HPI	up_HPI_down	HG	HPI	up_HPI_down	HG
<i>Escherichia coli</i> 536	12	68	4685	29	77	4378	59	50	52	75	60	62
<i>Escherichia coli</i> APEC01	12	77	4467	29	78	4445	59	53	52	75	61	62
<i>Escherichia coli</i> CFT073	12	66	5379	29	77	4743	59	50	51	75	59	62
<i>Escherichia coli</i> UTI89	12	70	5066	29	82	4616	59	53	52	75	65	62
<i>Yersinia enterocolitica</i>	12	62	4046	29	76	3900	59	51	48	75	62	55
<i>Yersinia pestis</i> 91001	12	63	3895	29	77	3748	59	53	49	75	63	56
<i>Yersinia pestis</i> Angola	12	69	3837	29	66	3461	59	53	49	75	64	56
<i>Yersinia pestis</i> Antiqua	12	84	4167	29	77	3954	59	52	49	75	62	56
<i>Yersinia pestis</i> CO92	12	74	3942	29	71	3795	59	53	49	75	63	56
<i>Yersinia pestis</i> KIM	12	72	4090	29	75	3835	59	52	49	75	63	56
<i>Yersinia pestis</i> Pestoides F	12	77	3850	29	77	3707	59	52	49	75	63	56
<i>Yersinia pseudotuberculosis</i>	12	62	3972	29	75	3966	59	52	49	75	64	57

6. Manuscrito de Artigo Científico

(Será submetido para a revista ***Genetics***)

Artigo 2 - Ancestrais da ilha de alta patogenicidade de *Yersinia pestis*

Título: Ancestrais da ilha de alta patogenicidade de *Yersinia pestis*

Autores: Nara S.A. Freitas*, Tiago A.E. Ferreira[✧], Manoel R. da Sena-Júnior^Ψ, César R.L. Costa-junior^Ω, Tereza C. Leal-Balbino[§], Valdir Q. Balbino^Ω, Alzira M.P. de Almeida[§]

Afilições: *Dep. Biologia/UFRPE - PE, Brasil; [✧]Dep. Estatística e Informática/UFRPE - PE, Brasil; ^ΨDep. Estatística /UFPE - PE, Brasil; ^ΩDep. Genética/UFPE - PE, Brasil; [§]Dep. Microbiologia/CPqAM/FIOCRUZ – PE.

Organização do cabeçalho: Seleção do uso de códon, elemento genético, HPI

Palavras chave: filogenômica, *Yersinia pestis*, conteúdo de GC, genômica comparativa

Autor correspondente: Valdir de Queiroz Balbino

Endereço: Departamento de Genética, Universidade Federal de Pernambuco, Av.

Prof. Moraes Rego S/N Cidade Universitária, 50732-970 Recife, PE – Brasil

Fone/Fax: (5581) 21268512

e-mail: vqbalbino@pq.cnpq.br

RESUMO

Os genomas de sete cepas de *Yersinia pestis* e de 11 genes do operon *ybt* foram avaliados quanto as suas relações ancestrais-descendentes. O operon *ybt* encontra-se na ilha de alta patogenicidade (HPI) das espécies de Yersiniaceae patogênicas e outras bactérias. As análises envolveram os conteúdos de GC de 65.567 mRNAs, os rearranjos cromossômicos dos genomas e as correlações entre as substituições sinônimas e não sinônimas (dN/dS) dos genes *ybt*. Os dados indicam que a inserção da HPI ocorreu de forma monofilética entre as yersínias podendo ter sido herdada das Vibrionaceae. A falta de genomas intermediários diminui o sinal evolutivo das HPIs, uma vez que alguns ancestrais de *Y. pestis* passaram por novos ciclos adaptativos após as divergências e alguns perderam a HPI. As correlações entre dN/dS e uso de GC1, 2 e 3 nos códons permitiram observar as relações homólogas dos genomas ancestrais, quando as similaridades dos alinhamentos deixavam dúvidas. O elevado número de rearranjos cromossômicos entre as cepas de *Y. pestis* revelou também que em algumas delas os genes *ybt* estão organizados de forma mais semelhante ao do seu ancestral direto, a *Y. pseudotuberculosis*. As dificuldades em caracterizar as relações ancestrais-descendentes da HPI decorrem da falta de considerações simultâneas de diferentes processos da evolução bacteriana, que parecem ser maiores porque eventos de transferência horizontal de genes são inferidos mais pela similaridade e menos pela homologia.

A *Yersinia pestis*, agente causador da peste, é uma bactéria Gram-negativa da família Enterobacteriaceae, sub-classe γ -Proteobacteria do filo Proteobacteria. A peste é uma doença primária dos roedores, transmitida geralmente por pulgas, que pode afetar o homem e outros mamíferos. Outras duas espécies do gênero, os enteropatógenos *Y. pseudotuberculosis* e *Y. enterocolitica*, são veiculados principalmente pela água e alimentos contaminados e são responsáveis por um quadro clínico denominado Yersiniose (PERRY e FETHERSTON, 1997) e onze espécies são ambientais, geralmente não patogênicas aos humanos (PERRY e FETHERSTON 1997; SULAKVELIDZE 2000; SPRAGUE *et al.* 2008; MERHEJ *et al.* 2008).

A presença de uma ilha de alta patogenicidade (High Pathogenicity Island) ou HPI é determinante para a patogenicidade das yersínias. A HPI contém o operon *ybt* formado por onze genes envolvidos na biossíntese, transporte e regulação do sideróforo Yersiniabactina (Ybt) e pode assim ser considerada como uma ilha de absorção de ferro (CARNIEL 2008). As sequências que codificam a Ybt estão presentes em vários outros membros da família Enterobacteriaceae (DE ALMEIDA *et al.* 1993; BACH *et al.* 2000; BUCHRIESER *et al.* 1998). A similaridade das sequências *ybt* de *Y. pestis* e de *Escherichia coli* levou a hipótese da transferência horizontal recente de genes da HPI de uma cepa patogênica de *E. coli* para *Y. pestis* (SCHUBERT *et al.* 1998; CARNIEL, 2008).

A aquisição das HPIs de *Y. enterocolitica* e *Y. pseudotuberculosis* é considerada como consequência de dois processos evolutivos de doações independentes, ocorridos muito após a divergência de *Y. enterocolitica* e *Y. pseudotuberculosis*, ancestral da *Y. pestis* (ACHTMAN *et al.* 2004; THOMSON *et al.* 2006; CARNIEL 2008). Outra possibilidade seria a HPI ter sido herdada

monofileticamente pelas yersínias patogênicas e a HPI de *E. coli* ser seu grupo irmão. Essas argumentações seriam sustentadas por diferentes tipos de análises (e.g. filogenética, filogenômica, assinaturas seletivas, conteúdos de GC3) e por *Y. enterocolitica* e *Y. pseudotuberculosis* possuírem sequências *cdtMPQ* (para os transportadores da família TRAP-T C4- dicarboxilato) próximas às HPIs, sugerindo a ancestralidade da inserção das ilhas. Apesar das novas evidências sobre a monofilia da HPI, a avaliação dos *ybt* colineares em sintenia limitou-se a observação do número de homólogos, a diversidade dos padrões de rearranjos das HPIs e das características genéticas e ecológicas de outros genomas hospedeiros.

O crescente número de novas sequências de *Y. pestis* permite a criação de modelos de avaliação sobre os genomas hospedeiros da HPI. A quantificação dos genes ricos em GC pode revelar a pressão da seleção em cada G/C dos genomas e os perfis filogenômicos dos hospedeiros. Por outro lado, as análises das sequências dos genes 16S e *tuf*, que codifica o fator de alongamento Tu (EF-Tu), têm sido exploradas nos estudos da história da monofília das Enterobacteriaceae e na identificação de novas espécies do gênero *Yersinia* (ISABEL *et al.* 2008; PARADIS *et al.* 2005; SULAKVELIDZE 2000; MERHEJ *et al.* 2008).

Outros grupos de genes que curiosamente foram observados próximos ao operon *ybt* são homólogos dos reguladores de *quorum-sensing* (*luxR*) extensamente distribuídos entre as protobactérias. Os genes associados à esse fenômeno são altamente similares e encontrados em regiões cromossômicas diferentes de grupos ecologicamente distintos, o que levantou a hipótese de que o segmento de DNA que carrega esses genes tenha sido transferido horizontalmente várias vezes (URBANCZYK *et al.* 2008). Os rearranjos cromossômicos também podem explicar a presença da HPI em diferentes locais dos genomas.

Segundo KASAI (2007) e URBANCZYK *et al.* (2008), a Enterobacteriaceae *Photorhabdus luminescens*, um simbiótico de nematóide, pode ter recebido os genes *lux* de um membro da família Vibrionaceae. *P. luminescens* também adquiriu outra ilha de patogenicidade, a YAPI, (*Yersinia adhesion pathogenicity island*) presente em alguns sorotipos de *Y. pseudotuberculosis* e *Y. enterocolitica* Biotipo 1B (CICH *et al.* 2003; COLLYN *et al.* 2006; CARNIEL 2008). Durante o processo de especiação as yersínias adquiriram horizontalmente outros elementos genéticos indispensáveis para a virulência, os plasmídeos pYV/pCD, pPla/pPCP e pFra/pMT. Os dois últimos são específicos de *Y. pestis* e foram adquiridos após a sua divergência de *Y. pseudotuberculosis* (CORNELIS 2002; WORSHAM e ROY 2003; CARNIEL 2008).

No presente trabalho realizamos uma avaliação filogenética das sequências consideradas como de origem horizontal e vertical de *Y. pestis*, exploramos a história da evolução da HPI partindo do ganho ou perda de sequências *ybt* e da organização cromossômica e apresentamos um método filogenômico para os hospedeiros da HPI, baseado nas frequências de GC dos códons.

MATERIAIS E MÉTODOS

Seleção de genes e de genomas

As sequências dos onze *ybt* da cepa *Y. pestis* CO92 (NC_003143) disponíveis no NCBI (<http://www.ncbi.nlm.nih.gov>) foram localizadas por palavra-chave e usadas como *query* para a seleção dos genomas (GenBank FTP/NCBI). Posteriormente todos os arquivos dos diretórios correspondentes foram compilados em um banco de dados local. Em seguida, foram identificadas no Blast/NCBI as sequências homólogas *luxR*, vizinha ao cluster *ybt* (*fyuA-ybtE-ybtT-ybtU-irp1-irp2-ybtA-ybtP-ybtQ-ybtX-ybtS*), 16S, *tufA* e *cdtMPQ*. O mesmo ocorreu com a sequência

ybtD, que codifica uma fosfopanteteína transferase, requerida para a síntese de Ybt (BOBROV *et al.* 2002; FOLEY e BURKART 2009), mas localizada fora da HPI.

Reconstrução filogenética da HPI

As sequências homólogas foram alinhadas individualmente no software ClustalW v.1.6, implementado no programa MEGA v.4.1 (TAMURA *et al.* 2007), usado para avaliação prévia dos modelos de dendrogramas. Após o alinhamento, um modelo evolutivo foi calculado usando o critério de informação Akaike (AIC), através do modeltest v.3,7 (POSADA *et al.* 2004). Para as análises filogenéticas foi usado o programa PAUP* v. 4.0b10 (SWOFFORD 2002) e como critério na reconstrução dos dendrogramas a Máxima Parcimônia (MP) pelo método de busca *Branch and Bound* e Máxima Verossimilhança (ML) com busca heurística de 100 replicações. A coerência entre os caracteres foi verificada pelo Índice Homoplasia (IH) e o índice de Consistência de Reescalamento (RC). Para a análise de MP a consistência dos ramos foi avaliada pelo teste *bootstrap* de 10.000 pseudo-repetições (FELSENSTEIN 1985).

Determinação das assinaturas seletivas dos genes que codificam Ybt

A determinação da razão (*R*) entre taxas de substituição não sinônimas por sinônimas (dN/dS) das sequências *ybt* das HPIs foi avaliada par a par pelo método de Pamilo-Bianchi-Li (1993), usando o software MatLab (<http://www.mathworks.com>). Os valores foram exportados para arquivo *.txt. Posteriormente, os dados foram importados para planilhas Microsoft Office ExcelTM e os valores de cada genoma foram sequencialmente organizados em uma única

matriz, que foi submetida ao software SPSS (SPSS for Windows, Rel. 13, Chicago, IL, USA), para obtenção das correlações de similaridades entre os vetores e valores.

Análise filogenômica

Inicialmente foi preparado um *script* em ambiente MatLab para calcular as frequências de GC em cada posição do códon (GC1, GC2 e GC3), que compõe os 20 aminoácidos, mais as do códon de terminação dos mRNAs. Todas as sequências gênicas dos genomas foram separadas em 19 classes (0.00 $\rightarrow$ 0.05 até 0.85 $\rightarrow$ 0.90), a partir das frequências dos 64 códons (arquivos *.ffn/GenBank/NCBI), exceto as que não apresentaram número de base múltiplo de três. As frequências das classes de GC foram programadas para serem importadas para planilhas (Excel™) e sequencialmente organizadas em uma única matriz. Posteriormente, as classes que obtiveram valores iguais a zero em todos os genomas foram excluídas e os valores analisados pelo método hierárquico de ligação entre grupo e a métrica da distância Euclidiana (software SPSS).

Rearranjos cromossômicos

Para visualizar os diversos rearranjos cromossômicos em blocos colineares locais (LCBs) dos genomas das cepas de *Y. pestis* e das sequências homólogas selecionadas foi usado o software Mauve.2.3 (DARLING *et al.* 2004). Os genomas selecionados foram alinhados par a par com o da cepa *Y. pestis* CO92 (*.ffn/FTP/NCBI). Os rearranjos entre as sequências dos *ybt* e as homólogas para *luxR*, *16S*, *tufA*, *ybtD* e *cdtMPQ* foram organizados pela posição e número da sequência gênica depositada no NCBI. Posteriormente os conjuntos de sequências

de cada genoma foram copiados para arquivos *.txt e alinhados como os genomas completos.

Na busca pela similaridade de sequências de YAPI de *Y. pseudotuberculosis* nos genomas de *Y. pestis* a sequência dessa ilha foi localizada por palavra-chave e usada como modelos de *query* no Blast/NCBI. Em seguida, suas sequências (*.ffn/FTP/NCBI) foram avaliadas.

RESULTADOS

Foram analisados os genomas das seguintes espécies de yersínia: *Y. pestis* (CO92, KIM, Microtus ou 91001, Angola, Antiqua, Nepal516 e Pestoides F), *Y. pseudotuberculosis* (IP32953, IP31758, PB1/+ e YPIII), *Y. enterocolitica* (8081), *Y. aldovae* (ATCC 35236), *Y. bercovieri* (ATCC 43970), *Y. frederiksenii* (ATCC 33641), *Y. intermedia* (ATCC 29909), *Y. kristensenii* (ATCC 33638), *Y. mollaretii* (ATCC 43969), *Y. rohdei* (ATCC 43380) e *Y. ruckeri* (ATCC 29473).

As outras espécies avaliadas foram *Escherichia coli* (536, APEC01 e UTI89), *Citrobacter koseri* (ATCC BAA-895), *Photobacterium luminescens* (TTO1), *Serratia proteamaculans* (568) da família Enterobacteriaceae e *Photobacterium profundum* (SS9), da família Vibrionaceae. Os genomas que não continham HPI (*Y. pestis* Nepal516, *Y. pseudotuberculosis* IP31758 e YPIII, *S. proteamaculans* 568 e oito espécies de *Yersinia* não patogênicas) foram avaliados filogeneticamente pelas sequências dos genes 16S (*rimM*) e *tufA* (Figuras 1A e B). Os genomas contendo HPis ou *ybt* (Figuras 2A, B e C) foram também avaliados filogenomicamente (Figura 3).

Os agrupamentos filogenéticos gerados a partir das sequências homólogas dos genes *tufA*, *dctPQM*, *ybtD*, *luxR*, 16S (*rimM*) e genes *ybt* da HPI (Figura 2A), apontam para uma evolução monofilética do gênero *Yersinia* e da sua HPI. A avaliação do gene 16S (*rimM*) sugere que a *Y. ruckeri* pode ser um representante ancestral do gênero (Figuras 1A e B). A cepa de *P. profundum*, que possui dois cromossomos, apresentou os homólogos dos genes *dctPQM* e *ybtP* no cromossomo II. O gene *ybtQ* não é sintênico aos demais homólogos *ybt*.

