

Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Pós-Graduação em Biologia Computacional e Sistemas

MICHEL FRANCISCO ABANTO MARIN

Epidemiologia Genômica de *Vibrio cholerae* da Epidemia de Cólera na
América Latina

Tese apresentada ao Instituto Oswaldo Cruz como
parte dos requisitos para obtenção do título de
Doutor em Ciências

Orientadores: Prof^a. Dr^a. Ana Carolina Paulo Vicente
Prof. Dr. Jeronimo Conceição Ruiz

RIO DE JANEIRO

2013

Ficha catalográfica elaborada pela
Biblioteca de Ciências Biomédicas/ ICICT / FIOCRUZ - RJ

M337 Marin, Michel Francisco Abanto

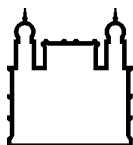
Epidemiologia genômica de *Vibrio cholerae* da Epidemia de Cólera na América Latina / Michel Francisco Abanto Marin. – Rio de Janeiro, 2013. xii,161 f. : il. ; 30 cm.

Tese (Doutorado) – Instituto Oswaldo Cruz, Pós-Graduação em Biologia Computacional e Sistemas, 2013.

Bibliografia: f. 91-105

1. Epidemiologia. 2. Genômica. 3. *Vibrio cholerae*. 4. Cólera. 5. América Latina. 6. Marcadores genômicos. 7. Genes ortólogos. 8. Pangenoma. 9. Genoma core. 10. Genes únicos. I. Título.

CDD 614.514



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Pós-Graduação em Biologia Computacional e Sistemas

MICHEL FRANCISCO ABANTO MARIN

**EPIDEMIOLOGIA GENÔMICA DE *Vibrio cholerae* DA EPIDEMIA DE
CÓLERA NA AMÉRICA LATINA**

ORIENTADORES: Prof. Dr^a. Ana Carolina Paulo Vicente

Prof. Dr. Jeronimo Conceição Ruiz

Aprovada em: ____ / ____ / ____

EXAMINADORES:

Prof. Dr. Antonio Basílio de Miranda - Presidente

Prof^a. Dr^a. Claudia Torres Codeço

Prof. Dr. Luis Caetano Martha Antunes

Prof. Dr. Alberto Martín Rivera Dávila

Prof. Dr. David Eduardo Barroso

Rio de Janeiro, 06 de junho de 2013

À minha família

AGRADECIMENTOS

À minha orientadora Dra. Ana Carolina Paulo Vicente pelo exemplo de dedicação à pesquisa, pela orientação, ensino e discussões acadêmicas que não estiveram sujeitas a horários, pelas incontáveis contribuições durante a realização desta tese e os artigos dela resultantes, pela compreensão e confiança.

Ao Dr. Antonio Basílio de Miranda por aceitar e ser o Revisor desta tese, pelas críticas e sugestões, assim como pelas palavras de estímulo e confiança.

A todos os integrantes do Laboratório de Genética Molecular de Microrganismos, aos que conheci e se engajaram em outros desafios, aos atuais amigos e amigas que, com seu carinho e disposição, preenchem todos vazios das saudades familiares.

A todos os co-autores dos trabalhos relacionados a esta tese, pela colaboração e ajuda, pelas discussões e crescimento mútuo.

Aos colegas e professores da BCS pelos momentos de amizade e ensino.

Aos Coordenadores e pessoal administrativo do Programa de Pós-graduação de Biologia Computacional e Sistemas pelo apoio sempre diligenciado.

Ao Instituto Oswaldo Cruz e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES pelo auxílio financeiro.

Finalmente à minha família e amigos por seu apoio e carinho que mesmo não citados nominalmente foram muito importantes na minha motivação de enfrentar os desafios com prazer e determinação.

Seja a mudança que você deseja ver no mundo

Mahatma Gandhi

ÍNDICE

1	INTRODUÇÃO	2
1.1	CÓLERA	2
1.2	ASPECTOS HISTÓRICOS DA CÓLERA	2
1.3	<i>VIBRIO CHOLERAE</i>	4
1.4	VIRULÊNCIA E PATOGENIA	5
1.5	GENOMA E ELEMENTOS GENÔMICOS DO <i>VIBRIO CHOLERAE</i>	5
1.5.1	O BACTERÍOFAGO CTX	10
1.5.2	ILHAS DE PATOGENICIDADE EM <i>V. CHOLERAE</i> (VPI-1 E VPI-2)	11
1.5.3	VSP-I E VSP-II	12
1.5.4	O SUPERINTEGRON	12
1.5.5	OS ELEMENTOS INTEGRATIVOS E CONJUGATIVOS	14
1.6	EPIDEMIOLOGIA DA SÉTIMA PANDEMIA DE CÓLERA	15
1.6.1	A CÓLERA NA AMÉRICA LATINA	16
1.6.2	SITUAÇÃO ATUAL	20
2	JUSTIFICATIVA	22
3	OBJETIVOS	25
3.1	OBJETIVO GERAL	25
3.2	OBJETIVOS ESPECÍFICOS	25
4	METODOLOGIA	27
4.1	TRABALHO EXPERIMENTAL	27
4.1.1	ISOLADOS DE <i>V. CHOLERAE</i>	27
4.1.2	EXTRAÇÃO DE DNA, AMPLIFICAÇÃO E SEQUENCIAMENTO	28
4.1.3	ANÁLISE DE SEQUÊNCIAS MULTILOCUS (MLSA - <i>MULTILOCUS SEQUENCE ANALYSIS</i>)	28
4.2	ANÁLISES GENÔMICAS	30
4.2.1	BUSCA POR MARCADORES EPIDEMIOLÓGICOS	30
4.2.2	COMPARAÇÃO DE GENOMAS DA AMÉRICA LATINA	32
4.2.3	IDENTIFICAÇÃO DE GENES ORTÓLOGOS, DETERMINAÇÃO DO PANGENOMA, GENOMA CORE E GENES ÚNICOS DE <i>V. CHOLERAE</i>	32
5	RESULTADOS	36
5.1	ALELOS DO GENE <i>CTXB</i>	36
5.2	CARACTERIZAÇÃO GENÉTICA DE <i>V. CHOLERAE</i> O1 DE GANA/ÁFRICA	38

5.3	CARACTERIZAÇÃO GENÉTICA DE ISOLADOS CLÍNICOS DE <i>V. CHOLERA</i> E DA NIGÉRIA	43
5.4	ESTRUTURA DO SUPERINTEGRON DE <i>V. CHOLERA</i> E	48
5.5	CARACTERIZAÇÃO DAS DIFERENTES LINHAGENS DE <i>V. CHOLERA</i> E BASEADA NO CONTEÚDO GÊNICO DO SUPERINTEGRON	54
5.6	CARACTERIZAÇÃO GENÉTICA E GENÔMICA DE <i>V. CHOLERA</i> E DA AMÉRICA LATINA	57
5.6.1	ANÁLISES FILOGENÉTICAS E CARACTERIZAÇÃO GENÉTICA	57
5.6.2	ANÁLISES GENÔMICAS	61
5.7	IDENTIFICAÇÃO E CARACTERIZAÇÃO DO ELEMENTO INTEGRATIVO E CONJUGATIVO DE <i>V. CHOLERA</i> E PANDÊMICO	72
5.8	PRESENÇA DE REGIÕES GENÔMICAS DE INTERESSE EPIDEMIOLÓGICO NA POPULAÇÃO DE <i>V. CHOLERA</i> E	76
6	<u>DISCUSSÃO</u>	80
7	<u>CONCLUSÕES</u>	89
8	<u>REFERÊNCIAS BIBLIOGRÁFICAS</u>	91
9	<u>ANEXOS</u>	107
9.1	GENOMAS E ISOLADOS DE <i>VIBRIO CHOLERA</i> E USADOS NAS ANÁLISES DO MLSA	107
9.2	INICIADORES USADOS NO ESTUDO	115
9.3	GENES DO SUPERINTEGRON DE GENOMAS DE <i>V. CHOLERA</i> E USADOS PARA A CARACTERIZAÇÃO DE LINHAGENS	116
9.4	ESTUDO DO EFEITO DE NUCLEOTÍDEOS E GAPS RANDOMICAMENTE INSERIDOS NA PREDIÇÃO E ANOTAÇÃO DE GENES	123
9.5	COMPRIMENTO DOS GENES CODIFICADORES DOS GENOMAS DE <i>V. CHOLERA</i> E	124
9.6	ÁRVORE DE MÁXIMA VEROSSIMILHANÇA DE HOMÓLOGOS DA INTEGRASE VC0516 DE UMA POPULAÇÃO DE <i>V. CHOLERA</i> E	125
9.7	PUBLICAÇÕES GERADAS DURANTE TESE	126

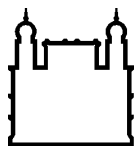
ÍNDICE DE FIGURAS

<i>Figura 1. Genoma de V. cholerae composto pelo cromossomo maior e o cromossomo menor..</i>	7
<i>Figura 2. Genomas completos e montados em contigs disponíveis no GenBank.</i>	8
<i>Figura 3. Pangenoma e genoma core de V. cholerae.</i>	9
<i>Figura 4. Estrutura do profago CTX e do fago satélite RS1 em diferentes variantes de V. cholerae O1.</i>	10
<i>Figura 5. Organização esquemática do genoma de V. cholerae e a localização do superintegron.</i>	13
<i>Figura 6. Organização genética do elemento integrativo e conjugativo (ICE) de um isolado do Haiti de 2010.</i>	15
<i>Figura 7. Dispersão e número de casos de cólera detectados no mundo no período de 1950-2004.</i>	16
<i>Figura 8. Dispersão de Vibrio cholerae O1 na América do Sul e Central, Janeiro 1991–Novembro 1994.</i>	17
<i>Figura 9. Número de casos por ano de cólera na América Latina, 1991-2011.</i>	19
<i>Figura 10. Eventos de transmissão de cólera correspondente a sétima pandemia.</i>	20
<i>Figura 11. Relações filogenéticas dos isolados de V. cholerae de Gana.</i>	40
<i>Figura 12. Dendrograma construído a partir dos perfis de PFGE de isolados representativos de V. cholerae O1.</i>	41
<i>Figura 13. Relações filogenéticas dos isolados de V. cholerae da Nigéria.</i>	44
<i>Figura 14. Dendrograma dos perfis de macro-restrição de V. cholerae O1 e não O1/não O139.</i>	44
<i>Figura 15. (A) Plot do pangenoma da região do superintegron considerando sete genomas de V. cholerae e um de V. mimicus. (B) Nuvem de palavras mostrando a abundância de clusters segundo sua função COG para o total e o core dos clusters.</i>	49
<i>Figura 16. Sintenia dos superintegrans das sete linhagens de V. cholerae e um V. mimicus.</i>	51
<i>Figura 17. Árvores filogenéticas baseadas nos SNPs dos genes do core do superintegron de V. cholerae construídos pelos métodos de máxima verossimilhança (A), Neighbor-Joining (B) e UPGMA (C).</i>	52
<i>Figura 18. Conteúdo gênico do superintegron de diferentes linhagens de V. cholerae.</i>	56
<i>Figura 19. Árvore filogenética de máxima verossimilhança de V. cholerae baseada na análise do MLSA.</i>	59
<i>Figura 20. Árvore consenso ClonalFrame para as sequências concatenadas do conjunto de dados do MLSA.</i>	60
<i>Figura 21. Distribuição das razões do comprimento do ramo externo/interno das árvores resultantes da análise do ClonalFrame comparada a árvores simuladas sob modelo coalescente.</i>	61
<i>Figura 22. Comparação dos genomas da América Latina usando o genoma N16961 como referência.</i>	62
<i>Figura 23. Organização genética e comparação da ilha VSP-II em V. cholerae El Tor N16961 com a VSP-II do genoma da IEC224.</i>	63
<i>Figura 24. Comparação da VSP-II dos genomas N16961 e 2010EL1786.</i>	64
<i>Figura 25. Relações filogenéticas da integrase da VSP-II.</i>	65
<i>Figura 26. Comparação dos genomas da América Latina usando o IEC224 como referência.</i>	67
<i>Figura 27. Pangenoma e conjunto de genes ortólogos da linhagem da América Latina e genomas referenciais.</i>	68
<i>Figura 28. Genes core exclusivos identificados pela análise de ortólogos na linhagem da América Latina.</i>	69
<i>Figura 29. Split network para o conteúdo gênico dos genomas de V. cholerae da linhagem da América Latina e genomas referenciais.</i>	70
<i>Figura 30. Distribuição espaço/temporal de membros da família ICEVchHai1.</i>	75

Figura 31. Presença de marcadores genéticos em V. cholerae.. 78

LISTA DE TABELAS

<i>Tabela 1. Pandemias de cólera registradas no mundo.....</i>	<i>3</i>
<i>Tabela 2. Número e origem dos V. cholerae da América Latina analisados por MLSA.</i>	<i>29</i>
<i>Tabela 3. Genes e regiões alvo da busca por BLAST nos genomas de V. cholerae</i>	<i>31</i>
<i>Tabela 4. Genomas de V. cholerae da epidemia da América Latina</i>	<i>34</i>
<i>Tabela 5. Polimorfismos associados aos alelos de ctxB em V. cholerae.....</i>	<i>37</i>
<i>Tabela 6. Características fenotípicas e genotípicas de V. cholerae O1 de Gana</i>	<i>39</i>
<i>Tabela 7. Características genéticas e fenotípicas de V. cholerae da Nigéria.....</i>	<i>46</i>
<i>Tabela 8. Genes core do superintegron de V. cholerae.</i>	<i>50</i>
<i>Tabela 9. Genomas de V. cholerae positivos para ICEVchHai1.....</i>	<i>74</i>



Ministério da Saúde

FIOCRUZ

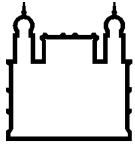
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Epidemiologia Genômica de *Vibrio cholerae* da Epidemia de Cólera na América Latina

RESUMO

Vibrio cholerae é o agente causal da cólera, uma infecção ancestral, epidêmica e pandêmica, que é um dos principais problemas de saúde pública no mundo. A humanidade já passou por, pelo menos, seis pandemias de cólera e, atualmente, vive no contexto da sétima pandemia. *V. cholerae* é classificado em mais de 200 sorogrupos e alguns biotipos. Linhagens de *V. cholerae* O1 do biotipo clássico e do biotipo El Tor estão relacionadas à sexta (1899–1923) e à sétima (1960 -....) pandemias de cólera, respectivamente. A sétima pandemia chegou à América Latina em 1991 sendo que, inicialmente, sua origem foi relacionada ao Sudoeste Asiático, região endêmica e epidêmica de cólera. Uma segunda hipótese sugere uma origem ambiental local. A partir do início deste século, a epidemia se extinguiu, restando apenas eventuais relatos de casos isolados. Sob a ótica da genética e genômica do *V. cholerae*, estudamos a linhagem da epidemia da América Latina (AL). Nossos objetivos eram inferir sua origem e identificar marcadores genômicos que possibilitassem seu monitoramento. Utilizamos informações obtidas *in vitro* e *in silico* para estabelecer relações genéticas entre isolados de *V. cholerae* de antes, durante e depois da epidemia que atingiu a América Latina. Analisamos regiões do genoma *core* e do genoma acessório. Realizamos análises de genômica comparativa com informação de 355 *V. cholerae*, clínicos e ambientais, de diferentes anos e regiões geográficas. Com base em três genes do genoma *core*, nossos resultados mostram que a linhagem da epidemia da AL pertence ao clado El Tor, e por sua vez, os genótipos dos principais determinantes de virulência *ctxB* e *tcpA*, (genoma acessório), os agrupa com isolados do início da sétima pandemia, incluindo o canônico N16961. As análises genômicas e filogenéticas revelaram a presença de um profago (WASA1) e de uma variante exclusiva da ilha genômica VSP-II. A VSP-II da linhagem da AL é caracterizada por seu gene da integrase e uma inserção de 7 Kb, que contém três genes putativos. Curiosamente, os dois marcadores, que fazem parte do genoma acessório desta linhagem, WASA1 e VSP-II, apresentam relação de similaridade com elementos genéticos identificados em *Vibrio vulnificus/Vibrio parahaemolyticus* e *Vibrio vulnificus/V. cholerae do ambiente*, respectivamente. Estes marcadores seriam, portanto, uma fração do mobiloma que evoluiu e se fixou na linhagem de *V. cholerae* da epidemia da América Latina.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Genomic Epidemiology of the *Vibrio cholerae* from the Latin American Cholera Epidemic

ABSTRACT

Vibrio cholerae is the causative agent of cholera, an ancestral, epidemic and pandemic infection, which is a major public health concern worldwide. Mankind has gone through at least six cholera pandemics and currently lives in the context of the seventh pandemic. *V. cholerae* is classified into more than 200 serogroups and some biotypes. *V. cholerae* O1 classical and El Tor biotype are related to the sixth (1899 to 1923) and seventh (1960 -) cholera pandemics, respectively. The seventh pandemic reached Latin America (LA) in 1991 and its origin was related, immediately, to Southwest Asia, where cholera is endemic and epidemic. A second hypothesis emerged suggesting a regional environmental origin. Since the beginning of this century, the LA epidemic ended, occurring only occasional case reports. Here, we studied, from the perspective of *V. cholerae* genetics and genomics, the lineage of the LA cholera epidemic. Our objectives were to infer their origin and identify genomic markers that would enable its monitoring. In order to establish the genetic relationships among *V. cholerae* strains isolated before, during and after the epidemic in Latin America, we used *in vitro* and *in silico* approaches. Comparative analyzes, targeting the core and accessory genome of 355 clinical and environmental *V. cholerae* strains, from different years and geographical regions, were performed. MLSA (multi locus sequence analysis), based on three genes of the core genome, revealed that the LA epidemic lineage belongs to the El Tor clade. Additionally, the genotypes of the major virulence determinants, *ctxB* and *tcpA* (accessory genome), show that the LA lineage is related with strains from the beginning of the seventh pandemic, including the canonical N16961. Considering this large set of genomes, we confirm the prophage WASA1 as a LA lineage marker. The LA lineage is also characterized by an unique VSP-II genomic island. The VSP-II AL harbors a particular integrase gene and a 7.0 kb insert which contains three putative genes. Interestingly, the two LA markers, WASA1 and VSP-II, which are part of its accessory genome, have similarity with genetic elements identified in *Vibrio vulnificus* / *Vibrio parahaemolyticus* and *V. vulnificus* / environmental *V. cholerae*, respectively. Therefore, these two elements are a fraction of the *V. cholerae* mobilome that evolved and fixed in the LA lineage.

Introdução

1 INTRODUÇÃO

1.1 Cólera

A cólera é uma doença diarréica aguda, que começa subitamente com diarreia volumosa e aquosa podendo levar a perda de um litro de líquido por hora, a qual tem aspecto de água de arroz, sem muco ou sangue. Sem a reposição de líquido e eletrólitos leva rapidamente a desidratação, colapso circulatório e morte. O quadro clínico vem acompanhado de cólicas abdominais, náuseas e vômitos (Sack *et al.* 2004; Chaudhuri e Chatterjee 2009). A infecção se dá, principalmente, pela ingestão de alimentos e água contaminada com *V. cholerae* patogênico (Kaper *et al.* 1995).

1.2 Aspectos históricos da cólera

As ocorrências de cólera têm sido registradas associadas a surtos, epidemias e pandemias. Uma pandemia é caracterizada por epidemias que podem durar anos ou décadas e que atingem diversos países em distintos continentes (Kaper *et al.* 1995). Embora se tenha descrições de casos semelhantes à cólera na antiguidade, são reconhecidas sete pandemias de cólera na literatura científica, a primeira delas no início do século XIX (Tabela 1).

Todas as pandemias de cólera tiveram origem na Ásia, e a partir daí espalharam-se para outros continentes. Um fenômeno diferente aconteceu com a sétima pandemia, que se originou na ilha Célebes, na Indonésia, alcançando logo depois o subcontinente indiano. De acordo com os registros, a primeira pandemia iniciou-se no delta do Ganges (Bangladesh e Índia) em 1817 e teve uma duração de seis anos, alcançando vários países europeus (Chaudhuri e Chatterjee 2009; Morris 2011). Foi durante a segunda pandemia, que começou em 1826 e alcançou as principais cidades européias, que William Brooke O'Shaughnessy, um famoso médico irlandês, demonstrou que os pacientes com cólera tinham fezes com aspecto característico de água de arroz contendo alto grau de eletrólitos, e sugeriu uma

terapia intravenosa de reposição de fluidos e eletrólitos (Chaudhuri e Chatterjee 2009).

Tabela 1. Pandemias de cólera registradas no mundo [Adaptado de (Chaudhuri e Chatterjee 2009)].

Pandemia	Período	Origem	Principais áreas afetadas
1	1817–1823	Índia	China, Japão, sudeste Asiático, leste da África,
2	1829–1851	Índia	Europa, América do Norte
3	1852–1859	Índia	Ásia, Europa, América do Norte e África
4	1863–1879	Índia	Europa, África, América do Norte
5	1881–1896	Índia	Ásia, Europa, América do Sul
6	1899–1923	Índia	Sul da Ásia, Índia, Europa, América do Norte
7	1961–presente	Indonésia	Mais de 100 países afetados nos cinco continentes

A segunda pandemia teve um efeito devastador na Europa e América e pouco tempo depois de seu fim emergiu a terceira pandemia, afetando principalmente Índia e Rússia, mas também a Inglaterra em 1848-1849. John Snow, um médico inglês considerado o precursor da epidemiologia moderna, demonstrou que a transmissão da cólera estava associada a uma fonte de água e traçou a origem de um surto de cólera no bairro londrino do Soho em 1854 (Sack *et al.* 2004; Chaudhuri e Chatterjee 2009).

Foi só durante a quinta pandemia (1881–1896), que Robert Koch isolou, a partir de fezes de aspecto de água de arroz de um paciente com cólera, uma bactéria à qual se referiu como “bacilos curvos” devido a sua forma quando observado ao microscópio óptico: a *Vibrio cholerae* (Sack *et al.* 2004; Chaudhuri e Chatterjee 2009).

A sétima e atual pandemia apresenta notáveis características em relação às anteriores como sua extensão geográfica, dispersão e duração (Salim *et al.* 2005). Esta pandemia começou na Indonésia e, diferentemente da origem das outras pandemias que começaram no delta do Ganges e que tinham como agente

etiológico o *V. cholerae* sorogrupo O1 biotipo clássico, esta vem sendo desencadeada pelo biotipo chamado de “El Tor”. *V. cholerae* deste biotipo foi isolado pela primeira vez em 1905, de peregrinos indonésios que estavam viajando para Meca, em uma estação de quarentena na aldeia de El Tor, no Egito. Este biotipo foi encontrado novamente em 1937, na ilha Célebes (Sulawesi), na Indonésia. Então, em 1960, esta linhagem começou a se espalhar pelo mundo, invadindo a Índia em 1964, a África em 1970, o sul da Europa em 1970 e a América do Sul em 1991. A cólera é atualmente endêmica em muitas destas regiões, particularmente no Sul da Ásia e na África (Sack *et al.* 2004).

1.3 *Vibrio cholerae*

O *Vibrio cholerae* pertence à família *Vibrionaceae*, é um bastonete Gram negativo levemente curvo, com único flagelo polar, e é uma bactéria autóctone de ambientes aquáticos, podendo estar colonizando crustáceos do zooplâncton, em especial os copépodes (Colwell 1996), assim como outros organismos aquáticos como crustáceos (Senderovich *et al.* 2010).

O *V. cholerae* pode ser classificado dentro de sorogrupos determinados pela estrutura do antígeno O da membrana externa da bactéria. Atualmente são conhecidos mais de 200 sorogrupos (Chaudhuri e Chatterjee 2009; Aydanian *et al.* 2011). Até a década de 90, acreditava-se que apenas *V. cholerae* O1 causasse cólera, mas uma epidemia na Índia e em Bangladesh, causada por um novo sorogrupo (O139), numa população conhecida como imune a *V. cholerae* O1 (Ryan 2011), mudou este paradigma. Outros sorogrupos não-O1 de *V. cholerae* causam casos ou surtos de diarreia similar à cólera, mas não têm sido relacionados como determinantes de epidemias de cólera (Dalsgaard *et al.* 1995; Kaper *et al.* 1995; Chatterjee *et al.* 2009; Mohapatra *et al.* 2009).

Os *V. cholerae* do sorogrupo O1, relacionados a epidemias e pandemias, são caracterizado em dois biotipos: ‘clássico’ e ‘El Tor’, os quais são definidos por uma combinação de características bioquímicas e microbiológicas (Safa *et al.* 2010). As primeiras seis pandemias de cólera que aconteceram entre 1817 e 1923 foram, possivelmente, causadas pelo *Vibrio cholerae* O1 do biotipo clássico. Embora os

isolados do biotipo El Tor estejam relacionados a infecções esporádicas e surtos de cólera desde 1910, foi em 1961 que este biotipo emergiu, substituindo o biotipo clássico, como o determinante da sétima pandemia (Dziejman *et al.* 2002).

1.4 Virulência e patogenia

Os principais determinantes da virulência do *V. cholerae* são a toxina colérica (CT) e o TCP (*toxin co-regulated pilus*). Esta última é constituída de duas subunidades, A e B, e é a responsável pela diarreia secretória característica da cólera. A CT é uma enterotoxina que atua sobre os enterócitos através da ligação de cinco frações da subunidade B que formam um canal por onde a subunidade A, enzimaticamente ativa, é introduzida. A subunidade A deflagra uma cascata de eventos que resulta na ativação da adenilato ciclase, acumulação de cAMP e incremento na secreção de cloro e água. A liberação aumentada de água dentro do lúmen do intestino leva a diarreia secretória (Vanden Broeck *et al.* 2007). O “pilus” TCP é responsável pela adesão do vibrião ao epitélio intestinal e também pela auto-aglutinação dos vibriões, o que sinaliza a expressão da toxina pela bactéria.

1.5 Genoma e elementos genômicos do *Vibrio cholerae*

O genoma do *V. cholerae* compreende dois cromossomos (Figura 1). O sequenciamento completo de um isolado da linhagem El Tor (N16961) mostrou que o genoma contém 4.033.460 pares de bases (pb) distribuídos em dois cromossomos, um com 2.961.146 pb (cromossomo 1) e o segundo com 1.072.314 pb (cromossomo 2), com um conteúdo médio G+C de 46,9% e 47,7%, respectivamente. No total, para ambos os cromossomos foram preditas 3,885 ORFs (*open reading frames*¹) (Heidelberg *et al.* 2000).

¹ A fase de leitura aberta (*open reading frame* - ORF), é a sequência de DNA compreendida entre um códon de início da tradução e um códon de terminação. Existem seis possíveis fases de leitura aberta para qualquer molécula de DNA de cadeia dupla.

No cromossomo maior está localizada uma grande quantidade de genes reconhecidos como essenciais para as funções celulares (como replicação do DNA, transcrição, tradução e biossíntese da parede celular), patogenicidade (como toxinas, antígenos de superfície e adesinas) e oito cópias do operon ribossomal. Em contraste, o cromossomo menor contém uma fração maior (59%) de genes hipotéticos quando comparado com o cromossomo maior (42%), assim como vários genes que parecem ter origem em outras gama-proteobactérias. O cromossomo menor também carrega um sistema de captura de genes conhecido como superintegron ou integron cromossômico, de cerca de 150 Kpb (Heidelberg *et al.* 2000).

Atualmente existem 76 genomas de *V. cholerae* disponíveis no GenBank (setembro de 2012) dos quais 68 se encontram 'montados' em *contigs*² e oito são genomas completos (Figura 2). Destes 76 genomas, 35 são de isolados da linhagem da epidemia de cólera que emergiu no Haiti em 2010.

² *Contig* é uma sequência composta por um conjunto de *reads* (leituras) sobrepostas.

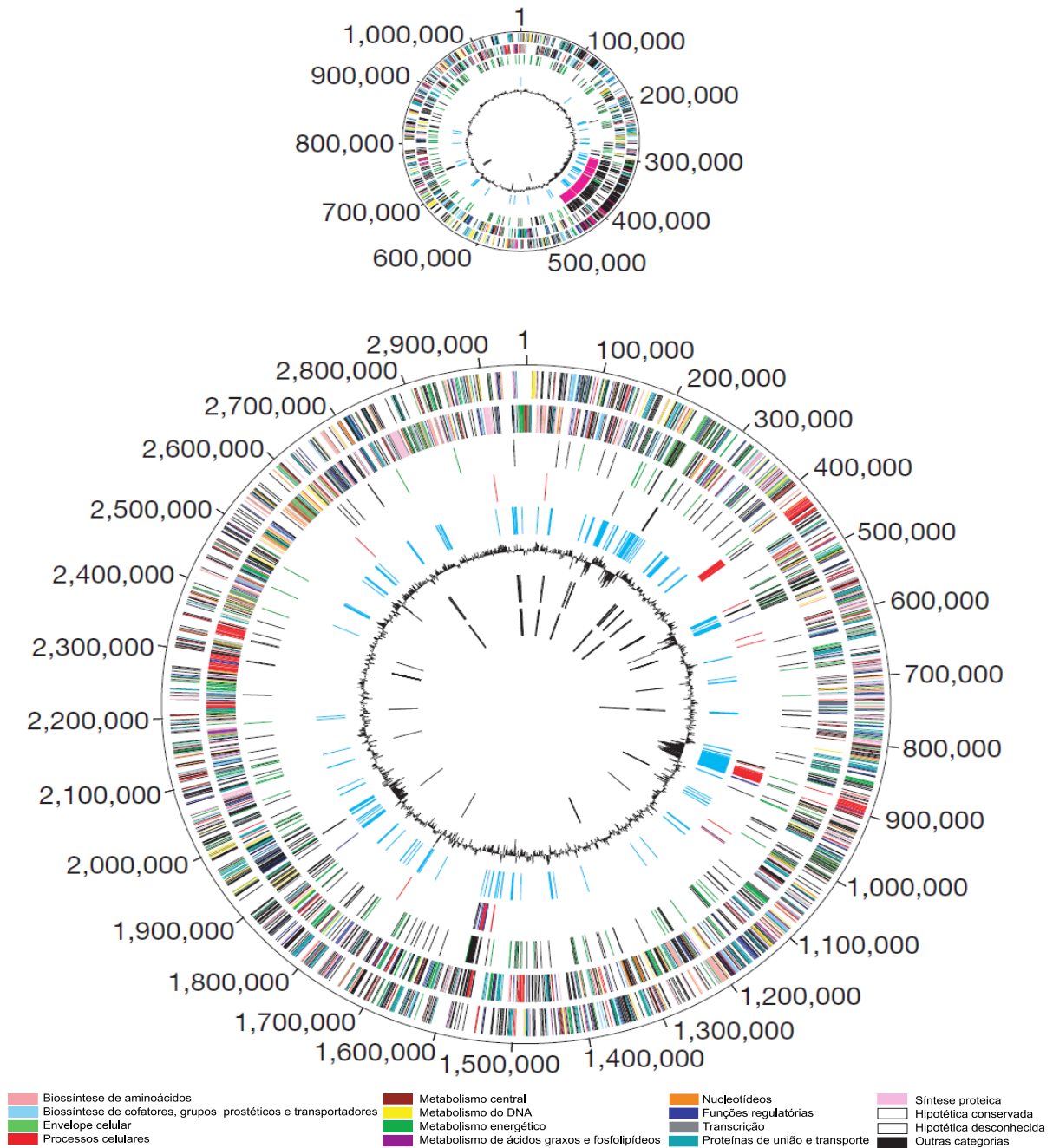


Figura 1. Genoma de *V. cholerae* composto pelo cromossomo maior (cromossomo 1) e o cromossomo menor (cromossomo 2). Descrevendo os círculos de fora para dentro: o primeiro e segundo círculo correspondem às regiões codificadoras de proteínas na fita positiva e negativa do DNA, representando sua função (segundo o código de cores mostrado na parte inferior). No terceiro círculo são mostrados genes duplicados no mesmo cromossomo (preto) e em diferentes cromossomos (verde). O quarto círculo mostra genes relacionados a transposons (preto), relacionados a fagos (azul), seqüências repetitivas (rosa) e patogênese (vermelho). O quinto círculo mostra regiões com composição de trinucleotídeos diferenciada do genoma. O sexto círculo mostra as porcentagens G+C. O sétimo e oitavo círculos são os tRNA e rRNAs, respectivamente. (Adaptado de (Heidelberg *et al.* 2000).