Os filogramas construídos com onze, dez ou nove genes do operon *ybt* (sem *ybtP* e *ybtQ*) não evidenciaram de forma clara a monofilia das HPIs e revelaram uma maior variação da cepa de *Y. pestis* Pestoides F (Figuras 2B e C). Os clusters repetidos foram apenas os das cepas de *P. luminescens* e *P. profundum*, membros externos e separados por ramos mais longos, devido ao aumento do número de sítios dos alinhamentos e diminuição da sinonímia das bases.

A cepa de *C. koseri* apresentou-se como tipo intermediário entre as yersínias patogênicas e as de *E. coli* avaliadas (Figuras 2A, B e C). Todavia, os *ybt* da cepa de *C. koseri* não estão na mesma ordem que nesses patógenos. A cepa da espécie intermediária localizada mais próxima das yersínias foi a *S. proteamaculans* 568 que não possui a HPI (Figuras 1A e B).

As correlações entre dN/dS indicam relações positivas ou diretamente proporcionais entre os *ybt* (57-60% de GC) das espécies *Y. pestis*, *Y. pseudotuberculosis*, *Y. enterocolitica*, *E. coli* e *C. koseri*, quando comparados com os valores das correlações entre os pares dos *ybt* de *P. luminescens* e *P. profundum*. Ambas as espécies apresentaram correlações negativas ou inversamente proporcionais ($R < dN/dS$) dos seus conjuntos dos pares de

sequências. Os valores de R variaram individualmente entre os pares dos onze *ybt* (Tabela 1).

Entre as yersínias as correlações dos dN/dS foram maiores para *Y. enterocolitica* e *Y. pestis* Pestoides F. A cepa de *C. koseri* apresentou valores superiores (~0,1) em relação as de *E. coli*. Os mais altos valores de correlação entre os pares de sequências dos *ybt* foram detectados entre *Y. pestis* e *Y. pseudotuberculosis*, com uma conservação entre os sítios superior a 80%.

Os agrupamentos dos genomas (Figura 3) obtidos a partir das 18 classes de frequências de GC1, GC2 e GC3 (0.00 $\rightarrow$ 0.05_0.80 $\rightarrow$ 0.85) revelaram também a monofília das yersínias, como foi observado na análise com as sequências gênicas. No entanto, a maior variação em relação às frequências de GC das demais cepas deixou todas as yersínias em um único grupo. *P. luminescens* (Enterobacteriaceae) se comportou como um grupo mais distante das yersínias e *C. koseri* apareceu como um ancestral de *E. coli*, que por sua vez comportou-se como grupo irmão das yersínias. Porém a posição de *P. luminescens* no agrupamento foi junto aos dois cromossomos de *P. profundum* (Vibrionaceae).

Um total 65.567 genes foram avaliados, sendo o maior número de genes ricos em GC1 e GC3 encontrado nas classes de frequências nas cepas de *C. koseri* e *E. coli* (0.50 $\rightarrow$ 0.65), seguido das cepas de *Y. pestis*, *Y. pseudotuberculosis*, *Y. enterocolitica* e *P. luminescens* (0.45 $\rightarrow$ 0.55). Por outro lado, o genoma mais pobre em GC1 e GC3 foi o de *P. profundum*, tanto no cromossomo 1 (2.23794 Mbp) quanto no cromossomo 2 (4.1 Mbp). Esse genoma apresentou mais genes ricos em GC1 na classe de frequência entre 0.50 $\rightarrow$ 0.55 e em GC3 na classe 0.35 $\rightarrow$ 0.40.

Os conteúdos de GC2 variaram em 15 classes de frequências (0.00 $\rightarrow$ 0.05_0.70 $\rightarrow$ 0.75). Os genomas avaliados apresentaram uma grande parte dos

genes na classe 0.40 $\rightarrow$ 0.45, exceto para *P. luminescens* e *P. profundum* que tiveram a maioria dos genes na classe de frequências de 0.35 $\rightarrow$ 0.40. Porém de forma geral em todas as três posições do códon os conteúdos de GC tenderam a uma distribuição normal. Entre as cepas de yersínia o conteúdo de GC2 foi maior nas classes acima da média de *Y. enterocolitica* e menor nas suas classes de GC1 e GC3.

Os alinhamentos entre os sete genomas de *Y. pestis* revelaram 3015 LCBs. Os números de LCBs variaram entre cada par formado com a cepa *Y. pestis* CO92, revelando diversos rearranjos cromossômicos intraespecíficos (Figura 4), indicados pelas linhas transversais aos LCBs. Os genomas com maior número de regiões em sintenia foram os das cepas *Y. pestis* 91001, Angola e Antiqua, enquanto as cepas Nepal516, Kim e Pestoides F apresentaram um menor número de regiões em sintenia. No início da região referente à sequência de nucleotídeo 1800000 encontram-se os LCBs das HPIs das cepas, exceto de *Y. pestis* Nepal516, que não possui o loco *ybt*.

Os genomas das yersínias e de *E. coli* apresentaram todos os onze *ybt* (*fyuA* até *ybtS*) na mesma ordem (Figura 5). Os *ybt* homólogos de *P. luminescens* estão em sintenia, porém em ordem diferente, enquanto os de *C. koseri* *ybtPQXS* estão separados dos demais *ybt* por 27 ORFs. No caso do homólogo de *ybtQ* de *P. profundum* foram identificadas 714 ORFs até os demais *ybt*. O homólogo de *ybtP* foi encontrado no cromossomo II. O número de LCBs menor indica que as sequências dos *ybt* e as homólogas dos genes *luxR*, *16S*, *tufA* e *cdtMPQ* estão na mesma ordem. A distância entre os pares de base foi conservada apenas entre as cepas de *Y. pseudotuberculosis* (dados não mostrados) em relação as cepas de *Y. pestis* e *E. coli*.

Entre as cepas de *Y. pestis* a Pestoides F foi a que apresentou o menor número de ORFs (180) entre os *cdtMPQ* e os *ybt*. Os homólogos do gene *ybtD* foram encontrados em sintonia com os *cdtMPQ* em todas as yersínias e a quatro e 24 ORFs de distância dos *ybt* de *Y. pseudotuberculosis* e *Y. enterocolitica*, respectivamente. Nos genomas que são colineares aos *ybt* os *cdtMPQ* estão a mais de 1000 ORFs de distância.

O blast da sequência de YAPI revelou similaridades com diferentes regiões genômicas de cepas de *Y. pestis* e *E. coli* e com sequências dos plasmídeos pMT e pPCP de *Y. pestis*, pYPTS01 de *Y. pseudotuberculosis*, pVM01 e pUTI89 de *E. coli*, e do pNF1358 de *Salmonella enterica* (Tabela 2). A sequência de YAPI da cepa de *Y. pseudotuberculosis* 32777 (Nº de acesso AJ627388) foi a maior isoladamente sequenciada (102,752 bp) localizada no NCBI.

DISCUSSÃO

Individualmente os homólogos dos genes *ybt* de *Y. pestis* em perfeita sintonia e colineares ou não em conjunto de genes transferidos verticalmente (*rimM*, *tufA*, *ybtD* e *cdtMPQ*) e *luxR* confirmam a história da ancestralidade da HPI monofilética entre as yersínias. Adicionalmente nossos resultados sugerem que a ancestralidade dos *ybt* pode ter uma origem anterior a formação da família Enterobacteriaceae. A presença de homólogos dos genes *ybt* no genoma de *P. profundum* (Vibrionaceae) fortalece essa nova versão da história evolutiva da HPI (Figuras 1 e 2). A perda dos sinais evolutivos e da similaridade das unidades responsáveis pela síntese da Ybt pode ser compreendida pelas avaliações simultâneas das assinaturas seletivas dos *ybt* (Tabela 1), reorganização cromossômica das suas unidades gênicas, dos seus

genomas completos (Figuras 4 e 5), dos seus conteúdos de GCs (Figura 3) e de seus plasmídeos (Tabela 2).

A posição filogenética das HPIs das três yersínias patogênicas tem como ancestral intermediário a *C. koseri*. Essa cepa não apresenta os onze genes *ybt* em sintenia, o que indica que passou por uma maior pressão seletiva de códons em relação aos *ybt* de *E. coli* (Tabela 1), mas sua HPI não perdeu o sinal filogenético.

C. koseri apresenta características de um membro das Enterobacteriaceae que se diversificou em período anterior a origem de *Y. enterocolitica* (Figura 3). Segundo THOMSON *et al.* (2006) e CARNIEL (2008), *Y. enterocolitica* e *Y. pseudotuberculosis* divergiram há cerca de 200 milhões de anos. Assim houve um maior tempo para a diferenciação e seleção dos *ybt* em *C. koseri* após a doação desses genes aos seus descendentes. As diferenças de *C. koseri* também foram favorecidas por seu genoma possuir um maior número de genes ricos em GC quando comparados com os de *E. coli*, *Y. enterocolitica*, *Y. pseudotuberculosis* e *Y. pestis*.

Uma avaliação filogenética do gene *tufA*, realizada por ISABEL *et al.* (2008), evidenciou *Y. ruckeri* como um representante de um clado ancestral das yersínias. O cladograma construído com base nas nossas análises do gene 16S *rimM* também mostra a ancestralidade dessa espécie (Figura 1A e B). A hipótese dessa ancestralidade está sujeita a oscilações entre os valores de confiança dos filogramas de ambos os genes.

A posição de *Y. ruckeri* no gênero *Yersinia* é controversa (CARNIEL 2008; GIBELLO *et al.* 2004). Essa espécie, assim como *P. profundum* é um patógeno de peixes (NOGI *et al.* 1998). A patogenicidade de *Y. ruckeri* e de *P. profundum* e suas preferências ecológicas podem sugerir que a HPI veio do ambiente aquático e foi

perdida em alguns ancestrais de *Yersinia* e em espécies intermediárias entre *P. profundum*, permanecendo apenas nas espécies de yersínia patogênicas. Essa hipótese precisa ser reavaliada após a produção de novos genomas completos contendo HPIs homólogas de espécies de regiões marinhas, mas em princípio os padrões de posições das cepas são semelhantes nos filogramas e filogenoma (Figuras 1, 2 e 3).

Os conjuntos de dados avaliados neste trabalho são uma forte indicação da ancestralidade comum entre as espécies avaliadas e suas HPIs. Nesse sentido, mesmo que membros do gênero *Serratia* possuam HPI (BACH *et al.* 2000; CARNIEL 2008) na cepa analisada (*S. proteomaculans* 568) a ilha está ausente, assim como na cepa *Y. pestis* Nepal516.

A ancestralidade comum da HPI baseia-se no fato de que os *ybt* de *Y. pestis*, *Y. pseudotuberculosis*, *Y. enterocolitica* e *E. coli* são vizinhos dos genes reguladores do sistema *quorum sensing*, incluindo homólogos de *luxR*. Assim, se os genes *lux* de um membro das Vibrionaceae foram passados para um membro de Enterobacteriaceae (KASAI 2007; URBANCZYK *et al.* 2008), todos os genes filogeneticamente avaliados no presente trabalho (*rimM*, *tufA*, *ybtD*, *cdtMPQ*, *ybt* e *luxR*) foram portanto, herdados verticalmente nas espécies avaliadas.

Outra evidência é que apesar de grande heterogeneidade os genes regulados por *quorum sensing*, como os responsáveis pela produção de biofilmes (*hms*) em *Y. pestis*, estão presentes também em membros do gênero *Serratia* e são responsáveis tanto por atividades simbióticas em plantas quanto por patogenicidade em animais (VAN HOUDT *et al.* 2005; DODD *et al.* 2006; BOBROV *et al.* 2008). No entanto, são geralmente considerados como de origem horizontal, mesmo que os genomas hospedeiros sejam de membros dos gêneros *Yersinia* e *Serratia* evolutivamente

relacionados (Figuras 1A e B). Em *Y. pestis* o operon *hms* encontra-se vizinho ao *ybt* na HPI.

A alta similaridade entre as ilhas cromossômicas e plasmídeos de gêneros conhecidos classicamente como pertencentes à espécies evolutivamente próximas, pode indicar que alguns genomas hospedeiros tenham passado por maior pressão seletiva que diversas transferências horizontais independentes. No caso da HPI de *Y. pestis*, *Y. pseudotuberculosis*, *Y. enterocolitica* e *E. coli* as preferências por hospedeiros e mecanismos de transmissão podem facilitar a compreensão da manutenção de uma HPI conservada e ao mesmo tempo homóloga.

Um fato importante que deve ser considerado é a alta similaridade de DNA dos *ybt* entre *Y. pestis* e *E. coli*, ao mesmo tempo que os genes de *Y. enterocolitica* estão sob uma maior pressão seletiva de substituições não sinônimas de uso de códons. Segundo SHAPIRO e ALM (2008), uma reconstrução filogenética pode não ser significativa com valores muito divergentes de dN/dS . Porém a falta de um maior número de sequências genômicas é o que de fato aumenta a dificuldade para compreender os processos genéticos evolutivos das populações.

Na ausência de uma avaliação das alterações das frequências gênicas ao longo das gerações as alterações dos códons, rearranjos cromossômicos, ganho e perda de GC e análise filogenética são fatos que quando avaliados simultaneamente com os elementos móveis facilitam a compreensão entre similaridades, homologies e evolução, principalmente porque a similaridade pode ser perdida, mas não a homologia.

As relações entre as correlações positivas ou negativas (dN/dS) indicam variações de sítios conservados, porém ambas as situações não excluem a homologia entre as sequências ancestrais. Ilhas genômicas diferentes tendem a ter

sequências tanto para a integrase quanto para *tRNA* diferentes em sentença e não se mobilizam facilmente, ao mesmo tempo que algumas possuem maior dificuldade em manter os genes *int* e o *tRNA* presentes ou funcionais (INGE-VECHTOMOV 2002; LESIC e CARNIEL 2005; BOYD *et al.* 2009) e assim podem perder as características clássicas das ilhas.

A cepa de *C. koseri* (53% GC) não apresenta uma HPI padrão, ou seja, seus 11 *ybt* não estão em sentença, assim como os de *P. profundum* (~41% GC), que estão em cromossomos diferentes. Mesmo com maior variação suas posições ancestrais não perturbaram o filograma, como ocorreu com as posições de *E. coli* e *Y. pestis* (GC de ~50 e 47%, respectivamente). No primeiro caso as distâncias evolutivas favoreceram a fidelidade dos filogramas. No segundo caso, a similaridade dos *ybt* de *E. coli* e *Y. pestis* perturbaram os filogramas.

O fato de quase não haver alterações entre os códons dos *ybt* de *Y. pestis* e as substituições sinônimas serem maiores na *E. coli* são a razão para as perturbações dos filogramas, uma vez que em conjunto com os demais genes os sinais filogenéticos são recuperados (Figura 2A). Apesar da escassez de dados sobre a HPI no gênero *Photobacterium* a cepa *P. damsela* subsp. *piscicida* foi relatada com estrutura semelhante a HPI de *Y. pestis* (OSORIO *et al.* 2006), mas a espécie ainda não tem genoma completo conhecido.

O maior conteúdo de GC genômico encontrado na cepa de *C. koseri* pode ter favorecido sua adaptação em uma ampla gama de ambientes. Segundo LIOLIOS *et al.* (2006), esse patógeno pode ser encontrado no solo, água e alimentos. A maior similaridade da HPI de *C. koseri* com a de *Y. enterocolitica* (Figuras 2 e 5) poderia ser explicada por um maior número de substituições não sinônimas nessas espécies, apesar de *C. koseri* ser um ancestral mais distante das Yersínias do que

E. coli (Figuras 1 e 3). Assim os valores similares das correlações das substituições não sinônimas por sinônimas dos *ybt* de ambas foram análogos.

Por pressões semelhantes os valores das correlações entre as substituições não sinônimas nos *ybt* de *Y. pseudotuberculosis* e de *Y. pestis* foram análogos aos de *E. coli* e a homologia das relações evolutivas das HPIs das três yersínias patogênicas ficou menos perceptível (Figuras 2B e C). Entretanto as relações foram evidenciadas quando a construção do filograma foi baseado nas sequências dos genes 16S, *tufA*, *luxR*, *ybtD* e os *dctMPQ* (Figura 2A).