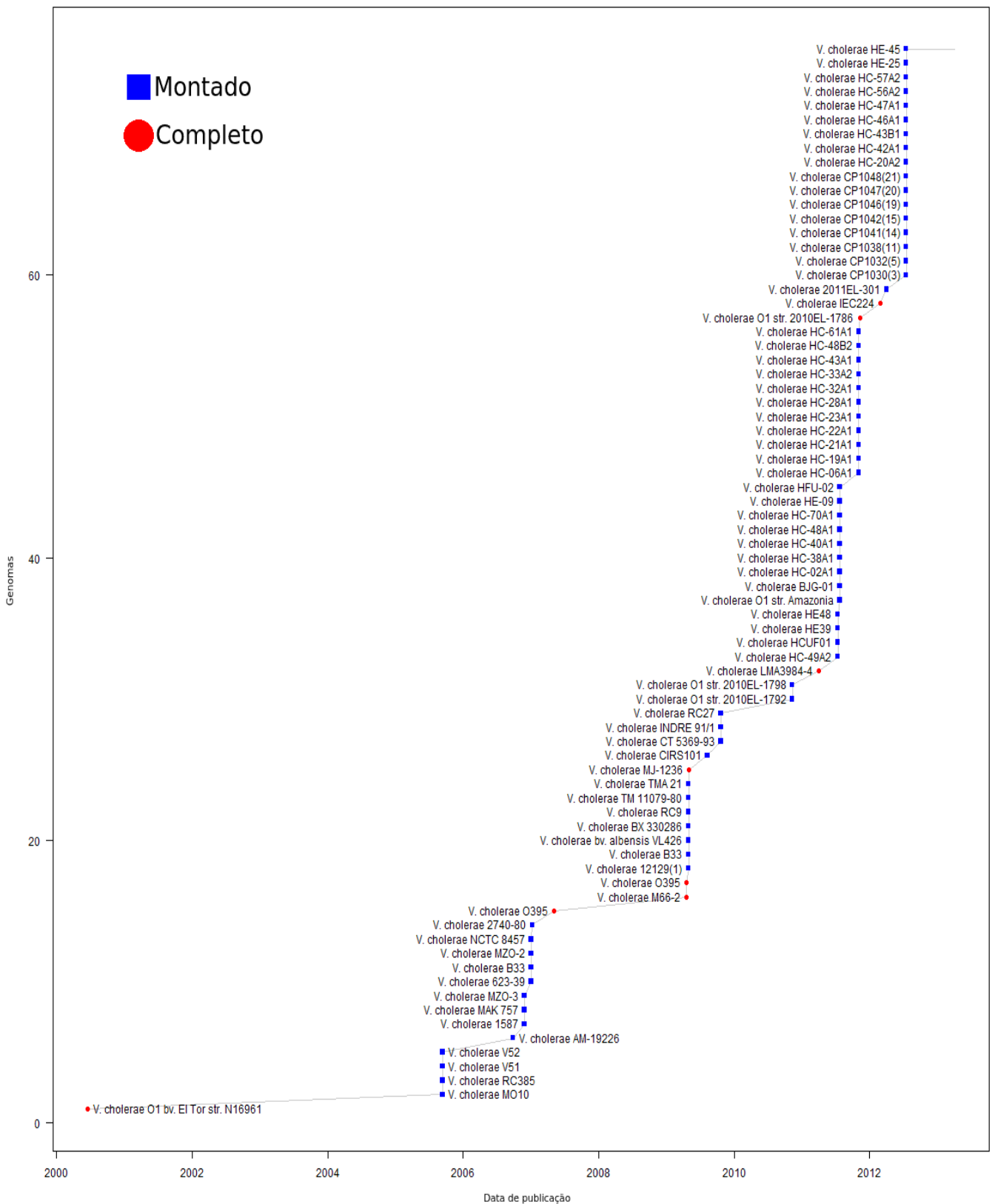


Figura 2. Genomas completos e montados em *contigs* disponíveis no GenBank (Informação disponível em 17 de setembro de 2012).

Um aspecto importante no estudo da genômica comparativa entre organismos é a definição de pangenoma, genoma *core* e genes únicos. O pangenoma descreve o repertório completo de genes identificados em isolados de uma espécie, que inclui o genoma *core*, o qual contém os genes compartilhados por isolados desta espécie (genes em comum), o genoma dispensável ou acessório, representado por genes presentes em dois ou mais isolados, e os genes únicos, exclusivos de um isolado (Medini *et al.* 2005). No caso de *V. cholerae*, Thompson e colaboradores (Thompson *et al.* 2009) determinam que o pangenoma de *V. cholerae* estaria formado por 6923 genes e o genoma *core* estaria formado por 1520 genes (Figura 3).

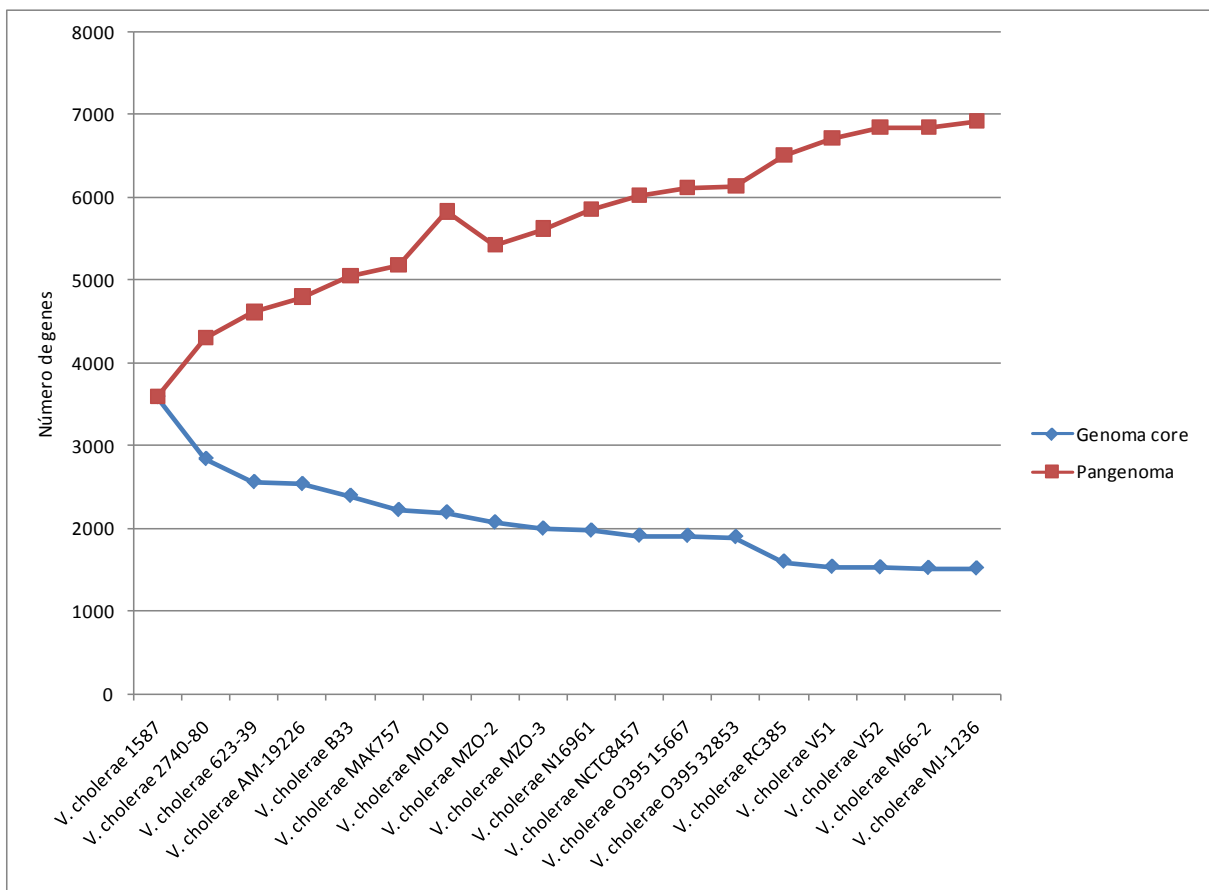


Figura 3. Pangenoma e genoma *core* de *V. cholerae*. Para o conjunto de genomas mostrados, o pangenoma foi calculado em 6923 genes e o genoma *core* em 1520 genes. Adaptado de (Thompson *et al.* 2009).

1.5.1 O bacteriófago CTX

O CTX é um bacteriófago filamentososo que se encontra em estado lisogênico, integrado ao genoma de linhagens toxigênicas de *V. cholerae*. Em contraste com outros fagos lisogênicos, tal como o bacteriófago λ , o CTX não entra em um ciclo lítico (Nickels 2009). Em lugar de isso, o CTX parece ter co-evoluído com seu hospedeiro (McLeod *et al.* 2005). Apesar do seu estado integrado estável no cromossomo bacteriano, eventualmente partículas virais são produzidas (Nickels 2009).

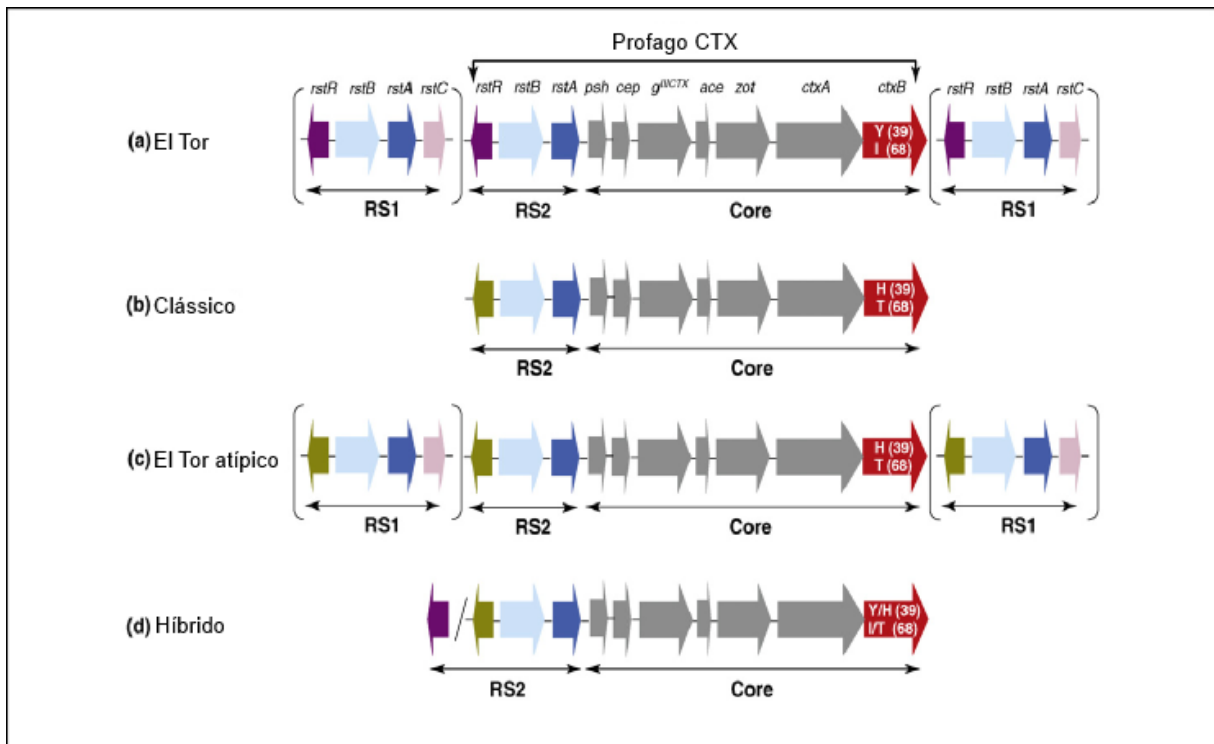


Figura 4. Estrutura do profago CTX e do fago satélite RS1 em diferentes variantes de *V. cholerae* O1. No canônico El Tor (a) e nos El Tor atípicos (c), os parênteses delineiam os elementos RS1 que flanqueiam o genoma do profago CTX integrado. O profago CTX de isolados do biótipo clássico é mostrado em (b). Em (d) é mostrado um profago híbrido hipotético. As posições polimórficas no *ctxB* estão indicadas para H (histidina), T (treonina), Y (tirosina) ou I (isoleucina). Os diferentes alelos $rstR^{El\ Tor}$ e $rstR^{clássico}$ são mostrados em diferentes cores. Adaptado de Safa e colaboradores (Safa *et al.* 2010)

O genoma do profago CTX, presente em linhagens de *V. cholerae* toxigênicas, consiste em dois grupos de genes: os do genoma *core* e os da região RS2 (Figura 4) (Waldor e Mekalanos 1996). Os genes do *core* incluem os genes *ctxAB* codificadores da toxina colérica, *ace* e *zot* – que codificam toxinas, *cep* - fator de colonização acessório e outros cinco genes codificadores de proteínas requeridas para a formação do fago. A região RS2 codifica proteínas que têm papel na replicação (RstA), integração (RstB) e na regulação (RstR) da recombinação sítio – específica do profago CTX no cromossomo de *V. cholerae*. RS1 é similar à RS2 exceto que esta última contém um gene *rstC* adicional, codificador de uma proteína anti-repressora, a qual promove a transcrição de genes do profago CTX que são requeridos para a geração de partículas infecciosas (Safa *et al.* 2010).

Embora o gene da toxina colérica CT seja conservado entre isolados do biotipo clássico e El Tor, existem diferenças sutis na seqüência do *ctxB* que são marcadores para a genotipagem dos biotipos, principalmente para fins epidemiológicos. Assim, o gene *ctxB* presente em isolados de *V. cholerae* do biotipo clássico possuem o genótipo 1, aqueles do biotipo El Tor e também alguns isolados da Austrália possuem o genótipo 2, e os El Tor do início da sétima pandemia apresentam o genótipo 3. Contudo, nos últimos anos novos genótipos do *ctxB* tem sido identificados, um dos quais (genótipo 7) estaria sendo o prevalente nas epidemias atuais de cólera, em uma linhagem nomeada como El Tor alterado (Safa *et al.* 2010). Até agora 11 genótipos do *ctxB* já foram identificados (Marin e Vicente 2012), sendo que seis deles em *V. cholerae* O1. Portanto, a identificação do genótipo *ctxB* que um isolado apresenta é mais um dos marcadores utilizados nos estudos de epidemiologia e evolução de *V. cholerae*.

1.5.2 Ilhas de patogenicidade em *V. cholerae* (VPI-1 e VPI-2)

A ilha VPI-1 (*Vibrio Pathogenicity Island 1*), associada a linhagens de *V. cholerae* epidêmicas e pandêmicas, é uma região de 41kb que codifica o pilus corregulado com a toxina CT (*toxin-coregulated pilus* - TCP), o fator de colonização acessório (*accessory colonization factor* - ACF) e os reguladores da virulência ToxT e TcpH e outros genes de função desconhecida. O TCP é um fator de colonização essencial de *V. cholerae* no epitélio intestinal humano, além de ser o receptor do bacteriófago CTX (Karaolis *et al.* 1998). O TCP, junto com a toxina colérica,

constituem os dois principais fatores de virulência em *V. cholerae* (Kaper *et al.* 1995).

Semelhante ao gene *ctxB*, o gene *tcpA* apresenta vários alelos que estão associados aos biotipos clássico e El Tor, os quais possuem alelos próprios, e outros alelos presentes em linhagens de *V. cholerae* não-O1 (Novais *et al.* 1999; Goel *et al.* 2011; Kumar *et al.* 2011; Talkington *et al.* 2011).

A ilha VPI-2 (*Vibrio Pathogenicity Island 2*) é uma região de 57,3 Kb que codifica genes para o metabolismo do ácido siálico e neuraminidase, além de possuir características de uma ilha de patogenicidade. Esta ilha se encontra predominantemente em cepas patogênicas de *V. cholerae* O1 e O139 (Murphy e Boyd 2008).

1.5.3 VSP-I e VSP-II

As ilhas VSP-I e VSP-II (*Vibrio Seventh Pandemic*) são ilhas genômicas que foram identificadas exclusivamente em linhagens de *V. cholerae* El Tor por análises de microarranjos (Dziejman *et al.* 2002); os genes que se encontram nestas ilhas têm sido associados com características únicas da linhagem El Tor da sétima pandemia (O'Shea *et al.* 2004a; Nusrin *et al.* 2009).

A ilha VSP-II presente no genoma da amostra N16961 compreende uma região de 27 Kb (VC0490-VC0516) integrada num sítio de tRNA-metionina, a qual codifica homólogos de RNaseH1, um pilus do tipo IV, uma proteína quimiotáctica e uma integrase (Murphy e Boyd 2008).

Murphy e Boyd (2008) demonstraram que VPI-2, VSP-I e VSP-II, podem ser liberadas como moléculas circulares extra-cromossômicas. Este é o primeiro passo para a transferência horizontal destas ilhas de patogenicidade (Murphy e Boyd 2008).

1.5.4 O superintegron

O superintegron (SI) é um extenso sistema de captura e excisão de genes de aproximadamente 120 Kb situados no cromossomo menor (Figura 5) (Heidelberg *et al.* 2000), caracterizado por conter 215 ORFs (a maioria delas não identificadas) em um arranjo de 179 cassetes gênicos que começam no gene codificador da integrase

(*Int14*) seguido do sítio de recombinação *attI* e um promotor Pc. Após esta região, que é conservada, está o arranjo de cassetes gênicos de composição variável e dinâmica. Dentro do SI, os cassetes gênicos geralmente consistem de uma ORF flanqueada por dois sítios de recombinação chamados *V. cholerae repeats* (VCRs) (Mazel 2006). Os cassetes podem ser clivados de qualquer posição do arranjo através da recombinação VCR x VCR mediada pela integrase. O intermediário circular resultante pode ser integrado pela integrase preferencialmente através da recombinação *attI* x VCR deixando o cassete sob o controle do Pc (Rowe-Magnus *et al.* 1999; Cambray *et al.* 2010). Devido aos cassetes gênicos serem usualmente desprovidos de promotor, somente os primeiros cassetes seriam expressos pelo Pc e o resto do arranjo atuaria como um reservatório de genes (Cambray *et al.* 2010).

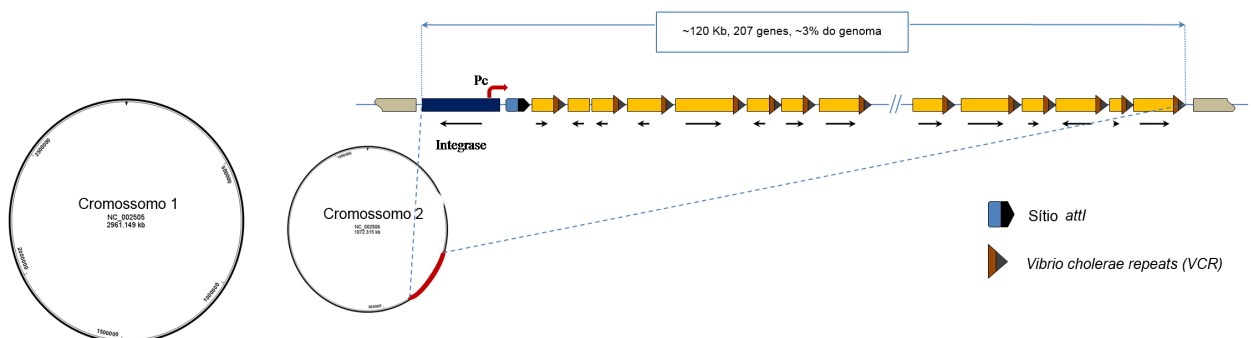


Figura 5. Organização esquemática do genoma de *V. cholerae* e a localização do superintegron (SI). A plataforma funcional do SI consiste de um gene da integrase, um promotor (Pc) e um sítio de recombinação (*attI*). A plataforma está composta de um arranjo de cassetes os quais geralmente consistem de uma ORF desprovida de promotor flanqueada por dois sítios de recombinação chamados VCR.

Um estudo filogeográfico considerando cepas de *V. cholerae* e sua espécie irmã *Vibrio metecus*, mostra que o conteúdo do SI apresenta uma forte relação com a origem geográfica destes Vibrios. Assim, baseados no conteúdo do SI, isolados de *V. cholerae* e *V. metecus* de uma mesma região geográfica se agrupam enquanto que isolados de *V. cholerae* de diferentes regiões geográficas não compartilham os mesmos conteúdos (Boucher *et al.* 2011). Isto sugere que a composição do SI é influenciada por interações com o ambiente.

1.5.5 Os elementos integrativos e conjugativos

Os elementos integrativos e conjugativos (*Integrative and Conjugative Element – ICE*) da família dos SXT/R391 são elementos móveis auto-transmissíveis envolvidos, entre outros aspectos, na dispersão de genes de resistência a antibióticos entre gama-proteobactérias, incluindo *V. cholerae*. O elemento SXT foi identificado pela primeira vez num isolado clínico de *V. cholerae* O139 em 1992; este elemento carrega genes de resistência para sulfametoxazol, trimetoprima e estreptomicina (Waldor *et al.* 1996).

Atualmente são conhecidos mais de 50 elementos relacionados a ICEs pertencentes a família SXT/R391, os quais tem entre 79 a 108 Kb e se caracterizam por seu gene da integrase, *int*_{SXT}, altamente conservado (Wozniak *et al.* 2009). A integrase permite a integração do ICE no extremo 5' do gene *prfC*. Este elemento é também caracterizado por apresentar uma região *core* de 52 genes presentes em todos os ICE SXT/R391 a qual é requerida para a excisão/integração, a transferência conjugativa e a regulação deste elemento. Outras regiões características dos ICEs são a região variável III onde se encontram os genes de resistência para sulfametoxazol (*sul2*), trimetoprima (*dfrA1*) e estreptomicina (*strAB*). Este elemento também apresenta cinco regiões conjugativas conhecidas como *hotspots*³ (HS) para a integração de DNA exógeno o qual dá características específicas a seu hospedeiro bacteriano (Figura 6)(Burrus *et al.* 2006; Wozniak *et al.* 2009).

³ Os *hotspots* de recombinação homóloga são pequenas regiões do genoma onde duas sequências de DNA que compartilham um trecho de identidade são unidas, numa reação catalisada por uma enzima, o que possibilita depois uma troca de fitas, gerando uma nova fusão das duas fitas.

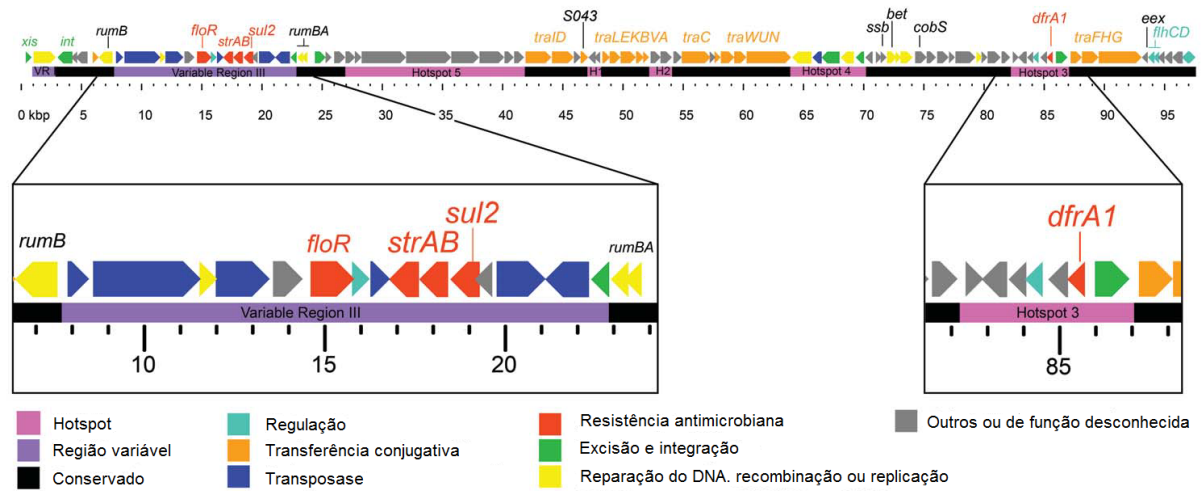


Figura 6. Organização genética do elemento integrativo e conjugativo (ICE) de um isolado do Haiti de 2010. Este ICE contém 97,915 pb e 95 ORFs. Regiões conservadas entre sequências de ICEs previamente sequenciados são mostradas em preto, regiões variáveis em roxo, e *hotspots* de recombinação homóloga previamente identificados em rosa. Genes conservados envolvidos em conjugação estão indicados em laranja. Genes associados com resistência a antimicrobianos (*floR* [cloranfenicol], *strAB* [estreptomicina], *sul2* [sulfametoxazol], e *dfrA1* [trimetoprima]) estão indicados em vermelho. Adaptado de Sjölund-Karlsson e colaboradores (Sjölund-Karlsson *et al.* 2011).

1.6 Epidemiologia da sétima pandemia de cólera

A sétima pandemia de cólera tem várias características importantes que a diferenciam das demais. Esta pandemia começou na ilha Célebes em 1961 onde, até hoje, se mantém. É também a mais extensa considerando tempo e espaço. Além disso, é determinada por *V. cholerae* O1 do biotipo El Tor e não mais do clássico, como as anteriores. O biotipo El Tor possui várias características que lhe conferem um alto grau de virulência epidêmica, permitindo que ele se espalhe pelo mundo. Por produzir menos toxina CT e persistir melhor no ambiente do que linhagens do biotipo clássico, se mantém em um número maior de portadores assintomáticos, por longos períodos e por extensas áreas (Kaper *et al.* 1995; Chaudhuri e Chatterjee 2009). Nos anos anteriores ao início da sétima pandemia, as epidemias de cólera estavam praticamente restritas ao subcontinente asiático (Figura 7).

A cólera atualmente é endêmica em aproximadamente 50 países no mundo, infectando 3-5 milhões de indivíduos e matando aproximadamente 100.000 pessoas/ano (Ryan 2011).

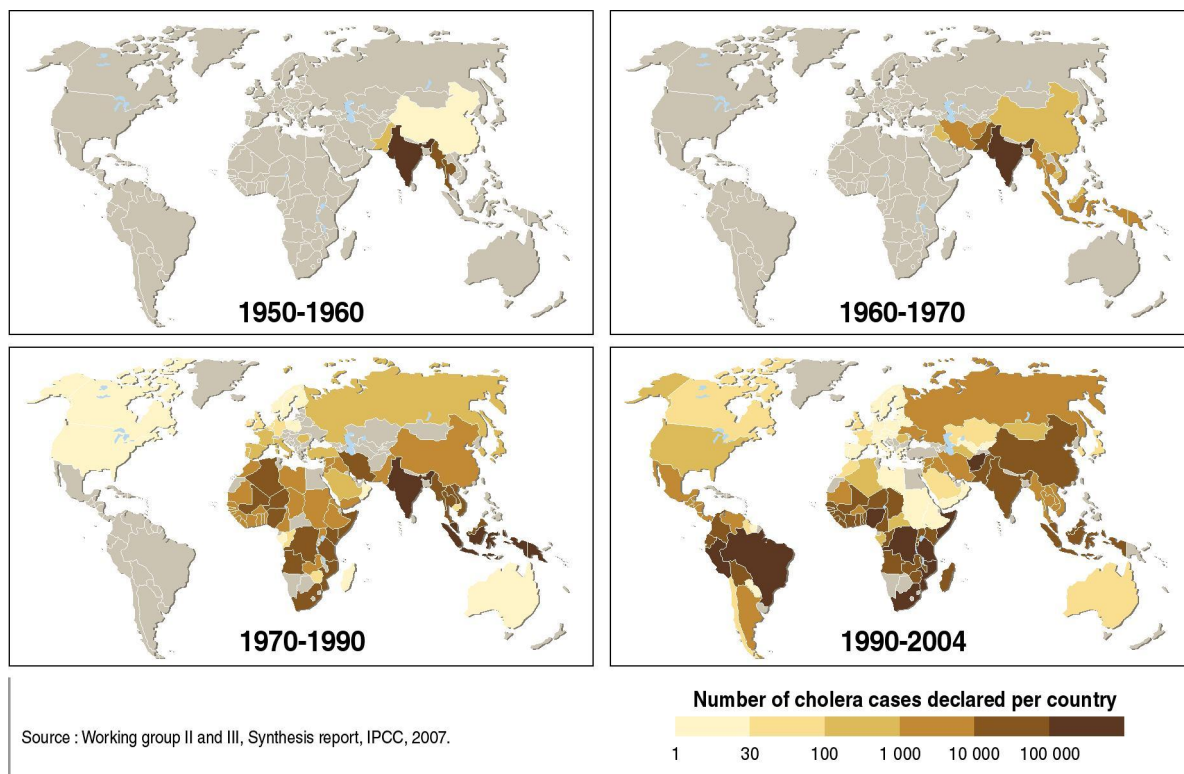


Figura 7. Dispersão e número de casos de cólera detectados no mundo no período de 1950-2004. Fonte: http://www.grida.no/graphicslib/detail/the-spread-of-cholera-1950-2004_1471#.

Nos início dos anos 90, epidemias de cólera surgiram primeiro na África e depois, no final de janeiro de 1991, no Peru. A cólera não ocorria de forma epidêmica na América do Sul e Central desde 1895 (Reeves e Lan 1998).

1.6.1 A cólera na América Latina

Durante os anos 80, a prevalência da cólera no mundo foi relativamente baixa, envolvendo principalmente países da Ásia e África. Já nos anos 90 epidemias importantes aconteceram na África e depois de mais de um século, a cólera e a sétima pandemia chegaram à América Latina, inicialmente no Peru, em 1991, quando mais de 340.000 casos aconteceram. De 1991 a 1994, mais de um milhão

de casos e 9600 óbitos ocorreram no continente (Figura 8)(Gil *et al.* 2004). Presumiu-se que a origem da epidemia teria sido a ingestão de frutos do mar contaminados pela água do lastro de navios carregados na Ásia e aportados no Peru (Seas *et al.* 2000; Tortora *et al.* 2004).

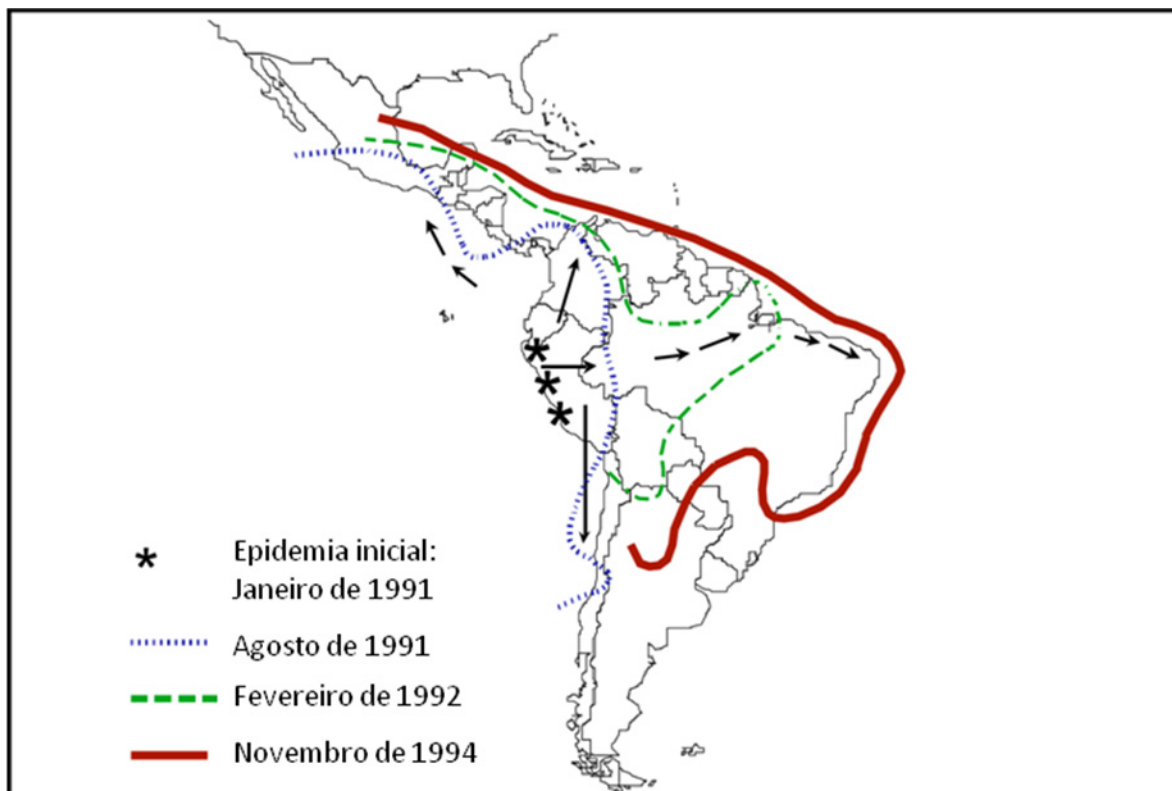


Figura 8. Dispersão de *Vibrio cholerae* O1 na América do Sul e Central, Janeiro 1991–Novembro 1994. Fonte: MMWR 44(11) (24-03-95).

Nos primeiros três anos da epidemia, a cólera alcançou quase todos os países da América Latina, com exceção do Uruguai e das ilhas do Caribe. A epidemia foi diminuindo progressivamente até 1998, quando se registrou um aumento significativo de casos e mortes (Harvez e Avila 2013).

A cólera atingiu o Brasil pela tríplice fronteira Brasil/Peru/Colômbia na região amazônica. Espalhou-se acompanhando as rotas dos barcos pelo sistema dos rios Solimões e depois Amazonas. Logo depois atingiu a região nordeste onde, devido à precariedade dos sistemas de água e esgoto, permaneceu por alguns anos. Surtos decorrentes do deslocamento humano foram registrados nas regiões sudeste e sul.

A linhagem de *V. cholerae* determinante desta epidemia foi identificada, com base nas técnicas e marcadores utilizados à época, como do biotipo El Tor, mas apresentando um novo alelo para a enzima LAP (Popovic *et al.* 1993; Freitas *et al.* 2002).