A ausência dos genes *ybtP* e *ybtQ* (ou de ambos) no alinhamento para formar um filograma (Figura 2C) evidenciou uma outra característica das cepas de *Y. pestis*. A cepa Pestoides F apresentou uma correlação de valores diferentes entre seus *ybt* em relação às outras cepas de *Y. pestis* e espécies avaliadas (Tabela 1). Essa variação entre os pares dos genes *ybt* foi acentuada com a ausência de *ybtP* e *ybtQ* nos alinhamentos. O comportamento atípico da cepa Pestoides F foi relatado inicialmente por ANISIMOV *et al.* (2004) a partir das suas característica bioquímicas. Nós também detectamos variação de sintonia dos LCBs, quando seu genoma completo foi alinhado com o da cepa CO92 (Figura 4), indicando que ela passou por um maior número de rearranjos cromossômicos, possivelmente devido a um maior período de divergência.

As cepas de *Y. pestis* possuem elevado número de elementos *IS* (PARKHILL *et al.* 2001; ACHTMAN *et al.* 2004; CHAIN *et al.* 2006), assim como foi elevado o número de LCBs detectados no nosso trabalho. A dinâmica dos movimentos desses elementos e das transposases provocaram rearranjos cromossômicos diferentes nas cepas de *Y. pestis*. Esses rearranjos originaram perfis individuais, o que poderia explicar a origem dos diferentes biotipos da espécie. A cepa *Y. pestis* KIM foi a que

apresentou maior variação: diferentes rearranjos são encontrados em toda sua sequência cromossômica quando comparada com a sequência da cepa CO92, exceto o último LCB.

Os *dctPQM* são responsáveis por unidades da família de transportadores TRAP (FORWARD *et al.* 1997). A descoberta de que as sequências de *dctPQM* são próximas às HPIs de *Y. pseudotuberculosis* e *Y. enterocolitica* permitiu compreender melhor a monofilia dos genes *ybt*. Adicionalmente observamos também que a cepa *Y. pestis* Pestoides F possui os genes *dctPQM* mais próximos aos *ybt*, evidenciando sua posição de membro mais antigo da espécie. No ramo externo dos filogramas, o fato de *ybtP* e *dctPQM* de *P. profundum* SS9 (cromossomo 2) e os demais genes *ybt* (cromossomo 1) não serem colineares pode ser explicado pelos rearranjos entre os cromossomos. Outra hipótese é se o ancestral possuísse os genes *ybtP* e *dctPQM* em cromossomos diferentes dos outros *ybt*, esses se teriam se unido em uma espécie intermediária que se diversificou.

Ainda existem limites para avaliar um número significativo de genomas completos de cada espécie e representativo de diferentes ecossistemas e da diversidade ecológica dos hospedeiros e seus ancestrais. No caso da cepa de *P. profundum*, patógeno de peixes, que pode viver a uma profundidade inferior a 5.110 m (NOGI *et al.* 1998), observamos que seu plasmídeo pPBPR1 e o plasmídeo pYV de *Y. enterocolitica* 8081 apresentam sete LCBs comuns. No entanto a similaridade entre eles é baixa (dados não mostrados) e semelhante aos padrões dos genes avaliados no alinhamento entre *Y. pestis* CO92 e *P. profundum* (final da Figura 5).

A ausência da YAPI em *Y. pestis* pode ser atribuída a sua disjunção do cromossomo da espécie ancestral, *Y. pseudotuberculosis*, e teria contribuído para a formação dos plasmídeos pMT e PCP (Tabela 2) específicos da *Y. pestis*

(WORSHAM e ROY 2003; CARNIEL 2008). Observamos que a YAPI apresenta GC inferior aos genomas hospedeiros da mesma forma que o conteúdo de GC dos plasmídeos pMT e pPCP é inferior ao do genoma da *Y. pestis*.

Também observamos que *Salmonella enterica* compartilha sequências de YAPI (Tabela 2). Por outro lado, ISABEL *et al.* (2008) ao analisar os genes *tuf* verificaram uma relação ancestral-descendente entre *Salmonella* e *Yersinia*. Portanto, uma reavaliação da homologia dessas sequências se faz necessário uma vez que grupos de genes similares aos plasmídeos de *Serratia* spp. e *Y. frederiksenii* são também considerados como resultado de aquisição horizontal (DODD *et al.* 2006).

Genes similares aos plasmídeos de *Y. pestis* PMT e pYV que codifica para um sistema de secreção Tipo III são encontrados em *P. asymbiotica* (WILKINSON *et al.* 2009) levando a hipótese de que uma espécie não patogênica da família Enterobacteriaceae poderia eventualmente ser virulenta.

As características ecológicas comuns de *P. luminescens* e *Y. enterocolitica* que também compartilham a YAPI (COLLYN *et al.* 2006; HEERMANN e FUCHS 2008) e a homologia encontrada nos genes *ybt* favorecem nossas conclusões sobre as relações ancestrais de *E. coli* e *Y. pestis* e que suas HPIs teriam a mesma origem. No entanto essa herança vertical apresenta diversas relações de parentesco intermediárias entre *Y. pestis* e as espécies ancestrais *Y. pseudotuberculosis* e *Y. enterocolitica*.

Em conclusão, a diversidade de bactérias patogênicas conhecida, seus diversos habitats e suas constantes migrações poderiam criar uma imensa quantidade de novas espécies em pouco espaço de tempo, se não houvesse limite para as transferências horizontais de genes de forma natural. Por outro lado, a

mobilidade interna promove grande variabilidade e está menos sujeita as pressões da seleção que uma transferência exógena de DNA, podendo promover tanto variação quanto o seu controle e protegendo a coesão das unidades genéticas, reprodutivas e ecológicas dessas espécies.

AGRADECIMENTOS

Os autores agradecem as seguintes instituições: Universidade Federal Rural de Pernambuco; Universidade Federal de Pernambuco; Centro de Pesquisas Aggeu Magalhães/Fundação Oswaldo Cruz; Fundação de Amparo a Ciência e Tecnologia do Estado de Pernambuco e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico.

LITERATURA CITADA

- ACHTMAN, M., MORELLI, G., ZHU, P., T. WIRTH and P. KEIM *et al.*, 2004 Microevolution and history of the plague bacillus, *Yersinia pestis*. Proc Natl Acad Sci USA. **101**: 17837-42.
- ANISIMOV, A. P., LINDLER, L. E. and G. B. PIER, 2004 Intraspecific diversity of *Yersinia pestis*. Clin Microbiol Rev. **17**: 434-64.
- BACH, S., A. M. P. DE ALMEIDA and E. CARNIEL, 2000 The *Yersinia* high-pathogenicity island is present in different members of the family *Enterobacteriaceae*. FEMS Microbiol Lett. **183**: 289-94.
- BOBROV, A. G., KIRILLINA, O., FORMAN, S., D. MACK and R. D. PERRY, 2008 Insights into *Yersinia pestis* biofilm development: topology and co-interaction of

- Hms inner membrane proteins involved in exopolysaccharide production. *Environ Microbiol.* **10**: 1419-32.
- BOYD, E. F., S. ALMAGRO-MORENO and M. A. PARENT, 2009 Genomic islands are dynamic, ancient integrative elements in bacterial evolution. *Trends Microbiol.* **17**: 47-53.
- BUCHRIESER, C., M. PRENTICE and E. CARNIEL, 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**: 2321-29.
- CARNIEL, E., 2008 Pathogenic *Yersinia*: Stepwise Gain of Virulence due to Sequential Acquisition of Mobile Genetic Elements. *In* Horizontal Gene Transfer in the Evolution of Pathogenesis (eds. H Hensel and M Schmidt). Cambridge University Press, Cambridge, 194-216.
- CHAIN, P. S., HU, P., MALFATTI, S. A., L. RADNEDGE and G. L. ANDERSEN *et al.* 2006 Complete genome sequence of *Yersinia pestis* strains Antiqua and Nepal516: evidence of gene reduction in an emerging pathogen. *J Bacteriol.* **188**: 4453-63.
- CICH, T. A., M. BLACKBURN, J. R. CARNEYAND and J. C. ENSIGN, 2003 Photobactin: a catechol siderophore produced by *Photorhabdus luminescens*, an entomopathogen mutually associated with Heterorhabditis bacteriophora NC1 nematodes. *Appl Environ Microbiol.* **69**: 4706-13.
- COLLYN, F., GUY, L., MARCEAU, M., M. SIMONET and C. A. ROTEN, 2006 Describing ancient horizontal gene transfers at the nucleotide and gene levels by comparative pathogenicity island genomics. *Bioinformatics.* **22**: 1072-79.

- CORNELIS, G. R., 2002 *Yersinia* type III secretion: send in the effectors. J Cell Biol. **158**: 401-8.
- DARLING, A. C., MAU, B., F. R. BLATTNER and N. T., PERNA, 2004 Multiple alignment of conserved genomic sequence with rearrangements. Genome Res. **14**: 1394-1403.
- DE ALMEIDA, A. M. P., GUILVOUT, I., I. ITEMAN and E. CARNIEL *et al.*, 1993 Chromosomal *irp2* gene in *Yersinia*: distribution, expression, deletion and impact on virulence. Microb. Pathog. **14**: 9-21.
- DODD S. J., HURST, M. R., GLARE, T. R., M. O'CALLAGHAN and C. W. RONSON, 2006 Occurrence of sep insecticidal toxin complex genes in *Serratia* spp. and *Yersinia frederiksenii*. Appl Environ Microbiol. **72**: 6584-92.
- FELSENSTEIN J. 1985 Confidence limits on phylogenies: an approach using the bootstrap. Evolution. **39**: 783-791.
- FOLEY, T. L. and M. D. BURKART, 2009 A homogeneous resonance energy transfer assay for phosphopantetheinyl transferase. Anal Biochem. **394**: 39-47.
- FORWARD, J.A., M. C. BEHRENDT, WYBORN N. R., R. Cross and D. J. KELLY, 1997 TRAP transporters: a new family of periplasmic solute transport systems encoded by the *dctPQM* genes of *Rhodobacter capsulatus* and by homologs in diverse gram-negative bacteria. J Bacteriol. **179**: 5482-93.
- GIBELLO, A., PORRERO, M. C., BLANCO, M. M., A. I. VELA and L. DOMÍNGUEZ, 2004 Analysis of the *gyrA* gene of clinical *Yersinia ruckeri* isolates with reduced susceptibility to quinolones. Appl Environ Microbiol. **70**: 599-602.
- HEERMANN, R. and T. M. FUCHS, 2008 Comparative analysis of the *Photorhabdus luminescens* and the *Yersinia enterocolitica* genomes: uncovering candidate genes involved in insect pathogenicity. BMC Genomics. **9**: 40.

- INGE-VECHTOMOV, S. G., 2002 A possible role of translation ambiguity in gene evolution (a hypothesis). *Mol Biol.* **36**: 202-208.
- ISABEL, S., LEBLANC. E., BOISSINOT, M., D. K. BOUDREAUAND and M. G. BERGERON, 2008 Divergence among genes encoding the elongation factor Tu of *Yersinia* species. *J Bacteriol.* **190**: 7548-58.
- KASAI, S., OKADA, K., HOSHINO, A., T. IIDA and T. HONDA 2007 Lateral transfer of the lux gene cluster. *J Biochem.* **141**: 231-7.
- LESIC, B. and E. CARNIEL, 2005 Horizontal Transfer of the High-Pathogenicity Island of *Yersinia pseudotuberculosis*. *J Bacteriol.* **187**: 3352–3358.
- LI, W. H. 1993 Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.* **36**: 96–99.
- LIOLIOS, K., TAVERNARAKIS, N., HUGENHOLTZ, P. and N. C. KYRPIDES 2006 The Genomes On Line Database (GOLD) v.2: a monitor of genome projects worldwide. *Nucleic Acids Res.* **34**: D332-4.
- MERHEJ V., ADÉKAMBI, T. PAGNIER, I., D. RAOULT and M. DRANCOURT, 2008 *Yersinia massiliensis* sp. nov., isolated from fresh water. *Int J Syst Evol Microbiol.* **58**: 779-84.
- NOGI, Y., N. MASUI and C. KATO, 1998 *Photobacterium profundum* sp. nov., a new, moderately barophilic bacterial species isolated from a deep-sea sediment. *Extremophiles.* **2**: 1-7.
- OSORIO, C. R., S. JUIZ-RÍO and LEMOS, M. L. 2006 A siderophore biosynthesis gene cluster from the fish pathogen *Photobacterium damsela* subsp. *piscicida* is structurally and functionally related to the *Yersinia* high-pathogenicity island. *Microbiology.* **152**: 3327-41.

- PAMILO, P. and N. O. BIANCHI 1993 Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. *Mol. Biol. Evol.* **10**: 271–281.
- PARADIS, S., BOISSINOT, M., PAQUETTE, N., S. D. BÉLANGER, and M. G. BERGERON *et al.* 2005 Phylogeny of the Enterobacteriaceae based on genes encoding elongation factor Tu and F-ATPase beta-subunit. *Int J Syst Evol Microbiol.* **55**: 2013-25.
- PARKHILL, J., WREN, B. W., THOMSON, N. R., R. W. TITBALL and B. G. BARRELL, *et al.*, 2001 Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature.* **413**: 523-27.
- PERRY, R. D. and J. D. FETHERSTON, 1997 *Yersinia pestis* - etiologic agent of plague. *Clin Microbiol Rev.* **10**: 35-66.
- POSADA D. and T. R. BUCKLEY 2004 Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. *Syst Biol.* **53**: 793-808.
- SCHUBERT, S., RAKIN, A., KARCH, H., E. CARNIEL and J. HEESEMANN, 1998 Prevalence of the 'high-pathogenicity island' of *Yersinia* species among *Escherichia coli* strains that are pathogenic to humans. *Infect. Immun.* **66**: 480–85.
- SHAPIRO, B. J. and E. J. ALM, 2008 Comparing patterns of natural selection across species using selective signatures. *PLoS Genet.* **4**: e23.
- SPRAGUE, L. D., SCHOLZ, H. C., AMANN, S., H. J. BUSSE and H. NEUBAUER 2008 *Yersinia similis* sp. nov. *Int J Syst Evol Microbiol.* **58**: 952-8.
- SULAKVELIDZE, A., 2000 *Yersinia* other than *Y. enterocolitica*, *Y. pseudotuberculosis*, and *Y. pestis*: the ignored species. *Microbes Infect.* **2**: 497-513.

- SWOFFORD, D. L. 2000 PAUP*: Phylogenetic Analysis Using Parsimony (* and Other Methods) (Sinauer, Sunderland, MA/USA), Version 4.0b10.
- TAMURA, K., DUDLEY, J., M. NEI and S. KUMAR 2007 MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. **24**: 1596-99.
- THOMSON, N. R., HOWARD, S., WREN, B. W., M. T. HOLDEN and M. B. PRENTICE *et al.* 2006 The complete genome sequence and comparative genome analysis of the high pathogenicity *Yersinia enterocolitica* strain 8081. PLoS Genet. **2**: e206.
- URBANCZYK, H., AST, J. C., KAEDING, A. J., J. D. OLIVER and P. V. DUNLAP 2008 Phylogenetic analysis of the incidence of lux gene horizontal transfer in *Vibrionaceae*. J Bacteriol. **190**: 3494-504.
- VAN HOUTT, R., MOONS, P., JANSEN, A., K. VANOIRBEEK and C. W. MICHIELS, 2005 Genotypic and phenotypic characterization of a biofilm-forming *Serratia plymuthica* isolate from a raw vegetable processing line. FEMS Microbiol Lett. **246**: 265-72.
- WILKINSON P., N. R. WATERFIELD, CROSSMAN, L., C. CORTON and M. SANCHEZ-CONTRERAS, 2009 Comparative genomics of the emerging human pathogen *Photobacterium asymbiotica* with the insect pathogen *Photobacterium luminescens*. BMC Genomics. **10**: 302.
- WORSHAM, P. L. and C. ROY, 2003 Pestoides F, a *Yersinia pestis* strain lacking plasminogen activator, is virulent by the aerosol route. Adv Exp Med Biol. **529**: 129-31.

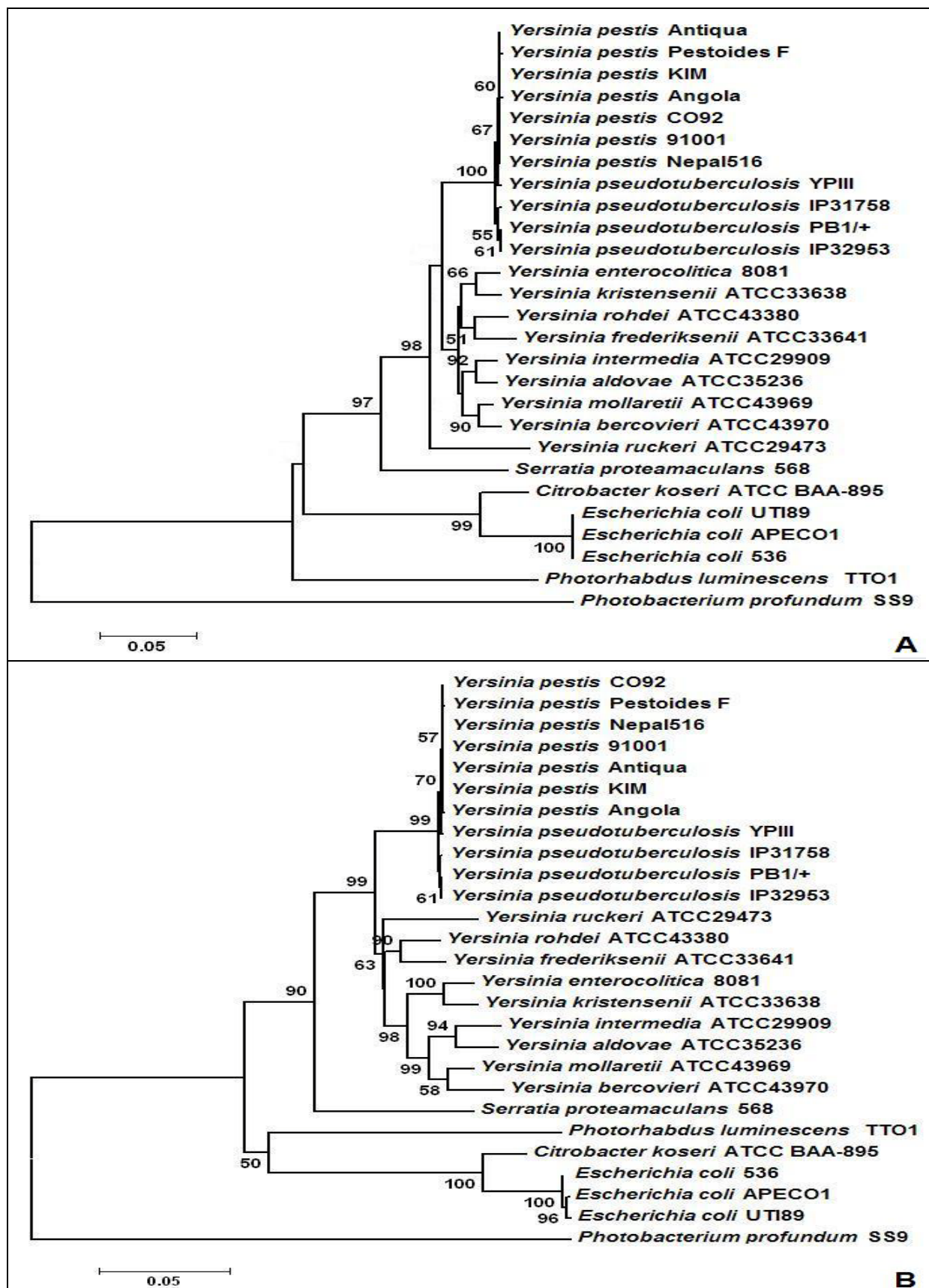


Figura 1 - Reconsruição filogenética baseada no método Neighbor-Joining e teste Bootstrap (4000 réplicas), software MEGA4.1: A) gene 16S (*rimM*); e B) genes *tufA* e 16S (*rimM*) concatenados.

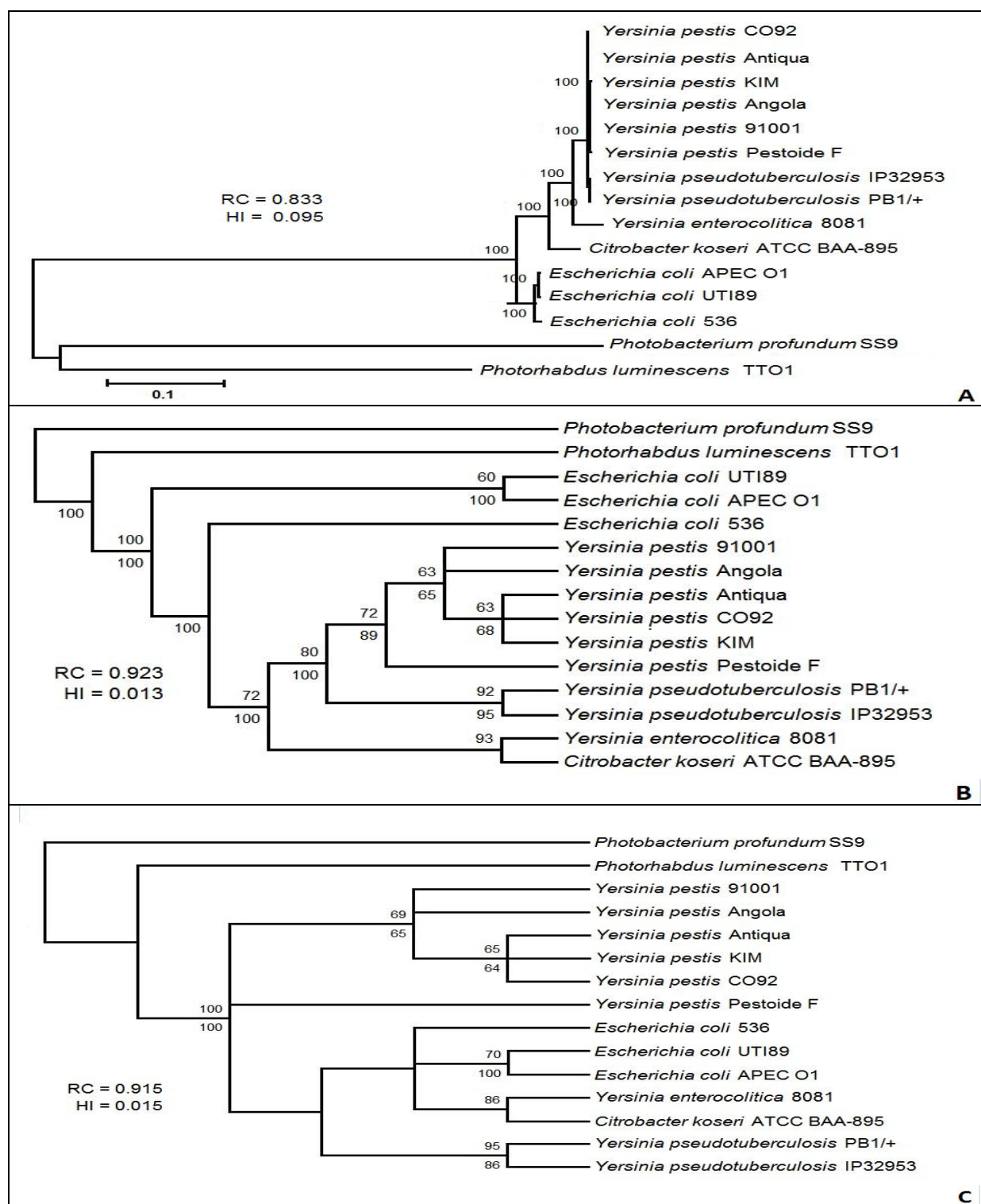


Figura 2 - Reconstruções filogenéticas baseadas nos genes concatenados: A) 11 genes *ybt* (*fyuA-ybtE-ybtT-ybtU-irp1-irp2-ybtA-ybtP-ybtQ-ybtX-ybtS*), *16S*, *tufA*, *cdtMPQ*, *luxR* e *ybtD*; B) 11 genes *ybt*; e C) *fyuA-ybtE-ybtT-ybtU-irp1-irp2-ybtA-ybtX-ybtS*; usando o método de Máxima Parcimônia (software PAUP* v. 4.0b10), busca exaustiva (bootstrap com 10.000 pseudo-replicações) e índice de homoplasia (HI) e consistência de reescalamento (RC).

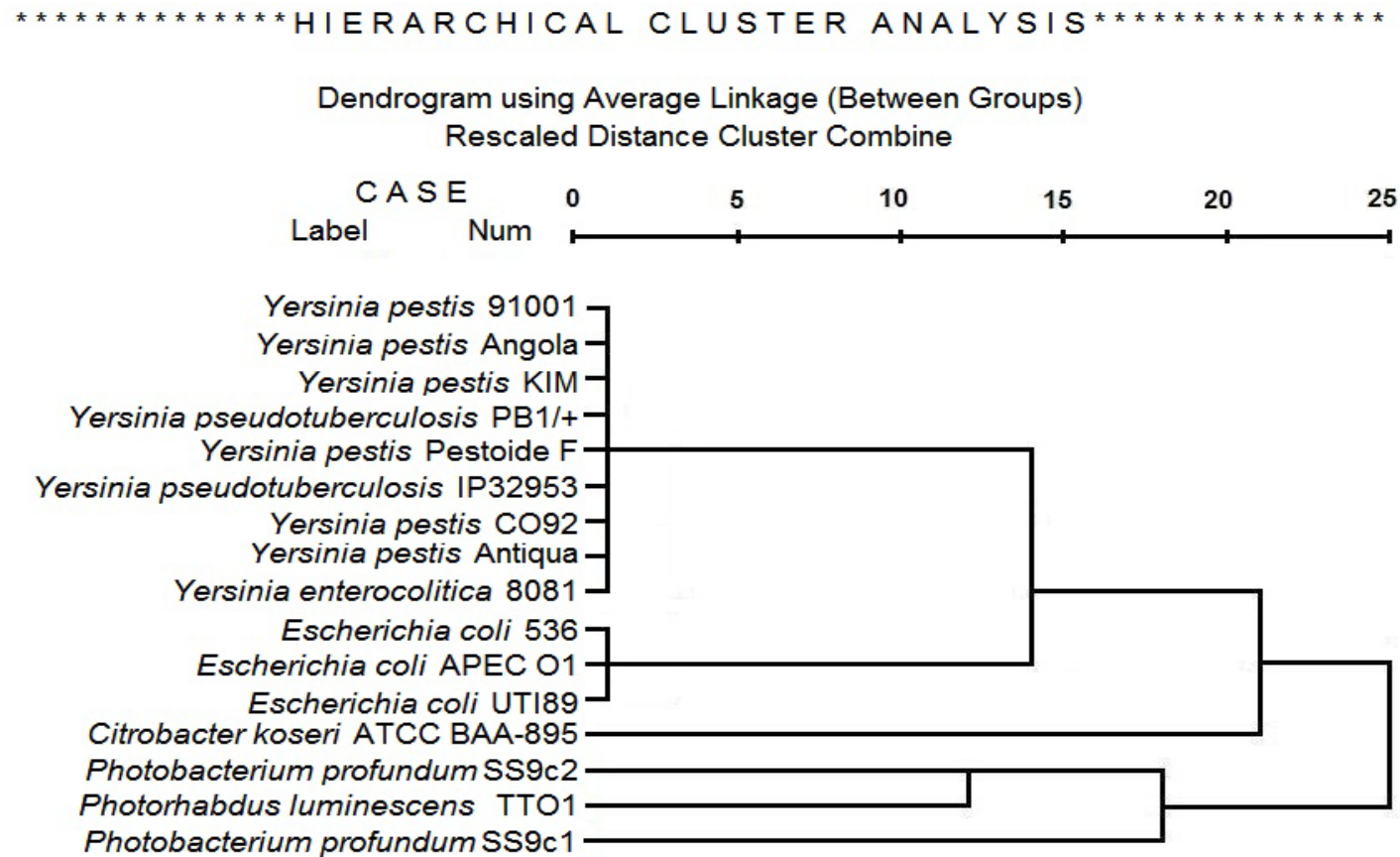


Figura 3 – Filograma construído a partir das 18 classes de frequências de GCs em cada posição do códon dos mRNAs de 15 genomas hospedeiros dos genes *ybt*, usando os softwares MatLab/ SPSS.13.

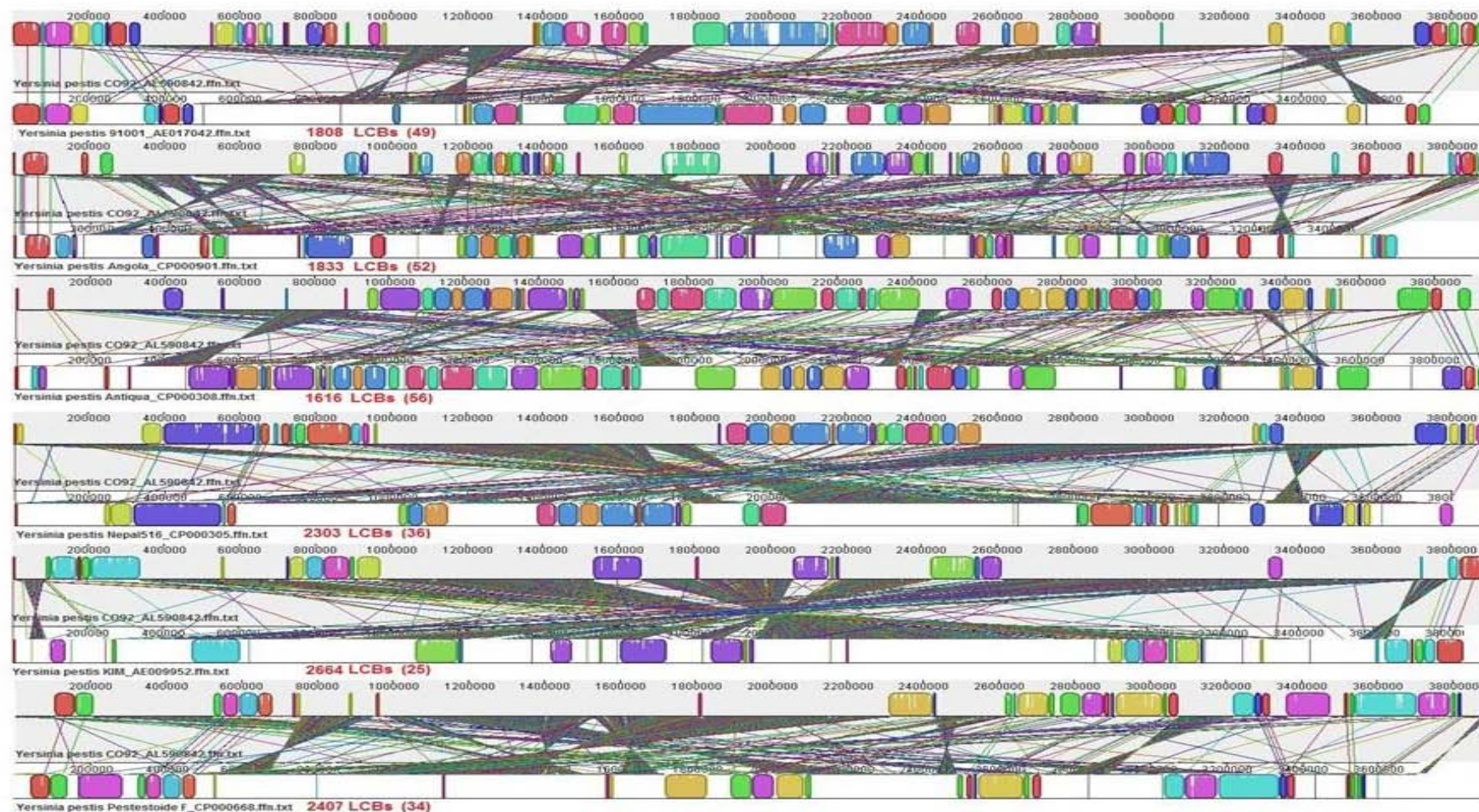


Figura 4 – Rearranjos dos Blocos Colineares Local (LCBs), formados pelas sequências cromossômicas das cepas de *Yersinia pestis* (91001, Angola, Antiqua, CO92, KIM e Pestoides F), produzidos pelos alinhamentos par a par com a cepa CO92 (software MAUVE 2.3). Os LCBs nas linhas horizontais são homólogos a *Y. pestis* CO92 e nas linhas verticais indicam as posições dos LCBs no cromossomo alinhado. A baixo de cada par de cepas, em vermelho, encontram-se os números totais de LCBs e os apresentados nos alinhamentos.