No decorrer da epidemia, observou-se o surgimento de um variante de *V. cholerae* O1 epidêmico defectivo na fermentação de sacarose. A fermentação de sacarose, que é visualizada no meio seletivo (TCBS) utilizado no isolamento e identificação de *V. cholerae*, é um dos marcadores fenotípicos mais amplamente aplicados no diagnóstico da cólera. Esta variante foi responsável por casos desde a Guiana Francesa até a região Centro Oeste do Brasil, passando pelo Pará. Só recentemente, com o sequenciamento completo do genoma desta variante, é que foi identificada a inserção de um único nucleotídeo no gene codificador da proteína transportadora de sacarose (De Sá Morais *et al.* 2012; Garza *et al.* 2012).

Durante o período de 1991-2011, o maior número de casos de cólera notificados na América Latina ocorreu durante os primeiros dois anos, com 391.733 casos em 1991 e 383.457 em 1992 (Figura 9). Nos cinco anos seguintes registrou-se uma diminuição drástica de casos, que caíram mais de cem vezes (17.764 casos em 1997), tendência que foi revertida em 1998, quando houve um aumento superior a 300%. Entre 1999-2009 as notificações de casos na América Latina caíram até desaparecer.

Em outubro de 2010, contudo, a cólera emergiu no Haiti, na região de Artibonite e pouco depois no resto do país. Em novembro desse ano chegaram as primeiras notificações no país vizinho, República Dominicana (Andrews e Basu 2011; Sack 2011). Estes países não tinham sido atingidos pela epidemia da América Latina de 1991. Assim, no ano 2010 são registrados 179.571 casos e no ano seguinte o número dobra, chegando a 361.214 casos notificados (Figura 9) (Harvez e Avila 2013). Esta epidemia, conhecida como Hispaniola, gera uma corrida mundial, envolvendo epidemiologistas e geneticistas, para a determinação de sua origem. Evidências epidemiológicas atribuem a introdução da infecção a soldados nepaleses que vieram para socorrer o país assolado pelo terrível terremoto de 2010 (Piarroux *et al.* 2011). Um trabalho de epidemiologia genômica com o sequenciamento do genoma completo de três isolados do Haiti mostram sua relação com isolados do

sudoeste asiático, o que afastou por completo a possibilidade de uma relação com a epidemia da América Latina (Chin *et al.* 2011).

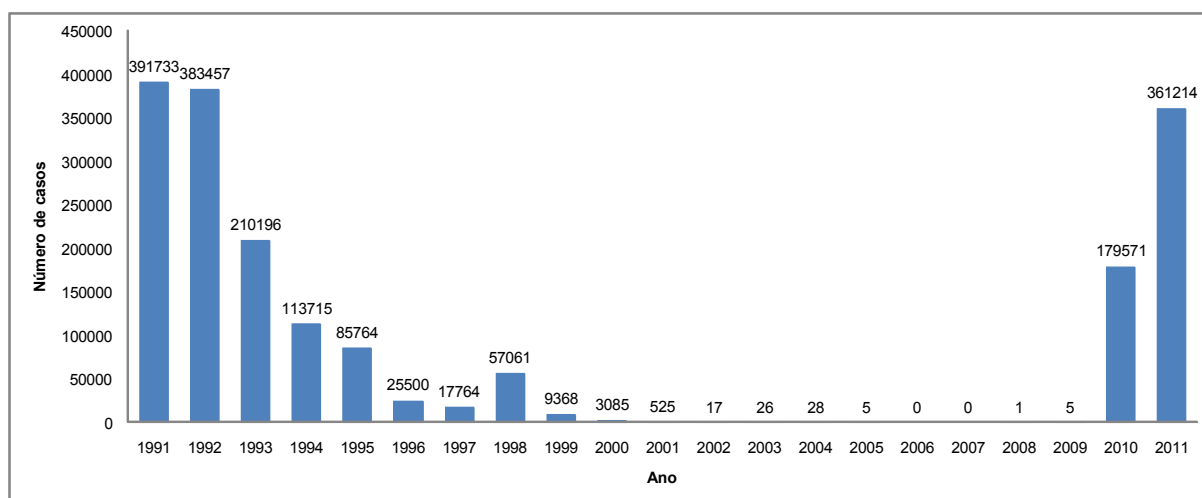


Figura 9. Número de casos por ano de cólera na América Latina, 1991-2011. Adaptado de Harvez e Avila (Harvez e Avila 2013).

Em uma recente publicação, pesquisadores estudaram a epidemiologia genômica da sétima e atual epidemia de cólera, a caracterizando por três sucessivas ondas, determinadas por distintas linhagens de *V. cholerae* do biotipo El Tor (Mutreja *et al.* 2011). Os autores mostram que os isolados da América Latina formam um agrupamento, juntamente com um único isolado de Angola, que é nomeado WASA (*West African-South American*), sugerindo a África como a origem da epidemia da América Latina (Figura 10). Os isolados da América Latina são diferenciados dos demais, de outras latitudes geográficas, por conter uma ilha genômica chamada de WASA1, recentemente caracterizada como um profago (Garza *et al.* 2012) e por apresentar a ausência de vários genes da ilha genômica VSP-II, que caracteriza as linhagens da sétima pandemia (Nusrin *et al.* 2009).

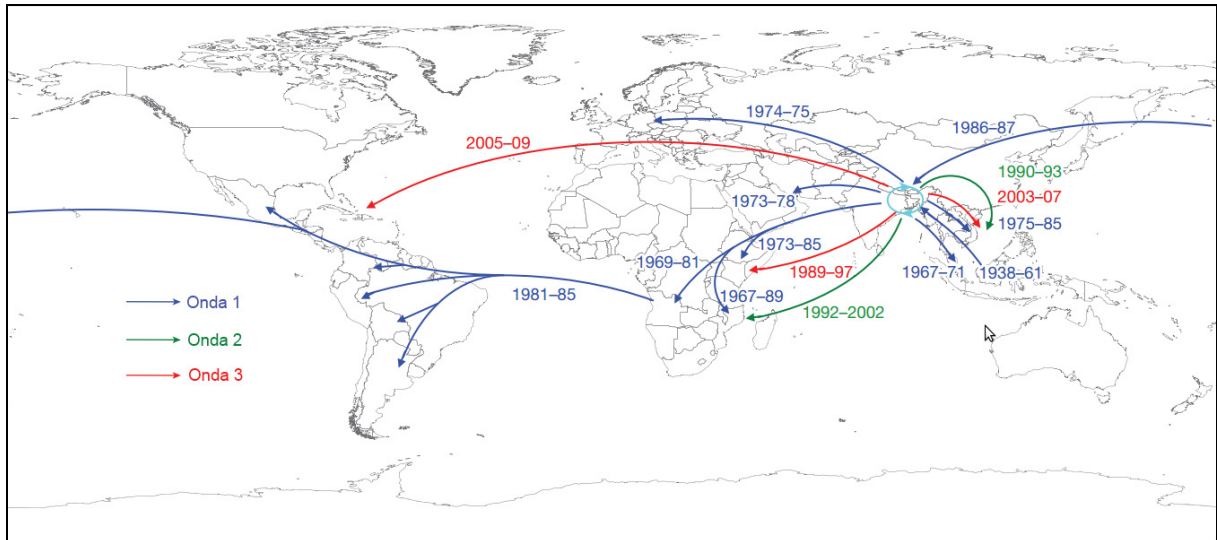


Figura 10. Eventos de transmissão de cólera correspondente a sétima pandemia como inferido por análises filogenéticas (Mutreja *et al.* 2011).

1.6.2 Situação atual

Variantes da linhagem El Tor tem surgido e, de acordo com a frequência de isolamento de casos clínicos de cólera, acredita-se que tenham substituído o protótipo da sétima pandemia na Ásia e na África (Safa 2008; Piarroux e Faucher 2012).

V. cholerae recentemente isolados de surtos ou epidemias de cólera no mundo tem sido identificados como híbridos, alterados ou variantes. Estes apresentam o fenótipo do El Tor, mas contém alelos de *ctxB* diferentes do El Tor canônico. Muitos deles apresentam perfil fenotípico de multi-resistência a antibióticos devido a mutações em genes *housekeeping*⁴ ou a aquisição de elementos integrativos e conjugativos (ICEs) (Mutreja *et al.* 2011).

Estas variantes de *V. cholerae* tem sido relacionados a casos clínicos mais severos, além de evidenciar a substituição do El Tor tradicionalmente conhecido em várias regiões geográficas (Chin *et al.* 2011; Mutreja *et al.* 2011; Ryan 2011; Piarroux e Faucher 2012). A prevalência desta linhagem emergente poderia explicar

⁴ Genes *housekeeping* são genes constitutivos que são requeridos para a manutenção das funções celulares básicas.

a maior taxa de fatalidade (de 1-5%) em recentes surtos, em comparação à taxa histórica de <1% (Ryan 2011).

Embora isolados de *V. cholerae* do sorogrupo O1 sejam predominantes entre isolados epidêmicos, tem sido registradas outras linhagens clínicas também virulentas, como o *V. cholerae* O1 Amazônia, que apresenta características genótípicas diferentes (Coelho *et al.* 1995; Thompson *et al.* 2008), e também isolados não O1/não O139 (Chen *et al.* 2007; Kondo e Ajawatanawong 2009).

Mais de 10 anos transcorreram desde que o primeiro genoma de um *V. cholerae* do biotipo El Tor foi sequenciado mediante a tecnologia de sequenciamento de Sanger (Heidelberg *et al.* 2000); na época o projeto demorou três anos até o genoma ser completamente sequenciado. Em 2010, depois do início da epidemia no Haiti, dois genomas completos de *V. cholerae* responsáveis pela epidemia foram sequenciados em menos de 24 horas através da tecnologia PacBio (Chin *et al.* 2011). Mais recentemente, 136 genomas de *V. cholerae* foram sequenciados mediante a tecnologia Illumina (Mutreja *et al.* 2011). Este crescimento vertiginoso da disponibilidade de novos genomas completos e o surgimento de novas tecnologias de sequenciamento em alta escala (*high-throughput sequencing*⁵) está mudando o paradigma de como estudar as doenças infecciosas e sua epidemiologia. Cada vez é mais importante obter sequências suficientes e de qualidade destas novas tecnologias, que permitam estabelecer de forma robusta as associações entre patógenos, assim como estudar características genéticas ou elementos genômicos característicos vinculados a determinados surtos ou epidemias.

⁵ O sequenciamento em alta escala (*high-throughput sequencing* ou *next-generation sequencing*) é uma metodologia que se baseia em paralelizar o processo de sequenciamento, produzindo milhares ou milhões de sequências de uma vez só.

2 JUSTIFICATIVA

A epidemia de cólera ocasionada por *V. cholerae* na América Latina, considerada como uma extensão da sétima pandemia, oferece uma oportunidade para a compreensão da natureza e do surgimento das epidemias de cólera, assim como da biologia e evolução do patógeno. *V. cholerae* e cólera são temas amplamente estudados e ainda assim várias questões permanecem em aberto. Uma delas é a epidemia da América Latina: enquanto alguns trabalhos apontam para sua origem na água do lastro de navios vindos da Ásia e aportados no Peru, outros sugerem que eventos climáticos como *El Niño*⁶ teriam provocado o reaparecimento de *V. cholerae* do ambiente (Colwell 1996, 2004; Constantin de Magny *et al.* 2008). Contudo, um recente estudo a relacionou, pela primeira vez, com um isolado de *V. cholerae* de Angola, no oeste da África (Mutreja *et al.* 2011).

Quando iniciamos nosso projeto, no ano de 2009, os estudos sobre a epidemia de cólera da América Latina tinham abrangência limitada a um único país e os resultados, na maioria dos casos, não permitiam extrapolações mais amplas para o contexto geral da sétima pandemia de cólera. Portanto, nosso objetivo inicial foi conhecer a diversidade genômica considerando tanto o genoma estável, aplicando MLSA (*Multilocus Sequence Analysis*), quanto regiões do genoma “flexível” e determinar as relações filogenéticas de isolados de *V. cholerae* obtidos antes, durante e depois da epidemia que atingiu a América do Sul. Naquela ocasião não havia nenhuma informação completa do genoma de *V. cholerae* da epidemia de cólera da América Latina. No ano de 2010, Lam e colaboradores (Lam *et al.* 2010) publicam estudo sobre a evolução da sétima pandemia e a origem da epidemia na América Latina dos anos 1990s. Eles analisam trinta polimorfismos de nucleotídeo único (*single-nucleotide polymorphisms – SNPs*) para rastrear a propagação da sétima pandemia, e assim é estabelecida a relação de isolados da epidemia de 1991 da América Latina com isolados de 1970s da África. Em 2011, Mutreja e

⁶ O *El Niño* é um fenômeno climático, de caráter atmosférico-oceânico, em que ocorre o aquecimento fora do normal das águas superficiais e sub-superficiais do Oceano Pacífico Equatorial. Este fenômeno está associado ao incremento da população de *V. cholerae* e emergência de surtos de cólera em regiões endêmicas.

colaboradores (Mutreja *et al.* 2011), fazendo uma reconstrução filogenética baseada em genomas completos, mostram a relação dos isolados da América Latina com um isolado de 1989 da Angola, sugerindo, mais uma vez, que a linhagem da América Latina teria uma relação direta com a África. Assim, ampliamos nossas análises para considerar também isolados epidêmicos de *V. cholerae* do oeste da África (Gana e Nigéria), almejando obter mais evidências sobre a ligação América Latina-África.

Estes dois estudos foram os mais importantes e representativos sobre a epidemiologia da cólera na América Latina. Contudo, o estudo de Lam e colaboradores analisou um número limitado de polimorfismos (30 SNPs); por sua vez, no trabalho de Mutreja e colaboradores não foram estudados genomas do Brasil. Além disso, a ligação da epidemia de cólera da América Latina com a África foi inferida a partir de um único isolado da Angola, o que é ainda uma evidência pouco robusta.

No ano 2012, é publicado o primeiro genoma de *V. cholerae* da epidemia no Brasil (De Sá Morais *et al.* 2012). Este *V. cholerae* apresentava a particularidade de ser defectivo na fermentação de sacarose, que é um dos marcadores fenotípicos mais amplamente considerados na identificação de *V. cholerae*. Logo depois, outro trabalho focando em duas particularidades deste genoma, mostra que a presença de um SNP é o determinante do fenótipo defectivo para a fermentação de sacarose e que neste genoma está também presente o profago WASA1 (Garza *et al.* 2012), que seria único e característico da linhagem da América Latina, como sugerido previamente (Mutreja *et al.* 2011).

Um surto ou uma epidemia são caracterizados pela incidência, em curto período, de grande número de casos de uma doença. As causas deste fenômeno estão relacionadas principalmente à patogenicidade e virulência do patógeno. Com a epidemiologia genômica pretende-se buscar associações não-aleatórias entre as variáveis respostas (características do surto ou epidemia) e potenciais variáveis explicativas numa doença (características genéticas ou genômicas dos patógenos envolvidos). Portanto, o presente trabalho teve como finalidade estabelecer a epidemiologia genômica de *V. cholerae* da América Latina baseado na caracterização genética e informação genômica dos isolados obtidos desta epidemia.

Objetivos

3 OBJETIVOS

3.1 Objetivo Geral

Estabelecer a epidemiologia genômica de *V. cholerae* na América Latina.

3.2 Objetivos Específicos

1. Caracterizar a linhagem de *V. cholerae* da América Latina.
2. Caracterizar as linhagens de *V. cholerae* do oeste da África.
3. Identificar genótipos de *ctxB* em *V. cholerae*.
4. Determinar a estrutura genômica do superintegron e relacioná-la as linhagens de *V. cholerae*.
5. Caracterizar marcadores genéticos das linhagens de *V. cholerae* da sétima pandemia.
6. Realizar genômica comparativa de *V. cholerae* da América Latina.

Metodologia

4 METODOLOGIA

Com o propósito de enriquecer nosso conhecimento sob isolados de surtos ou epidemias de cólera que, por sua vez, ajudasse a uma melhor caracterização dos isolados da América Latina, foram seguidas diferentes abordagens. Primeiro foi feita uma caracterização dos genótipos do *ctxB*, gene importante na patologia e epidemiologia da cólera. Depois, devido à possível relação da cólera da América Latina com isolados da África, nosso segundo objetivo foi caracterizar isolados de *V. cholerae* de países da África ocidental, os quais poderiam representar marcos de referência para nosso estudo. Estudar a cólera no ocidente da África representa um desafio; diferentemente de outros países da África, o lado ocidental é o menos acessível para a disponibilidade de amostras, o que se reflete nas poucas publicações sobre cólera desta região. Tivemos acesso a isolados de surtos de cólera de Gana e Nigéria, sendo estes isolados caracterizados geneticamente.

Do ponto de vista de composição genômica, foi analisada a região do superintegron de *V. cholerae*, sua estrutura e composição, assim como a distribuição do seu conteúdo gênico e sua relação com as diferentes linhagens de *V. cholerae*.

Outro ponto de referência para nossas análises foram os genomas de *V. cholerae* da recente epidemia no Haiti; eles foram incluídos nas nossas análises, mas também foi caracterizado um elemento integrativo e conjugativo (ICE) cuja associação e importância nas epidemias atuais da cólera no mundo foram estudadas.

A metodologia destes estudos está descrita nos artigos anexados no final do texto (ver Anexos). A presente seção descreve e explica a metodologia usada no estudo da epidemiologia genômica da América Latina.

4.1 Trabalho experimental

4.1.1 Isolados de *V. cholerae*

Os 188 isolados de *V. cholerae* pertencem a Coleção de Pesquisa em Vibrio do Laboratório de Genética Molecular de Microorganismos (LGMM). Todos os

isolados foram reativados em caldo BHI (*Brain Heart Infusion*), e para posterior cultivo em TCBS (*Tiosulfato citrato bile sacarose*) e TSA (*Agar tripton de soja*) com 1.5% adicional de NaCl e incubados à 37 °C de 18 a 24 horas.

4.1.2 Extração de DNA, amplificação e sequenciamento

A extração do DNA genômico, amplificação por PCR (*Polymerase Chain Reaction* – Reação em Cadeia da Polimerase) e sequenciamento foram realizadas segundo Thompson *et al* (2008). Os produtos de PCR foram purificados utilizando kit comercial (*GFX PCR DNA and Gel Band Purification Kit* – *GE Healthcare*) para o posterior sequenciamento segundo as recomendações do kit de sequenciamento *ABI Prism BigDye terminator cycle sequencing ready reaction mix* (Applied Biosystem). O sequenciamento foi realizado no sequenciador *ABI PRISM 3730 Genetic Analyzer* (Applied Biosystem) da plataforma de sequenciamento do Instituto Oswaldo Cruz (IOC).

4.1.3 Análise de sequências multilocus (MLSA - Multilocus Sequence Analysis)

Com base na literatura (Byun *et al.* 1999; Mohapatra *et al.* 2009; Thompson *et al.* 2009; Pascual *et al.* 2010; Boucher *et al.* 2011), foram selecionados os genes: *recA* (*recombination/repair protein*), *pyrH* (*uridylate kinase*), *mdh* (*malate dehydrogenase*). A relação dos iniciadores para estes genes e outros utilizados no presente estudo está no Anexo II. O alinhamento das seqüências concatenadas foram realizadas usando o programa MAFFT v6.850b (Kato e Toh 2008).

Um conjunto com 355 seqüências pertencentes aos genes *recA*, *pyrH* ou *mdh* obtidas de *V. cholerae* da coleção de *Vibrios* do Laboratório de Genética Molecular de Microrganismos e de seqüências de bancos de dados públicos (BD) foi considerado para a MLSA (Tabela 2).

Tabela 2. Número e origem dos *V. cholerae* da América Latina analisados por MLSA.

	Clínicos	Ambientais	Desconhecidos
	Isolados (genomas)	Isolados (genomas)	Isolados (genoma)
LGMM	48(08)	23(02)	23(0)
BD	22(22)	04(04)	0(0)
Total	70(30)	27(06)	23(0)

LGMM: Isolados do Laboratório de Genética Molecular de Microrganismos; BD: Banco de dados públicos. (Mais detalhes no Anexo I)

A construção das árvores filogenéticas foi feita no programa SeaView v4 (Gouy *et al.* 2010) usando o método de máxima verossimilhança (*maximum likelihood*⁷) usando PhyML v3.0.1 (Guindon e Gascuel 2003) com os parâmetros pré-estabelecidos: busca da topologia inicial mediante intercâmbios de vizinho mais próximo (NNIs), árvore inicial construída com BioNJ, modelo evolutivo GTR (*General time reversible model*, que assume a presença dos quatro nucleotídeos em diferentes frequências) e aLTR (*approximate likelihood-ratio test*) foi usado para o teste de robustez dos ramos (Anisimova e Gascuel 2006).

Considerando que o conjunto de nossos dados pode conter informações vindas de organismos estreitamente relacionados ou não ou que resultem de eventos de recombinação, usamos o programa *ClonalFrame*, que estima relações a partir de um amplo conjunto de variáveis numa única análise.

O programa *ClonalFrame* v1.1 (Didelot e Falush 2007) foi usado para inferir as relações filogenéticas entre os 283 isolados a partir das seqüências concatenadas dos genes *recA*, *pyrH* e *mdh* (MLSA). Duas análises independentes

⁷ A abordagem de *maximum likelihood* (ML) pode inferir relações evolutivas e fazer testes de *bootstrap* para nucleotídeos e aminoácidos. ML é um método que demanda muito gasto computacional; um novo método heurístico é o PhyML, que recupera mais frequentemente a árvore correta quando comparado a outras abordagens.

foram feitas, cada uma com 400.000 interações (incluindo 200.000 interações *burn-in*) para um conjunto de 283 organismos, usando valores pré-determinados em todos os parâmetros. A construção da árvore principal consenso foi feita com SplitsTree4 (Huson e Bryant 2006) a partir da amostra posterior (as primeiras 200.000 interações foram consideradas o período de *burn-in* e foram descartadas. As interações remanescentes foram amostradas a cada 100 gerações para produzir 2000 topologias na amostragem posterior. A edição e visualização das árvores finais foi feita usando a ferramenta web iTOL (Letunic e Bork 2011).

4.2 Análises genômicas

4.2.1 Busca por marcadores epidemiológicos

Analisando genomas completos de *V. cholerae*, incluindo os da América Latina (Tabela 2), foram feitas buscas das regiões genômicas usadas em estudos da epidemiologia da cólera e outras que poderiam ser únicas e, portanto, características da linhagem da epidemia da América Latina. Estas regiões genômicas estão listadas na Tabela 3; a busca destas seqüências foi feita com BLAST (Altschul *et al.* 1997) utilizando *scripts* customizados para a busca em paralelo.

Tabela 3. Genes e regiões alvo da busca por BLAST nos genomas de *V. cholerae*.

Gene/região alvo	Descrição	Genoma referência	Número de acesso	Coordenadas
IntI1	Integrase de classe 1	<i>V. cholerae</i> B33	AAWE01000022.1	15348..16283
IntI2	Integrase de classe 2	<i>V. cholerae</i> VC97	GU570569	395..931
IntI3	Integrase de classe 3	<i>Klebsiella pneumoniae</i> p22K9	AY219651	8..1048
IntI4	Integrase de classe 4	<i>V. cholerae</i> N16961	AE003853	309750..310712
aadA6	Aminoglicosídeo adenililtransferase	<i>V. cholerae</i> VC622	ND	1..991
ctxB	Subunidade B da toxina colérica	<i>V. cholerae</i> N16961	AE003852	1566967..1567341
TcpA	Pilina	<i>V. cholerae</i> N16961	AE003852	890449..891123
VC2346	Proteína smp	<i>V. cholerae</i> N16961	NC002505	2502825..2503511
RfbN	Proteína acyl synthase/acyl-CoA	<i>V. cholerae</i> N16961	AE003852	257292..259769
VC1786ICE	Integrase do SXT	<i>V. cholerae</i> VC1786ICE	JN648379	3010..4251
WASA	Profago WASA	<i>V. cholerae</i> IEC224	CP003330	1601813..1651018
UniqueLA	Região única da linhagem da AL	<i>V. cholerae</i> IEC224	CP003330	540080..547156
VC0511	Região variável da VSP-II	<i>V. cholerae</i> N16961	AE003852	540156..540540
VC0513	Região variável da VSP-II	<i>V. cholerae</i> N16961	AE003852	544362..545177

AL: América Latina, ND: Não disponível

4.2.2 Comparação de genomas da América Latina

Foram selecionados genomas de *V. cholerae* O1 clínicos do período e região da epidemia da cólera da América Latina dos anos 1990s e o genoma de referência N16961 como descrito na Tabela 3. A primeira análise feita foi usando o programa BRIG (*Blast Ring Image Generator*) (Alikhan *et al.* 2011) o qual permite fazer comparação entre genomas baseado em regiões comuns identificadas por BLAST. Nesta análise foram usados os genomas completos N16961 e IEC224 como genomas de referência para encontrar regiões comuns e exclusivas.

4.2.3 Identificação de genes ortólogos, determinação do pangenoma, genoma core e genes únicos de *V. cholerae*

A identificação de genes ortólogos entre os genomas selecionados foi feita usando o método de família gênica, implementado no programa PGAP-1.01 (Zhao *et al.* 2012). Para esta análise foram empregados 24 genomas, dos quais 21 pertencem à epidemia da América Latina. Os isolados R-18246, R-18304, R-18316, R-18317, R-18327, R-18332, R-18338, R-18348 e R-18356 foram sequenciados mediante a tecnologia *Ion Torrent*⁸. A montagem dos genomas foi feita usando *gsAssembler* (Margulies *et al.* 2005). Os genomas foram anotados mediante o servidor web RAST (*Rapid Annotations using Subsystems Technology*) (Aziz *et al.* 2008). Observamos que as predições de genes pelo RAST, a partir de sequências vindas do *IonTorrent* continham distorções que levavam a uma super estimacão do número de genes. Observamos que esta distorção poderia ser devido à presença de erros no sequenciamento ou presença de Ns inseridos na sequência montada. Para testar esta hipótese, realizamos a seguinte análise: foram simuladas diferentes taxas de erros no genoma referência de *Escherichia coli* str. K-12 substr. W3110 (NC_007779.1) mediante um *script* no programa R que substituiu randomicamente

⁸ A tecnologia de sequenciamento de alto desempenho *Ion Torrent* forma parte dos chamados PGM (*Personal Genomic Machine*), este método se baseia na detecção de íons de hidrogênio que são liberados durante a polimerização do DNA. Este é um método de sequenciamento por síntese, durante o qual a fita complementar é produzida baseada na sequência da cadeia molde.

os nucleotídeos originais por um set de caracteres (A, T, G, C, N, -). Os caracteres randômicos foram incorporados em uma proporção de 1%, 2%, 5% e 10% do total do tamanho do genoma. O genoma original e as quimeras foram submetidos ao RAST. Os resultados mostraram que o número de ORFs preditas está correlacionado com a proporção de erros inseridos (Ver Anexo IV). Portanto, na análise de ortólogos foram incluídos apenas os genomas cujo tamanho médio dos genes estava dentro do intervalo médio relativo aos dos genomas referenciais (Ver Anexo V). Os genomas C6706, CP1030, CP1044, CP1046 e CP1047, recentemente disponíveis no *GenBank*, também foram incluídos (Tabela 4).

Com a finalidade de identificar os genes comuns (*core*), os genes acessórios e genes únicos, os *clusters* de ortólogos identificados pelo PGAP foram convertidos em uma matriz de presença e ausência (0/1) usando o programa R.

A árvore e a rede filogenética baseadas no conteúdo gênico foram construídas com o programa SplitsTree4 (Huson e Bryant 2006), usando o método de distância por conteúdo gênico desenvolvido por Huson e Steel (Huson e Steel 2004).

Tabela 4. Genomas de *V. cholerae* da epidemia da América Latina.

Organismo	Sorogrupo	Origem*	Fonte	Ano [‡]	Número de acesso [‡]
N16961	O1	Banladesh	Clínica	1975	NC 002505/NC 002506
IFC224	O1	Pará, Brasil	Clínica	1994	NC 016944/NC 016945
ERS013174	O1	Colômbia	Clínica	1992	ERS013174
ERS013175	O1	Colômbia	Clínica	1992	ERS013175
ERS013182	O1	Colômbia	Clínica	1992	ERS013182
ERS013183	O1	Colômbia	Clínica	1992	ERS013183
ERS013184	O1	Argentina	Clínica	1992	ERS013184
ERS013185	O1	Bolívia	Clínica	1992	ERS013185
ERS013186	O1	México	Clínica	1991	ERS013186
ERS013188	O1	Argentina	Clínica	1992	ERS013188
ERS013189	O1	Argentina	Clínica	1992	ERS013189
ERS013195	O1	México	Clínica	1991	ERS013195
ERS013200	O1	Argentina	Clínica	1993	ERS013200
ERS013201	O1	Argentina	Clínica	1993	ERS013201
ERS013251	O1	Peru	Clínica	1991	ERS013251
ERS013274	O1	Peru	Clínica	1991	ERS013274
ERS013275	O1	Peru	Clínica	1991	ERS013275
ERS013276	O1	Peru	Clínica	1991	ERS013276
ERS013280	O1	Argentina	Clínica	1993	ERS013280
INDRE-91/1	O1	México	Clínica	1991	ADAK01000000
R-18246	O1	Ceará, Brasil	Clínica	1992	ND
R-18304	O1	Ceará, Brasil	Clínica	1993	ND
R-18316	O1	Rio de Janeiro, Brasil	Clínica	1993	ND
R-18317	O1	Rio de Janeiro	Clínica	1994	ND
R-18327	O1	Paraná, Brasil	Clínica	1999	ND
R-18332	O1	Amazonas, Brasil	Clínica	1991	ND
R-18338	O1	Ceará, Brasil	Clínica	1992	ND
R-18348	O1	Pernambuco, Brasil	Clínica	1999	ND
R-18356	O1	Amazonas, Brasil	Clínica	1991	ND
C6706	O1	Peru	Clínica	1991	AHGQ01000000
CP1044	O1	Peru	Clínica	1991	ALDG01000000
CP1046	O1	Peru	Clínica	1995	ALDH01000000
CP1047	O1	Peru	Clínica	1995	ALDI01000000
CP1030	O1	México	Clínica	2008	ALCZ01000000

*: Origem geográfica, ‡: ano de isolamento, ‡: número de acesso no *GenBank* e no *European Nucleotide Archive (ENA)*. ND: Não disponível.

Resultados

5 RESULTADOS

Nesta seção são apresentados por capítulos, e de acordo com os objetivos, nossos resultados, alguns consubstanciados em publicações. Para atingir nosso objetivo final de estudar a epidemiologia genômica de *V. cholerae* na América Latina, realizamos também alguns trabalhos adicionais como por exemplo a caracterização de linhagens de surtos/epidemias em países do oeste da África, o que contribuiu na construção de um cenário mais completo sobre a epidemia da América Latina.

5.1 Alelos do gene *ctxB*

O gene *ctxB* codifica a subunidade da toxina colérica, a qual forma um pentâmero que se liga às células do epitélio intestinal e permite que a outra subunidade, a enzimaticamente ativa, possa atuar. Alelos deste gene caracterizam os biotipos de *V. cholerae*, sendo usados como marcador epidemiológico e de patogenicidade.

Recentemente, visando caracterizar os alelos deste gene típicos dos biotipos clássico e El Tor, foi desenvolvida uma abordagem baseada na técnica de PCR conhecida como MAMA-PCR (*mismatch amplification mutation PCR assay*) (Morita *et al.* 2008). Através desta abordagem mostrou-se a presença do alelo característico do biotipo clássico em linhagens do biotipo El Tor (Bhuiyan *et al.* 2012).

Objetivando caracterizar o alelo deste gene na linhagem da América Latina comparamos todas as sequências completas *ctxB* disponíveis no GenBank, relacionadas ou não a trabalhos, assim como as obtidas por nós. Considerando as sequências de nucleotídeos e aminoácidos deste conjunto foi possível identificar um total de 11 alelos. A linhagem da América Latina foi caracterizada pelo alelo 3, característico de El Tor do início da sétima pandemia; além disto, os novos alelos, 10 e 11, foram caracterizados. Interessantemente, as sequências nucleotídicas destes novos alelos faziam parte de um trabalho (Bhuiyan *et al.* 2012) cujos autores, aplicando a caracterização pela MAMA (que é parcial), os haviam identificado erroneamente como alelo 1 ou seja, genótipo clássico (Tabela 5).

Tabela 5. Polimorfismos associados aos alelos de *ctxB* em *V. cholerae*.

Sorogrupo	Genótipo <i>ctxB</i>	Posição nucleotídica (posição do aminoácido)								Referências
		58 (20)	72 (24)	83 (28)	101 (34)	115 (39)	138 (46)	165 (55)	203 (68)	
O1	1	C (His)	A (Gln)	A (Asp)	A (His)	C (His)	T (Phe)	A (Lys)	C (Thr)	(Olsvik <i>et al.</i> 1993; Goel <i>et al.</i> 2008)
	2	C (His)	A (Gln)	A (Asp)	A (His)	C (His)	G (Leu)	A (Lys)	C (Thr)	
	3	C (His)	A (Gln)	A (Asp)	A (His)	T (Tyr)	T (Phe)	A (Lys)	T (Ile)	
	7	A (Asn)	A (Gln)	A (Asp)	A (His)	C (His)	T (Phe)	A (Lys)	C (Thr)	Este estudo
	10	C (His)	A (Gln)	A (Asp)	C (Pro)	T (Tyr)	T (Phe)	A (Lys)	T (Ile)	
	11	C (His)	A (Gln)	A (Asp)	C (Pro)	C (His)	T (Phe)	A (Lys)	C (Thr)	
O139	3	C (His)	A (Gln)	A (Asp)	A (His)	T (Tyr)	T (Phe)	A (Lys)	T (Ile)	(Olsvik <i>et al.</i> 1993)
	4	C (His)	A (Gln)	A (Asp)	A (His)	T (Tyr)	T (Phe)	A (Lys)	C (Thr)	
	5	C (His)	A (Gln)	C (Ala)	A (His)	C (His)	T (Phe)	A (Lys)	C (Thr)	
	6	C (His)	A (Gln)	A (Asp)	C (Pro)	T (Tyr)	T (Phe)	A (Lys)	C (Thr)	
O27	8	C (His)	C (His)	C (Ala)	A (His)	C (His)	T (Phe)	A (Lys)	C (Thr)	(Li <i>et al.</i> 2002)
O37	9	C (His)	A (Gln)	A (Asp)	A (His)	C (Leu)	G (Leu)	C (Asn)	C (Thr)	(Yamamoto <i>et al.</i> 1995)

Associação do genótipo a linhagens: 1 (Classical, 569B); 2 (El Tor, Austrália); 3 (El Tor, N16961 e América Latina) e 7 (Alterado El Tor).