Figura 5 – Rearranjos entre os Blocos Colineares Local (LCBs) de 15 cepas, formados pelos alinhamentos dos genes *tufA* *cdtMPQ*, *ybtD*, *luxR*, onze *ybt* e 16S (*rimM*) homólogos da cepa de *Y. pestis* CO92 (software MAUVE 2.3). Maior intensidade de cor nos LCBs indica maior similaridade de DNA.

Tabela 1 - Matriz de similaridade das correlações das taxas da razão entre as substituições não sinônimas e sinônimas das sequências dos genes *ybt* das cepas de *Escherichia coli* (536, APEC01 e UTI89), *Yersinia pestis* (91001, Angola, Antiqua, CO92, KIM e Pestoides F), *Yersinia pseudotuberculosis* (IP32953 e PB1/+), *Yersinia enterocolitica* 8081, *Citrobacter koseri* ATCC BAA-895, *Photobacterium luminescens* TTO1 e *Photobacterium profundum* SS9, baseada no método de Pamilo-Biachi-Li (1993), softwares SPSS.13/MatLab.

Proximity Matrix																
Correlation between Vectors of Values																
	Genome	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	<i>C. koseri</i> ATCC BAA 895	1,000	0,513	0,452	0,569	-0,415	-0,525	0,638	0,589	0,603	0,545	0,545	0,629	0,612	0,603	0,600
2	<i>E. coli</i> 536		1,000	0,825	0,979	-0,185	-0,757	0,272	0,401	0,409	0,364	0,364	0,532	0,419	0,412	0,403
3	<i>E.coli</i> APEC01			1,000	0,910	0,057	-0,536	0,207	0,543	0,474	0,585	0,585	0,785	0,395	0,298	0,469
4	<i>E. coli</i> UTI89				1,000	-0,118	-0,758	0,326	0,502	0,490	0,484	0,485	0,668	0,475	0,440	0,483
5	<i>P. luminescens</i> TTO1					1,000	-0,129	-0,353	-0,163	-0,183	-0,084	-0,084	-0,078	-0,230	-0,273	-0,182
6	<i>P. profundum</i> SS9c1						1,000	-0,330	-0,377	-0,444	-0,317	-0,317	-0,413	-0,497	-0,527	-0,431
7	<i>Y. enterocolitica</i> 8081							1,000	0,400	0,426	0,358	0,358	0,373	0,448	0,456	0,423
8	<i>Y. pestis</i> 91001								1,000	0,988	0,988	0,988	0,901	0,952	0,888	0,989
9	<i>Y. pestis</i> Angola									1,000	0,956	0,956	0,845	0,987	0,947	1,000
10	<i>Y. pestis</i> Antiqua										1,000	1,000	0,925	0,896	0,811	0,957
11	<i>Y. pestis</i> CO92											1,000	0,925	0,896	0,811	0,957
12	<i>Y. pestis</i> KIM												1,000	0,770	0,669	0,843
13	<i>Y. pestis</i> Pestoides F													1,000	0,986	0,986
14	<i>Y. psdotuberculosis</i> PB1/+														1,000	0,946
15	<i>Y. psdotuberculosis</i> IP32953															1,000

Tabela 2 – Resultados do blast de nucleotídeos das sequências completas da Ilha de Aderência de *Yersinia* (YAPI) da cepa de *Yersinia pseudotuberculosis* 32777 (NCBI).

Accession	Description	Total score	E value
AJ627388.1	<i>Yersinia pseudotuberculosis</i> (YAPI), complete sequence, strain 32777	1,89E+08	0.0
CP000720.1	<i>Yersinia pseudotuberculosis</i> IP 31758, complete genome	1,26E+08	0.0
AM286415.1	<i>Yersinia enterocolitica</i> subsp. <i>enterocolitica</i> 8081 complete genome	7,54E+07	0.0
BX936398.1	<i>Yersinia pseudotuberculosis</i> IP32953 genome, complete sequence	5,21E+07	0.0
CP001048.1	<i>Yersinia pseudotuberculosis</i> PB1/+, complete genome	5,61E+07	0.0
CP000950.1	<i>Yersinia pseudotuberculosis</i> YPIII, complete genome	3,13E+07	0.0
CP000901.1	<i>Yersinia pestis</i> Angola, complete genome	2,57E+08	0.0
CP000668.1	<i>Yersinia pestis</i> Pestoides F, complete genome	1,54E+08	0.0
AE017042.1	<i>Yersinia pestis</i> biovar <i>Microtus</i> str. 91001, complete genome	1,83E+08	0.0
CP000308.1	<i>Yersinia pestis</i> Antiqua, complete genome	3,16E+08	0.0
CP000305.1	<i>Yersinia pestis</i> Nepal516, complete genome	1,82E+08	0.0
AE009952.1	<i>Yersinia pestis</i> KIM, complete genome	1,90E+08	0.0
AL590842.1	<i>Yersinia pestis</i> CO92 complete genome	2,26E+08	0.0
AE017045.1	<i>Yersinia pestis</i> biovar <i>Microtus</i> str. 91001 plasmid pMT1, complete sequence	1,31E+07	0.0
CP000900.1	<i>Yersinia pestis</i> Angola plasmid pMT-pPCP, complete sequence	1,61E+07	0.0
CP000669.1	<i>Yersinia pestis</i> Pestoides F plasmid CD, complete sequence	6630	0.0
CP000311.1	<i>Yersinia pestis</i> Antiqua plasmid pCD, complete sequence	4270	0.0
CP000310.1	<i>Yersinia pestis</i> Antiqua plasmid pPCP, complete sequence	3526	0.0
CP000309.1	<i>Yersinia pestis</i> Antiqua plasmid pMT, complete sequence	1,31E+07	0.0
AF074612.1	<i>Yersinia pestis</i> KIM plasmid pCD1, complete sequence	4656	0.0
AF053946.1	<i>Yersinia pestis</i> KIM plasmid pCD1, complete sequence	4656	0.0
DQ017661.1	<i>Salmonella enterica</i> plasmid pNF1358, complete sequence	3730	0.0
AL117189.1	<i>Yersinia pestis</i> CO92 plasmid pCD1	4224	0.0
CP000306.1	<i>Yersinia pestis</i> Nepal516 plasmid pMT, complete sequence	9572	0.0
AE017043.1	<i>Yersinia pestis</i> biovar <i>Microtus</i> str. 91001 plasmid pCD1, complete sequence	4269	0.0
AE017046.1	<i>Yersinia pestis</i> biovar <i>Microtus</i> str. 91001 plasmid pPCP1, complete sequence	3525	0.0
AF053945.1	<i>Yersinia pestis</i> KIM plasmid pPCP1, complete sequence	3525	0.0
AF074611.1	<i>Yersinia pestis</i> KIM plasmid pMT-1, complete sequence	9566	0.0
AL117211.1	<i>Yersinia pestis</i> CO92 plasmid pMT1	9572	0.0
AL109969.1	<i>Yersinia pestis</i> CO92 plasmid pPCP1	3525	0.0
AF053947.1	<i>Yersinia pestis</i> KIM plasmid pMT1, complete sequence	9561	0.0
CP000670.1	<i>Yersinia pestis</i> Pestoides F plasmid MT, complete sequence	8392	0.0
CP000307.1	<i>Yersinia pestis</i> Nepal516 plasmid pPCP, complete sequence	3514	0.0
EU330199.1	<i>Escherichia coli</i> plasmid pVM01, complete sequence	3913	0.0
CP000247.1	<i>Escherichia coli</i> 536, complete genome	4,48E+07	0.0
CP000244.1	<i>Escherichia coli</i> UTI89 plasmid pUTI89, complete sequence	3722	0.0
CP000243.1	<i>Escherichia coli</i> UTI89, complete genome	7129	0.0
CP001049.1	<i>Yersinia pseudotuberculosis</i> PB1/+ plasmid pYPTS01, complete sequence	3057	0.0

7. Informações complementares

Os alinhamentos das sequências cromossômicas não apresentados no artigo I (Item 5), referentes as regiões 30kb antes e depois dos genes *ybt* das HPIs ou regiões up_HPI_down das cepas de *E. coli*, apresentaram 15 LCBs (Blocos Colineares Local). As cepas de *E. coli* 536, CFT073 e UTI89 são uropatogênicas ao homem, enquanto a cepa APEC01 é patógeno de aves. Apesar das variações de especificidade ecológicas as cepas apresentaram as regiões avaliadas mais similares entre si que as cepas de *Y. pestis*.

Os alinhamentos entre as sequências dos plasmídeos pPBPR1 de *P. profundum* e pYV de *Y. enterocolitica* 8081 apresentam 7 LCBs e foram discutidos no artigo 2 (Item 6). As sequências avaliadas foram obtidas nos arquivos *ffn/NCBI (Anexo 10.3).

Os LCB na horizontal correspondem as unidades em sintenia de cada cepa, e os na vertical correspondem aos possíveis genes homólogos nas demais cepas, enquanto as linhas vermelhas transversais marcam os espaços entre as ORFs (Open Reading Frames) avaliadas.

7.1. Resultados dos alinhamentos dos LCBs nas regiões up_HPI_down das cepas de *E. coli*

No início da região up_HPI, tendo como referência o LCB1 de *E. coli* 536, estão as sequências que codificam a proteína chaperona HchA e uma suposta sensor-like histidina quinase YedV homólogas em *E. coli* CFT073 e UTI89. O LCB1 dessas cepas é o LCB2 de *E. coli* APEC01 e contém ORFs que estão invertidas entre suas regiões up_HPI e HPI_down. No LCB2 de *E. coli* 536 encontram-se os homólogos da proteína reguladora YedW (Figura 7.1.1). Na cepa *E. coli* APEC01 o LCB1 invertido apresenta baixa identidade (áreas sem cor) e corresponde ao LCB6 das demais cepas de *E. coli*, com ORFs para YedZ and YodA.

As regiões 5HPI de *E. coli* 536, CFT073 e UTI89 são similares de HchA a YodA e homólogas por mais 34,4kb (grande bloco verde ou os LCB7, onde estão localizados os *ybt*), com menor similaridade apenas em trechos de *irp1* e *irp2* e na região terminal do gene *ybtE* de CFT073 (poção branca maior no LCB7).

A primeira ORF dos LCB7 codifica YeeI (proteína hipotética), em sintonia com ORFs que codificam Int e Ybt, e a última ORF codifica YeeL (LPS heptosiltransferase). Após os LCB7 há um grupo de sequências para proteínas hipotéticas, entre as quais YeeN, membro dos LCB10 e homólogo do gene tRNA_{asn} .

Os LCB11, 12 e 13 correspondem respectivamente à proteína reguladora da assimilação de nitrogênio Nac, Cbl da família LysR de reguladores e a ORF prevista no sistema de efluxo de multidrogas que se encontra em sintonia com outra sequência de tRNA_{asn} . As ORFs dos LCB11, 12 e 13 estão na mesma ordem nas cepas de *E. coli* 536 e CFT073, mas invertidas nas cepas de *E. coli* APEC01 e UTI89. Os LCB14 (HPI3) correspondem à proteína YeeO, membro da família MOP (multidrogas/Oligossacarídeos-lipídio), e seu mRNA foi localizado também na região up_HPI_down de *Y. pseudotuberculosis*. Entre os LCB14 e LCB15 há uma sequência homóloga para o gene *int*, o LCB15 contém sequências para prováveis tioesterases, antiporter hipotético para droga/sódio e policetídeo amidase sintetase. Esse LBC possui cerca de 12,5 kb e baixa similaridade, mas quando a cepa *E. coli* APEC01 é subtraída do alinhamento todas regiões HPI_down apresentam homologia.

As HPIs de *Y. pestis* e *E. coli* estão envolvidos também em uma rede de sequências dos sistemas de efluxo de multidrogas e de proteínas reguladoras da assimilação de nitrogênio. Ambos os sistemas também são regulados por *quorum-sensing* e envolvem diversos polissacarídeos.

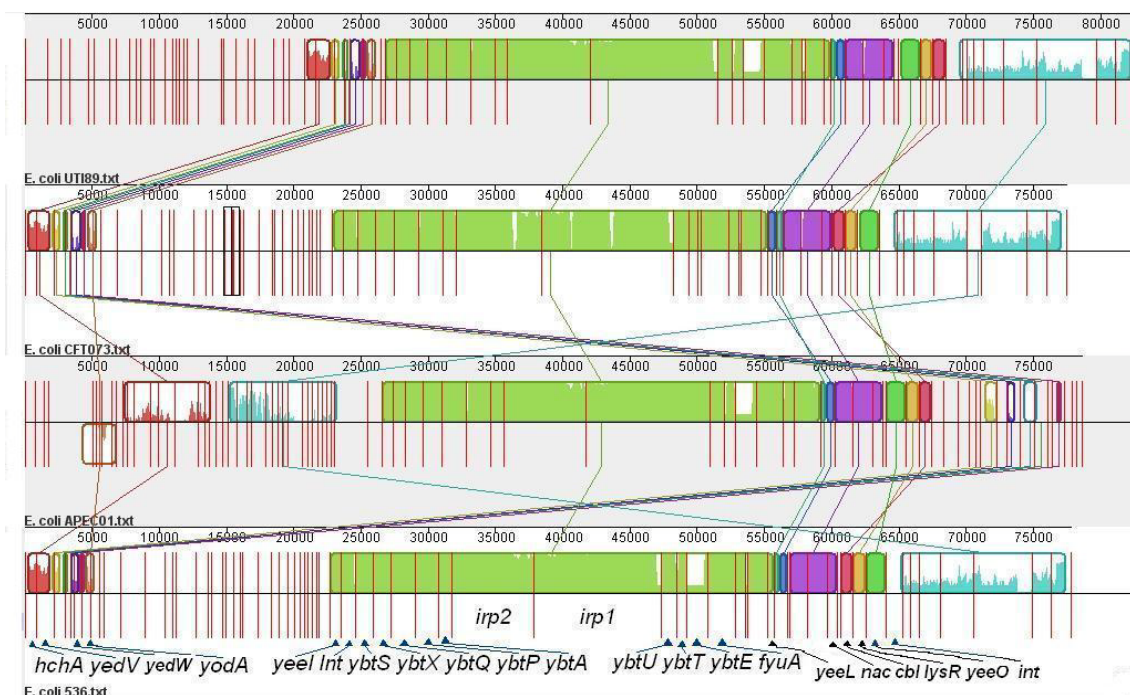


Figura 7.1.1. - 15 Blocos Colineares Local (LCBs), formado pelo alinhamento das sequências 30kb antes e depois da HPI das cepas de *Escherichia coli* 536, APEC01, CFT073 e UTI89 (software MAUVE 2.1).

7.2. Resultados dos alinhamentos dos LCBs entre o plasmídeo pPBPR1 de *P. profundum* e o plasmídeo pYV de *Y. enterocolitica* 8081

A evolução dos plasmídeos das cepas com genomas conhecidos, assim como dos elementos genéticos móveis dos cromossomos bacterianos, pode ter em breve sua história evolutiva reavaliada. No momento, parece que a luta para se perpetuar envolve aproveitar parcelas de DNA herdadas verticalmente. O alinhamento entre as sequências dos plasmídeos pPBPR1 e pYVe8081 de *P. profundum* e *Y. enterocolitica*, respectivamente (Figura 7.2.1), sugere que ambos compartilham semelhanças ao longo de suas sequências, apesar das diferença de bases de DNA (regiões sem coloração nos LCBs).

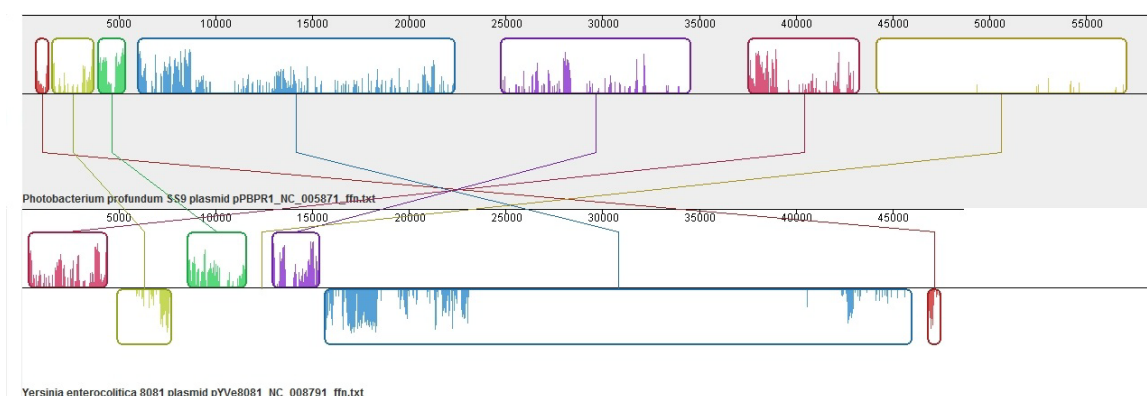


Figura 7.2.1. - 7 Blocos Colineares Local (LCBs) formados pelos alinhamentos entre os plasmídeos pPBPR1 de *Photobacterium profundum* SS9 e o plasmídeo pYV de *Yersinia enterocolitica* 8081 (software MAUVE 2.3).

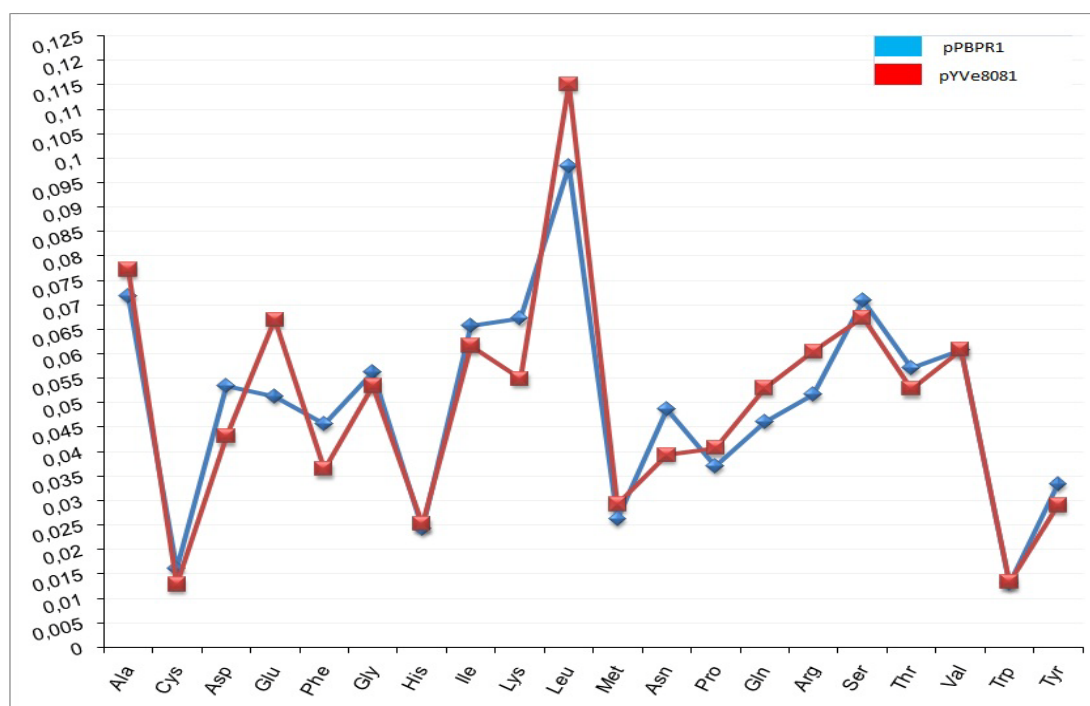


Figura 7.2.2. – Variações das frequências de aminoácidos entre os plasmídeos pPBPR1 de *Photobacterium profundum* SS9 e o plasmídeo pYV de *Yersinia enterocolitica* 8081, baseadas nos mRNAs dos arquivos *ffn/NCBI (software INCA).

8. Conclusões Gerais

Os genomas hospedeiros da Ilha de Alta Patogenicidade (HPI) das yersínias tendem a estar sujeitos a diferentes pressões seletivas, da mesma forma que cada *ybt* adaptou-se aos padrões de frequência de uso de aminoácidos do hospedeiro.

Os valores seletivos dos genes *ybt*, intra e interespécies, variaram ao longo da sua história evolutiva. Na cepa *Y. enterocolitica* 8081 o sinal filogenético dos genes *ybt* pode ser obscurecido pela assinatura seletiva negativa dos genes *ybt*, a alta similaridade com os homólogos das outras yersínias, o menor número de genes ricos em GC3 e a pressão do processo de homogeneização de sua HPI.

A vizinhança dos genes *dctPQM* e *ybtD* aos genes *ybt* de *Y. enterocolitica* e *Y. pseudotuberculosis* e na cepa *Y. pestis* Pestoides F sugere uma organização gênica semelhante no ancestral de *Y. pestis* (derivada de *Y. pseudotuberculosis*). As análises simultâneas desses genes e dos genes *ybt* revelaram que a HPI da família Vibrionaceae é ancestral da HPI da família Enterobacteriaceae.

O *fitness* adaptativo da HPI pode ser observado nos genomas de *Y. pseudotuberculosis* e *Y. pestis*, porque essas espécies mantêm quase inalteradas as sequências dos seus *ybt*, mesmo após sua divergência evolutiva.

As HPIs de yersínia e de *E. coli* estão envolvidas em uma rede de sequências reguladas pelo sistema *quorum sensing* e de sequências responsáveis pela produção de polissacarídeos, que até o momento eram desconhecidas.

9. Abstract

Evolution of the bacterial virulence mechanisms is directly related to the acquisition of transposable elements. The production of complete genome sequences enhanced the understanding of these phenomena. At present, seven complete sequences of *Yersinia pestis* strains are available. Our analysis of these sequences revealed new aspects of the evolution of the high-pathogenicity island (HPI), which harbors sequences coding for the siderophore yersiniabactin (Ybt). The evidence for adaptive cycles, even in the absence of intermediate genomes, provided the information that the *Yersinia* HPI has a monophyletic origin and the Vibrionaceae family could be its ancestral. The methodology was based on the analysis of sequences adjacent to HPI, *ybt* genes and its homologous selective signatures, and codon usage frequency of the host genomes revealing new aspects of the evolution of this genomic island. Patterns and differences in GC-content and synonymous and non-synonymous substitutions indicate that the *ybt* genes and their genomes suffered different selective pressures. Therefore different ancestral groups found diverse ways to preserve their HPIs. We observed that the HPI neighboring sequences are regulated by *quorum sensing*, indicating a general regulatory network involving *ybt* genes. The analysis of the chromosomal organization of the seven strains, representing different regions of the world, revealed a rearrangement dynamic that could justify the recognition of new variants of *Y. pestis* until recently considered very homogeneous specie.

Key words: *Yersinia pestis*; codon usage, evolution, HPI, comparative genomics.

10. Anexos

10.1. Instruções para Autores

- *Genetics and Molecular Research*

Instructions for Authors

Genetics and Molecular Research (GMR) publishes research articles, research reports, technical notes, scientific commentaries, news, views, and review articles with regard to genetics, evolution, molecular biology, and bioinformatics. Review articles are normally received by invitation only. If you would like for us to consider a review article, please consult the editor first; send a proposed title, a brief outline and a list of papers relevant to the review submitted by the author(s). GMR is an exclusively online journal.

The journal is maintained by the not-for-profit scientific foundation **Ribeirão Preto Foundation for Scientific Research (FUNPEC-RP)**. **The fee per accepted submission is R\$ 800.00 for Brazilian authors and US\$450.00 for authors from other countries. The US dollar amount reflects the approximate current foreign exchange rate and is subject to change.** This fee covers part of the expenses for final language and technical revision, for page setup, and for publishing online.

Payment of the publishing fee should be made by the authors only after receiving a letter of acceptance. After payment is received by our office, the manuscript will be processed further for publication.

Payment, both from within or outside Brazil, should be made by bank transfer (Banco do Brasil or Commerce Bank) or by credit card (Visa). Instructions for payment of the publishing fee will be sent together with the letter of acceptance for publication.

Please contact the editorial office [gmr@funpecrp.com.br] if you have any questions.

All GMR articles must meet the highest standards of scientific quality, both in terms of originality and significance, and the research findings reported should make substantial advances. As it is a journal serving a wide and varied scientific community, article abstracts, introductions and conclusions should be comprehensible to the non-specialist, stressing any wider implications of the work. However, the papers should not compromise on the scientific rigor and detail demanded by an international research journal. The broad readership that GMR attracts gives authors an opportunity to convey to a large audience, as well as to specialists, the importance of their work. The journal is currently indexed in over 64 services; see [<http://www.funpecrp.com.br/gmr/indexers.htm>].

Contributions should be sent either by e-mail as attachments to [gmr@funpecrp.com.br] or by regular mail to: Prof. Dr. Francisco A. Moura Duarte, Editor. Av. Presidente Vargas, 2627, 2º andar - Condomínio Itamarati, 14020-260 Ribeirão Preto, SP, Brazil.

It is a fundamental condition that submitted manuscripts have not been previously published and will not be simultaneously published elsewhere. With the acceptance of a manuscript for publication, the publishers acquire full and exclusive copyright for all languages and countries. The use of registered names, trademarks, etc., in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

All papers should be prepared in U.S. English. An initial evaluation of the language will be made upon receipt of each manuscript. Those that are considered inadequate for initial review will be returned or sent out for correction, at the discretion of the author. The manuscript will be considered officially received when the corrected version is ready to be sent to the referees.

Before final acceptance, a submission letter with the title of the article and names and signatures of all the authors should be posted to the above address or faxed to the journal at 55 (16) 3621-1991. Galley proofs will be sent in "pdf" form via e-mail for final revision. All authors are co-responsible for their submissions and they should make every effort to check the paper before this final step to avoid costly reformatting and possible introduction of new errors.

GMR articles have no rigid length restrictions. They should contain sufficient technical detail for an expert reader to understand and assess the methods and results. There is no page limit for GMR articles, but authors should still be concise, for two main reasons. First, our electronic refereeing system relies on e-mail, and very large files occasionally cause problems. Second, lengthy manuscripts can be cumbersome to read and study. Referees tend to dislike them, and they take longer to process. In addition, readers of electronic journals often print articles to read them. Remember that a 10,000-word article takes up around 11 pages.

How many pages would you be willing to read on-screen or print out?

Editorial policies: GMR is a refereed journal. Only original manuscripts will be considered for publication. Manuscripts will be reviewed by at least two independent reviewers before a decision is made on publication. The whole process is conducted electronically to speed progress and final publication. Papers will be published (placed online) within two months after final acceptance. Papers accepted in their final form from January 1 to March 31 constitute the first issue of each volume, and so on. There are four issues per year.

Manuscripts (in U.S. English), together with a cover letter from the author responsible for all correspondence, should be submitted to the Editor at [gmr@funpecrp.com.br] in electronic format as .doc files saved in Microsoft Word 97 for Windows, or later version. Do not use formatting such as Word's "Heading" or "Style Sheets". Spelling, punctuation, sentence structure, spacing, length, and consistency of usage in form and descriptions should be checked before submission. Please also check references for accuracy. Ensure that all figures and tables are mentioned in the text, and that all references are cited in the text. Figure and table files (see below) should be separate.

Submission information

Authors are required to provide the following information with their electronic submissions:

Author submitting the article; article title; authors (full list); article type and session; status of article (e.g., new, revised, etc.); postal address; e-mail address; phone number; fax number; names and types of the files sent.

Brazilian authors should not translate their institutional addresses. These should remain in the original (Portuguese) language.

Revised versions: Authors submitting a revised version of an article, must remember to include a list of changes, and replies to the referees (or technical editor). All the files, not just those revised, for the final draft of paper should be sent.

Acknowledgment of electronic submissions: Successful receipt and processing of the author's submission will be acknowledged by e-mail when the submitted manuscript has been

checked. If no reply has been received within one week, the author should contact the editor at [gmr@funpecrp.com.br].

Review: Articles are reviewed anonymously by independent referees. Authors are encouraged to suggest names of expert reviewers, but selection remains the prerogative of the editors. To facilitate the review process, the authors can send supplementary material, such as cited accepted but not yet published papers, which may be important for assessment of the manuscript.

A review article should contain: an abstract of 250 words or less, no more than 6 key words, a running title and no more than 60 references. It should be divided into sections with appropriate titles and subtitles.

Preparation of the manuscript

Order the sections comprising the manuscript as follows: title, running title, author, address, abstract, key words, introduction, material and methods, results, discussion, acknowledgments, and references.

Title Page: The title page should include the title of the article, authors' names, and authors' affiliation. The affiliation should comprise the department, institution (usually university or company), city, and state (or nation). The title page should include the name and complete mailing address, telephone number, fax number, and e-mail address of the author designated to review proofs. The title page should start below the top margin, be single-spaced, and have no space left before the Summary/Abstract. A running title of no more than 60 characters (including spaces) should be provided.

Abstract: An abstract of up to 250 words, single-spaced, is required of research articles and reports and should be arranged in one paragraph. The following information (without headings) should be included: purpose, methods, results, and conclusions. Review articles also require an abstract, which need not include all of these items.

Key words: A list of key words or indexing terms (up to six) should be included.

Text

Format: Headings should be bold, and first letters capitalized and left-aligned. All text should be set in Times New Roman font, 12 point, left-aligned, single-spaced. Do not justify the right margin. Leave only one (1) space after periods. Paragraphs should not be indented; there should not be any blank lines between them. Use line returns only at the end of paragraphs. Do not use tabs or spaces to create indents. Use the Symbol font for symbols and special characters. Do not use equation editors or footnoting utilities. Save equations as images. Equations should be numbered consecutively with Arabic numerals in parentheses on the right hand side of the page.

Footnotes: Footnotes should be avoided. When their use is absolutely necessary, footnotes should be numbered consecutively using Arabic numerals and should be placed at the bottom of the page to which they refer. Place a line above the footnote, so that it is set off from the text.

Tables/Charts: Special care should be taken to ensure that all tables are properly formatted. Scientific symbols used should be in Symbol or Times New Roman. Tables should be on a separate page, numbered consecutively (with Arabic numerals) referred to by number in the text and designed to fit the column or page size of the journal. Use tables with cells to separate columns. Do not use spaces, tabs or vertical lines. Left justify the title above the table. Indicate each table's location within the manuscript.

Illustrations: Illustrations/figures (photographs, drawings, diagrams, and charts) should each be in a single file, numbered in a consecutive series of Arabic numerals in the order in which they are cited in the text. Illustrations must be submitted as separate files. All illustrations are to be supplied in JPEG (jpg) format in either color or black and white. Images must be saved as separate, stand-alone files. The image resolution should be 300 dpi. Do not embed images within the text file. The placing of graphics in the paper should be indicated in the text and should include the captions for the figures. The authors should also send, by mail, a printed version of the figures. These should be at least 10 x 15 cm, up to US letter size, so that figures can be scanned to guarantee good quality for publishing online.

Abbreviations: Try to use abbreviations in the text sparingly. Write out abbreviations in full before the first time they are used in the text. Use the metric system for all measurements without periods (cm, mL, s). Define all symbols used in equations and formulas. Do not abbreviate the word "Figure" or "Table" in titles or text.

Acknowledgments: All acknowledgments (including those for grant and financial support) should be typed in one paragraph directly preceding the reference section. Authors of manuscripts submitted to GMR are requested to state the source of all funding that enabled the described research to be undertaken.

Reference: References in the text should include the name of the author and the year in parentheses, e.g. (Searle, 1961) or (King and Wilson, 1975). When a reference with more than two authors is cited, only the first author is named, e.g. (Comstock et al., 1958). The references must be cited in the text in chronological order, e.g. (Ideber, 2001; Uetz, 2002; Ottavai, 2004). References to "unpublished results" and "submitted papers" should appear in the text in parentheses following the name(s) of the individual(s). Example: (Pereira KS, Martins PK and Silva TM, unpublished results). **No more than 40 references should be cited in a Full-length paper, 20 references in a Short Communication and 60 references in a Review article.**

Reference, under the heading "References", should include only works referred to in the text. It should be arranged in alphabetical order under the first author's last name. References should be cited as follows: journal papers - names and initials of the first four authors (after that using et al.), year, full title, journal abbreviated according to PubMed or Web of Science, volume number, first and last page numbers; books - names of authors, year, full title, edition, publishers, address (city); articles published in symposia - names of authors, year, full title of book, name(s) of editor(s) in parentheses, publisher, address (city), first and last page numbers.

The references should consist of indexed articles only. References for techniques that are essential for understanding or repeating the methods should always be in easily accessible journals.

Reference style: The list of references at the end of the paper should follow the format requested by GMR. The link below can be accessed to see how the references should appear.