Estes resultados fazem parte do artigo “Variants of *Vibrio cholerae* O1 El Tor from Zambia showed new genotypes of *ctxB*” publicado na revista *Epidemiology and Infection* 140(8):1386–7 (Marin e Vicente 2012) (Ver Anexos).

5.2 Caracterização genética de *V. cholerae* O1 de Gana/África

Lam e colaboradores, com base em 30 SNPs, estabeleceram uma possível relação/origem da linhagem da epidemia da América Latina com a África (Lam *et al.* 2010). Neste trabalho poucas amostras foram consideradas: 04 da América Latina: Peru/Brasil/Guiana Francesa e 07 da África: Etiópia/Malawi/Algéria/Nigéria/Senegal/Marrocos. Portanto, a fim de robustecer esta hipótese e devido a oportunidade de analisar uma coleção de *V. cholerae* isolados do início da sétima pandemia e de epidemias atuais em Gana, realizamos este outro trabalho. As amostras de Gana foram caracterizadas mediante análise de sequência de vários locus (*multilocus sequence analysis, MLSA*), pela caracterização de genes de virulência e perfil de resistência aos antibióticos (Tabela 6).

Pela análise por MLSA os *V. cholerae* O1 de Gana se encontram em dois principais *clusters* (grupos), um compartilhado pela linhagem El Tor da sétima pandemia e o outro Amazônia/Gana (Figura 11). O *cluster* El Tor contém isolados do período 1970-2006, enquanto que o *cluster* Amazônia/Gana contém isolados de Gana dos anos 1978-1980, e isolados da linhagem Amazônia dos anos 1991 e 1992. O mesmo agrupamento foi obtido utilizando a comparação de perfis de macro restrição resolvidos em PFGE (eletroforese de campo pulsado) (Figura 12).

Tabela 6. Características fenotípicas e genotípicas de *V. cholerae* O1 de Gana.

<i>V. cholerae</i> O1	Isolamento (local, ano)	Perfil de resistência a antibióticos*	SXT/integron de classe 2	GCV ^a / <i>ctxB</i> / <i>tcpA</i>	VSP-II (VC0511/VC0513)
0395	Índia, 1965	ND	-/-	+ / + / +	- / -
LMG 21698 ^T	Ásia	ND	ND	ND	ND
MAK757	Ilhas Célebes, 1937	ND	-/-	+ / + / +	- / -
N16961	Bangladesh, 1975	ND	-/-	+ / + / +	+ / +
TM 11079-	Brasil, 1980	ND	-/-	+ / - / -	- / -
V33	Gana, 2006	SXT, AMP, NAL	+ / -	+ / + / ND	ND
V34	Gana, 2006	SXT, NAL	+ / +	+ / + / ND	ND
V35	Gana, 2006	SXT, NAL	+ / +	+ / + / ND	ND
V42	Gana, 2006	SXT	+ / +	+ / + / ND	ND
V47	Gana, 2006	SXT	+ / +	+ / + / ND	ND
V51	Gana, 2006	SXT	+ / +	+ / + / ND	ND
V52	Gana, 2006	SXT	+ / +	+ / + / ND	ND
V53	Gana, 2006	SXT, AMP, NAL	+ / +	+ / + / ND	ND
V78	Gana, 2006	SXT, AMP, NAL, CAZ, CRO, CHL, CXM	+ / +	+ / + / ND	ND
V84	Gana, 2006	SXT, AMP, TET, CXM	+ / +	+ / + / ND	ND
V90	Gana, 2006	SXT, AMK, CAZ, CRO, CTXCHL, GEN	+ / +	+ / + / ND	ND
V97	Gana, 2006	SXT	- / +	+ / + / ND	ND
V98	Gana, 2006	SXT, AMP	+ / +	+ / + / ND	ND
VC33	Tanzânia, 1979	Susceptível	- / -	+ / +	
VC34	Gana, 1970	Susceptível	- / -	+ / + / +	+ / +
VC37	Gana, 1980	Susceptível	- / -	+ / - / +	- / -
VC92	Gana, 1980	STP, SPT	- / +	+ / - / -	- / -
VC94	Gana, 1979	STP, SPT	- / +	+ / - / -	- / -
VC95	Gana, 1978	STP, SPT	- / +	+ / - / -	- / -
VC96	Gana, 1978	STP, SXT	- / -	+ / + / +	+ / +
VC97	Gana, 1979	STP, SPT	- / +	+ / - / -	- / -
VC98	Gana, 1979	STP, SPT	- / +	+ / - / -	- / -
VC102	Gana, 1979	STP	- / -	+ / - / +	+ / +
VC104	Algeria, 1972	STP	- / -	+ / - / +	+ / +
VC106	Gana, 1980	Susceptível	- / -	+ / + / +	+ / +
VC107	Gana, 1980	Susceptível	- / -	+ / + / +	+ / +
VC121	Índia, 1973	Susceptível	- / -	+ / + / +	+ / +
VC192	Gana, 1980	Susceptível	- / -	+ / + / +	+ / +
VC193	Gana, 1981	Susceptível	- / -	+ / - / +	+ / +
VC225	Gana, 1976	Susceptível	- / -	+ / + / +	+ / +
VC355	Amazônia (Brasil), 1992	Susceptível	- / -	+ / - / -	+ / +
VC356	Amazônia (Brasil), 1991	STP, SPT	- / +	+ / - / -	ND

+ positivo; - negativo; ND não determinado

*: Antibióticos testados: SXT, Sulfametoxazol/Trimetoprima; AMK, amicacina; AMP, ampicilina; CAZ, ceftazidima; CRO, ceftriaxona; CTX, cefotaxima; CHL, cloranfenicol; GEN, gentamicina; TET, tetraciclina; NAL, ácido nalidixico; CXM, cefuroxima; STP, estreptomicina; SPT, espectinomicina; TRM, trimetoprima.

^a: CVG: Genes core de virulência.

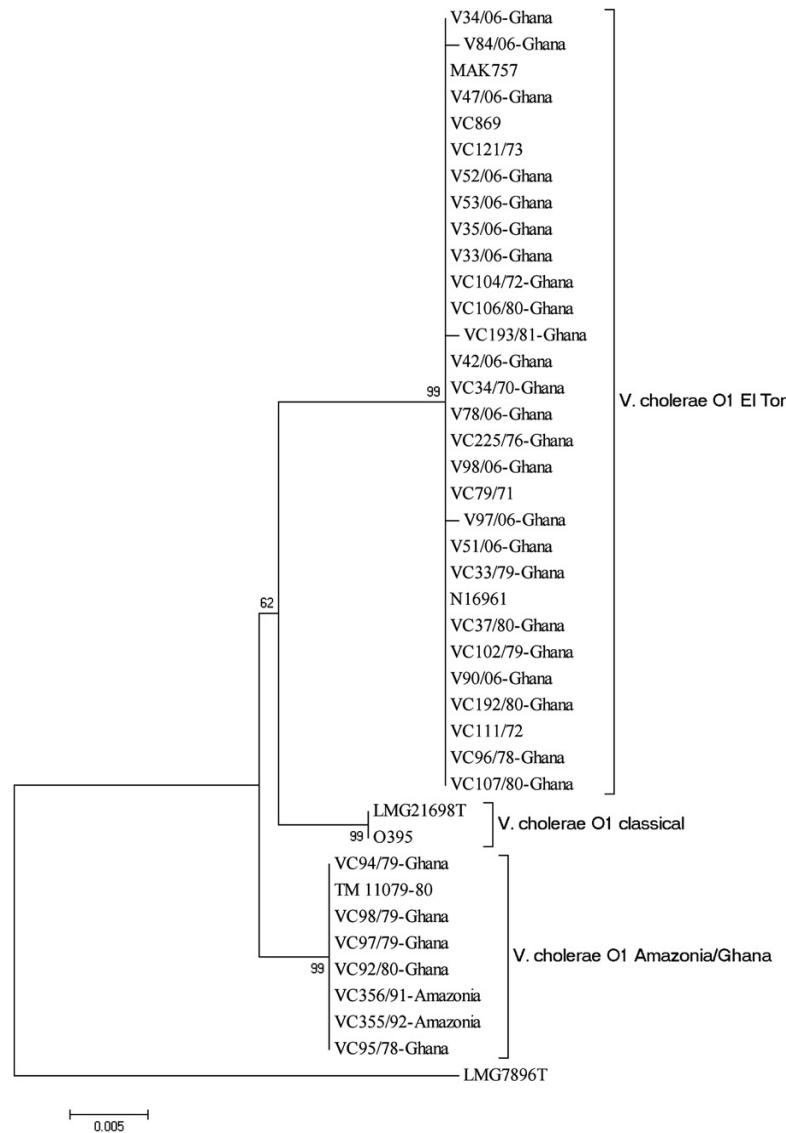


Figura 11. Relações filogenéticas dos isolados de *V. cholerae* de Gana. Árvore baseada nas sequências concatenadas dos genes *pyrH*, *recA* e *rpoA* usando o modelo *Kimura dois-parâmetros*⁹ e o método *Neighbor-joining*¹⁰. Na árvore são indicadas as linhagens de *V. cholerae* O1 El Tor, clássica e a linhagem Amazônia/Gana. *Vibrio mimicus* LMG7896T foi usado como grupo externo.

⁹ *Kimura dois-parâmetros* é um modelo evolutivo que considera substituições múltiplas e a razão de transições e transversões.

¹⁰ *Neighbor-joining* (NJ) ou agrupamento de vizinhos é um método de construção de árvore filogenética que se inicia com uma árvore semelhante a um arbusto (sem ramos internos), o algoritmo vai introduzindo ramos internos e calculando o tamanho das árvores resultantes, finalmente recupera uma única árvore, totalmente bifurcada, com a menor soma total de ramos.

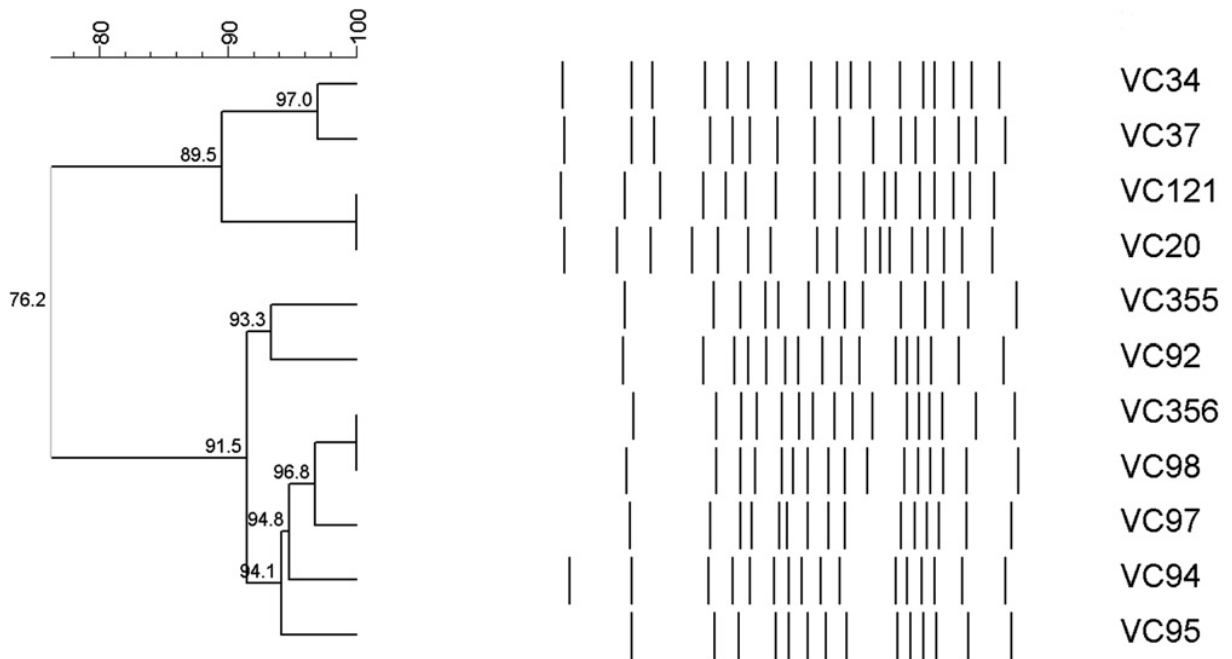


Figura 12. Dendrograma construído a partir dos perfis de PFGE de isolados representativos de *V. cholerae* O1. O lado direito mostra os perfis de macro restrição resolvidos em eletroforese de campo pulsado (PFGE). A similaridade entre os isolados, mostrada no lado esquerdo, foi determinada usando o coeficiente *Dice* e a clusterização foi feita usando UPGMA.¹¹.

Os outros marcadores também corroboraram os dois *clusters*. Segundo a caracterização dos elementos associados com resistência a antimicrobianos, os resultados mostram que todos os isolados de Gana que pertencem ao *cluster* Amazônia/Gana foram positivos para o integron de classe 2 (Tabela 7) da mesma forma que o isolado VC356 da Amazônia. Os isolados do início da sétima pandemia em Gana foram susceptíveis a todos os antimicrobianos testados enquanto que os

¹¹ UPGMA (*Unweighted Pair Group Method with Arithmetic Mean*) é o método mais simples para a construção de árvores, foi originalmente criado para construir árvores que refletissem semelhanças fenotípicas. O processo consiste em que, dada uma matriz de distâncias entre pares de taxa, a análise de *cluster* tenta representar essas informações em um diagrama chamado de fenograma que expressa as semelhanças gerais entre os táxons. O Método do Grupo Par usa o seguinte algoritmo: (1) Identifica a distância mínima entre dois taxa, (2) Junta estes dois taxa como um único par, (3) Re-calcula a distância média entre este par e todos os outros taxa para formar uma nova matriz, (4) identifica o par mais próximo na nova matriz (5), e assim por diante, até que os dois últimos grupos estão unidos.

isolados atuais apresentam um perfil de multiresistência e presença do elemento SXT (Tabela 6).

Os dois principais fatores de virulência *ctxAB* e *tcpA* não estão presentes no *cluster* Amazônia/Gana. Quanto ao *ctxB*, presente nos isolados atuais, este apresenta o alelo clássico que caracteriza a linhagem como El Tor 'Alterado' ou 'Variante'. Assim, não conseguimos detectar em amostras de Gana nenhum marcador compatível com a linhagem da América Latina. O interessante foi mostrar uma relação Brasil/Gana no que se refere à linhagem Amazônia.

Estes resultados fazem parte do artigo "*Vibrio cholerae* O1 lineages driving cholera outbreaks during seventh cholera pandemic in Ghana" publicado na revista *Infection, Genetics and Evolution* 11(2011):1951–6 (Thompson *et al.* 2011a) (Ver Anexos).

5.3 Caracterização genética de isolados clínicos de *V. cholerae* da Nigéria

Embora a cólera seja endêmica na África apresentando surtos e epidemias frequentes, há poucos estudos de epidemiologia molecular caracterizando os determinantes destes episódios. Atualmente, a Nigéria é um dos três principais focos da cólera no mundo (Piarroux e Faucher 2012). Este capítulo descreve a caracterização genética de isolados de *V. cholerae* de surtos de cólera na Nigéria em 2009 e 2010.

Foram analisados 19 isolados de *V. cholerae* vindos de casos clínicos de cólera e um isolado ambiental, todos relacionados a surtos de cólera na Nigéria durante o período 2009-2010; foram incluídos também três isolados da Nigéria do início da sétima pandemia.

As relações filogenéticas observadas na árvore de MLSA construída com as sequências concatenadas dos genes *recA*, *pyrH* e *mdh* mostrou três principais *clusters*: (i) O1 El Tor/O139, onde se encontraram os isolados clínicos de *V. cholerae* O1 atuais e do início da sétima pandemia da Nigéria, (ii) o cluster do biotipo clássico, e (iii) não-O1/não-O139 da Nigéria (Figura 13). O mesmo agrupamento é definido pela análise de perfis de macrorestrição por PFGE (Figura 14).

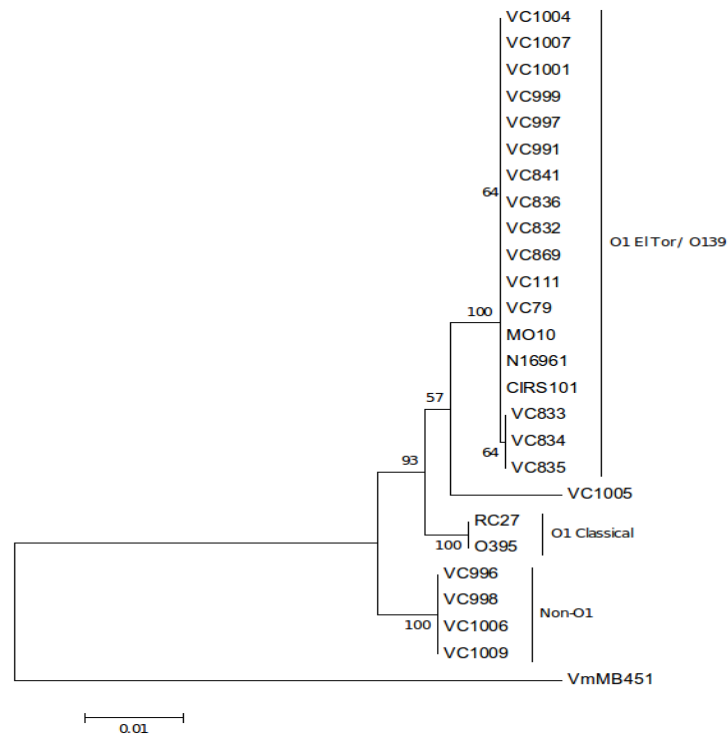


Figura 13. Relações filogenéticas dos isolados de *V. cholerae* da Nigéria. Os isolados epidêmicos O1 estão no clado El Tor/O139 e isolados não O1/não O139 estão localizados em outro clado. A árvore foi construída por *Minimal Evolution*, baseado em *Kimura-dois parâmetros* usando as sequências concatenadas dos genes *recA*, *pyrH* e *mdh*. 2000 réplicas *bootstrap* foram feitas. *Vibrio mimicus* MB451 foi usado como grupo externo.

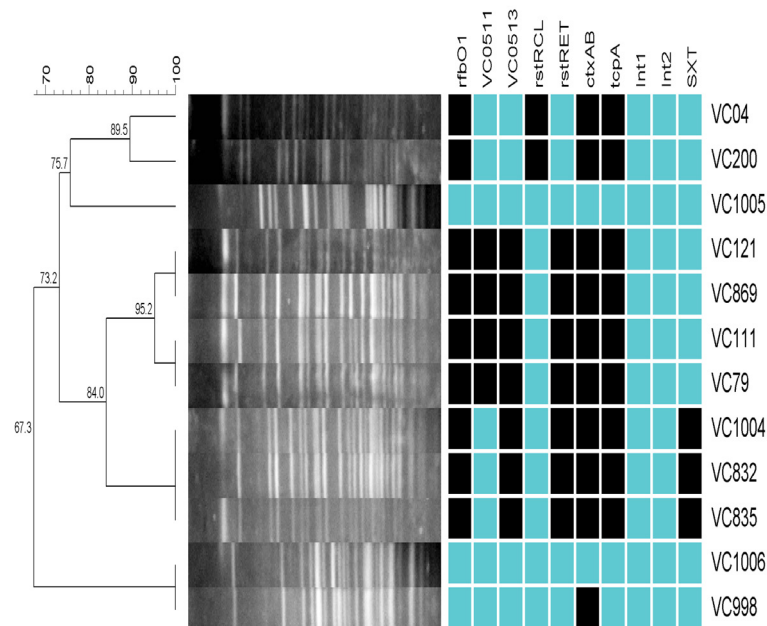


Figura 14. Dendrograma dos perfis de macro-restrição de *V. cholerae* O1 e não O1/não O139. A similaridade foi determinada usando o coeficiente *Dice* e a análise de *cluster* for feita usando UPGMA. O painel da direita mostra a presença (caixas pretas) e ausência (caixas claras) de algumas características genômicas. VC04 e VC200 pertencem à linhagem clássica, VC121 é uma cepa da Índia representativa do início da sétima pandemia representante da linhagem El Tor.

Quanto ao *ctxB*, todos os O1 El Tor atuais da Nigéria apresentam o genótipo 7, caracterizado por uma mudança His20Asn quando comparado com os El Tor do início da sétima pandemia e com a canônica N16961 (Marin e Vicente 2012). O *tcpA* dos O1 atuais da Nigéria foram caracterizados como *tcpA*^{CIRS} por apresentar alelo semelhante ao encontrado na amostra CIRS101, um *V. cholerae* O1 isolado no Bangladesh em 2002 (Talkington *et al.* 2011) enquanto que os demais apresentam *tcpA* El Tor. Baseados nestes resultados, os isolados atuais de *V. cholerae* O1 da Nigéria foram classificados como El Tor atípicos e os do início da sétima pandemia como El Tor típicos, ou seja nenhum deles apresenta relação direta com a linhagem da epidemia da América Latina.

A linhagem da epidemia da América Latina apresentava susceptibilidade à maioria dos antibióticos enquanto que os isolados atuais da Nigéria foram resistentes à estreptomicina, trimetoprim e sulfonamidas, resistências associadas à presença de integrons de classe 1 ou 2 (Opintan *et al.* 2008) e aos elementos integrativos e conjugativos (SXT/ICEs) (Wozniak *et al.* 2009). Foi possível detectar a presença do SXT/ICE nos isolados atuais da Nigéria. Além disto, estes isolados apresentavam suscetibilidade reduzida à ciprofloxacina que foi associada a mutações nos genes *gyrA* e *parC*, que não são encontradas nas linhagens do início da sétima pandemia nem nas amostras da América Latina. (Tabela 7).

Tabela 7. Características genéticas e fenotípicas de *V. cholerae* da Nigéria.

Cepa	Sorogrupo (biotipo)	Lugar de isolamento	Ano de isolamento	Genes relacionados a virulência e biotipo					Relacionados a resistência a antibióticos						
				VSP-II		<i>ctxB</i> *	<i>tcpA</i>	<i>rstR</i>	IntI1	IntI2	SXT	Resistência a quinolonas		Sensibilidade a antibióticos	
				VC0511	VC0513							<i>gyrA</i>	<i>parC</i>	MIC ^{CIP} (µg/ml)	MIC ^{CM} (µg/ml)
Cepas Referenciais [‡]															
N16961	O1 (El Tor)	Bangladesh	1975	+	+	3	ET	ET	-	-	-	Ser83	Ser85	ND	ND
CIRS101	O1 (El Tor)	Bangladesh	2002	-	+	1	ET ^{CIRS}	ET	-	-	+	Ser83Ile	Ser85	ND	ND
MO10	O139	Madras, India	1992	-	+	3	ET	ET	-	-	+	Ser83	Ser85	ND	ND
O395	O1 (Classical)	India	1965	-	-	1	CL	CL	-	-	-	Ser83	Ser85	ND	ND
RC27	O1 (Classical)	Indonesia	1991	-	-	1	CL	CL	-	-	-	Ser83	Ser85	ND	ND
Cepas da Nigéria															
VC79	O1	Nigéria	1971	+	+	-	-	ET	-	-	-	Ser83	Ser85	0.008	ND
VC111	O1	Nigéria	1972	+	+	3	ET	ET	-	-	-	Ser83	Ser85	0.008	ND
VC869	O1	Nigéria	1971	+	+	3	ET	ET	-	-	-	Ser83	Ser85	0.008	ND
VC832	O1	Maiduguri	2009	-	+	7	ET ^{CIRS}	ET	-	-	+	Ser83Ile	Ser85Leu	0.380	8.0
VC841	O1	Maiduguri	2009	-	+	7	ET ^{CIRS}	ET	-	-	+	Ser83Ile	Ser85Leu	0.250	8.0
VC833	O1	Ile Ife	2010	-	+	7	ET ^{CIRS}	ET	-	-	+	Ser83Ile	Ser85Leu	0.250	6.0
VC834	O1	Ile Ife	2010	-	+	7	ET ^{CIRS}	ET	-	-	+	Ser83Ile	Ser85Leu	0.250	6.0
VC835	O1	Ile Ife	2010	-	+	7	ET ^{CIRS}	ET	-	-	+	Ser83Ile	Ser85Leu	0.380	6.0
VC991	O1	Ile Ife	2010	-	+	7	ET ^{CIRS}	ET	-	-	+	Ser83Ile	Ser85Leu	0.380	6.0
VC997	O1	Ile Ife	2010	-	+	7	ET ^{CIRS}	ET	-	-	+	Ser83Ile	Ser85Leu	0.380	6.0
VC1001	O1	Ile Ife	2010	-	+	7	ET ^{CIRS}	ET	-	-	+	Ser83Ile	Ser85Leu	0.380	6.0
VC1007	O1	Ile Ife	2010	-	+	7	ET ^{CIRS}	ET	-	-	+	Ser83Ile	Ser85Leu	0.380	8.0
VC836	O1	Bauchi	2010	-	+	7	ET ^{CIRS}	ET	-	-	+	Ser83Ile	Ser85Leu	0.380	8.0
VC996	Não-O1	Bauchi	2010	-	-	-	-	-	-	-	-	Ser83	Ser85	0.023	0.5
VC998	Não -O1	Bauchi	2010	-	-	-	-	-	-	-	-	Ser83	Ser85	0.008	0.5
VC999	O1	Bauchi	2010	-	+	7	ET ^{CIRS}	ET	-	-	+	Ser83Ile	Ser85Leu	0.380	6.0
VC1004	O1	Bauchi	2010	-	+	7	ET ^{CIRS}	ET	-	-	+	Ser83Ile	Ser85Leu	0.500	6.0
VC1006	Não -O1	Bauchi	2010	-	-	-	-	-	-	-	-	Ser83	Ser85	0.012	0.5
VC1009	Não -O1	Bauchi	2010	-	-	-	-	-	-	-	-	Ser83	Ser85	0.008	0.5
VC1005 [§]	Não -O1	Ile Ife	2010	-	-	-	-	-	-	-	-	Ser83Ile	Ser85Leu	0.250	ND

CL, clássico; ET, El Tor; CIP, ciprofloxacina; CM, Cloramfenicol; ND, Não determinado; *, Genótipo *ctxB* (Marin e Vicente 2012); ‡, As características genéticas para estas cepas foram determinadas por análises *in silico*; §, VC1005 é um isolado ambiental.

Resumindo, os nossos resultados mostram que as epidemias de cólera na Nigéria estão associadas à linhagem de *V. cholerae* O1 atípico e multi-resistente, que é a determinante das epidemias atuais de cólera, como a do Haiti. Além disto, surtos estariam associados a outra linhagem não-O1/não-O139. Os detalhes deste estudo estão no artigo “Cholera Outbreaks in Nigeria Are Associated with Multidrug Resistant Atypical El Tor and Non-O1/Non-O139 *Vibrio cholerae*” publicado em PLoS Neglected Tropical Diseases 7(2):e2049 (Marin *et al.* 2013) (Ver Anexos).

5.4 Estrutura do superintegron de *V. cholerae*

O superintegron (SI) é uma extensa região genômica de aproximadamente 120 kbp localizada no cromossomo menor de *V. cholerae*. Esta região atua como um sistema dinâmico de captura e excisão de genes produzindo inserções, deleções e re-arranjos dos cassetes gênicos. Portanto, seria uma região de evolução rápida e possível carreadora de marcadores para linhagens. Alguns estudos têm associado a diversidade e plasticidade deste elemento ao nicho ecológico explorado pelo *Vibrio*.

Este capítulo descreve a caracterização do superintegron de *V. cholerae* quanto à sintenia, aos genes do *core* e genes únicos. Para isto utilizamos os dados recuperados dos genomas completos de *V. cholerae*.

A partir de sete genomas completos de *V. cholerae* (pre-El Tor M66-2 da Indonésia 1937, clássico O395 da Índia 1965, El Tor N16961 de Bangladesh 1975, MJ-1236 El Tor de Bangladesh 1994, IEC224 El Tor do Brasil 1994, 2010EL1786 El Tor do Haiti 2010, LMA-3984 ambiental do Brasil 2007) e um genoma de *V. mimicus* (isolado clínico de Bangladesh), e considerando as localizações de regiões estruturais e características do superintegron, foram recuperados e identificados os genes em comum (genes do *core*) e genes únicos de cada linhagem.

Foram extraídos 1285 genes que foram agrupados em 408 *clusters* (Figura 15) que correspondem aos grupos de ortólogos identificados pelo método MultiParanoid implementado no programa PGAP (Zhao *et al.* 2012). Oito genes correspondem ao genoma *core*, 196 são genes distribuídos ou dispensáveis (presentes em dois ou mais organismos) e 204 genes únicos. Portanto o pangenoma destes superintegrons possui 408 genes/*clusters*. Vários estão em múltiplas cópias como, por exemplo, seis dos oito genes do *core* (Tabela 8). O perfil do pangenoma mostra que, quando adicionados novos SI nas análises, o número de *clusters* do genoma *core* é mantido enquanto que o pangenoma continua aumentando (Figura 15A). Portanto, se infere que o SI do *V. cholerae* tem um pangenoma aberto, isto significa que *V. cholerae* tem a capacidade de importar novos cassetes gênicos mostrando sua plasticidade e diversidade.

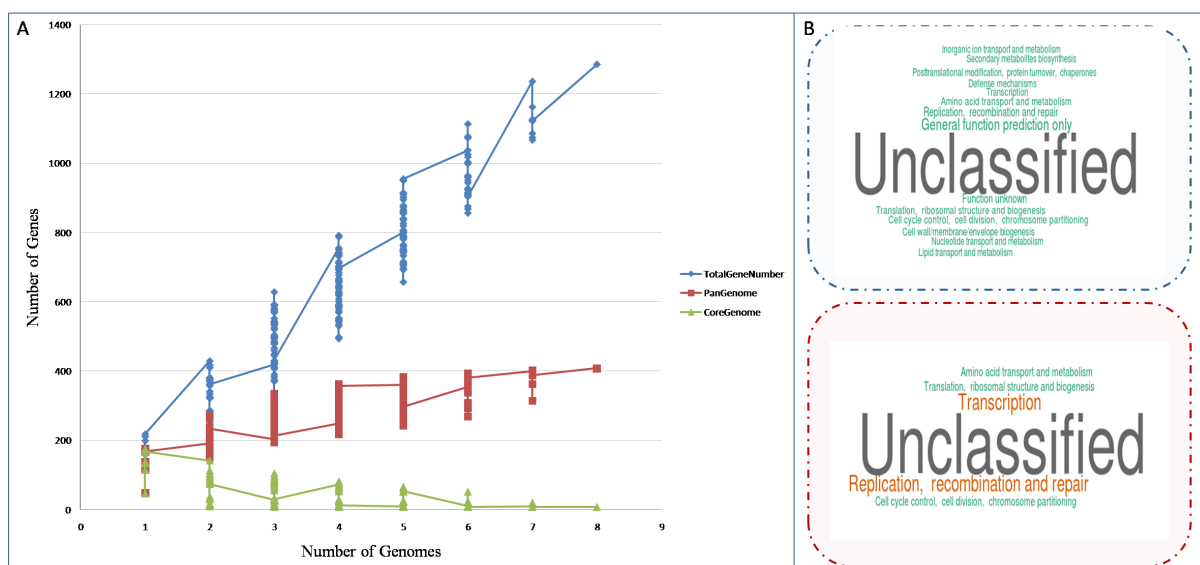


Figura 15. (A) Plot do pangenoma da região do superintegron considerando sete genomas de *V. cholerae* e um de *V. mimicus*. De um total de 1285 genes foram identificados 408 *clusters* correspondentes ao pangenoma e oito *clusters* core do SI. (B) Nuvem de palavras mostrando a abundância de *clusters* segundo sua função COG (*cluster of orthologous group*¹²) para o total (acima) e o core (embaixo) dos *clusters*. *Clusters* que não tem uma classificação COG foram excluídos da figura.

Considerando o pangenoma, de 408 *clusters*, 329 são hipotéticos não caracterizados; semelhante perfil se observa também com os genes *core* (Figura 15B).

Foram identificados 204 genes únicos, 94 pertencem a *V. mimicus* MB451, nove a LMA3984, 45 a O395, nove a 2010EL1786, 14 a MJ1236, sete IEC224

¹²COG (*Cluster of Orthologous Groups*): Também conhecido como *clusters* de genes ortólogos, é uma base de dados que é usada para descrever as características funcionais de um microrganismo ou uma comunidade de microrganismos. Foi gerada pela comparação de proteínas conhecidas e preditas em todos os genomas microbianos completamente sequenciados, dos que se inferiu os conjuntos de ortólogos. Cada COG consiste em um grupo de proteínas identificadas como ortólogas em pelo menos três linhagens. Os COGs estão classificados em 23 categorias funcionais, relacionados à sinalização e processos celulares, armazenamento e processamento da informação, metabolismo e os pobremente caracterizados (<http://www.ncbi.nlm.nih.gov/COG>).

(pertence à linhagem da América Latina), 20 a M66, e seis a N16961 (Figura 16). Considerando somente o SI de *V. cholerae*, há 21 genes no *core*, a maioria deles presentes em múltiplas cópias e associados com as categorias funcionais de transcrição, replicação, recombinação e reparo, tradução, estrutura ribossomal e biogênese.

Tabela 8. Genes core do superintegron de *V. cholerae*.

ClusterID	Nível de conservação [‡]	COG*	Descrição da anotação	Locus_tag em N16961
1	8	-	Proteína hipotética	VCA0407,VCA0353,VCA0336,VCA0297,VCA0302
2	8	COG0456R	Acetiltransferase	VCA0470
3	8	-	Lipoproteína	VCA0425,VCA0414
4	8	-	Proteína hipotética	VCA0381,VCA0435,VCA0357,VCA0306
5	8	-	Proteína hipotética	VCA0434,VCA0411
7	8	COG4974L	Recombinase sítio-específica IntI4	VCA0291
8	8	-	Proteína relB	VCA0349,VCA0504
9	8	COG1670J	Acetiltransferase	VCA0505,VCA0436,VCA0417,VCA0316
24	7	COG0110R	Acetiltransferase	VCA0473
25	7	COG3668R	Elemento de estabilização de plasmídeos ParE	VCA0359
27	7	COG2944K	Gene repressor de virulência RsaL	VCA0469
31	7	-	Proteína hipotética	VCA0497
32	7	COG1694R	Proteína relacionada a mazG	VCA0485
33	7	-	Repressor traducional citotóxico do sistema de estabilização toxina-antitoxina	VCA0468
34	7	COG0346E	Proteína de resistência a glioxalase/bleomicina	VCA0506,VCA0347
35	7	-	Proteína hipotética	VCA0486
37	7	COG2161D	Antitoxina do sistema de estabilização toxina-antitoxina	VCA0477
40	7	COG0456R	N-acetiltransferase relacionada a GCN5	VCA0382
41	7	COG1943L	Transposase IS1004	VCA0493
43	7	COG3668R	Proteína do sistema de estabilização de plasmídeo	VCA0489
44	7	COG3636K	Proteína hipotética	VCA0498

[‡]O nível de conservação se refere ao número de genomas onde são encontrados estes *clusters*. *COG: *Cluster* de grupo de ortólogos descreve a classificação por categorias funcionais de proteínas ortólogas.

No SI, excisões randômicas ocorrem através de um arranjo do cassete para uma forma intermediária circular não replicativa contendo um ou vários cassetes, eventos de integração ocorrem preferencialmente no sítio *attI* pela ação da integrase; devido a posição do promotor da expressão dos cassetes gênicos, é esperado que os genes *core* estejam dispostos e mantidos juntos; no entanto, na nossa análise os genes *core* foram encontrados distribuídos ao longo do SI (Figura 16), aparentemente sem nenhum efeito na posição.

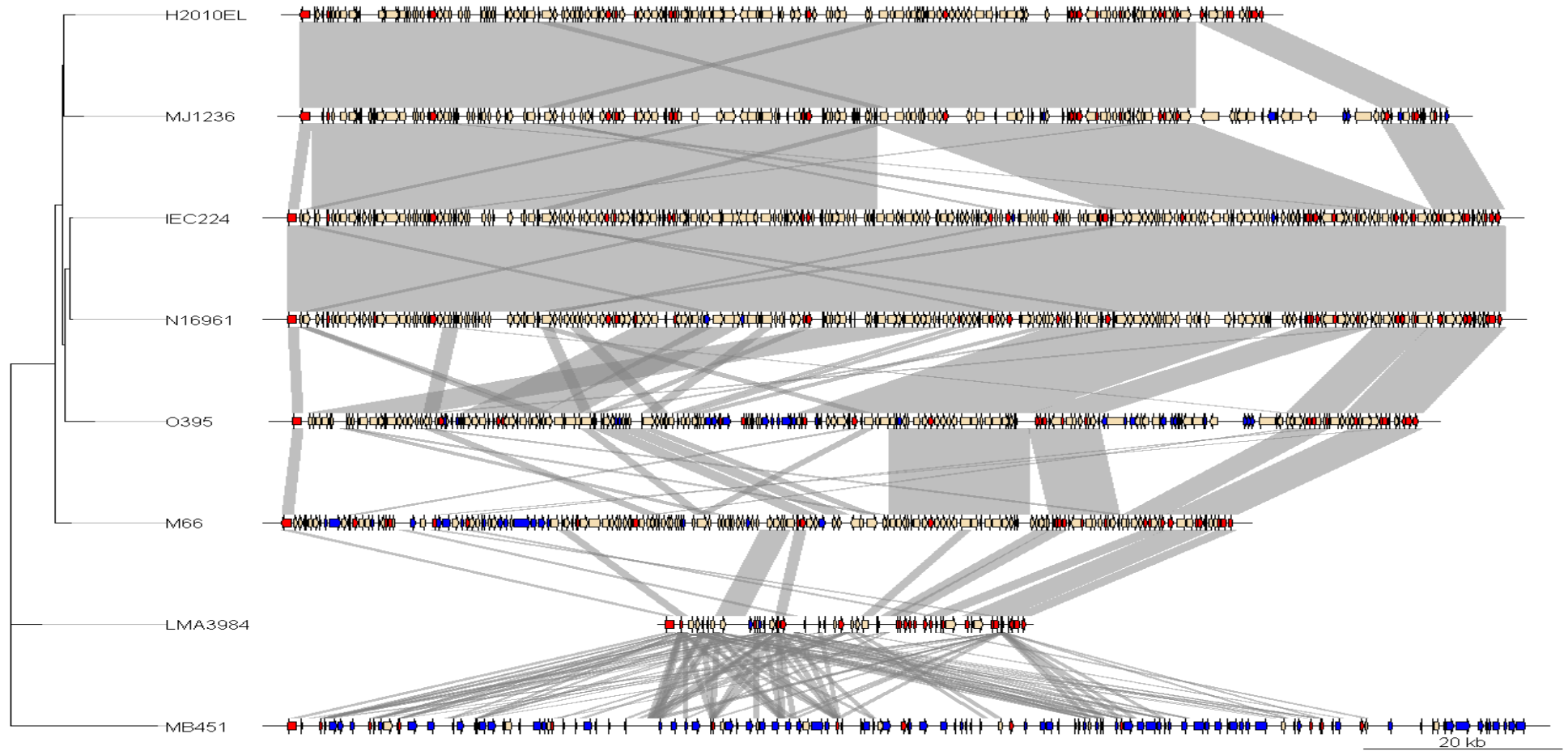


Figura 16. Sintenia dos superintegrons das sete linhagens de *V. cholerae* e um *V. mimicus*. Os genes core, genes dispensáveis e genes únicos são indicados como setas em vermelho, creme, e azul, respectivamente. Os blocos verticais cinza entre as seqüências indicam regiões que compartilham similaridade maior a 1kb de acordo com o BLASTn. A árvore filogenética da esquerda é baseada no conteúdo gênico do SI.

Também foram identificados seis genes do sistema Toxina-Antitoxina como parte dos genes *core* do SI, isto sugere como essencial a função destes genes, sendo elementos de resposta ao estresse, ajudando na sobrevivência das células (Christensen-Dalsgaard e Gerdes 2006), assim como também estaria atuando na estabilização dos arranjos dos cassetes do SI, tal como tem sido proposto (Rowe-Magnus *et al.* 2003).

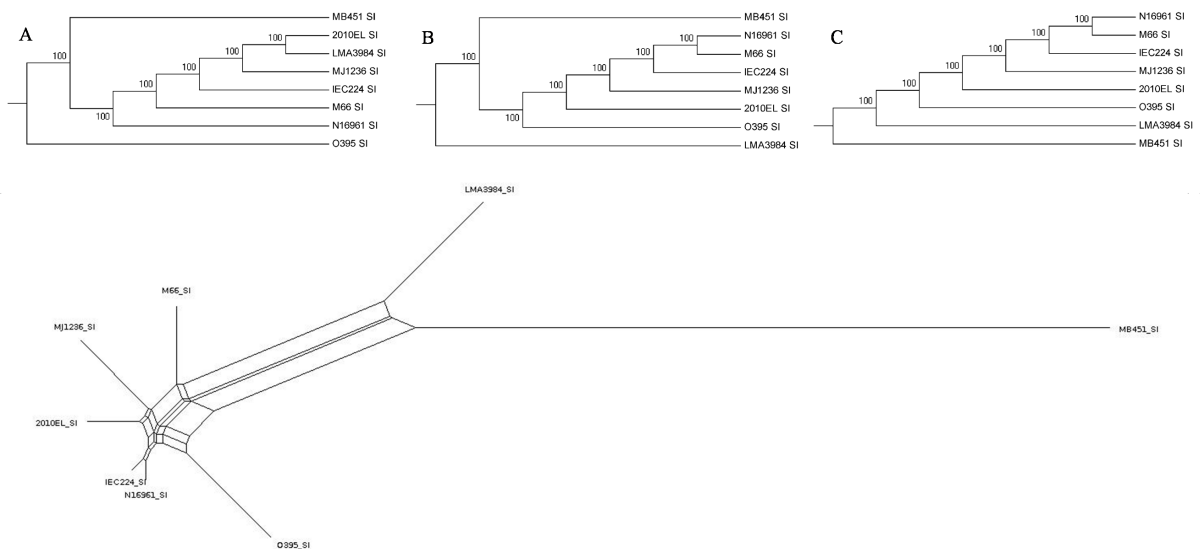


Figura 17. Árvores filogenéticas baseadas nos SNPs dos genes do core do superintegron de *V. cholerae* construídos pelos métodos de máxima verossimilhança (A), Neighbor-Joining (B) e UPGMA (C). Os números indicam os valores de bootstrap. Na parte inferior é mostrado a rede filogenética construída com SplitsTree4 usando o método de conteúdo gênico.

Finalmente, baseados nos polimorfismos de nucleotídeos únicos (*Single Nucleotide Polymorphism, SNP*) do conjunto de genes *core* e no conteúdo gênico do pangenoma, foram construídas árvores filogenéticas (Figura 17). Destas árvores a que apresentou melhor resolução para a distinção das diferentes linhagens de *V. cholerae*, da sua natureza clínica ou ambiental e da evolução da espécie foi a que usava a informação do conteúdo gênico. Isto indicaria que a evolução de *V. cholerae* em diferentes linhagens se reflete na diversidade do seu SI, o qual poderia estar sendo influenciada pela transferência lateral de genes nesta região genômica.

Os resultados destas análises se encontram no artigo “Architecture of the superintegron in *Vibrio cholerae*: identification of core and unique genes” enviado para publicação na *F1000Research* (<http://f1000research.com/articles/2-63/v1>) atualmente aguardando revisão (Marin e Vicente 2013) (Ver Anexos).

5.5 Caracterização das diferentes linhagens de *V. cholerae* baseada no conteúdo gênico do superintegron

O superintegron (SI) é um sistema de captura e excisão de genes de aproximadamente 120 kb, o que corresponde a mais de 10% do tamanho do cromossoma 2 de *V. cholerae*. Estudar o SI a partir de genomas sequenciados apresenta limitações metodológicas devido, principalmente, a falhas na recuperação de sequências e montagem desta região por causa da presença de sequências repetitivas, os VCRs (*Vibrio cholerae repeats*), que flanqueiam os cassetes gênicos. A maioria dos genomas sequenciados não se encontram completos (~90% dos genomas montados de *V. cholerae* se encontram em *contigs*), o que impede a recuperação de todo o conteúdo desta região. Por isto, nosso próximo objetivo foi estabelecer uma metodologia que possibilitasse o estudo do conteúdo do SI de um conjunto de genomas, independentemente de estarem montados ou não, e com isto tentar caracterizar ou distinguir diferentes linhagens a partir deste conteúdo.

Para comparar o conteúdo gênico dos SIs das diferentes linhagens de *V. cholerae*, foram extraídas as regiões do SI de cinco genomas completos de *V. cholerae* (N16961, O395, MJ-1236, M66-2 e LMA3984-4). As leituras do sequenciamento (*reads*) de todos os genomas disponíveis foram mapeadas considerando as regiões dos SIs dos genomas completos. Para isto usou-se BLASTn (Altschul *et al.* 1997) aplicando os parâmetros pré-determinados. Depois foi calculada a profundidade média de *reads* por gene. Foram considerados como genes presentes aqueles que tinham um valor acima do calculado para o gene da integrase *intI4* do organismo correspondente, que é o gene constitutivo de um SI. Um cladograma foi construído a partir de uma matriz euclidiana e de ligação máxima usando a função *heatmap.2* do pacote *gplots* no R (*The R Project for Statistical Computing*; <http://www.r-project.org>).

O conteúdo gênico considerando os cinco SIs mostram uma estratificação da população de *V. cholerae* (Figura 18) na qual pode se apreciar 15 clusters (SI.1-15), onze dos quais (SI.1-11) correspondem ao *V. cholerae* El Tor da sétima pandemia,

incluindo o *cluster* da América Latina (SI.11). Interessantemente, são distinguíveis nestes *clusters* isolados que correspondem ao início da sétima pandemia (SI.11 e SI.12) com um conteúdo no SI similar ao N16961, e isolados contemporâneos onde podemos observar diferenças significativas (grandes deleções e algumas inserções) no conteúdo gênico do SI, que poderiam ser características das novas linhagens dos El Tor atípicos ou alterados, os quais estão sendo os determinantes das atuais epidemias.

O *cluster* da América Latina contém o isolado de Angola (1989) (ERS013245), isolados da Ásia do início da sétima pandemia e isolados da AL. O conteúdo gênico do SI deste *cluster* mostra semelhança com o conteúdo de SIs do início da sétima pandemia (início anos 70) (SI.12) e do próprio N16961 representante El Tor da sétima pandemia do ano 1975 (Figura 18).

SIs de genomas da linhagem clássica se encontram no *cluster* SI.14. O *cluster* SI.15 é o menos consistente; neste se encontram a linhagem Amazônia-Gana, e várias outros isolados não pandêmicos (Figura 18).

Os genes *core* do superintegron, assim como o gene da integrase *intl4* e genes de acetiltransferase foram encontrados em todos os genomas avaliados. Os diferentes *clusters* identificados apresentam vários genes únicos e característicos (Figura 18B). Por exemplo, o *cluster* SI.5 tem uma região característica mas que se encontra também no clado clássico SI.14 (genes 234-251, segundo a referência N16961 ou 558-577 na referência MJ-1236; os números dos genes estão referidos no Anexo III). Estes contêm vários genes que estão amplamente distribuídos no SI.5 codificando proteínas hipotéticas (234, 239, 242, 244, 246, 248, 251, 559, 560, 563, 570, 572, 573, 574, 576, 577 e 578). Além disso, os genes 115, 128, 311, 327 e 348 (relacionados com sistema toxina-antitoxina, biosíntese de sideróforo e proteínas hipotéticas) são específicos para isolados clássicos no *cluster* SI.14.

Os *clusters* do SI.1 ao SI.10 têm perdido duas grandes regiões, uma de 45 genes (98-143, a maioria proteínas hipotéticas) e uma de 20 genes (233-253, todas elas proteínas hipotéticas) que estão presentes no genoma de *V. cholerae* N16961 isolada em 1975.

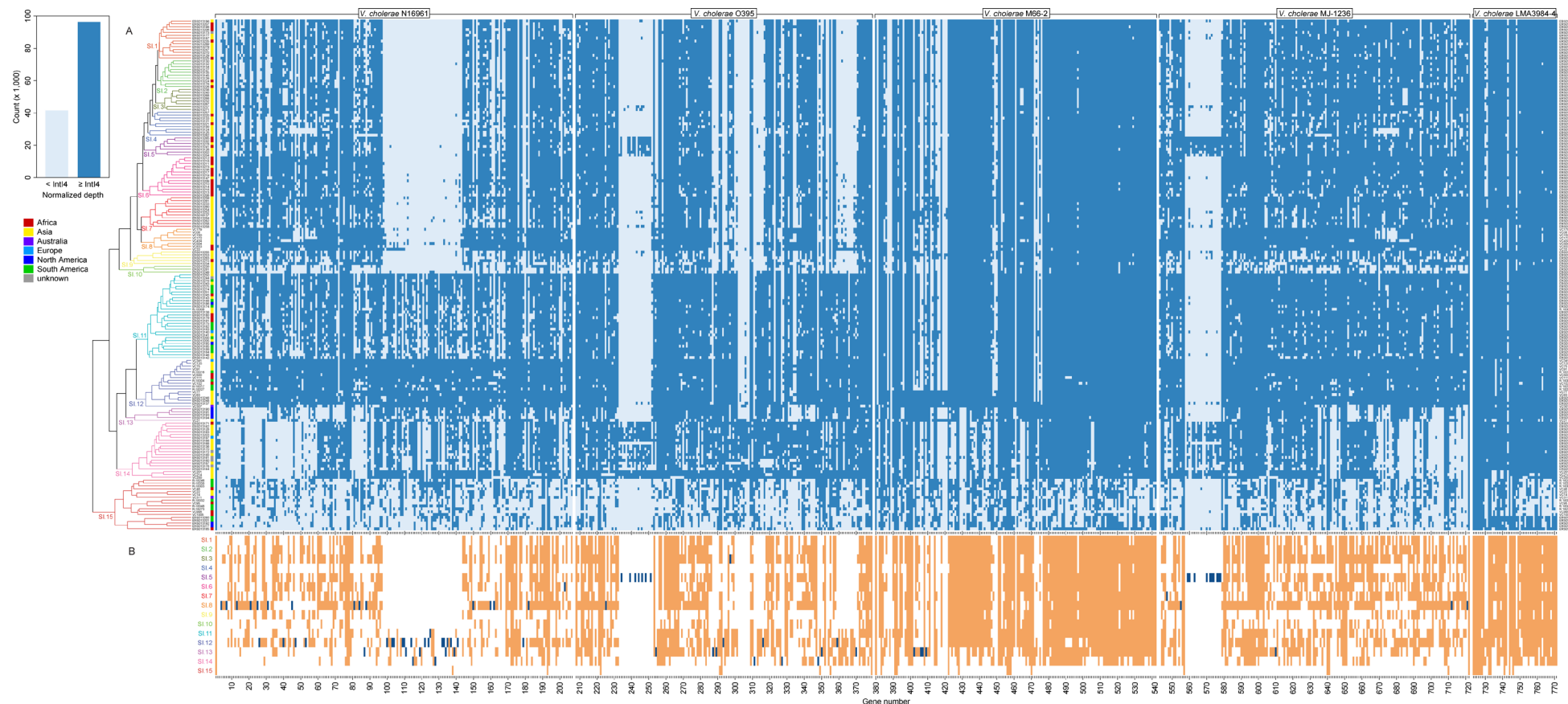


Figura 18. Conteúdo gênico do superintegron de diferentes linhagens de *V. cholerae*. (A) Presença (azul escuro) e ausência (azul claro) dos genes referentes a cinco superintegrons completos (N16961, O395, MJ-1236, M66-2 e LMA3984-4) em 179 genomas de *V. cholerae*. O cladograma mostra a *clusterização* dos organismos baseado na matriz do conteúdo gênico. (B) Os genes *core* (laranja) presentes em todos os isolados de um *cluster* e os genes distintivos (azul) não presentes fora do *cluster*. Os números dos genes estão referidos no Anexo III.

5.6 Caracterização Genética e Genômica de *V. cholerae* da América Latina

5.6.1 Análises filogenéticas e caracterização genética

A heterogeneidade genética de *V. cholerae* foi inicialmente estimada por MLSA. Ao longo de nossos trabalhos, observamos que o gene *recA* teria maior resolução filogenética. Sendo assim, nossa análise filogenética aplicando máxima verossimilhança considerou sempre a sequência deste gene (Figura 19). Os 355 isolados de *V. cholerae* epidêmicos estão localizados em dois principais clados: clássico e El Tor. No clado clássico estão isolados do sudoeste asiático; não existem evidências de linhagens atuais que pertenceriam a este clado. No clado El Tor estão isolados de diferentes tempos e regiões do mundo de *V. cholerae* O1 e isolados de *V. cholerae* O139.

Devido à possibilidade da ocorrência de recombinação envolvendo os genes do MLSA (Didelot e Falush 2007; Paauw *et al.* 2008) e com a finalidade de aplicar uma outra abordagem para inferência filogenética, submetemos o conjunto completo de sequências dos genes do MLSA de 283 dos 355 organismos à análise por ClonalFrame. ClonalFrame determina a relação filogenética de um conjunto de organismos baseado nas mutações pontuais e em eventos de recombinação homóloga (Didelot e Falush 2007) (Figura 20). De um modo geral, a mesma topologia foi mantida nas duas análises, ou seja, eventos de recombinação não estariam afetando de uma forma significativa a aferição da evolução destas linhagens.

O teste da razão do comprimento do ramo externo/interno, na análise pelo ClonalFrame, demonstrou que este conjunto de *V. cholerae* apresenta uma razão comprimento ramo externo/interno significativamente menor que o esperado no caso de um modelo coalescente (Figura 21). Isto sugere que a população contemporânea de *V. cholerae* experimentou uma expansão recente, o que pode significar ou um evento de “gargalo populacional recente” ou epidemias (Smith *et al.* 2000).

Análises por PCR na busca de possíveis marcadores da LA evidenciaram que todos os isolados dos clados clássico e El Tor (com exceção dos O139) são positivos para o locus *rfbN*, característico do sorogrupo O1 (Hoshino *et al.* 1998), e todos são carreadores dos genes *ctxB* e *tcpA*. Nusrin e colaboradores (Nusrin *et al.* 2009) mostram que isolados do Peru da epidemia da AL apresentavam perda de regiões da ilha VSP-II, que caracteriza a linhagem El Tor da sétima pandemia e que conteria 24 ORFs. Duas regiões, VC0511 e VC0513, teriam sido perdidas em todos os isolados testados por eles. Assim, fomos verificar se esta seria uma característica da linhagem da AL. Isolados do início da sétima pandemia de cólera (anos 1970s assim como o prototípico N16961) foram positivos para a presença de VC0511 e VC0513, enquanto que isolados mais recentes só contém VC0513. Os isolados da América Latina (1991-1999), incluindo isolados do Brasil/Bolívia/México/Argentina, foram negativos para as duas regiões. Com estes resultados podemos considerar que a ausência destes dois loci seria uma característica da linhagem AL.

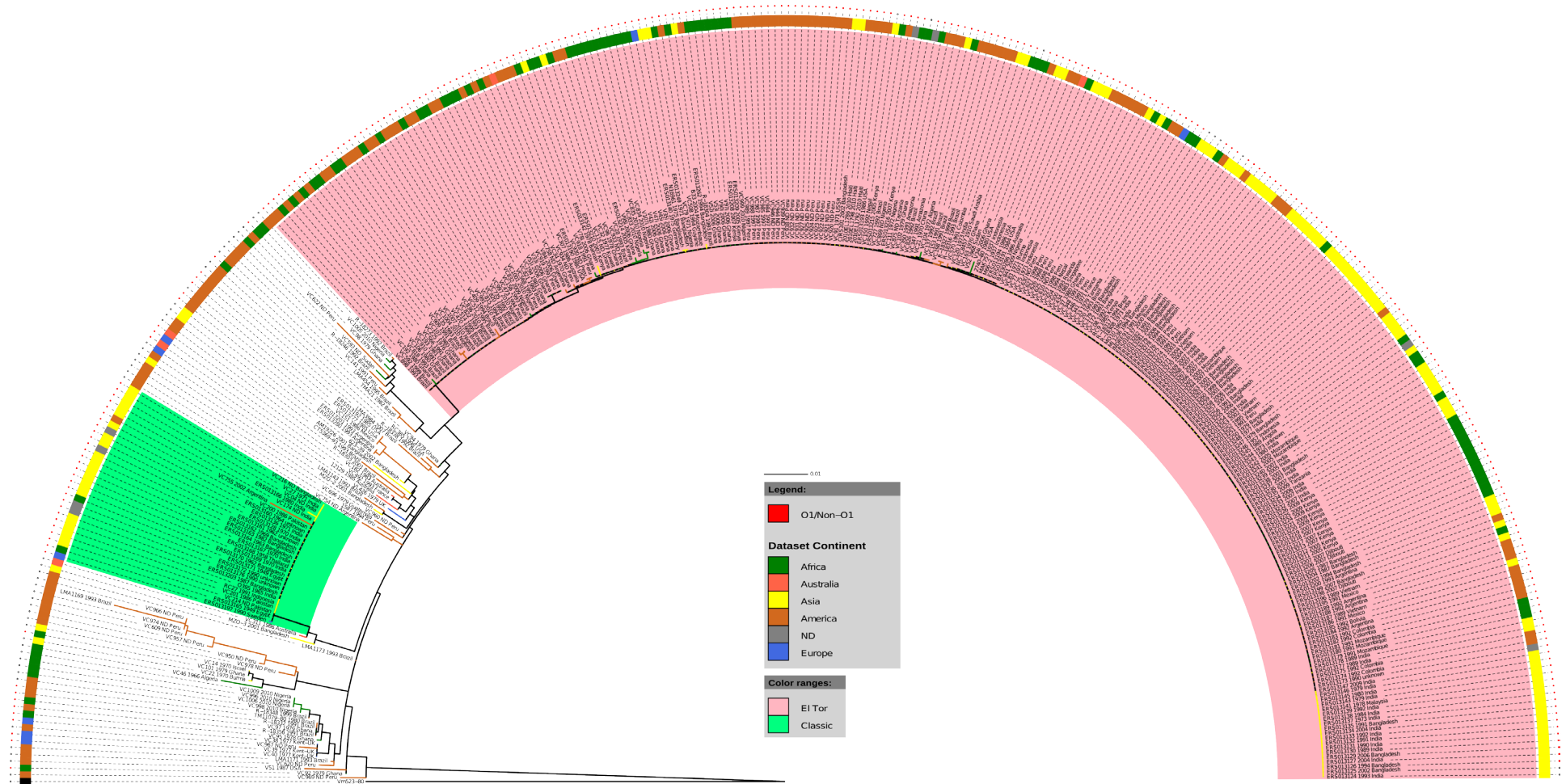


Figura 19. Árvore filogenética de máxima verossimilhança de *V. cholerae* baseada na análise do MLSA. O biotipo clássico e El Tor são mostrados em verde e rosa, respectivamente. Os ramos estão coloridos de acordo com a região de isolamento. A presença do locus *rfbN*, característico do sorogrupo O1, é mostrada como pontos vermelhos nas pontas dos ramos.

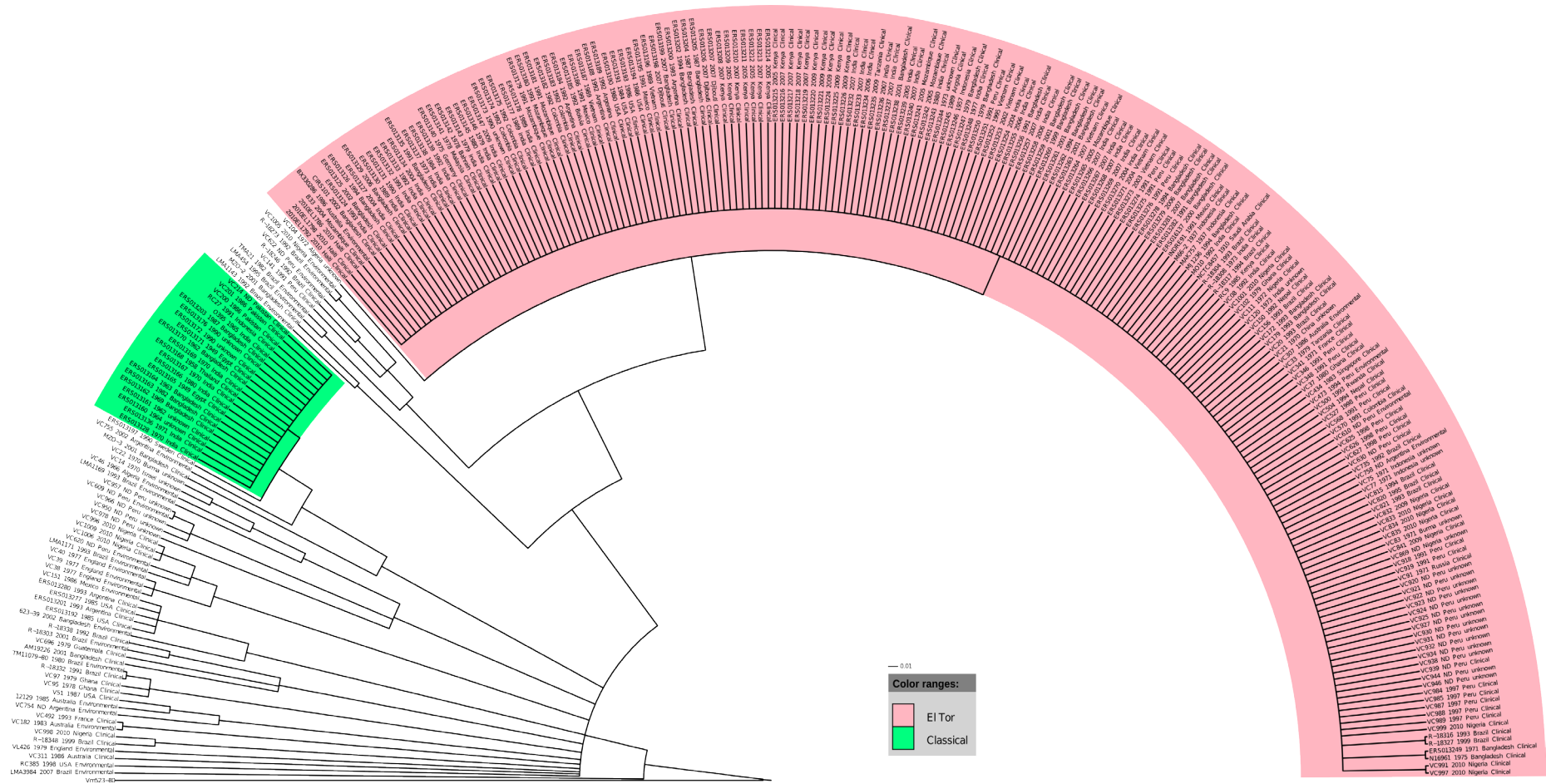


Figura 20. Árvore consenso ClonalFrame para as sequências concatenadas do conjunto de dados do MLSA. A árvore consenso incorpora possíveis eventos de recombinação na reconstrução filogenética. O clado clássico e El Tor são mostrados em cores verde e rosa. *V. mimicus* foi utilizado como grupo externo.

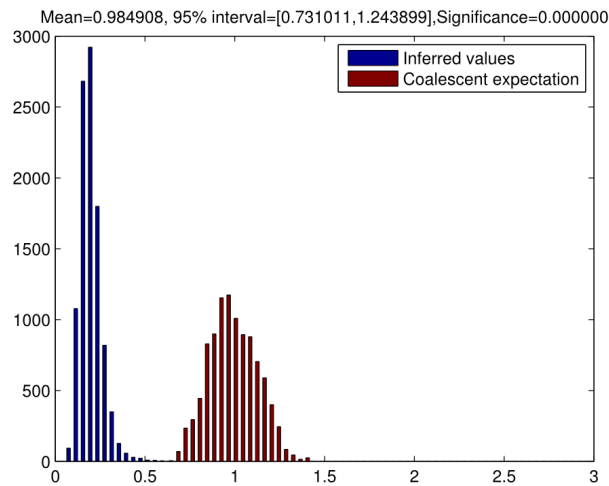


Figura 21. Distribuição das razões do comprimento do ramo externo/interno das árvores resultantes da análise do ClonalFrame comparada a árvores simuladas sob modelo coalescente. *V. cholerae* apresenta uma razão significativamente menor (0.2, $P=0.000$) do que o esperado no modelo coalescente, sugerindo que essa população passou por uma recente expansão do tamanho da população.

5.6.2 Análises genômicas

A linhagem de *V. cholerae* da epidemia da América Latina (AL) pertence ao clado El Tor (Figura 20), onde se encontram os El Tor da sétima pandemia. Isolados desta linhagem também compartilham os mesmos alelos do *ctxB* e *tcpA* que estão presentes nos isolados do início da sétima pandemia. Como já observado a linhagem AL teria como marcador a ausência de duas regiões da ilha VSP-II. Assim, para realizarmos uma análise global, considerando o conteúdo genômico desta linhagem AL, foram selecionados genomas disponíveis nos bancos de dados públicos que foram analisados usando os programas BRIG e PGAP¹³.

Usando o genoma da N16961 como referência, evidenciou-se que todos os genomas da América Latina têm uma deleção em comum na ilha VSP-II (Figura 22). A ilha VSP-II em *V. cholerae* O1 El Tor N16961 é uma região de aproximadamente 26,9 kb abrangendo as ORFs de VC0490 a VC0516 (O'Shea *et al.* 2004a). A deleção de 8,7 kb que foi identificada pelo BRIG e inclui os loci VC0511 a VC0515 do genoma de N16961 (Figura 22), corresponde àquela já observada na caracterização genética.