Examples of reference style

10.2. Instruções para Autores

– *Genetics*

PREPARING MANUSCRIPTS FOR SUBMISSION

Before submitting a manuscript at <http://submit.genetics.org/> make sure that it adheres to the guidelines laid out in this document. The version of the manuscript that was submitted will be published-ahead-of-print, verbatim, soon after it is accepted for publication. Submission of a manuscript to GENETICS implies that the corresponding author has obtained the permission of all authors. The corresponding author has the authority to act on behalf of all authors, but all authors are responsible for the article's content. By submitting a manuscript to GENETICS, all authors guarantee that the manuscript represents original research, that the data are available to the Associate Editor, Senior Editor, and Editor-in-Chief if necessary, and that the manuscript (and its substance) has not been previously published and is not currently being considered for publication by another journal. The corresponding author has full responsibility to ensure that all figures, tables, or text, and supplementary material is properly cited and, where appropriate, to obtain permission for reprinting other copyrighted material.

Authorship data entered during manuscript submission is the data that appears in the published-ahead-of-print version (for accepted articles), so be sure to complete all fields.

EDITORIAL STYLE

GENETICS editorial style conforms to the The Chicago Manual of Style (The University of Chicago Press) and Harbrace College Handbook (Harcourt) as interpreted and modified by the editors and Dartmouth Journal Services. The copyeditors rely on a variety of supplementary resources due to the diverse content of the journal:

- The Merriam-Webster Dictionary of English Usage
- Mathematics into Type (American Mathematical Society)
- Words into Type (Prentice Hall)
- A Dictionary of Genetics (Oxford University Press)
- Encyclopedia of Genetics (Academic Press)
- Scientific Style and Format (Council of Science Editors)

The editors reserve the privilege of editing manuscripts to conform with the stylistic conventions established in recent volumes of GENETICS, in these Instructions, and in the above-mentioned reference materials, as well as current terminology and accepted usage and practice.

MANUSCRIPT TEXT

Manuscripts must be:

- written in English with American spelling and correct grammar and punctuation,
- in 12-point type, double-spaced throughout, including the Literature Cited section appendices, tables, and legends,
- marked with consecutive page numbers, beginning with the cover page.

The first page should contain:

- a concise, informative title that includes the name of the organism under study,
- the authors' names,
- the authors' institutional affiliations, including department; institution; and city, state or province, country if outside the United States, and postal code (do not include districts or street addresses),
- sequence, expression or other reference numbers for public repositories.

For the authors' names, indicate different affiliations with the superscript symbols: *, §, **, § § .

Use a numbered footnote to indicate an author's present address. Indicate a dedicatory footnote (if desired) with boldfaced type.

The second page contains:

- a short running title of about 35 characters, including spaces,
- up to five key words or phrases,
- the corresponding author's name, mailing address including street name and number, phone and fax numbers, and email address.

Abstract

The abstract serves many purposes. It should be written for people who may not read the entire paper, so it must stand on its own. The impression it makes usually determines whether the reader will go on to read the article, so the abstract must be engaging, clear, and concise. In addition, the abstract may be the only part of the article that is indexed in databases, so it must accurately reflect the content of the article. A well-written abstract is THE most effective way to reach intended readers, leading to more robust search, retrieval, and usage of the article.

The abstract is a synopsis of the entire article. It should begin with the broad context of the study, followed by specific background for the study, its purpose, methods and procedures, core findings and results, and the conclusions. The abstract should emphasize new or important aspects of the research. It should engage the broad readership of GENETICS and be understandable to a diverse audience; jargon must be avoided.

Abstracts are a single paragraph, must be fewer than 225 words, and may not contain citations or abbreviations. The abstract must contain the full name of the organism studied.

Introduction

For the introduction, authors should be mindful of the broad readership of the journal. The introduction should set the stage for the importance of the work to a generalist reader and draw the reader in to the specific study. The scope and impact of the work should be clearly stated. In individual organisms where a mutant is being studied, the rationale for the study of that mutation must be clear to a geneticist not studying that particular organism. Similarly, study of particular phenotypes should be justified broadly and not on the basis of interest for that organism alone. General background on the importance of the genetic pathway and/or phenotype should be provided in a single, well-reasoned paragraph near the beginning of the introduction.

Authors are encouraged to cite the supporting literature completely rather than select a subset of citations. Additionally, authors are encouraged to provide important background citations, including relevant review papers [to help orient the non-specialist reader] and to cite similar work in other organisms.

Materials and Methods

Manuscripts submitted to GENETICS should contain a clear description of the experimental design in sufficient detail so that the experimental analysis could be repeated by another scientist. For example, indicate how many individuals were used, and if applicable how individuals or groups were combined for analysis. If working with mutants indicate how many independent mutants were isolated. If working with populations indicate how samples were collected and whether they were random with respect to the target population. If the level of detail necessary to explain the protocol goes beyond two paragraphs, give a short description in the main body of the paper and prepare a detailed description for a supplement.

It is important to indicate what statistical analysis has been performed; not just the name of the software and options selected, but the method and model applied. In the case of many genes being examined simultaneously, or many phenotypes, a multiple comparison correction should be used to control the type I error rate, or a rationale for not applying a correction must be provided. The type of correction applied should be clearly stated. It should also be clear whether the p-values reported are raw, or after correction. Corrected p-values are often appropriate, but raw p-values should be available in the supplementary materials so that others may perform their own corrections.

Results and Discussion

The results and discussion should not be repetitive. The results section should give a factual presentation of the data and all tables and figures should be referenced. The discussion should not summarize the results but provide an interpretation of the results, and should clearly delineate between the findings of the particular study and the possible impact of those findings in a larger context. Present and discuss results only once, not in both the Results and Discussion sections. It is sometimes acceptable to combine results and discussion.

The text should be as succinct as possible. Heed Strunk and White's dictum: "Omit needless words!"

The text should conform to the style in recent issues of GENETICS.

With in text citations include both names for citations with two authors. Citations with three or more authors name the first author with *et al.* for the other authors. Cite only articles that are published or in press. To cite personal communications or unpublished results, list all contributors by initials and last name; do not use *et al.*

In the text, write out numbers nine or less except as part of a date, a fraction or decimal, a percentage, or a unit of measurement. Use Arabic numbers for those larger than nine, except as the first word of a sentence; however, try to avoid starting a sentence with such a number.

Use abbreviations of the customary units of measurement only when they are preceded by a number: "3 min" but "several minutes". Write "percent" as one word, except when used with a number: "several percent" but "75%." To indicate temperature in centigrade, use ° (for example, 37°); include a letter after the degree symbol only when some other scale is intended (for example, 45°K).

Italicize names of organisms only when the species is indicated: *Neurospora*, but *Neurospora crassa* or *N. crassa*. Italicize the first three letters of the names of restriction enzyme cleavage sites, as in *HindIII*. Write the names of strains in roman except when incorporating specific genotypic designations.

Italicize genotype names and symbols, including all components of alleles, but not when the name of a gene is the same as the name of an enzyme. Do not use "+" to indicate wild type. Carefully distinguish between genotype (italicized) and phenotype (*not* italicized) in both the writing and the symbolism.

To divide the text into sections, only these four levels are allowed:

- ☐ Level 1: Centered, all capital letters. Use this level for major sections of text, such as Materials and Methods and the Discussion.
- ☐ Level 2: Freestanding flush-left boldface. Use this level only to group two or more closely related Level 3 headings in long article.
- ☐ Level 3: Paragraph-initiating boldface, followed by a colon.
- ☐ Level 4: Paragraph-initiating italic. Level 4 headings may be numbered, but only when the numbers must be cited in the text.

FIGURES

To prevent publication delay, these guidelines should be followed when submitting figure files. A summary is included below and detailed information is available in the individual sections. Figures include drawings, graphs, mating-type charts, complex chemical formulas, and other sketches.

File Format

Acceptable file formats are:

- ☐ TIFF (.tiff)
- ☐ JPEG (.jpg)
- ☐ Microsoft Word (.doc)
- ☐ Microsoft PowerPoint (.ppt)
- ☐ EPS (.eps)
- ☐ Adobe Illustrator (.ai)

Formats not supported include: LaTeX/Tex, Bitmap (.bmp), PICT (.pict), Excel (.xls), Photoshop (.psd), Canvas (.cnv), Corel Draw (.cdr), and PDFs (.pdf).

Resolution, Contrast, and Size

All files must conform to these resolution ranges:

- ☐ Halftone (color or black & white): 200-300 dpi
- ☐ Grayscale and color combination images: 300-500 dpi
- ☐ Line art images: 500-1000 dpi

RGB, CMYK, and Grayscale are all acceptable. Halftones should be high contrast with sharp detail, because some loss of detail and contrast is inevitable in the printing process. Good contrast is particularly important for chromatographs. Each file must be no larger than 10 MB; figures should be 10-20 cm in width and 1-25 cm in height. Graph axes must be exactly perpendicular and all lines of equal density (.5 point minimum).

Note that while there is a nominal charge for color figures in the main text (\$100.00 for GSA members per figure), there is no charge for color figures in supplementary material.

Labels and Legends

For figures with multiple parts use A, B, etc. in non-bolded type to label the parts for easy reference within the text. If a figure has important areas to highlight, arrows or numbers can be used to draw attention to these areas. If the figure is submitted in unattached parts, include a scan of how these should be arranged in the printed version.

Label all figures in numerical order with Arabic numerals, and cite figures in the text in their numerical order. When preparing the manuscript for publication, the copyeditors will insert or relocate a citation, or will renumber figures as needed. Within the label, be sure to distinguish between similar characters, such as the letter l and the number 1 or the letter O and the number 0. Labels and legends should be Helvetica 10 point bold font. The label heading (Figure #) should be in all Caps.

Figure legends should be a self-contained description of the content of the figure and should provide enough detail to fully understand the data presented. Figure legends should start with a brief title. All conventional symbols used to indicate figure data points are available for typesetting; unconventional symbols should not be used. Mathematical variables (both in the figure legend and figure) should be italicized; all genotypes must be italicized. Any additional symbols that are normally italicized should continue to use italic rather than underlined roman characters. Note that Figures can be automatically converted into PowerPoint-ready slides available for download in the printed article.

Video and Movie Files

GENETICS supports video and movie files that can be linked from any portion of the article - including the abstract. Acceptable formats include .asf, avi, .wav, and all types of Windows Media files.

TABLES

Minimize shading, color type, line drawings, graphics, or other illustrations within tables. If necessary, prepare a separate, numbered figure to accompany the table. Histograms should not be used to present data that can be captured easily in text or small tables, as they take up much more space.

Format

The only acceptable file format is Microsoft Word (.doc), and each table must start on a new page, with long tables presented on several pages.

Layout

Title (required) Ⓓ Should be concise and list the table number in Arabic numerals. Tables should not be numbered 1A, 1B, etc., but if necessary, interior parts of the table can be labeled A, B, etc. for easy reference in the text.

Footnotes Ⓓ If used, these should be typed directly below the table, with lowercase, superscript italic letters (a, b, c, etc.). Use *, **, and *** to indicate conventional levels of statistical significance, explained below the table.

Tables should have clearly defined boxheads and bottom lines that are .5 point in weight. Use shorter horizontal rules within the boxhead to indicate unambiguously which subheadings are subordinate to a higher-level heading. Do not use vertical or diagonal rules and do not use horizontal rules between the boxhead and the closing rule. Each column in the boxhead should have a title, and each boxhead entry should refer to material beneath it and not to material to the right.

Labels and Legends

Labels for tables should be in all Caps and in Helvetica 10 point font. Material within the table should also be in Helvetica 10 point font and double-spaced. Any totals and sums should be in boldface type. A legend, if present, must precede any footnotes and should follow the font requirements for the labels and table text. Legends should be a self-contained description of the content and should provide enough detail to fully understand the data presented.

REFERENCES

The References section lists only articles that are published or in press. References should be formatted as they appear in a recent issue of GENETICS. Parenthetical references are cited in text chronologically. For author names, order references alphabetically by first author with a space between an author's initials. For multiple citations with the same first author, first list single-author entries by year using 1996a, 1996b, etc., then list two-author entries alphabetically by second author. Finally, list entries by three or more authors (cited in the text as "FIRST et al. 1996") only by year and without regard to number of authors or alphabetical rank of authors beyond the first. For articles with more than five authors, list the first five names and then "et al."

Sample journal article citation:

Bridges, C. B., and E. G. Anderson, 1925 Crossing over in the X chromosomes of triploid females of *Drosophila melanogaster*. *Genetics* **10**: 418–441. (Note spaces between authors' initials and after the boldface colon.)
(Note spaces between authors' initials and after the boldface colon.)

Sample book citation:

Sturtevant, A. H., and G. W. Beadle, 1939 *An Introduction to Genetics*. W. B. Saunders, Philadelphia.

Sample chapter-in-book citation:

Beadle, G. W., 1957 The role of the nucleus in heredity, pp. 3Ⓓ22 in *The Chemical Basis of Heredity*, edited by W. D. McElroy and B. Glass. Johns Hopkins Press, Baltimore.

SUPPORTING INFORMATION

We encourage authors to submit supporting information that enhances the published manuscript and which may be of interest to other investigators. Examples include tables, figures, movies, and extended Materials and Methods sections.

Enough detail must be provided between the text and the supporting information for the complete replication of the work by an independent scientist. In addition, authors are encouraged to provide information that may aid in the use of the results of the paper for additional studies. For example, in eQTL studies, the data needed to produce the summary *cis* and *trans* plot provides detailed information that another scientist seeking to understand what genes may be regulating their favorite gene may find extremely useful. Any information which enhances the impact and usefulness of the work presented is appropriate for the supplement. Other examples include complete analysis results (including those not statistically significant) for microarray analysis, metabolomic data or phenotypic data.

All supporting information is uploaded to the site by the editorial staff as one combined PDF and also as individual files. We provide hyperlinks within the published MS directly to each individual supporting file and one link to the combined supporting information file. This combined file is also appended to the published PDF for offline accessibility.

File Submission

You can submit files online at <http://submit.genetics.org> during the initial manuscript submission, or when the final manuscript source files are uploaded. Files are checked prior to advanced online publication to ensure conformity with GENETICS' style guidelines. If any changes are required, we will return the file(s) to you for corrections. The editorial office will publish all supporting information as it is submitted, without editing, proofreading, or alteration.

Please refer to each piece of supporting information at least once in the manuscript so that we can provide a hyperlink to the file. In-text citations belong at the end of the sentence in which the supporting information is referenced and should say (see Figure S#, Table S#, or File S#).

If uploading a file that contains a Supporting Methods section with figures and tables, please let the editorial office know if this section should be uploaded as one file. Unless otherwise notified, the editorial office will upload each figure and table as a separate file.

File Format

Submit supporting information combined as one non-PDF file—preferably a Microsoft Word file. This file will be converted to a PDF prior to being uploaded. Please number each page with # SI, and format the numbering to start at 2. Page numbers should be located at the top, outside edge of the page and use New Baskerville type, 9 point in size. Additionally, please center your author information at the top of each page. For papers with three or more authors, use the first author's first and middle initials and full last name followed by *et al.* For papers with only two authors, include both names in the format of first and middle initials and full last name.

example

J.R. Smith *et al.*

J.R. Smith and S. Doe

If possible, the document should use a header margin of .25 inches and a 1 inch margin for all sides.

Please contact the editorial office if you have any questions about submitting your files. If necessary, individual files may be submitted in addition to the combined file (for example, videos, computer programs, or compressed data folders). In this case, please refer to the file in your combined document. An example the combined supporting information document is available [here](#). An example of the webpage for the supporting information files is available [here](#).

Supporting Figures

Supporting figures display a title and legend below the figure (FIGURE S#.—). "Figure" should use small caps and be followed by a capital S, Arabic number, period, and em dash. The Legend belongs on the same line as the title and both should use New Baskerville type in a regular weight, 9 point in size and single-spaced.

Multiple parts of a figure can be labeled with A, B, C, etc. Any references in the legend to a part of the figure should place the letter within parentheses at the beginning of the sentence. Please use New Baskerville type in a regular weight.

Supporting Tables

Supporting tables display a centered number and title at the top of the table with the title below the number (TABLE S#). The first word of the title should use initial caps and no period is necessary unless the title is a complete sentence. Type is New Baskerville in a bold weight, 9 point in size and double-spaced.

Type for the table body is New Baskerville in a regular weight, 9 point in size and spaced one and a half points. Please display totals and sums in a bold weight.

Any legend(s) or footnote(s) below the table use New Baskerville type in a regular weight, 9 point in size. The first line of the legend should be indented 0.15 inches with double-spaced text. Footnote text is single-spaced. Where applicable, use *, **, and *** to indicate conventional levels of statistical significance.