¹³ BRIG - *BLAST Ring Image Generator* (<http://sourceforge.net/projects/brig/files/>)
PGAP - *Pan-Genomes Analysis Pipeline* (<http://sourceforge.net/projects/pgap/files/>)

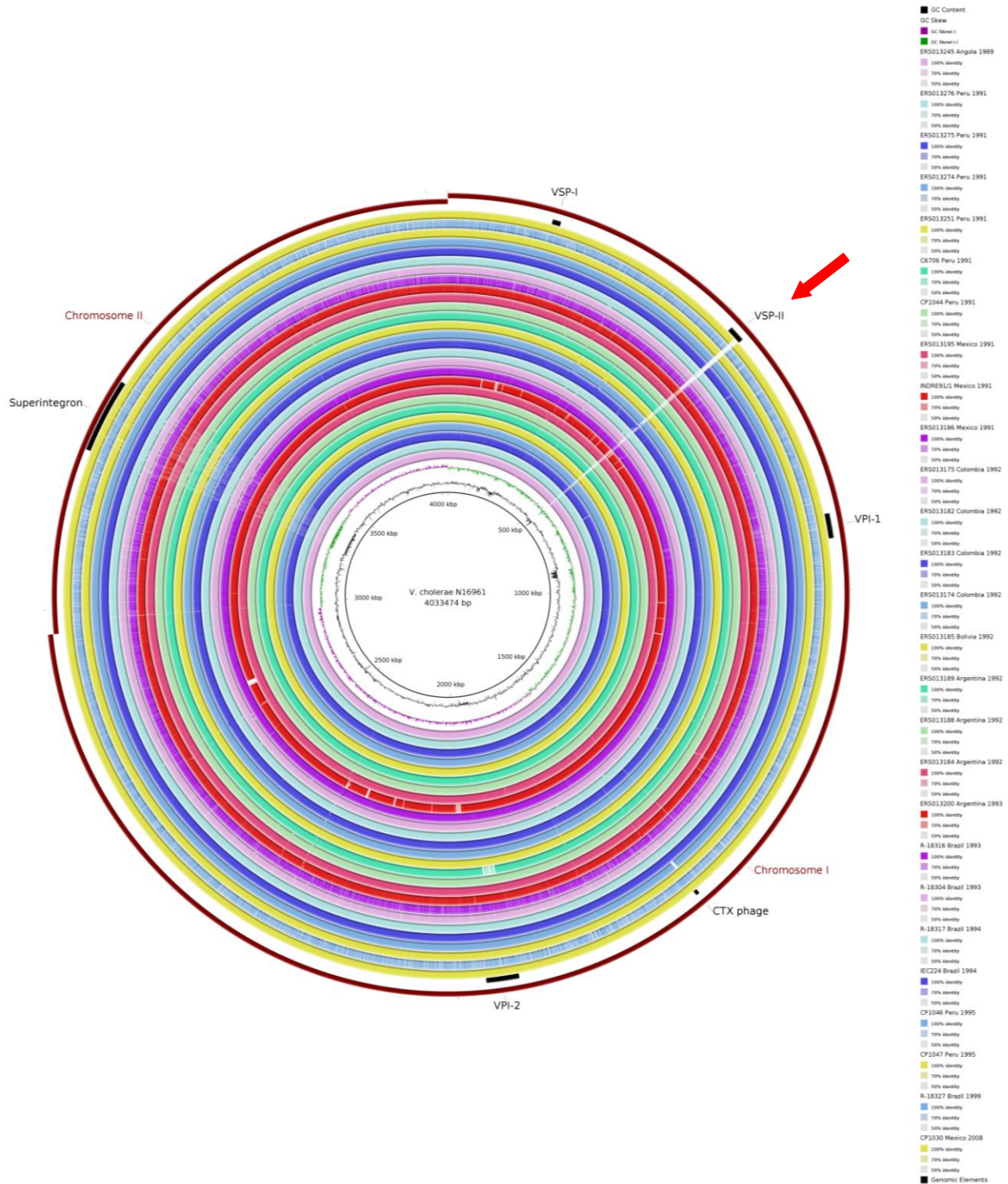


Figura 22. Comparação dos genomas da América Latina usando o genoma N16961 como referência. A seta vermelha indica a ausência de uma região na ilha VSP-II nos genomas da AL. Os círculos mais internos mostram o conteúdo GC (preto) e GC skew (roxo/verde). Os seguintes círculos mostram as comparações por BLAST de 27 genomas contra o genoma de N16961. Os círculos mais externos mostram alguns dos principais elementos genômicos (preto) e as regiões do genoma que corresponderiam ao cromossomo I e II.

Para confirmar a análise feita pelo BRIG, recuperamos a VSP-II de um representante da epidemia da América Latina (neste caso o genoma IEC224) e do canônico N16961 e as mesmas foram comparadas. Observou-se que, de fato, a região que compreende as ORFs VC0511-VC0515 não está presente na IEC224; no seu lugar encontra-se uma sequência de 7077 pb, com 39,2% de conteúdo GC (próximos aos 40% de conteúdo médio de GC da VSP-II). Esta região, que seria exclusiva da linhagem da América Latina, contém três ORFs codificadoras de proteínas hipotéticas (Figura 23). Fazendo buscas pelo NCBI-BLAST, até o momento não foram encontrados *hits* com outros organismos.

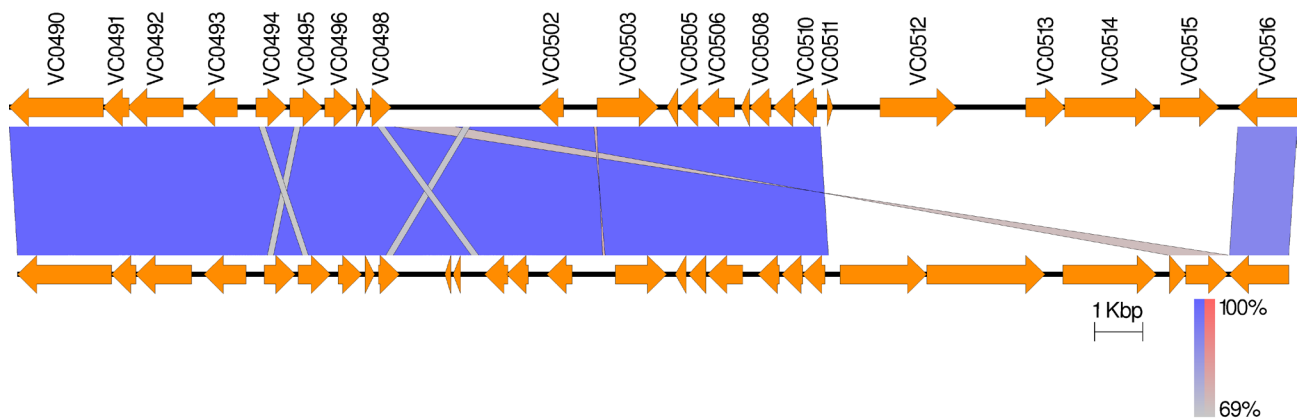


Figura 23. Organização genética e comparação da ilha VSP-II em *V. cholerae* El Tor N16961 (parte superior) com a VSP-II do genoma da IEC224 (parte inferior). IEC224 foi usada como referência dos genomas da América Latina. Os blocos verticais entre as sequências indicam as regiões semelhantes sombreadas segundo a similaridade do BLASTn (escala indicada na parte inferior direita). Uma região que abrange os loci VC0511-VC0515 da N16961 não é compartilhada na IEC224, a qual possui, nesta região, três ORFs diferentes.

Com a finalidade de conhecer a variabilidade da VSP-II, comparou-se a VSP-II canônica com um genoma atual, neste caso o 2010EL1786, que é do Haiti (2010). Observou-se que a VSP-II presente no genoma 2010EL1786 tinha 13.747 pb enquanto que a VSP-II da canônica possui 26.107 pb; estas deleções envolvem a perda de 15 ORFs em relação a canônica N16961 (Figura 24), o que demonstra a plasticidade e diversidade desta ilha genômica e seu potencial papel como marcador das epidemias de cólera.

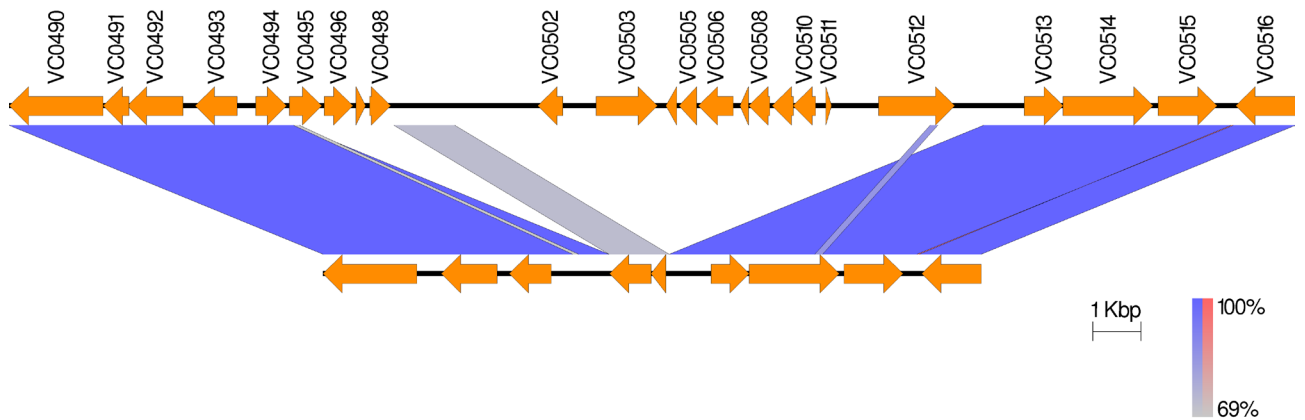


Figura 24. Comparação da VSP-II dos genomas N16961 (parte superior) e 2010EL1786 (parte inferior). Os blocos verticais entre as sequências indicam as regiões semelhantes sombreadas segundo a similaridade do BLASTn. A escala e variação de similaridade são mostradas na parte inferior direita. As deleções na 2010EL1786 envolvem a perda de 15 ORFs em relação à canônica N16961.

Um locus interessante para ser explorado é o VC0516 da N16961, que codifica uma integrase que possui 91% de identidade nucleotídica com a mesma região na linhagem AL. Então, realizamos uma análise filogenética com este gene (Figura 25A). A árvore mostra que os isolados da América Latina formam um clado coeso e distante daquele que contém os demais isolados de *V. cholerae* da sétima pandemia (Figura 25A). Paralelamente, usando o BLAST do NCBI, foi feita uma busca por sequências similares à integrase do IEC224 (O3Y_02430). Surpreendentemente, uma integrase de *Vibrio vulnificus* (YJ016) apresentou 98% de identidade nucleotídica com a IEC224 enquanto que a similaridade com N16961 (*V. cholerae* da sétima pandemia) é de 91%. As relações filogenéticas destas sequências estão representadas na Figura 25B. A comparação filogenética das integrases da VSP-II incluindo *V. vulnificus* YJ016 está mostrada no Anexo VI.

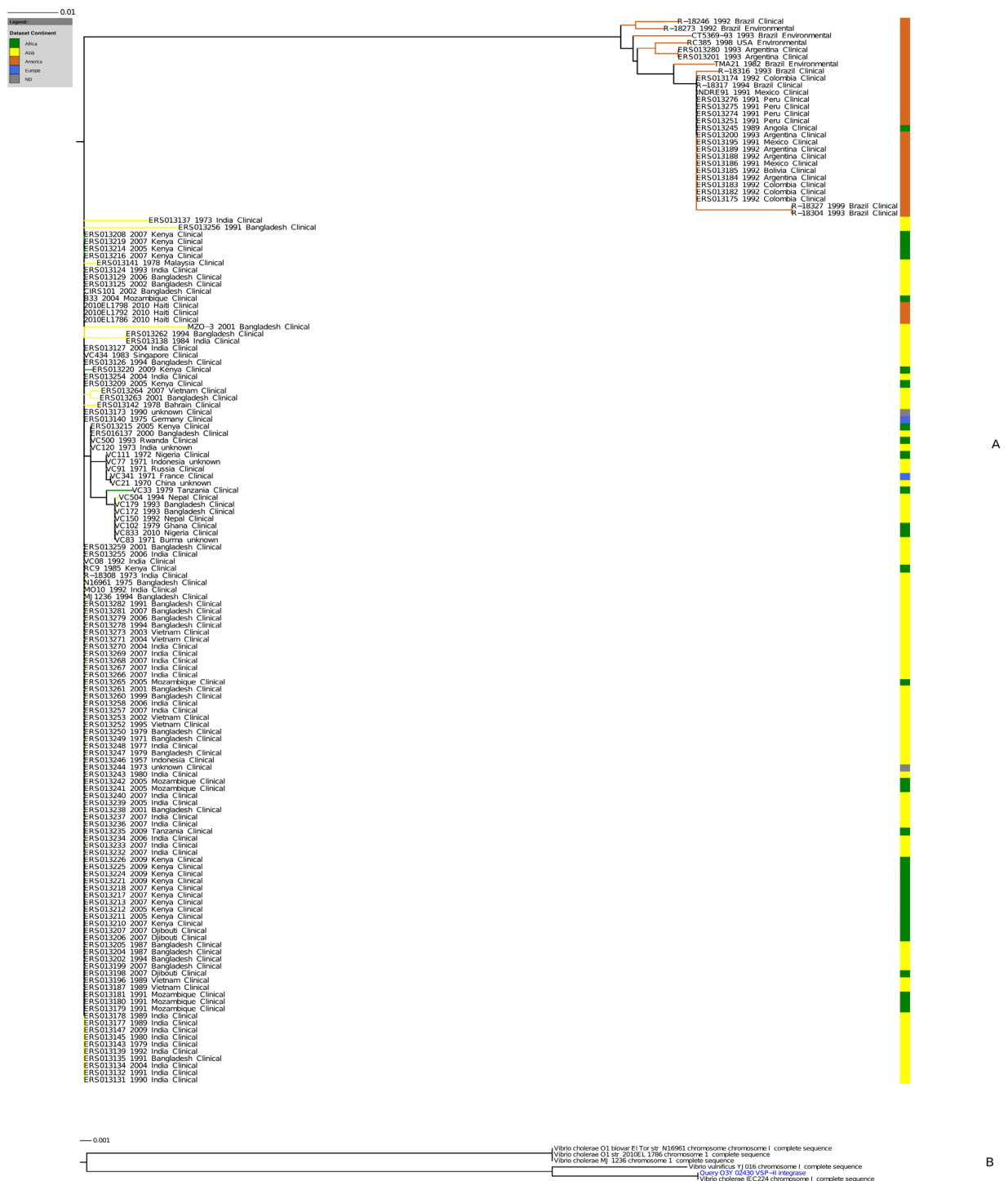


Figura 25. Relações filogenéticas da integrase da VSP-II. (A) Árvore de máxima verossimilhança de homólogos de VC0516 entre a população de *V. cholerae*. Os isolados da América Latina estão posicionados em um clado coeso. (B) Árvore de evolução mínima construída com seqüências homólogas de VC0516 encontradas no *Genome-BLAST* no NCBI, entre os *V. cholerae* está o *V. vulnificus*. A seqüência consulta representativa O3Y_02430 do genoma da IEC224 se mostra em azul.

Comparamos todos os genomas da América Latina, usando o IEC224 como referência, ao N16961, buscando identificar regiões exclusivas e comuns aos genomas da América Latina (Figura 25). Identificamos duas regiões, uma correspondente ao profago WASA1, que havia sido recentemente caracterizado (Garza *et al.* 2012). Realizando buscas com BLAST no [NCBI](#) (*Microbial Nucleotide BLAST*) e [ENA](#) (*European Nucleotide Archive*), encontramos *hits* positivos para sequências deste profago, além do genoma IEC224, em *Vibrio vulnificus* MO6-24/O, 88% de cobertura (dos 49.206 pb e 96% de identidade máxima), e *Vibrio parahaemolyticus* AN5034, 48% de cobertura e 84% de identidade máxima. Em relação à região exclusiva da VSP-II da linhagem da América Latina, não foram encontrados *hits* positivos, além dos com os genomas da linhagem da América Latina.

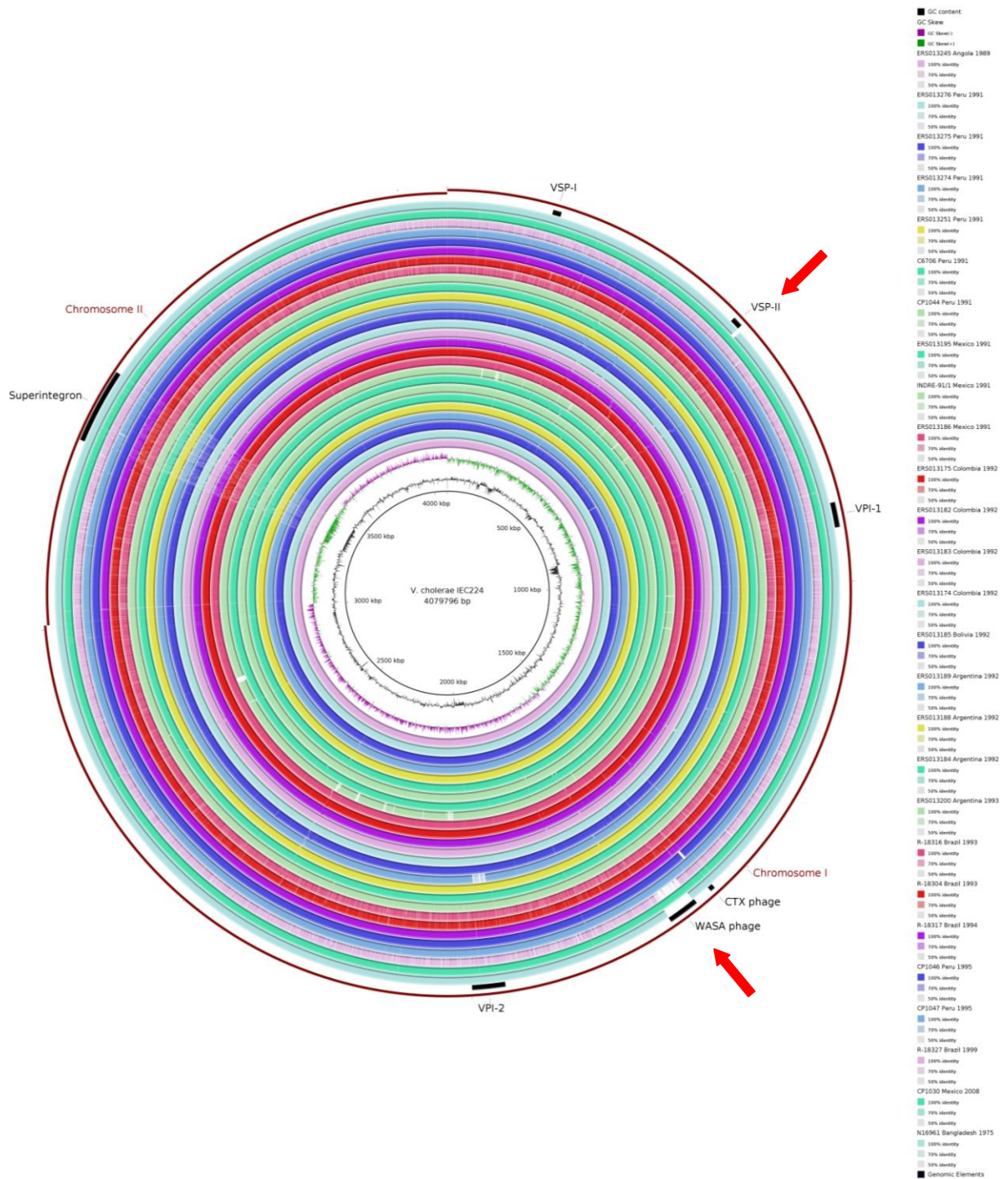


Figura 26. Comparação dos genomas da América Latina usando o IEC224 como referência. Os anéis mais internos mostram o conteúdo GC (preto) e GC skew (roxo/verde). Os anéis restantes mostram as comparações por BLAST de 24 outros genomas de *V. cholerae* frente ao genoma do IEC224. O anel externo mostrado em azul claro corresponde ao genoma de N16961. Os anéis mais externos destacam alguns dos principais elementos genômicos (preto) e as regiões que correspondem aos cromossomos I e II. As setas vermelhas indicam duas regiões comuns para os genomas da AL e diferentes da N16961: uma região distintiva na ilha VSP-II e a presença do profago WASA1.

As duas últimas análises foram realizadas considerando os genomas N16961 (Figura 24) e IEC224 (Figura 26); esta abordagem permitiu identificar, na linhagem da América Latina, marcadores genômicos como o profago WASA1 e a região exclusiva na VSP-II, mas também a integrase da VSP-II a qual poderia ser usada como marcador filogenético. Realizamos uma terceira análise independente de uma referência. Extraímos genes ortólogos (genes homólogos compartilhados por dois ou mais genomas) a partir dos genomas da América Latina e genomas de linhagens referenciais, objetivando definir genes do *core*, genes dispensáveis e genes únicos.

Foram considerados 24 genomas, dos quais três de linhagens referenciais (N16961, BX330286 e M66-2). As sequências codificadoras extraídas foram *clusterizadas* e um total de 7009 *clusters* (ortólogos) foi obtido. Neste conjunto foram definidos para a linhagem da América Latina conjuntos de genes mostrados na Figura 27.

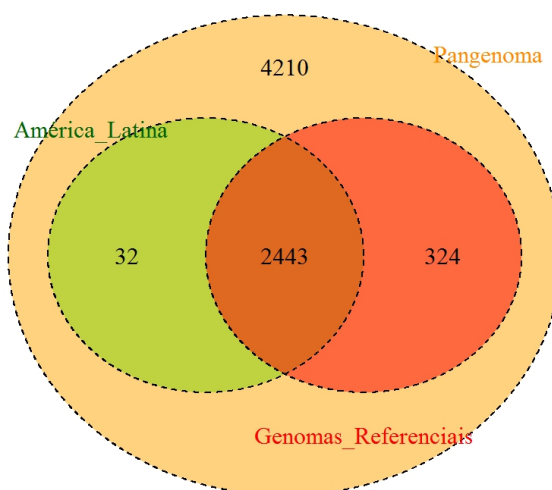


Figura 27. Pangenoma e conjunto de genes ortólogos da linhagem da América Latina e genomas referenciais. Do total de 7009 *clusters* de ortólogos obtidos (pangenoma), 2443 foram comuns para o conjunto de genomas avaliados (genoma *core*), 324 não foram compartilhados por nenhum genoma da AL e 32 foram exclusivos dos genomas da AL.

Como validações, foram feitas buscas pelo NCBI-BLAST com os 32 genes *core* exclusivos e encontramos que os loci O3Y_02350, O3Y_02360 e O3Y_01320, tinham *hits* positivos no genoma da N16961. Contudo, estes *hits* correspondem aos loci VC0284, VC0500 e VC0501, que apresentavam mudança de quadro de leitura (*authentic frame shift*), o que ocasionou este viés. Todos estes falsos positivos

(falsos genes *core*) foram excluídos do set de genes *core* da linhagem da América Latina.

Os 29 genes *core* exclusivos da linhagem da América Latina foram mapeados no genoma da IEC224 (Figura 28), mostrando que pertencem às regiões do profago WASA1 e à ilha VSP-II. Estes resultados confirmam que estas duas regiões seriam marcadores da linhagem da América Latina.

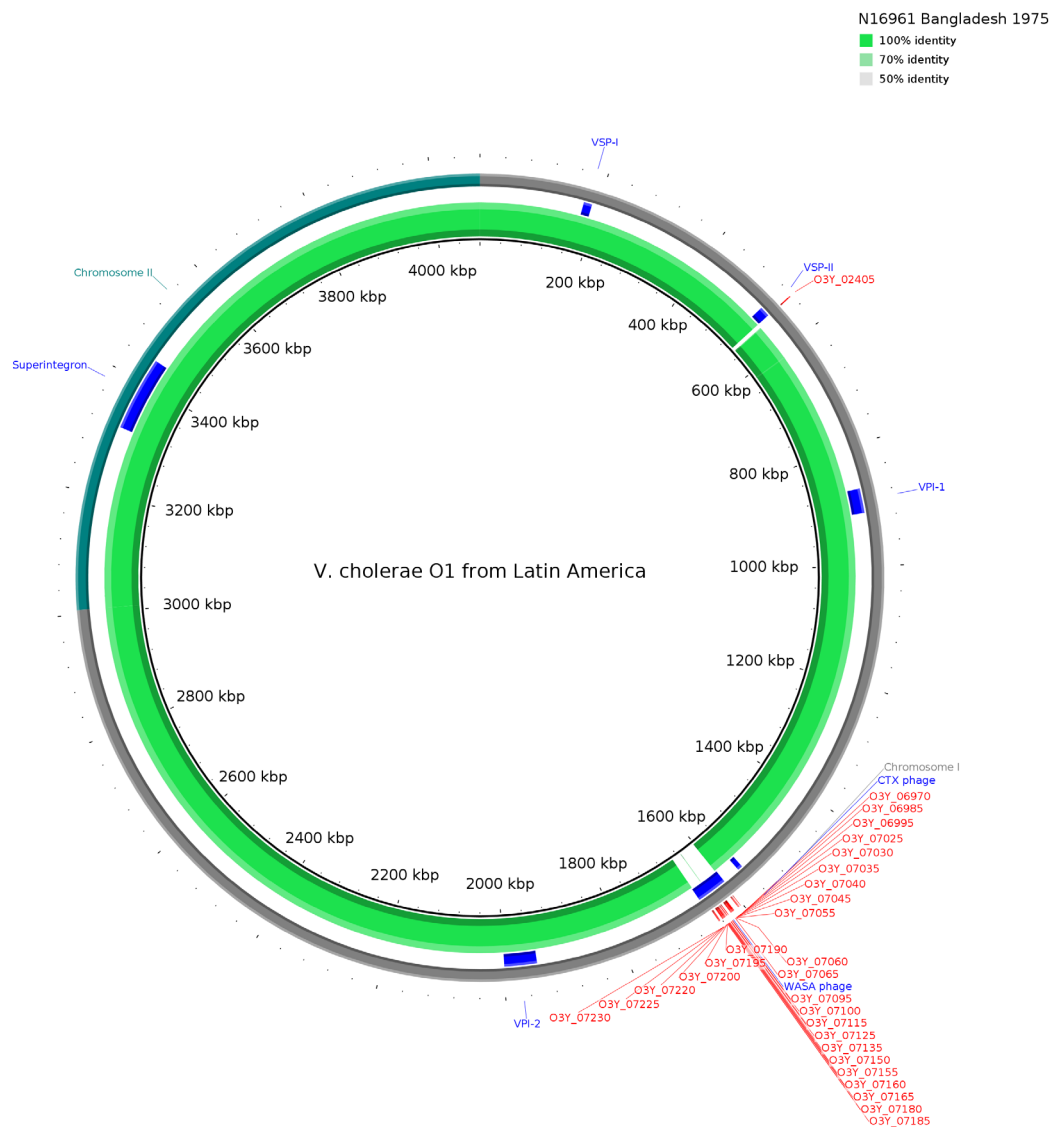


Figura 28. Genes *core* exclusivos identificados pela análise de ortólogos na linhagem da América Latina. A figura mostra a comparação do genoma da América Latina IEC224 (círculo interno) com o genoma canônico N16961 (círculo verde). As etiquetas do locus (*locus tag*) dos genes exclusivos são mostradas em vermelho.

Com a finalidade de conhecer as relações filogenéticas a partir do conteúdo gênico, foi construída uma rede de divisão (*split network*) com os 7009 genes ortólogos do conjunto de genomas de *V. cholerae* (Figura 29). Devido à grande parte de nosso *set* de genomas encontrar-se em *contigs*, esta análise permitiu avaliar, além do conteúdo gênico dos genes cromossomais, o conteúdo gênico de genes extra-cromossomais como os plasmídeos, que poderiam estar associados a estes genomas e que poderiam estar caracterizando dois ou mais genomas. Esta abordagem é importante para conhecer as relações dos isolados, baseada em genes compartilhados, e poderia ajudar na compreensão do componente ecológico ou patogênico dos isolados de *V. cholerae*.

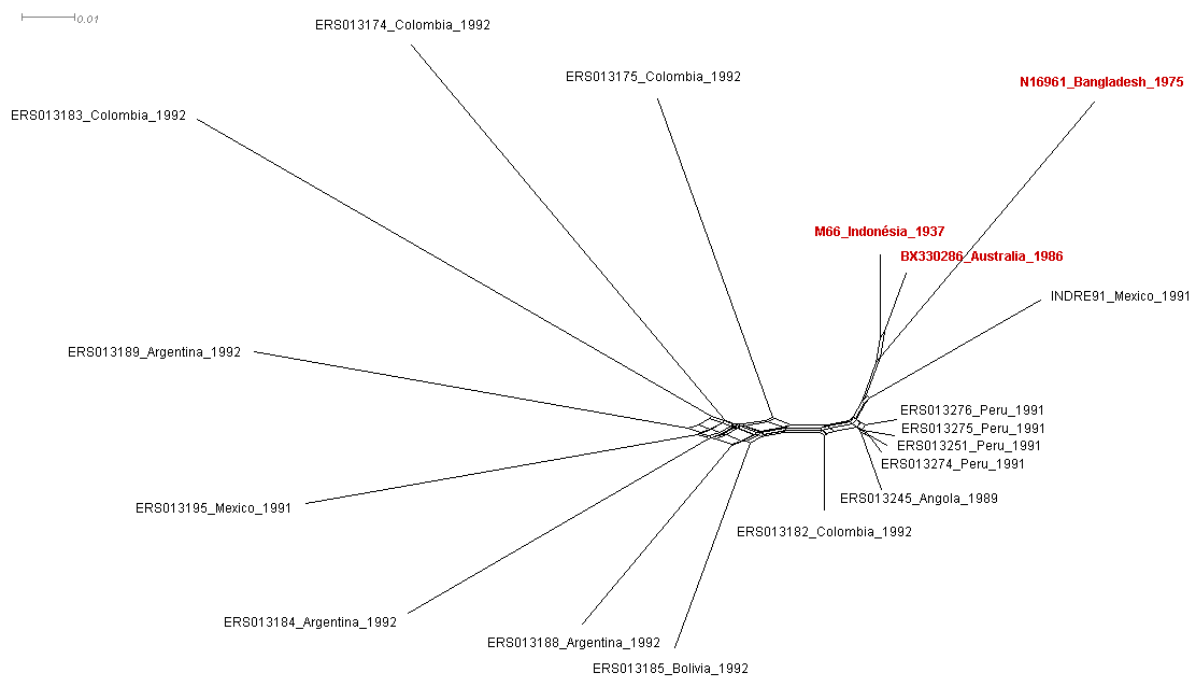


Figura 29. *Split network* para o conteúdo gênico dos genomas de *V. cholerae* da linhagem da América Latina e genomas referenciais. A rede filogenética foi construída com 7009 genes usando o método de distância por conteúdo gênico (Huson e Steel 2004) implementado no programa SplitsTree4 (Huson e Bryant 2006). Os genomas referenciais são mostrados em vermelho.

A Figura 29 mostra que a linhagem da América Latina está mais próxima do genoma BX330286, que corresponde a uma cepa de *V. cholerae* O1 El Tor isolada na Austrália em 1986, do que com os outros dois genomas referenciais, que são pré-El Tor (M66 asiático) e do início da sétima pandemia asiática (N16961).

A Figura 29 mostra também uma estreita relação entre os genomas ERS013251, ERS013274, ERS013275 e ERS013276, todos do início da epidemia da América Latina (Peru 1991). Por outro lado mostra a evolução individual de isolados da AL, causada por aquisição de elementos extra-core, já que encontramos genomas de 1991 e 1992 apresentando sequências únicas.

5.7 Identificação e caracterização do Elemento Integrativo e Conjugativo de *V. cholerae* pandêmico

Nossas análises até aqui mostraram que as características genéticas da linhagem da América Latina são similares às das linhagens do início da sétima pandemia (clonais pelo MLSA, perfil de PFGE relacionado geneticamente, genótipo 3 do *ctxB*, *tcpA* canônico, sensibilidade aos antimicrobianos); já os isolados atuais se caracterizam por terem outro perfil alélico para *ctxB* e *tcpA* e apresentar resistência a alguns antimicrobianos. A emergência de *V. cholerae* com fenótipo de multi-resistência é contemporânea à aquisição dos elementos integrativos e conjugativos (*Integrative and Conjugative Elements - ICEs*), os quais foram identificados em isolados de epidemias de cólera na Ásia, África e América Central (Mutreja *et al.* 2011).

Nossa seguinte motivação partiu com a emergência de uma nova epidemia de cólera na América Latina, desta vez no Haiti, em outubro de 2010 (Chin *et al.* 2011). Como o Haiti não estava entre os países atingidos pela epidemia de cólera da AL dos anos 90, pensou-se que a linhagem da AL seria a responsável pela epidemia.

Um estudo aplicando a epidemiologia genômica (janeiro de 2011) demonstrou que a linhagem da epidemia do Haiti não estava relacionada à da AL; além disso, mostrou que esta linhagem estava relacionada às linhagens do sudoeste da Ásia (Chin *et al.* 2011). Além de SNPs que as relacionavam evolutivamente, estes isolados apresentavam um fenótipo de multi-resistência aos antibióticos que não era algo comum aos *V. cholerae* epidêmicos do início da sétima pandemia (Reimer *et al.* 2011; Sjölund-Karlsson *et al.* 2011). Mutreja e colaboradores (Mutreja *et al.* 2011) identificaram que elementos genômicos do tipo SXT/ICE caracterizam as linhagens das epidemias atuais. Assim, fomos buscar em *V. cholerae* O1 da Coleção de Bactérias do Ambiente e Saúde (CBAS) pela presença de elementos genéticos associados com resistência a antibióticos, como os ICEs, integrons de classe 1 e 2.

Identificamos em 12 isolados da Nigéria de 2010 (Marin *et al.* 2013) e um do Nepal de 1994 o gene da integrase *int*_{SXT}, indicando a presença do elemento SXT/ICE. Os integrons de classe 1 e 2 não foram encontrados nestes isolados.

Também foram identificadas mutações nos genes *gyrA* e *parC* associadas com a resistência a quinolonas.

Um isolado da linhagem da Nigéria e um do Nepal foram selecionados para a caracterização de seus ICEs. Em um estudo prévio, o conteúdo gênico de vários elementos da família SXT havia sido minuciosamente analisado, mostrando que estes elementos apresentam regiões comuns e variáveis e, em alguns deles, conjuntos de genes de resistência presentes nas regiões variáveis (Wozniak *et al.* 2009). Baseados nesta informação, recuperamos por sequenciamento o conteúdo gênico dos SXT/ICE dos isolados do Nepal e da Nigéria que foram nomeados como ICEVchNep1 e ICEVchNig1 (GenBank KC886257 e KC886258, respectivamente). Comparando estes elementos a outros ICEs, identificamos outros ICEs que tem o mesmo conteúdo, as mesmas 94 ORFs, incluindo genes do *core*, as regiões variáveis e as cinco regiões HS, previamente identificadas em outros três ICEs: ICEVchHai1, ICEVchInd5 e ICEVchBan5, identificados e caracterizados em linhagens de *V. cholerae* epidêmicas multi-resistente do Haiti (Ceccarelli *et al.* 2011b; Sjölund-Karlsson *et al.* 2011), Índia (Ceccarelli *et al.* 2011a) e Bangladesh (Wozniak *et al.* 2009), respectivamente. Os ICEVchNig1 e ICEVchNep1 se diferenciam do ICEVchHai1 por dois e oito SNPs, respectivamente. De acordo com nossas análises, estes ICEs compartilham o mesmo conteúdo gênico e sintonia; portanto, definimos estes ICEs como ‘irmãos’ do ICEVchHai1, devido ao impacto epidemiológico da cólera no Haiti em 2010.

As regiões mais variáveis dos elementos ICEs estão presentes nos cinco HS (Wozniak *et al.* 2009). Comparando as sequências HS do ICEVchHai1 usando NCBI-BLASTn (outubro de 2012), foi verificado que o conteúdo das regiões HS4 (5394 pb) e HS5 (15040 pb) são únicos e característicos dos ICEs ‘irmãos’ do ICEVchHai1. Portanto, estas regiões, assim como o gene *int_{SXT}* (marcador da presença do ICE), poderiam ser usadas como marcadores epidemiológicos do ICEVchHai1 e seus irmãos.

Aplicamos estes dois marcadores, junto com o gene da integrase do SXT (*int_{SXT}*), para avaliar 202 genomas de *V. cholerae* disponíveis no GenBank e os descritos por Mutreja e colaboradores (Mutreja *et al.* 2011) e Hendriksen e colaboradores (Hendriksen *et al.* 2011). Identificamos o ICEVchHai1 em 62 genomas: Quênia (2005-2009, n=17), Índia (1984-2009, n=15), Bangladesh (1991-

2007, n=13), Nepal (1994 e 2010, n=9), Djibouti (2007, n=3), Haiti (2010, n=3), Nigéria (2010, n=1) e Tanzânia (2010, n=1) (Tabela 9 e Figura 30).

Tabela 9. Genomas de *V. cholerae* positivos para ICEVchHai1.

País	Ano	Isolados	Número de acesso
Índia	1984	1	ERS013138
	1989	3	ERS013130, ERS013177, ERS013178
	1990	1	ERS013131
	2004	4	ERS013127, ERS013134, ERS013254, ERS013270
	2005	1	ERS013239
	2006	2	ERS013255, ERS013258
	2007	2	ERS013237, ERS013269
	2009	1	ERS013147
Bangladesh	1991	2	ERS013135, ERS013282
	1994	1	ERS013262
	1999	1	ERS013260
	2001	4	ERS013238, ERS013259, ERS013261, ERS013263
	2002	1	ACVW00000000
	2006	2	ERS013129, ERS013279
	2007	2	ERS013199, ERS013281
Quênia	2005	5	ERS013209, ERS013211, ERS013212, ERS013214, ERS013215
	2007	7	ERS013208, ERS013210, ERS013213, ERS013216, ERS013217, ERS013218, ERS013219
	2009	5	ERS013220, ERS013221, ERS013224, ERS013225, ERS013226
Nepal	1994	1	VC504
	2010	8	SRX082837, SRX082838, SRX082839, SRX082940, SRX082941, SRX082942, SRX082943, SRX082944
Haiti	2010	3	AELH00000000, AELJ00000000, AELI00000000
Nigéria	2010	1	VC 833
Tanzânia	2010	1	ERS013235
Djibouti	2007	3	ERS013198, ERS013206, ERS013207

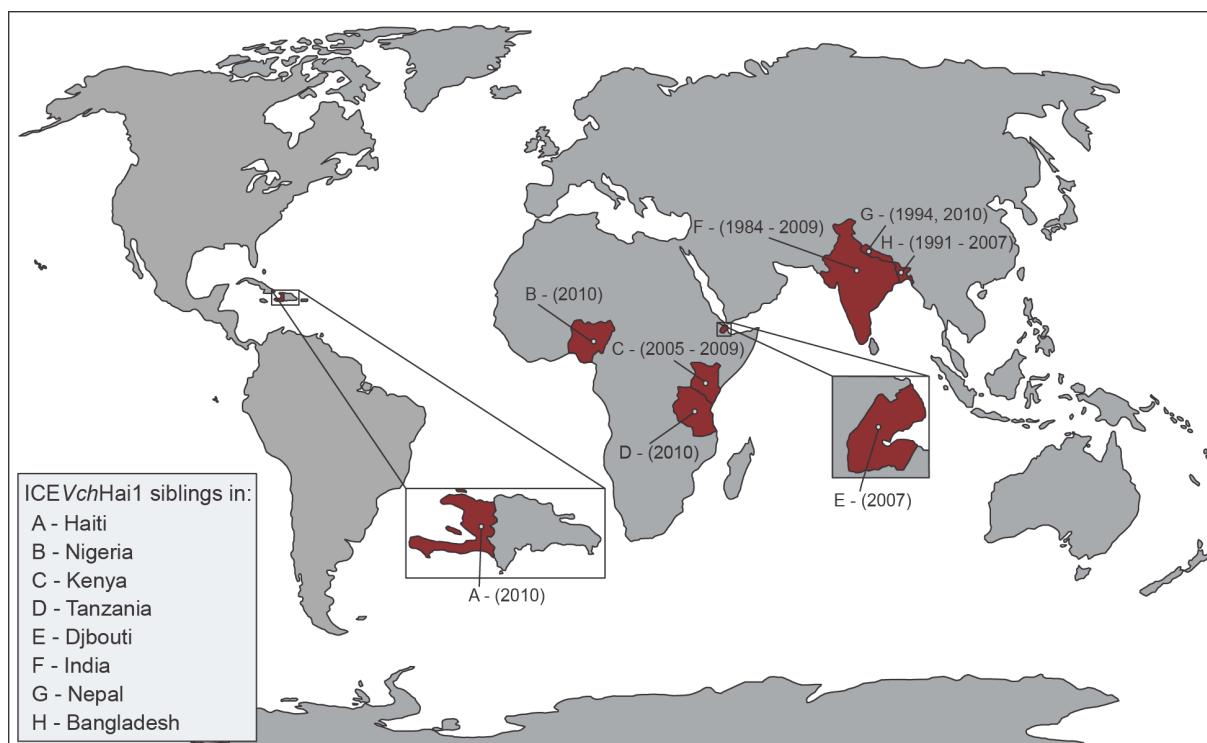


Figura 30. Distribuição espaço/temporal de membros da família ICEVchHai1. Os países e as datas dos isolamentos são mostrados.

Considerando que os membros da família ICEVchHai1, caracterizados pela presença das regiões HS4 e HS5, estão espalhados em diversos continentes, e que sequências das regiões características HS4 e HS5 não tem sido encontradas em nenhum outro contexto, tivemos como finalidade rastrear as possíveis origens destes elementos no ambiente. Para isto, estabelecemos buscas destas regiões na base de dados metagenômicos do CAMERA (<http://camera.calit2.net/>, Outubro 2012) por BLASTn com um limiar de similaridade de 10^{-20} . Embora não tenhamos encontrado *hits* para a região HS4, oito *hits* relacionados às ORFs da região HS5 estão presentes em dois metagenomas: dois *hits* correspondem as ORFs 4 e 5 de um manguezal na ilha Fernandina no Galápagos, e seis *hits* correspondentes às ORFs 1, 3 e 4 foram encontrados na Antártida. Nestes metagenomas também foram encontrados *hits* positivos para o gene da integrase *int*_{SXT}.

5.8 Presença de regiões genômicas de interesse epidemiológico na população de *V. cholerae*

Considerando os resultados até aqui encontrados baseados nas análises de diferentes marcadores epidemiológicos de *V. cholerae* tradicionalmente conhecidos e outros aqui identificados, nossa última análise consistiu em fazer uma busca destes marcadores (Tabela 3) na população de genomas de *V. cholerae* (Anexo I).

Os integrons são elementos de captura de genes caracterizados pela integrase que possuem. Inicialmente fizemos buscas por BLASTn das integrases características de integrons de resistência *IntI*, *IntI2*, *IntI3*, e do superintegron, *IntI4*. Genes da integrase do tipo 1 (*intI*) foram encontrados em isolados clínicos de Moçambique (2004-2005), Bangladesh (1979,1991), Índia (1991) e um isolado da Argentina (1993). Estudos de caracterização genética de determinantes de resistência presentes em isolados da América Latina identificaram um isolado ambiental do Peru (VC622), do período da epidemia, carreando um integron de classe 1 contendo o cassete gênico *aadA1*. Observou-se que todos os genomas que foram positivos para o gene da integrase do tipo 1, também foram positivos para este arranjo gênico, com exceção do isolado RC9 da África (1985), que apresentou somente o gene da integrase. *intI4* foi um controle positivo, já que é característica do SI, que é constitutivo de *V. cholerae*. Entretanto, a integrase de tipo 3 (*intI3*), encontrada em linhagens de *Klebsiella pneumoniae*, não está presente neste conjunto de genomas de *V. cholerae*. A integrase de tipo 2 (*intI2*) ocorreu em um único O1 clínico de Gana de 1978. A integrase do elemento conjugativo e integrativo ICE (*int_{SXT}*) foi identificada em vários isolados de epidemias recentes de cólera no mundo.

Os genes VC0511 e VC0513 da VSP-II estão ausentes nos genomas da linhagem da América Latina, e VC0511 não foi encontrado no genoma de El Tor atípico CIRS101.

Nossos resultados constataram, que de fato, os genes *ctxB* e *tcpA* definem linhagens epidêmicas; assim como a ausência da região VSP-I é característica da linhagem clássica. Por sua vez, considerando o conteúdo gênico como um todo das ilhas VSP-II, VPI-1 e VPI-2, observou-se que quase toda a população avaliada continha regiões dessas ilhas, indicando que a presença das mesmas não seria

definidora de linhagem. Contudo, o gene VC0516, que codifica a integrase desta VSP-II, caracterizou isolados da sétima pandemia e alguns outros variantes da América Latina e da Ásia, sendo este gene, portanto, um marcador mais definidor de linhagens da sétima pandemia considerando a VSP-II. Constatou-se também que o gene VC2346 é constitutivo da população de *V. cholerae*, porém apresenta um alelo específico que caracteriza a linhagem da AL, podendo ser usado como outro marcador filogenético para distinguir esta linhagem.

Finalmente, observamos uma correspondência da presença do WASA1 e da região única do VSP-II, previamente identificada, em todos os isolados da linhagem da América Latina. Contudo, foi observada a presença da região única do VSP-II em dois genomas de isolados de linhagens diferentes daquela da América Latina (R-18273, Brasil 1992 e VC21, China 1970) que não continham WASA1. Portanto, a presença destes dois marcadores (WASA e região única da VSP-II) definiria a linhagem da América Latina.

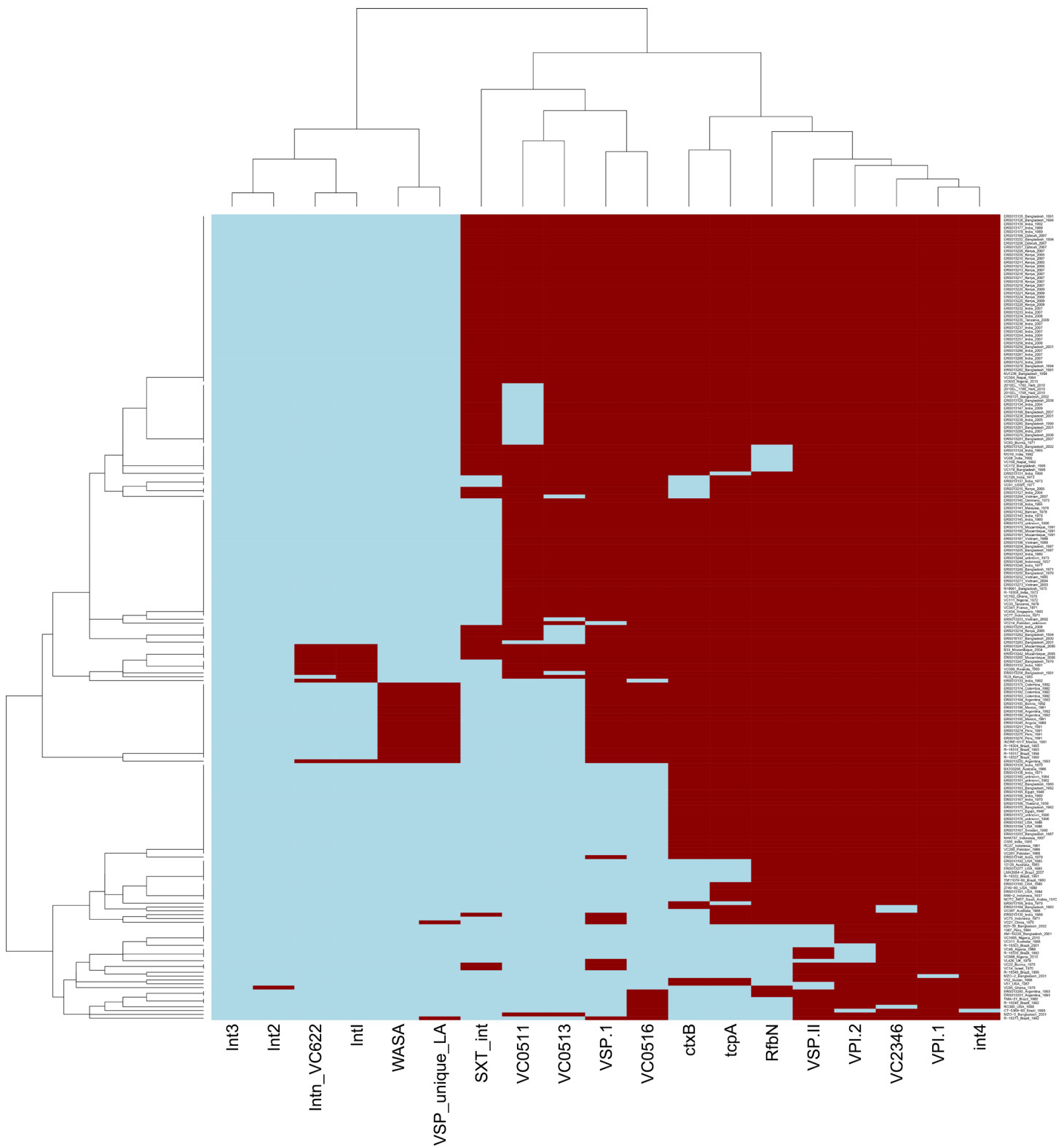


Figura 31. Presença de marcadores genéticos em *V. cholerae*. O *heatmap* foi feito no ambiente R. A *clusterização* esta baseada no compartilhamento dos marcadores genéticos (descritos no eixo horizontal) nos genomas de *V. cholerae* avaliados (descritos no eixo vertical, n=210). A presença é representada em vermelho e a ausência em azul claro. Genomas da epidemia da América Latina possuem os dois marcadores que caracterizam a linhagem, o profago WASA e a região única da ilha VSP-II.

Discussão

6 DISCUSSÃO

A cólera foi uma das primeiras infecções a serem estudadas sob a lógica da epidemiologia quando John Snow, ícone desta especialidade no mundo, estabeleceu o mapa da cólera na cidade de Londres, no final do século XIX. Dois séculos depois, já considerando todo o conhecimento sobre a biologia desta, a partir de dados e abordagens genômicas foi possível estabelecer, após um enorme debate, a origem da atual epidemia de cólera no Haiti (2010) (Chin *et al.* 2011).

Estudar o *V. cholerae* é uma tarefa desafiadora, mas ao mesmo tempo empolgante; afinal, esta bactéria tem sido o foco de estudos de muitos grupos de pesquisa no mundo por mais de um século. É usada como modelo em estudos de epidemiologia, patogênese, genética, evolução, desenvolvimento de vacinas e drogas (Codeço 2001; Emch *et al.* 2008; Okeke 2009; Safa *et al.* 2010; Fazil e Singh 2011; Azman *et al.* 2013).

Quando iniciamos este projeto, em meados de 2009, pouco se conhecia da epidemiologia da cólera na América Latina (AL). Nos anos 1990s, quando esta epidemia começou, como uma extensão da sétima pandemia, que era determinada por *V. cholerae* O1 do biotipo El Tor, os isolados eram caracterizados por MLEE e pelos genótipos dos dois principais determinantes de virulência, a toxina colérica e o “pilus” TCP. Portanto, com base nestes marcadores, a linhagem da epidemia da AL apresentava uma única diferença em relação à linhagem causadora das epidemias asiáticas, o alelo de uma leucina aminopeptidase (Freitas *et al.* 2002). Assim, as duas hipóteses principais sobre sua origem, i) a origem ambiental, devido a eventos climáticos como El Niño, que teria provocado a ressurgência de *V. cholerae* das águas costeiras do oceano Pacífico no Peru (Colwell 1996), e ii) a introdução a partir do fluxo humano de regiões endêmicas/epidêmicas (Wachsmuth *et al.* 1993; Kaper *et al.* 1995), se mantiveram.

Nos anos que se seguiram, a abordagem de MLEE evoluiu para uma abordagem genômica através do MLSA e nosso grupo passou a aplicar esta abordagem em espécies do gênero *Vibrio*, mas também em linhagens de *V. cholerae*.

O Laboratório de Genética Molecular de Microrganismos possui uma coleção rara de *V. cholerae*, composta por isolados dos últimos 50 anos de diferentes

regiões do mundo, incluindo exemplares da epidemia da AL. Assim, era factível estudar a epidemiologia do *V. cholerae* da AL utilizando a genômica, contribuindo desta forma para o esclarecimento da sua origem e evolução. Durante o desenvolvimento destes estudos, dois fatos impactaram, de forma direta e indireta, nosso projeto: a re-emergência da cólera nas Américas e a publicação do trabalho sobre a epidemiologia genômica de *V. cholerae* na sétima pandemia (Mutreja *et al.* 2011).

Em outubro de 2010, o Haiti, que havia sido destruído por um terremoto em janeiro do mesmo ano, foi assolado por uma epidemia de cólera. Este país não havia sido acometido pela cólera durante a epidemia da AL, e os últimos relatos de cólera eram do final do século XIX (Piarroux *et al.* 2011). Da mesma forma que na epidemia da AL, duas hipóteses principais sobre a origem do *V. cholerae* no Haiti surgiram: a emergência de uma linhagem de origem ambiental e a emergência da linhagem responsável pela epidemia da AL. Imediatamente, em janeiro de 2011, um grupo sequenciou os genomas completos de isolados da AL (Peru), do Haiti e de regiões endêmicas, e mostrou que a linhagem do Haiti não estava relacionada a linhagem da AL (Chin *et al.* 2011). A linhagem do Haiti teria relação com isolados do sudoeste asiático, o que corroborou as evidências epidemiológicas que mostravam que a cólera no Haiti teria sido introduzida por soldados nepaleses, das Nações Unidas, que estavam no país dando suporte (Piarroux *et al.* 2011).

O trabalho sobre a epidemiologia genômica de *V. cholerae* na sétima pandemia publicado na Nature (Mutreja *et al.* 2011), demonstra que esta pandemia, que se iniciou nos anos sessenta e até hoje continua, é caracterizada por linhagens que definem três principais ondas de transmissão: a primeira, causada pelo El Tor típico, afetou todos os continentes, inclusive a AL; a segunda, caracterizada pelo surgimento do O139 e aquisição de elementos da família SXT e a terceira, e atual, por El Tor atípico. Naquele trabalho 139 genomas completos foram sequenciados e, pela primeira vez, genomas completos da AL foram tornados públicos.

A aplicação do MLSA a isolados de *V. cholerae* definiu três grandes grupos: o clado da sétima pandemia, caracterizado pelo biotipo El Tor, o clado do biotipo clássico e o clado com linhagens diversas, tanto ambientais quanto clínicas. Portanto, o MLSA só teve resolução em relação aos biotipos, o que seria importante para caracterizar a sexta e sétima pandemias e linhagens ambientais. Quanto a discriminação de linhagens dentro destes biotipos, que é nosso objetivo em relação

a AL, esta abordagem não teria sensibilidade. Este conjunto de genes foi aplicado com sucesso na discriminação de linhagens ambientais (O'Shea *et al.* 2004b). Como uma das hipóteses para a origem da epidemia da AL seria a emergência do *V. cholerae* a partir do ambiente e a tipagem molecular aplicando MLEE discriminava a linhagem da AL (Freitas *et al.* 2002), aplicamos esta abordagem. Além disto, um trabalho com enfoque na epidemia do México mostrou *V. cholerae* do biotipo clássico co-circulando com *V. cholerae* El Tor durante a epidemia da AL naquele país (Alam *et al.* 2010). Portanto, em nossa amostragem, que contém isolados de vários países da América do Sul, ficou clara a relação com a linhagem principal da sétima pandemia, a El Tor.

Trabalhos aplicando a ribotipagem demonstraram a relação genética entre El Tor da AL e amostras asiáticas (Faruque e Albert 1992; Popovic *et al.* 1993). Contudo, a ribotipagem, que tem como referência a localização dos operons ribossomais para a definição de perfis, é limitada para a avaliação da diversidade em *V. cholerae*. O genoma desta espécie está distribuído em dois cromossomos, mas todos os oito operons ribossomais encontram-se no cromossomo 1. Assim, a ribotipagem teria o viés de não avaliar o conteúdo do cromossomo 2 e portanto, poderia estar subestimando a variabilidade dos isolados.

No cromossomo 2 (menor) do *V. cholerae* encontra-se o superintegron (SI) ou integron cromossomal, região sujeita a intensa perda e incorporação de conteúdo gênico. Tem sido mostrado que o conteúdo do SI varia em função do nicho e da localização geográfica dos isolados (Boucher *et al.* 2011). Portanto, o SI seria um potencial marcador de linhagens de epidemias dentro de pandemias.

V. cholerae dos biotipos clássico e El Tor eram caracterizados por distintos genótipos do gene *ctxB*, *ctxB-1* e *ctxB-3*, respectivamente. A linhagem de *V. cholerae* da epidemia da América Latina, assim como os demais isolados El Tor, do início da sétima pandemia, apresentam o genótipo *ctxB-3*. Recentemente, trabalhos com isolados obtidos a partir dos anos 1990s de epidemias na Ásia e África mostraram a existência de novos genótipos *ctxB*, assim como a presença do genótipo *ctxB-1*, em isolados do biotipo El Tor. *V. cholerae* com estas características são denominados de El Tor atípico, alterado ou variante. Estas linhagens passaram a ser as determinantes das epidemias recentes de cólera (Quilici *et al.* 2010; Safa *et al.* 2010), inclusive da epidemia do Haiti (Chin *et al.* 2011). Até o momento, 11 alelos já foram identificados (Marin e Vicente 2012). Uma das explicações para o sucesso

destas novas linhagens é que as mesmas combinam a adaptabilidade ao ambiente à virulência do *ctxB* (características do biotipo El Tor e do biotipo clássico, respectivamente). De fato, as epidemias recentes têm apresentado taxas elevadas de mortalidade (Ryan 2011; Piarroux e Faucher 2012).

Estas análises indicavam que a epidemia da AL teria como determinante principal *V. cholerae* do biotipo El Tor relacionado geneticamente a isolados do sudoeste asiático do início da sétima pandemia. A existência de um fluxo comercial marítimo entre o sudoeste asiático (endêmico e epidêmico para cólera) e o Peru (onde eclodiu a epidemia da AL) robustece a hipótese desta região da Ásia ser a origem da epidemia de cólera na AL.

Contrariando esta hipótese, novos trabalhos com genomas completos apontam a África, e não o sudoeste asiático, como origem da linhagem que deflagrou a cólera em nosso continente. Contudo, apesar de analisar genomas completos, estes trabalhos não são definitivos. Um deles está baseado apenas em 30 SNPs (Lam *et al.* 2010) e o outro, que analisa 139 genomas completos de *V. cholerae* da sétima pandemia, chega a esta conclusão devido a uma única amostra de Angola (1989) (Mutreja *et al.* 2011). Portanto, já que dispúnhamos de amostras de surtos recentes e do início da sétima pandemia de cólera de países da África, fomos buscar outras evidências que corroborassem esta nova e interessante hipótese.

Nossas análises com isolados de epidemias de Ghana e Nigéria não estabeleceram uma relação direta entre a AL e a África, considerando a linhagem epidêmica principal do biotipo El Tor. Contudo, mostramos que havia uma ligação entre Ghana (anos 70s) e Brasil (anos 90s) que compartilhavam uma mesma linhagem de *V. cholerae* O1, que causou surtos de cólera na Amazônia brasileira, nos anos 90, e surtos em Ghana, nos anos 70 (Thompson *et al.* 2011a). Esta linhagem, que já havia sido descrita e caracterizada no contexto da epidemia da AL causando surtos na Amazônia brasileira (Coelho *et al.* 1995), não tem nenhuma relação com um El Tor e também não carrega os principais determinantes de virulência de um *V. cholerae* epidêmico. Mesmo assim, causava diarreia semelhante à cólera, e o sequenciamento de seu genoma completo mostrou a presença de vários genes associados à virulência (Thompson *et al.* 2011b). Ela foi por nós denominada Amazônia/Ghana.

Nossos trabalhos de Ghana (Thompson *et al.* 2011a) e da Nigéria (Marin *et al.* 2013) não trouxeram mais evidências para a hipótese da origem da epidemia da AL na África, mas há alguma via que é efetiva na disseminação de patógenos, como mostrado pela linhagem Amazônia/Ghana.

Recentemente, nosso grupo participou de projeto com sequenciamento genômico de um conjunto de isolados de *V. cholerae*. Neste conjunto havia amostras de vários países da África. Análises filogenômicas preliminares mostram que os genomas de isolados El Tor de Ghana (1970s), Nigéria (1980s) e países do oeste da África como Angola, estariam mais próximos ao *cluster* da América Latina, quando comparados a genomas do Bahrain (1978), Tanzânia (1979), Quênia (1985) e Ruanda (1993), que estão mais relacionados ao *cluster* do Sudoeste Asiático. De qualquer maneira, o número ainda limitado de isolados da África contemporâneos à epidemia da América Latina não permite uma conclusão definitiva sobre a origem da epidemia da AL a partir de Angola. De fato, epidemias de cólera ocorreram neste país, assim como migrações de angolanos para o Brasil nos anos que antecederam a epidemia de cólera na AL, e tais situações poderiam ter influenciado a disseminação do patógeno entre as duas regiões.

V. cholerae biotipo El Tor da sétima pandemia é caracterizado também por ilhas genômicas, que são elementos genéticos que formam parte do mobiloma¹⁴, transferidos horizontalmente entre bactérias e que eventualmente se fixam e passam a caracterizar linhagens (Chun *et al.* 2009). A ilha genômica VSP-II (*Vibrio Seventh Pandemic-II*) é um destes elementos (Dziejman *et al.* 2002). O conteúdo gênico e a organização da VSP-II da amostra N16961, que é um isolado da Índia do início da sétima pandemia e teve o genoma completamente sequenciado (Heidelberg *et al.* 2000), é o modelo canônico de VSP-II. Nusrin e colaboradores (Nusrin *et al.* 2009), avaliando por PCR a presença destes genes em isolados da epidemia da AL no Peru, encontram VSP-II modificadas, onde uma das regiões da canônica N16961 apresentava um outro conteúdo gênico.

Aqui, analisando o conteúdo genômico da VSP-II em mais de uma centena de genomas, observa-se uma particularidade quanto ao conteúdo gênico da VSP-II da

¹⁴ O mobiloma é o total de elementos genéticos móveis em um genoma.

linhagem da América Latina que é a ausência dos genes VC0511 e VC0513 e a presença de três novas ORFs em substituição a estes genes.

As ilhas VSP-I e VSP-II são consideradas marcadores genéticos de *V. cholerae* da sétima pandemia (Dziejman *et al.* 2002; O'Shea *et al.* 2004b). O core deste tipo de elemento genético é a integrase, já o conteúdo gênico é variável. De fato, a diversidade do conteúdo gênico das VSP-II de *V. cholerae* já foi descrita (Dziejman *et al.* 2005; Nusrin *et al.* 2009; Taviani *et al.* 2010). Portanto, a integrase é a assinatura deste elemento.

A análise da VSP-II destas centenas de genomas nos trouxe uma evidência muito interessante e que viria a ser um marcador para a linhagem da AL. As análises filogenéticas mostraram que a linhagem da AL possui um alelo próprio da integrase da VSP-II. Este alelo compartilha 90% de identidade nucleotídica com a integrase dos isolados da sétima pandemia, as quais têm 100% de identidade entre si. A maior identidade da integrase da VSP-II da AL, 98%, é com integrases de *V. cholerae* ambientais e de um isolado clínico de *V. vulnificus* (YJ016) (Chen *et al.* 2003).

Identities entre ilhas genômicas de *V. cholerae* e *V. vulnificus* já tinham sido observadas por O'Shea *et al.* (2004a). Os autores mostraram semelhanças entre a VSP-II de *V. cholerae* e a VVI-1 (*Vibrio vulnificus island 1*). A VVI-1 do *V. vulnificus* YJ016 está inserida na mesma localização genômica que a VSP-II, no sítio da tRNA-metionina (*att*). Como o conteúdo gênico da VSP-II da AL é idêntico ao da canônica, exceto pela ausência de dois genes e uma inserção de 7 kb (3 genes), nossa hipótese é que tenha ocorrido recombinação entre os genes das integrases destas duas ilhas.

De fato, a ilha VSP-II pode ser excisada do cromossomo e formar intermediários circulares, através de recombinação sítio-específica mediada pela integrase codificada pela ilha (Murphy e Boyd 2008). Além disso, sabe-se que *V. cholerae* se torna naturalmente competente para a aquisição de DNA nu em presença de quitina, uma molécula abundante em ambientes aquáticos (Meibom *et al.* 2005). Isto sugere, portanto, um possível mecanismo de transferência horizontal de VSP-II entre isolados de *V. cholerae* ou de espécies próximas. Outra possível forma de transmissão da VSP-II é por transdução (Faruque e Mekalanos 2012), sendo a excisão da VSP-II do cromossomo o primeiro passo para a transferência horizontal posterior. Por exemplo, a transferência de ilha de patogenicidade entre

isolados mediada pelo fago CP-T1 tem sido documentada (O'Shea e Boyd 2002). É interessante ressaltar que, como descrito para o *V. cholerae* O1 El Tor N16961, a ilha VSP-II codifica uma pilina do tipo IV, duas proteínas quimiotáticas, um regulador transcricional semelhante ao AraC, uma proteína de reparo do DNA e uma integrase P4-like (Taviani *et al.* 2010). Alguns destes elementos podem estar relacionados, mesmo que indiretamente, com a transferência horizontal de material genético. Curiosamente, o outro marcador da linhagem da AL é também um elemento móvel, o bacteriófago WASA1 (Mutreja *et al.* 2011; Garza *et al.* 2012).

Os elementos genéticos móveis e/ou mobilizáveis são abundantes na espécie *V. cholerae* e tem papel importante nas linhagens colerogênicas, quer como marcadores ou como determinantes de fenótipos relacionados à virulência (Chun *et al.* 2009; Cho *et al.* 2010). Temos como exemplos o bacteriófago CTX, VPI (onde está o *cluster* TCP), e vários elementos que carregam genes de resistência aos antibióticos. Alguns deles se fixaram nos biotipos da sexta e sétima pandemias, como CTX e VPI, enquanto outros caracterizam linhagens destes biotipos, como o WASA1 da linhagem da AL.