To order the information, please use table lines that are 0.5 point in weight and have clearly defined boxheads. Use shorter horizontal rules within the boxhead to indicate which subheadings are subordinate to a higher-level heading. Do not use vertical or diagonal rules, and do not use horizontal rules between the boxhead and the closing rule. When possible, each column in the boxhead should have a title, and each boxhead entry should refer to material beneath it and not to material to the right.

Supporting Files

Supporting files refer to material other than figures and tables. Examples include additional Materials and Method sections, reference papers, computer programs, and videos.

For each file, please use a title if applicable. The title should be centered at the top of the page in initial and small caps and use New Baskerville type in a bold weight, 9 point in size and double-spaced (FILE S#). If included, a legend should follow the same type recommendations and may be centered below the title. Body text uses New Baskerville type in a regular weight, 9 point in size and double-spaced. References and other sections or information not discussed here should follow the style of recent issues of GENETICS.

When submitting videos or computer applications, please compress them as needed to reduce the file size. Although these files will not be included in the combined PDF, please refer to them in the combined document as shown [here](#)—the editorial office will provide the web address for the files.

Summary

	Supporting Figures	Supporting Tables	Supporting Files
Number	FIGURE S#.— Indent 0.15 inches, below figure New Baskerville, 9pt Regular weight Single-spaced	TABLE S# Center top of page New Baskerville, 9pt Bold weight Double-spaced	FILE S# Center top of page New Baskerville, 9pt Bold weight Double-spaced
Title/Legend	Follows number New Baskerville, 9pt Regular weight Single-spaced	Center below number New Baskerville, 9pt Bold weight Double-spaced	Center below number New Baskerville, 9pt Bold weight Double-spaced
Labels	Capital Letters (A-Z)	Not applicable	Not applicable
Body	Not applicable	New Baskerville, 9pt Regular weight 1.5 line spacing Bold weight for sums and totals	New Baskerville, 9pt Regular weight Double-spaced
Table Lines	Not applicable	0.5 wt. Defined boxheads No vertical or diagonal lines	Not applicable

CHOOSING A SECTION

Please choose the appropriate section for the paper.

- Genome Integrity and Transmission
- Gene Expression
- Cellular Genetics
- Developmental and Behavioral Genetics
- Population and Evolutionary Genetics

- Genetics of Complex Traits
- Genome and Systems Biology

ARTICLE TYPES

INVESTIGATIONS

Investigations are full-length research articles that present novel findings on a range of topics bearing on inheritance.

NOTES

Notes are intended for the presentation of observations that do not warrant full-length articles, but that are of sufficient interest to justify their timely publication. Notes are not considered preliminary communications but should be a complete study; they will receive the same review as full-length articles.

Each Note must have an abstract of 50 or fewer words. Only level 3 headings are permissible in the body of the Note. The Introduction, Results, and Discussion are in a single section. The text should be kept to a minimum, preferably not exceeding 1000 words; the number of figures and tables should also be kept to a minimum. Materials and methods should be described in the figure legends or table footnotes. Place acknowledgments at the end of the Note, without a heading. The Literature Cited section is identical to that of full-length articles.

PERSPECTIVES

Articles for the Perspectives section of the journal are by invitation only. Proposals for Perspectives articles are welcome, and should be directed to the Perspectives section Editor, Adam S. Wilkins.

REVIEWS

Reviews are by invitation only, managed by Reviews Editor Allan Spradling and are also peer reviewed. Page charges are waived. If you are interested in writing a review, please contact Dr. Spradling or one of the senior editors to discuss publishing your review.

EDUCATION

Genetics Education provides a scholarly forum for educators who have created new teaching instruments. Authors are invited to submit articles that focus on effective ways for students at all levels to learn principles of genetics and to appreciate the implications of research in genetics. The articles or notes should be sufficiently detailed to permit others to adapt the ideas to their own circumstances, and also to provide evidence concerning the value of the innovations. To view the articles that have been published in this section, please see the Education Collections available on the journal home page. The section's Associate Editor, Pat Pukkila, welcomes inquiries concerning potential submissions: pukkila@unc.edu

LETTERS TO THE EDITOR

Letters to the Editor dealing with research and theory in basic genetics or with social issues of particular interest to geneticists are welcomed. Constructive comments on the subjects of articles from recent issues of GENETICS are appropriate as well as thought provoking commentary about issues of concern to Geneticists. Figures, complex tables, and complex mathematical formulas should be avoided. The Editorial Board will choose to publish letters it considers most pertinent to the interests of the readers. Letters to the Editor should be submitted electronically.

COMPANION PAPERS

Papers that overlap in the use of primary data or share authors and consider a similar topic from a different angle are welcome to be submitted as companion papers for concurrent publication. Papers by different authors concerning the same topic may also be submitted as companions.

Companion manuscripts should be submitted within two weeks of each other, and the cover letter should clearly indicate which papers are being submitted as companions. Failure to indicate which papers are companions may cause a delay in their consideration, as the editors seek to understand the relationships between papers sharing data or authors.

Each companion paper should have a cover letter clearly explaining its unique and novel contributions, as well as the importance of the group of papers.

Papers intended to be a companion of a paper(s) submitted for publication in another journal require full disclosure upon submission. Responsibility for adhering to publication policies in consortium groups is the sole responsibility of the authors of the paper.

RESUBMISSIONS

Resubmission of manuscripts for which the decision was "reconsider pending revision" is permitted a single time (unless arranged otherwise with the Associate Editor). Authors are expected to use the reviewer's comments to revise rejected papers before submitting them anew, and should provide a detailed point-by-point response to reviewers. Manuscripts without a detailed response may be returned to the authors without review. When resubmitting a manuscript authors should fully disclose its history and indicate in the cover letter the ID number of the previous manuscript. Manuscripts that have been rejected can be resubmitted only with the permission of the corresponding associate editor.

REVIEW PROCESS

A submitted manuscript is assigned to the Senior Editor of the appropriate subject section. The Senior Editor, in consultation with other editors, either returns the manuscript without review or assigns it to an Associate Editor who manages and adjudicates its review.

The Senior Editor will return manuscripts that are judged to be outside the scope of the journal. Manuscripts can be returned without review for reasons that include:

- O Failure to follow the GENETICS style guide for authors;
- O Grammar and style that is not of the quality expected in a published article;
- O The topic or scope of the work is not within the scope of the journal.

At least two Editorial Board members collaborate in determining whether a manuscript will be returned without review. Papers rejected because of deficiencies in style or grammar may be resubmitted to GENETICS after suitable revision.

Manuscripts sent for review are examined by one or more reviewers selected for their expertise in the subject matter of the article. Reviewers will remain anonymous (unless they choose to reveal themselves).

The Associate Editor makes one of the following decisions on the manuscript:

- Accept
- Accept pending minor revision
- Reconsider upon revision
- Reject

The average time from submission until decision is currently 31 days.

CHECKING MANUSCRIPT STATUS

Questions or concerns regarding manuscripts should be sent to the editorial office at td2p@andrew.cmu.edu or msstatus-genetics@andrew.cmu.edu. You may also contact the Associate Editor directly.

AFTER YOUR ARTICLE HAS BEEN ACCEPTED

Detailed information will be mailed with your acceptance letter. Guidelines and the general process of what happens between article acceptance and publication are listed below.

UPLOADING ACCEPTED MANUSCRIPT FILES

Upon manuscript acceptance, authors are required to upload final manuscript and figure source files to the Genetics online manuscript processing system, Bench>Press. Until these files are received, your paper is not considered in production. Detailed instructions are found in the acceptance letter emailed by the associate editor handling your manuscript.

When uploading accepted manuscripts, Microsoft Word, TEX, and LaTeX files are acceptable. RTF format and PDF are not acceptable.

For TEX files, be sure to include also the BIB files associated with the manuscript source files.

Electronic images must be provided, separate from the text. Art files must conform to the following minimum resolution specifications:

- grayscale and color images: 300 dpi
- combination grayscale and color images: 500–900 dpi
- line art (Bitmap) images: 900–1100 dpi.

To learn more about preparing your digital figures for optimal appearance, please see these [image guidelines](#) provided by Sheridan Press and Dartmouth Journal Services.

COVER ART SUBMISSIONS

Authors are encouraged to submit original art related to either their article, a special issue, or other GENETICS-related topic. Figures may be uploaded along with the final source files, and may also be emailed to the editorial office. Include a legend for each figure.

When submitting cover art, please submit an image file (PDF, TIFF, JPEG, or PSD) with its size under 10MB. If the cover art candidate is chosen, the editorial office will contact you and request a TIFF (.tiff) file sized at 19.05 x 15.24 centimeters (cm).

PUBLISH-AHEAD-OF-PRINT

GENETICS publishes manuscripts online, as soon as possible after they have been accepted for publication. The manuscript PDFs published ahead of print are the last-reviewed and accepted versions, and have not been formatted, copyedited, or approved by the Editor-in-Chief and the final proofs have not been reviewed by the authors. In support of open access, the published-ahead-of-print version of an article will be freely accessible to all readers, until the time of print and online publication.

The published-ahead-of-print manuscripts appear online as PDFs approximately eight to ten weeks before publication in the printed and online versions of GENETICS. Each manuscript published online ahead-of-print is citable, with the official publication date being the date of the manuscript's first online posting. GENETICS' articles are assigned digital object identifiers (DOIs), providing a persistent, permanent way to identify manuscripts published in the online environment.

After a manuscript is first published online, it continues through the process of print publication, including final copyediting, typesetting, proofreading, and author review. Companion articles slated for print and online editions may or may not be published simultaneously as published-ahead-of-print papers.

PROOFS

Corresponding authors will receive an email from the editorial office with the article's scheduled publication month. In addition, Dartmouth Journal Services notifies the authors approximately one week before proofs are emailed. When proofs are ready, the corresponding author will receive an email with access to a PDF of page proofs, the reprint order form, and proofreading instructions.

Each proof will show the date that the original manuscript was submitted and the date that the final version was accepted by the Editors.

Please read the entire article, with extra attention to title page elements (names, addresses, etc.) equations and symbols.

Proof corrections must be faxed or emailed to Dartmouth Journal Services within 48-hours of the send-date, or the article may be delayed to a later issue. All corrections must be sent together; additional corrections may result in extra charges and a delay in publication.

Notes added in proof require editorial approval. Email a copy of your request to the handling Associate Editor and td2p@andrew.cmu.edu

PUBLICATION CHARGES

Page charges are \$65 per printed page for corresponding authors who are members of the Genetics Society of America, and \$80 per printed page for non-members. While no article will be refused because of an author's inability to pay page charges, authors citing grant support are assumed to have the resources to pay for publication charges. To request a page charge waiver form, contact the Executive Editor by email.

Members of the Genetics Society of America pay \$100 for each color figure in an article. Nonmembers will be charged \$400 for the first color figure and \$250 for each subsequent color figure in an article.

Authors are expected to pay all the extra costs of printing and handling an article that are made necessary by author alterations—changes in galley proof not due to printer's error—or by changes made necessary because the printer could not interpret the author's intention owing to poorly prepared copy, incorrect file formatting, extra rounds of corrections, or inadequately organized tables. Such charges will not be waived.

Offprint order forms are included with the proofs. A price schedule is included on the order blank.

All charges will be billed on one invoice soon after the article has been printed.

OBTAINING COPYRIGHT PERMISSIONS FOR GENETICS ARTICLES

Permissions to reuse or modify all or part of a published article, including figures, tables, and legends, can be obtained via the Copyright Clearance Center, or by emailing the full citation, reason for request, and deadline to permissions-genetics@andrew.cmu.edu.

Note that you do not need to obtain permission from the GSA to use an article published in GENETICS if you are reproducing an article (on which you are an author) for your dissertation.

REPRINTS

Reprints will be shipped (from Dartmouth Journal Services) to corresponding authors about four weeks after the issue has been published. Authors receive order forms via email along with the electronic proofs, and both the proofs and the order forms must be returned at the same time. All charges will be billed on one invoice.

10.3. Relação das cepas avaliadas e suas principais características.

Cepa	Nome	RefSeq	GenBank	Tamanho*	GC%	Proteínas	RNA
<i>Escherichia coli</i> 536	Cromossomo	NC_008253	CP000247	493.892	50.0	4620	103
<i>Escherichia coli</i> APEC O1	Cromossomo	NC_008563	CP000468	5.1	50.0	4428	115
	APEC-O1-R	NC_009838	DQ517526	0.24	46.0	224	--
	pColBM	NC_009837	DQ381420	0.17	49.0	199	--
<i>Escherichia coli</i> CFT073	Cromossomo	NC_004431	AE014075	1.523.143	50.0	5338	116
<i>Escherichia coli</i> UTI89	Cromossomo	NC_007946	CP000243	506.574	50.0	5021	110
	pUTI89	NC_007941	CP000244	0.11	51.0	145	--
<i>Citrobacter koseri</i> ATCC BAA-895	Cromossomo	NC_009792	CP000822	472.046	53.0	4980	142
	pCKO3	NC_009793	CP000823	0.0093	55.0	15	1
	pCKO2	NC_009794	CP000824	0.0056	51.0	13	1
<i>Photorhabdus luminescens</i> TTO1	Cromossomo	NC_005126	BX470251	568.899	42.0	4683	107
<i>Photobacterium profundum</i> SS9	Cromossomo1	NC_006370	CR354531	4.1	41.0	3416	190
	Cromossomo2	NC_006371	CR354532	223.794	41.0	2006	23
	pPBPR1	NC_005871	CR377818	0.08	44.0	67	--
<i>Yersinia enterocolitica</i> 8081	Cromossomo	NC_008800	AM286415	4.6	47.0	3979	144
	pYVe8081	NC_008791	AM286416	10.068	43.0	72	--
<i>Yersinia pestis</i> 91001	Cromossomo	NC_005810	AE017042	24.6	47.0	3891	101
	pCD1	NC_005813	AE017043	20.07	44.0	85	1
	pCRY	NC_005814	AE017044	20.021.742	49.0	30	--
	pMT1	NC_005815	AE017045	20.106.642	50.0	122	--
	pPCP1	NC_005816	AE017046	20.009.609	45.0	10	--
<i>Yersinia pestis</i> Angola	pnew_pCD	NC_010157	CP000902	0.068	44.0	88	1
	Cromossomo	NC_010159	CP000901	4.5	47.0	3832	115
	pMT-pPCP	NC_010158	CP000900	0.11	49.0	120	2
<i>Yersinia pestis</i> Antiqua	MT	NC_008120	CP000309	0.096	50.0	99	--
	Cromossomo	NC_008150	CP000308	4.7	47.0	4167	104
	pPCP	NC_008121	CP000310	0.011	45.0	9	--
	pCD	NC_008122	CP000311	0.07	44.0	89	--
<i>Yersinia pestis</i> CO92	Cromossomo	NC_003143	AL590842	465.373	47.0	3885	96
	pCD1	NC_003131	AL117189	0.070305	44.0	71	1
	pPCP1	NC_003132	AL109969	0.009612	45.0	9	--
	pMT1	NC_003134	AL117211	0.096	50.0	101	--
<i>Yersinia pestis</i> Nepal516	Cromossomo	NC_008149	CP000305	453.459	47.0	3981	108
	pMT	NC_008118	CP000306	0.1	50.0	104	--
	pPCP	NC_008119	CP000307	0.011	45.0	9	--
<i>Yersinia pestis</i> KIM	pMT-1	NC_004838	AF074611	0.10099	50.0	116	--
	Cromossomo	NC_004088	AE009952	4.6	47.0	4054	100
<i>Yersinia pestis</i> Pestoides F	pCD	NC_009377	CP000669	0.072	44.0	88	--
	Cromossomo	NC_009381	CP000668	4.5	47.0	3850	106
	pMT	NC_009378	CP000670	0.14	52.0	131	--
<i>Yersinia pseudotuberculosis</i> PB1/+	Cromossomo	NC_010634	CP001048	4.7	47.0	4150	97
	pYPTS01	NC_010635	CP001049	0.07	44.0	87	--
<i>Yersinia pseudotuberculosis</i> IP32953	Cromossomo	NC_006155	BX936398	474.467	47.0	3901	122
	pYptb32953	NC_006154	BX936400	0.027702	44.0	42	--
	pYV	NC_006153	BX936399	0.068525	44.0	95	--

*Mbp = milhões de pares de base; (Fonte: NCBI).