Mutreja (2011) mostrou que isolados a partir dos anos 1990s até a epidemia do Haiti (2010) passaram a ter como característica a presença de um elemento genômico da família SXT/ICE. Esta família de elementos é caracterizada por uma integrase conservada (*int*_{SXT}), um *core* e suas regiões variáveis, onde muitas vezes encontram-se genes de resistência a antibióticos. Os distintos ICEs são nomeados em função de todo o arranjo gênico (Wozniak *et al.* 2009). Sjölund-Karlsson e colaboradores caracterizaram na linhagem da epidemia do Haiti 2010 um ICE e o denominaram ICEVchHai1 (Sjölund-Karlsson *et al.* 2011). Em buscas nos genomas completos de *V. cholerae* encontramos ICEs idênticos ao da linhagem do Haiti, dispersos em três continentes e persistindo por mais de duas décadas. Ou seja, este elemento pode estar contribuindo para o sucesso das linhagens de *V. cholerae* epidêmicas atuais (Marin *et al.* 2013, trabalho submetido).

As principais linhagens de El Tor epidêmico são caracterizadas por elementos genéticos móveis que acabaram ou não sendo fixados, mas que, de qualquer maneira, têm papel importante no sucesso das linhagens. Boucher e colaboradores mostram que o genoma *core* do *V. cholerae* é muito conservado, e que o seu dinamismo epidemiológico seria consequência das novas informações contidas

nestes elementos (Boucher *et al.* 2011). O superintegron (SI) ou integron cromossômico seria um destes elementos móveis, que foi fixado ancestralmente, já que está presente em todo o gênero (Gu *et al.* 2009) e se especializou nas espécies. Em *V. cholerae* é caracterizado por uma integrase do tipo 4 (Rowe-Magnus *et al.* 1999). O conteúdo gênico destes elementos é principalmente definido pelo ambiente e pressões seletivas a que estão expostos (Boucher *et al.* 2011); poderiam portanto ser marcadores de linhagens (Labbate *et al.* 2007). Trabalhos analisando especificamente o conteúdo genômico de SIs de linhagens colerogênicas mostraram que elas possuem diferenças no conteúdo em função de inserções e deleções características (Labbate *et al.* 2007; Gao *et al.* 2011). Em nossas análises *in silico*, com dados brutos de sequenciamentos de nova geração, foi possível corroborar estes trabalhos. Assim mostramos que com a abordagem adequada é possível recuperar informação fidedigna de regiões com alta variabilidade como o SI. A informação aqui recuperada reflete a evolução do genoma *core*, já que reproduz a topologia filogenética baseada nos SNPs presentes nos genomas das diferentes linhagens (Mutreja *et al.* 2011).

Conclusões

7 CONCLUSÕES

- A epidemia da América Latina é uma extensão do início da sétima pandemia.
- A linhagem da América Latina tem como marcadores um profago (WASA1) e a ilha genômica VSP-II. Esta VSP-II é caracterizada por seu gene da integrase e uma inserção de 7 Kb, que contém três genes putativos.
- O conteúdo gênico do superintegron é um marcador para as diferentes linhagens de *V. cholerae*.
- A linhagem da epidemia de cólera do Haiti (2010), que é relacionada à da Nigéria (2010), é distinta da linhagem da América Latina. Ambas, Haiti e Nigéria são caracterizadas por mesmos alelos dos genes *ctxB*, *tcpA*, *gyrA*, *parC*, assim como pela presença do mesmo elemento ICE.
- Nossos resultados descartam a hipótese da origem ambiental para a linhagem determinante da epidemia da América Latina. Por outro lado, não é possível, com as amostras disponíveis, confirmar a origem a partir da África.
- O conteúdo extra cromossomal possui informação importante para estudos epidemiológicos e evolutivos.

Referências Bibliográficas

8 REFERÊNCIAS BIBLIOGRÁFICAS

- Alam M, Nusrin S, Islam A, Bhuiyan N a, Rahim N, Delgado G, et al. Cholera between 1991 and 1997 in Mexico was associated with infection by classical, El Tor, and El Tor variants of *Vibrio cholerae*. *Journal of clinical microbiology*. 2010 Oct;48(10):3666–74.
- Alikhan N-F, Petty NK, Ben Zakour NL, Beatson S a. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC genomics*. 2011 Jan;12(1):402.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*. 1997 Sep 1;25(17):3389–402.
- Andrews JR, Basu S. Transmission dynamics and control of cholera in Haiti: an epidemic model. *Lancet*. 2011 Mar 15;377(9773):1248–55.
- Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic biology*. 2006 Aug;55(4):539–52.
- Aydanian A, Tang L, Morris JG, Johnson J a, Stine OC. Genetic diversity of O-antigen biosynthesis regions in *Vibrio cholerae*. *Applied and environmental microbiology*. 2011 Apr;77(7):2247–53.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. *BMC genomics*. 2008 Jan;9:75.

- Azman AS, Rudolph KE, Cummings DAT, Lessler J. The incubation period of cholera: A systematic review. *The Journal of infection*. 2013 May;66(5):432–8.
- Bhuiyan NA, Nusrin S, Ansaruzzaman M, Islam A, Sultana M, Alam M, et al. Genetic characterization of *Vibrio cholerae* O1 strains isolated in Zambia during 1996-2004 possessing the unique VSP-II region of El Tor variant. *Epidemiology and infection*. 2012 Mar 31;140(3):510–8.
- Boucher Y, Cordero OX, Takemura A, Hunt DE, Schliep K, Baptiste E, et al. Local mobile gene pools rapidly cross species boundaries to create endemicity within global *Vibrio cholerae* populations. *mBio*. 2011 Jan;2(2):e00335–10.
- Burrus V, Marrero J, Waldor MK. The current ICE age: biology and evolution of SXT-related integrating conjugative elements. *Plasmid*. 2006 May;55(3):173–83.
- Byun R, Elbourne LD, Lan R, Reeves PR. Evolutionary relationships of pathogenic clones of *Vibrio cholerae* by sequence analysis of four housekeeping genes. *Infection and immunity*. 1999 Mar;67(3):1116–24.
- Cambray G, Guerout A-M, Mazel D. Integrons. *Annual review of genetics*. 2010 Jan;44:141–66.
- Ceccarelli D, Spagnoletti M, Bacciu D, Danin-Poleg Y, Mendiratta DK, Kashi Y, et al. ICEVchInd5 is prevalent in epidemic *Vibrio cholerae* O1 El Tor strains isolated in India. *International journal of medical microbiology: IJMM*. 2011 a Apr;301(4):318–24.
- Ceccarelli D, Spagnoletti M, Cappuccinelli P, Burrus V, Colombo MM. Origin of *Vibrio cholerae* in Haiti. *The Lancet infectious diseases*. 2011 b Apr;11(4):262.
- Chatterjee S, Ghosh K, Raychoudhuri a, Chowdhury G, Bhattacharya MK, Mukhopadhyay a K, et al. Incidence, virulence factors, and clonality among

- clinical strains of non-O1, non-O139 *Vibrio cholerae* isolates from hospitalized diarrheal patients in Kolkata, India. *Journal of clinical microbiology*. 2009 Apr;47(4):1087–95.
- Chaudhuri K, Chatterjee SN. Cholera toxins. *Microbial toxins*. Heidelberg, Germany: Springer-Verlag Berlin Heidelberg; 2009.
- Chen C-Y, Wu K-M, Chang Y-C, Chang C-H, Tsai H-C, Liao T-L, et al. Comparative genome analysis of *Vibrio vulnificus*, a marine pathogen. *Genome research*. 2003 Dec;13(12):2577–87.
- Chen Y, Johnson J a, Pusch GD, Morris JG, Stine OC. The genome of non-O1 *Vibrio cholerae* NRT36S demonstrates the presence of pathogenic mechanisms that are distinct from those of O1 *Vibrio cholerae*. *Infection and immunity*. 2007 May;75(5):2645–7.
- Chin C-S, Sorenson J, Harris JB, Robins WP, Charles RC, Jean-Charles RR, et al. The origin of the Haitian cholera outbreak strain. *The New England journal of medicine*. 2011 Jan 6;364(1):33–42.
- Cho YJ, Yi H, Lee JH, Kim DW, Chun J. Genomic evolution of *Vibrio cholerae*. *Current opinion in microbiology*. 2010;(Figure 1):1–6.
- Christensen-Dalsgaard M, Gerdes K. Two *higBA* loci in the *Vibrio cholerae* superintegron encode mRNA cleaving enzymes and can stabilize plasmids. *Molecular microbiology*. 2006 Oct;62(2):397–411.
- Chun J, Grim CJ, Hasan NA, Lee JH, Choi SY, Haley BJ, et al. Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. *Proceedings of the National Academy of Sciences of the United States of America*. 2009 Sep 8;106(36):15442–7.

- Codeço CT. Endemic and epidemic dynamics of cholera: the role of the aquatic reservoir. *BMC infectious diseases*. 2001 Jan;1:1.
- Coelho A, Andrade JR, Vicente AC, Salles CA. New variant of *Vibrio cholerae* O1 from clinical isolates in Amazonia. *Journal of clinical microbiology*. 1995 Jan;33(1):114–8.
- Colwell RR. Global climate and infectious disease: the cholera paradigm. *Science*. 1996 Dec 20;274(5295):2025–31.
- Colwell RR. Infectious disease and environment: cholera as a paradigm for waterborne disease. *International Microbiology*. 2004 Dec;7(4):285–9.
- Constantin de Magny G, Murtugudde R, Sapiano MRP, Nizam A, Brown CW, Busalacchi AJ, et al. Environmental signatures associated with cholera epidemics. *Proceedings of the National Academy of Sciences of the United States of America*. 2008 Nov 18;105(46):17676–81.
- Dalsgaard A, Albert MJ, Taylor DN, Shimada T, Meza R, Serichantalergs O, et al. Characterization of *Vibrio cholerae* non-O1 serogroups obtained from an outbreak of diarrhea in Lima, Peru. *Journal of clinical microbiology*. 1995 Oct;33(10):2715–22.
- De Sá Morais LLC, Garza DR, Loureiro ECB, Nunes KNB, Vellasco RS, Da Silva CP, et al. Complete genome sequence of a sucrose-nonfermenting epidemic strain of *Vibrio cholerae* O1 from Brazil. *Journal of bacteriology*. 2012 May;194(10):2772.
- Didelot X, Falush D. Inference of bacterial microevolution using multilocus sequence data. *Genetics*. 2007 Mar;175(3):1251–66.
- Dziejman M, Balon E, Boyd D, Fraser CM, Heidelberg JF, Mekalanos JJ. Comparative genomic analysis of *Vibrio cholerae*: genes that correlate with

- cholera endemic and pandemic disease. *Proceedings of the National Academy of Sciences of the United States of America*. 2002;99(3):1556.
- Dziejman M, Serruto D, Tam VC, Sturtevant D, Diraphat P, Faruque SM, et al. Genomic characterization of non-O1, non-O139 *Vibrio cholerae* reveals genes for a type III secretion system. *Proceedings of the National Academy of Sciences of the United States of America*. 2005 Mar 1;102(9):3465–70.
- Emch M, Feldacker C, Islam MS, Ali M. Seasonality of cholera from 1974 to 2005: a review of global patterns. *International journal of health geographics*. 2008 Jan;7:31.
- Faruque SM, Albert J. Genetic relation between *Vibrio cholerae* O1 strains in Ecuador and Bangladesh. *Lancet*. 1992 Mar 21;339(8795):740–1.
- Faruque SM, Mekalanos JJ. Phage-bacterial interactions in the evolution of toxigenic *Vibrio cholerae*. *Virulence*. 2012 Oct 17;3(7):1–10.
- Fazil MHUT, Singh D V. *Vibrio cholerae* infection, novel drug targets and phage therapy. *Future microbiology*. 2011 Oct;6(10):1199–208.
- Freitas FS, Momen H, Salles CA. The Zymovars of *Vibrio cholerae*: multilocus enzyme electrophoresis of *Vibrio cholerae*. *Memórias do Instituto Oswaldo Cruz*. 2002 Jun;97(4):511–6.
- Gao Y, Pang B, Wang HY, Zhou HJ, Cui ZG, Kan B. Structural variation of the superintegron in the toxigenic *Vibrio cholerae* O1 El Tor. *Biomedical and environmental sciences: BES*. 2011 Dec;24(6):579–92.
- Garza DR, Thompson CC, Loureiro ECB, Dutilh BE, Inada DT, Junior ECS, et al. Genome-wide study of the defective sucrose fermenter strain of *Vibrio cholerae* from the Latin American cholera epidemic. *PloS one*. 2012 Jan 25;7(5):e37283.

- Gil AI, Louis VR, Rivera ING, Lipp E, Huq A, Lanata CF, et al. Occurrence and distribution of *Vibrio cholerae* in the coastal environment of Peru. *Environmental microbiology*. 2004 Jul;6(7):699–706.
- Goel AK, Jain M, Kumar P, Bhadauria S, Kmboj D V, Singh L. A new variant of *Vibrio cholerae* O1 El Tor causing cholera in India. *The Journal of infection*. 2008 Sep;57(3):280–1.
- Goel AK, Jain M, Kumar P, Sarguna P, Bai M, Ghosh N, et al. Molecular characterization reveals involvement of altered El Tor biotype *Vibrio cholerae* O1 strains in cholera outbreak at Hyderabad, India. *Journal of microbiology* 2011 Apr;49(2):280–4.
- Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution*. 2010 Feb;27(2):221–4.
- Gu J, Neary J, Cai H, Moshfeghian A, Rodriguez S a, Lilburn TG, et al. Genomic and systems evolution in Vibrionaceae species. *BMC genomics*. 2009 Jan;10 Suppl 1:S11.
- Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology*. 2003 Oct;52(5):696–704.
- Harvez CB, Avila VS. [The cholera epidemic in Latin America: reemergence, morbidity, and mortality]. *Revista panamericana de salud pública = Pan American journal of public health*. 2013 Jan;33(1):40–6.
- Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, et al. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature*. 2000 Aug 3;406(6795):477–83.

- Hendriksen RS, Price LB, Schupp JM, Gillece JD, Kaas RS, Engelthaler DM, et al. Population genetics of *Vibrio cholerae* from Nepal in 2010: evidence on the origin of the Haitian outbreak. *mBio*. 2011 Jan 23;2(4):e00157–11.
- Hoshino K, Yamasaki S, Mukhopadhyay AK, Chakraborty S, Basu A, Bhattacharya SK, et al. Development and evaluation of a multiplex PCR assay for rapid detection of toxigenic *Vibrio cholerae* O1 and O139. *FEMS immunology and medical microbiology*. 1998 Mar;20(3):201–7.
- Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution*. 2006 Feb;23(2):254–67.
- Huson DH, Steel M. Phylogenetic trees based on gene content. *Bioinformatics* 2004 Sep 1;20(13):2044–9.
- Kaper JB, Morris JG, Levine MM. Cholera. *Clinical microbiology reviews*. 1995 Jan;8(1):48–86.
- Karaolis DKR, Johnson JA, Bailey CC, Boedeker EC, Kaper JB, Reeves PR. A *Vibrio cholerae* pathogenicity island associated with epidemic and pandemic strains. *Proceedings of the National Academy of Sciences of the United States of America*. 1998;95(6):3134.
- Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. *Briefings in bioinformatics*. 2008 Jul;9(4):286–98.
- Kondo S, Ajawatanawong P. Distribution and sequence analysis of virulence associated genes in *Vibrio cholerae* O1, O139 and non-O1/non-O139 isolates from Thailand. *The Southeast Asian journal of tropical medicine and public health*. 2009 Sep;40(5):1015–24.
- Kumar P, Thulaseedharan A, Chowdhury G, Ramamurthy T, Thomas S. Characterization of novel alleles of toxin co-regulated pilus A gene (*tcpA*) from

environmental isolates of *Vibrio cholerae*.
Mar;62(3):758–63.

Current microbiology. 2011

Labbate M, Boucher Y, Joss MJ, Michael C a, Gillings MR, Stokes HW. Use of chromosomal integron arrays as a phylogenetic typing system for *Vibrio cholerae* pandemic strains. *Microbiology*. 2007 May;153(Pt 5):1488–98.

Lam C, Octavia S, Reeves P, Wang L, Lan R. Evolution of seventh cholera pandemic and origin of 1991 epidemic, Latin America. *Emerging infectious diseases*. 2010 Jul;16(7):1130–2.

Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic acids research*. 2011 Jul;39:W475–8.

Li M, Shimada T, Morris JG, Sulakvelidze A, Sozhamannan S. Evidence for the emergence of non-O1 and non-O139 *Vibrio cholerae* strains with pathogenic potential by exchange of O-antigen biosynthesis regions. *Infection and immunity*. 2002 May;70(5):2441–53.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005 Sep 15;437(7057):376–80.

Marin M a, Vicente ACP. Architecture of the superintegron in *Vibrio cholerae*: identification of core and unique genes. *F1000Research*. 2013 Feb 27;1–9.

Marin MA, Thompson CC, Freitas FS, Fonseca EL, Aboderin a. O, Zailani SB, et al. Cholera outbreaks in Nigeria are associated with multidrug resistant atypical El Tor and non-O1/non-O139 *Vibrio cholerae*. *PLoS neglected tropical diseases*. 2013 Feb 14;7(2):e2049.

- Marin MA, Vicente ACP. Variants of *Vibrio cholerae* O1 El Tor from Zambia showed new genotypes of *ctxB*. *Epidemiology and infection*. 2012 Aug 23;140(8):1386–7.
- Mazel D. Integrons: agents of bacterial evolution. *Nature reviews. Microbiology*. 2006 Aug;4(8):608–20.
- McLeod SM, Kimsey HH, Davis BM, Waldor MK. CTXphi and *Vibrio cholerae*: exploring a newly recognized type of phage-host cell relationship. *Molecular microbiology*. 2005 Jul;57(2):347–56.
- Medini D, Donati C, Tettelin H, Massignani V, Rappuoli R. The microbial pan-genome. *Current opinion in genetics & development*. 2005 Dec;15(6):589–94.
- Meibom KL, Blokesch M, Dolganov N a, Wu C-Y, Schoolnik GK. Chitin induces natural competence in *Vibrio cholerae*. *Science*. 2005 Dec 16;310(5755):1824–7.
- Mohapatra SS, Ramachandran D, Mantri CK, Colwell RR, Singh D V. Determination of relationships among non-toxigenic *Vibrio cholerae* O1 biotype El Tor strains from housekeeping gene sequences and ribotype patterns. *Research in microbiology*. 2009;160(1):57–62.
- Morita M, Ohnishi M, Arakawa E, Bhuiyan N a, Nusrin S, Alam M, et al. Development and validation of a mismatch amplification mutation PCR assay to monitor the dissemination of an emerging variant of *Vibrio cholerae* O1 biotype El Tor. *Microbiology and immunology*. 2008 Jun;52(6):314–7.
- Morris JG. Cholera-modern pandemic disease of ancient lineage. *Emerging infectious diseases*. 2011 Nov;17(11):2099–104.

- Murphy RA, Boyd EF. Three pathogenicity islands of *Vibrio cholerae* can excise from the chromosome and form circular intermediates. *Journal of bacteriology*. 2008 Jan;190(2):636–47.
- Mutreja A, Kim DW, Thomson NR, Connor TR, Lee JH, Kariuki S, et al. Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature*. 2011 Sep 22;477(7365):462–5.
- Nickels BE. A new twist on a classic paradigm: illumination of a genetic switch in *Vibrio cholerae* phage CTX Phi. *Journal of bacteriology*. 2009 Nov;191(22):6779–81.
- Novais RC, Coelho A, Salles CA, Vicente AC. Toxin-co-regulated pilus cluster in non-O1, non-toxicogenic *Vibrio cholerae*: evidence of a third allele of pilin gene. *FEMS microbiology letters*. 1999 Feb 1;171(1):49–55.
- Nusrin S, Gil AI, Bhuiyan NA, Safa A, Asakura M, Lanata CF, et al. Peruvian *Vibrio cholerae* O1 El Tor strains possess a distinct region in the *Vibrio* seventh pandemic island-II that differentiates them from the prototype seventh pandemic El Tor strains. *Journal of medical microbiology*. 2009 Mar;58(Pt 3):342–54.
- O’Shea YA, Boyd EF. Mobilization of the *Vibrio* pathogenicity island between *Vibrio cholerae* isolates mediated by CP-T1 generalized transduction. *FEMS microbiology letters*. 2002 Sep 10;214(2):153–7.
- O’Shea YA, Finnan S, Reen FJ, Morrissey JP, O’Gara F, Boyd EF. The *Vibrio* seventh pandemic island-II is a 26.9 kb genomic island present in *Vibrio cholerae* El Tor and O139 serogroup isolates that shows homology to a 43.4 kb genomic island in *V. vulnificus*. *Microbiology*. 2004 Dec;150(Pt 12):4053–63.

- O'Shea YA, Reen FJ, Quirke AM, Boyd EF. Evolutionary genetic analysis of the emergence of epidemic *Vibrio cholerae* isolates on the basis of comparative nucleotide sequence analysis and multilocus virulence gene profiles. *Journal of clinical microbiology*. 2004 Oct;42(10):4657–71.
- Okeke IN. Cholera vaccine will reduce antibiotic use. *Science*. 2009 Aug 7;325(5941):674.
- Olsvik O, Wahlberg J, Petterson B, Uhlén M, Popovic T, Wachsmuth IK, et al. Use of automated sequencing of polymerase chain reaction-generated amplicons to identify three types of cholera toxin subunit B in *Vibrio cholerae* O1 strains. *Journal of clinical microbiology*. 1993 Jan;31(1):22–5.
- Opintan J a, Newman MJ, Nsiah-Poodoh OA, Okeke IN. *Vibrio cholerae* O1 from Accra, Ghana carrying a class 2 integron and the SXT element. *The Journal of antimicrobial chemotherapy*. 2008 Nov;62(5):929–33.
- Paauw A, Caspers MPM, Schuren FHJ, Leverstein-van Hall M a, Delétoile A, Montijn RC, et al. Genomic diversity within the *Enterobacter cloacae* complex. *PloS one*. 2008 Jan;3(8):e3018.
- Pascual J, Macián MC, Arahal DR, Garay E, Pujalte MJ. Multilocus sequence analysis of the central clade of the genus *Vibrio* by using the 16S rRNA, *recA*, *pyrH*, *rpoD*, *gyrB*, *rctB* and *toxR* genes. *International journal of systematic and evolutionary microbiology*. 2010 Jan;60(Pt 1):154–65.
- Piarroux R, Barraïis R, Faucher B, Haus R, Piarroux M, Gaudart J, et al. Understanding the cholera epidemic, Haiti. *Emerging infectious diseases*. 2011 Jul;17(7):1161–8.

- Piarroux R, Faucher B. Cholera epidemics in 2010: respective roles of environment, strain changes, and human-driven dissemination. *Clinical microbiology and infection*. 2012 Mar;18(3):231–8.
- Popovic T, Bopp C, Olsvik O, Wachsmuth K. Epidemiologic application of a standardized ribotype scheme for *Vibrio cholerae* O1. *Journal of clinical microbiology*. 1993 Sep;31(9):2474–82.
- Quilici ML, Massenet D, Gake B, Bwalki B, Olson DM. *Vibrio cholerae* O1 variant with reduced susceptibility to ciprofloxacin, Western Africa. *Emerging infectious diseases*. 2010 Nov;16(11):1804–5.
- Reeves PR, Lan R. Cholera in the 1990s. *British Medical Bulletin*. 1998;54(3):611–23.
- Reimer AR, Domselaar G Van, Stroika S, Walker M, Kent H, Tarr C, et al. Comparative Genomics of *Vibrio cholerae* from Haiti, Asia, and Africa. *Emerging infectious diseases*. 2011 Nov;17(11):2113–21.
- Rowe-Magnus D a, Guerout A-M, Biskri L, Bouige P, Mazel D. Comparative analysis of superintegrons: engineering extensive genetic diversity in the Vibrionaceae. *Genome research*. 2003 Mar;13(3):428–42.
- Rowe-Magnus DA, Guérout AM, Mazel D. Super-integrons. *Research in microbiology*. 1999;150(9-10):641–51.
- Ryan ET. The Cholera Pandemic, Still with Us after Half a Century: Time to Rethink. *PLoS Neglected Tropical Diseases*. 2011 Jan 25;5(1):e1003.
- Sack D a. How many cholera deaths can be averted in Haiti? *Lancet*. 2011 Apr 9;377(9773):1214–6.

- Sack D a, Sack RB, Nair GB, Siddique a K. Cholera. *Lancet*. 2004 Jan 17;363(9404):223–33.
- Safa A. *Vibrio cholerae* O1 Hybrid El Tor Strains, Asia and Africa. *Emerging Infectious Diseases*. Centers for Disease Control; 2008 Jun;14(6):987–8.
- Safa A, Nair GB, Kong RYC. Evolution of new variants of *Vibrio cholerae* O1. *Trends in microbiology*. 2010 Jan;18(1):46–54.
- Salim A, Lan R, Reeves PR. *Vibrio cholerae* pathogenic clones. *Emerging infectious diseases*. 2005 Nov;11(11):1758–60.
- Seas C, Miranda J, Gil AI, Leon-Barua R, Patz J, Huq A, et al. New insights on the emergence of cholera in Latin America during 1991: the Peruvian experience. *The American journal of tropical medicine and hygiene*. 2000 May;62(4):513–7.
- Senderovich Y, Izhaki I, Halpern M. Fish as reservoirs and vectors of *Vibrio cholerae*. *PloS one*. 2010 Jan;5(1):e8607.
- Sjölund-Karlsson M, Reimer A, Folster JP, Walker M, Dahourou GA, Batra DG, et al. Drug-resistance mechanisms in *Vibrio cholerae* O1 outbreak strain, Haiti, 2010. *Emerging infectious diseases*. 2011 Nov 5;17(11):2151–4.
- Smith JM, Feil EJ, Smith NH. Population structure and evolutionary dynamics of pathogenic bacteria. *BioEssays: news and reviews in molecular, cellular and developmental biology*. 2000 Dec;22(12):1115–22.
- Talkington D, Bopp C, Tarr C, Parsons MB, Dahourou G, Freeman M, et al. Characterization of Toxigenic *Vibrio cholerae* from Haiti, 2010-2011. *Emerging infectious diseases*. 2011 Nov;17(11):2122–9.

- Taviani E, Grim CJ, Choi J, Chun J, Haley B, Hasan N a, et al. Discovery of novel *Vibrio cholerae* VSP-II genomic islands using comparative genomic analysis. *FEMS microbiology letters*. 2010 Jul;308(2):130–7.
- Thompson CC, Freitas FS, Marin MA, Fonseca EL, Okeke IN, Vicente ACP. *Vibrio cholerae* O1 lineages driving cholera outbreaks during seventh cholera pandemic in Ghana. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2011 a Aug 30;11(8):1951–6.
- Thompson CC, Marin MA, Dias GM, Dutilh BE, Edwards R a, lida T, et al. Genome Sequence of the Human Pathogen *Vibrio cholerae* Amazonia. *Journal of bacteriology*. 2011 b Oct;193(20):5877–8.
- Thompson CC, Thompson FL, Vicente ACP. Identification of *Vibrio cholerae* and *Vibrio mimicus* by multilocus sequence analysis (MLSA). *International journal of systematic and evolutionary microbiology*. 2008;58(3):617.
- Thompson CC, Vicente ACP, Souza RC, Vasconcelos ATR, Vesth T, Alves N, et al. Genomic taxonomy of *Vibrios*. *BMC evolutionary biology*. 2009 Jan;9(1):258.
- Tortora, GJ. Funke, BR, Case C. *Microbiology: an Introduction*. 8th Editio. Pearson Education Inc.; 2004.
- Vanden Broeck D, Horvath C, De Wolf MJS. *Vibrio cholerae*: cholera toxin. *The international journal of biochemistry & cell biology*. 2007 Jan;39(10):1771–5.
- Wachsmuth IK, Evins GM, Fields PI, Olsvik O, Popovic T, Bopp CA, et al. The molecular epidemiology of cholera in Latin America. *The Journal of infectious diseases*. 1993 Mar;167(3):621–6.

- Waldor MK, Mekalanos JJ. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science*. 1996 Jun 28;272(5270):1910–4.
- Waldor MK, Tschäpe H, Mekalanos JJ. A new type of conjugative transposon encodes resistance to sulfamethoxazole, trimethoprim, and streptomycin in *Vibrio cholerae* O139. *Journal of bacteriology*. 1996 Jul;178(14):4157–65.
- Wozniak RAF, Fouts DE, Spagnoletti M, Colombo MM, Ceccarelli D, Garriss G, et al. Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. *PLoS genetics*. 2009 Dec;5(12):e1000786.
- Yamamoto K, Do Valle GR, Xu M, Miwatani T, Honda T. Amino acids of the cholera toxin from *Vibrio cholerae* O37 strain S7 which differ from those of strain O1. *Gene*. 1995 Sep 22;163(1):155–6.
- Zhao Y, Wu J, Yang J, Sun S, Xiao J, Yu J. PGAP: pan-genomes analysis pipeline. *Bioinformatics*. 2012 Feb 1;28(3):416–8.

Anexos

9 ANEXOS

9.1 Anexo I: Genomas e isolados de *Vibrio cholerae* usados nas análises do MLSA

Organismo	Ano	Continente	País	Fonte	Sorogrupo
12129	1985	Austrália	Austrália	Ambiental	O1
1587	1994	América	Peru	Clínica	O12
2010EL1786	2010	América	Haiti	Clínica	O1
2010EL1792	2010	América	Haiti	Clínica	O1
2010EL1798	2010	América	Haiti	Clínica	O1
2740-80	1980	América	USA	Ambiental	O1
623-39	2002	Ásia	Bangladesh	Ambiental	Não-O1/Não-O139
AM19226	2001	Ásia	Bangladesh	Clínica	O39
B33	2004	África	Moçambique	Clínica	O1
BX330286	1986	Austrália	Austrália	Ambiental	O1
CIRS101	2002	Ásia	Bangladesh	Clínica	O1
CT5369-93	1993	América	Brasil	Ambiental	desconhecido
ERS013124	1993	Ásia	Índia	Clínica	O139
ERS013125	2002	Ásia	Bangladesh	Clínica	O139
ERS013126	1994	Ásia	Bangladesh	Clínica	O1
ERS013127	2004	Ásia	Índia	Clínica	O1
ERS013128	1970	Ásia	Índia	Clínica	O1
ERS013129	2006	Ásia	Bangladesh	Clínica	O1
ERS013130	1989	Ásia	Índia	Clínica	O1
ERS013131	1990	Ásia	Índia	Clínica	O1
ERS013132	1991	Ásia	Índia	Clínica	O1
ERS013133	1992	Ásia	Índia	Clínica	O1
ERS013134	2004	Ásia	Índia	Clínica	O1
ERS013135	1991	Ásia	Bangladesh	Clínica	O1
ERS013136	1971	Ásia	Índia	Clínica	O1
ERS013137	1973	Ásia	Índia	Clínica	O1
ERS013138	1984	Ásia	Índia	Clínica	O1
ERS013139	1992	Ásia	Índia	Clínica	O1
ERS013140	1975	Europa	Alemanha	Clínica	O1
ERS013141	1978	Ásia	Malásia	Clínica	O1
ERS013142	1978	Ásia	Bahrain	Clínica	O1
ERS013143	1979	Ásia	Índia	Clínica	O1
ERS013145	1980	Ásia	Índia	Clínica	O1
ERS013146	1979	Ásia	Índia	Clínica	O1
ERS013147	2009	Ásia	Índia	Clínica	O1
ERS013160	1964	desconhecido	desconhecido	Clínica	O1
ERS013161	1962	desconhecido	desconhecido	Clínica	O1
ERS013162	1969	Ásia	Bangladesh	Clínica	O1
ERS013163	1982	Ásia	Bangladesh	Clínica	O1
ERS013164	1963	Ásia	Bangladesh	Clínica	O1
ERS013165	1949	África	Egito	Clínica	O1

ERS013166	1980	Ásia	Índia	Clínica	O1
ERS013167	1970	Ásia	Índia	Clínica	O1
ERS013168	1958	Ásia	Tailândia	Clínica	O1
ERS013169	1970	Ásia	Índia	Clínica	O1
ERS013170	1962	Ásia	Bangladesh	Clínica	O1
ERS013171	1949	África	Egito	Clínica	O1
ERS013172	1990	desconhecido	desconhecido	Clínica	O1
ERS013173	1990	desconhecido	desconhecido	Clínica	O1
ERS013174	1992	América	Colômbia	Clínica	O1
ERS013175	1992	América	Colômbia	Clínica	O1
ERS013176	1990	desconhecido	desconhecido	Clínica	O1
ERS013177	1989	Ásia	Índia	Clínica	O1
ERS013178	1989	Ásia	Índia	Clínica	O1
ERS013179	1991	África	Moçambique	Clínica	O1
ERS013180	1991	África	Moçambique	Clínica	O1
ERS013181	1991	África	Moçambique	Clínica	O1
ERS013182	1992	América	Colômbia	Clínica	O1
ERS013183	1992	América	Colômbia	Clínica	O1
ERS013184	1992	América	Argentina	Clínica	O1
ERS013185	1992	América	Bolívia	Clínica	O1
ERS013186	1991	América	México	Clínica	O1
ERS013187	1989	Ásia	Vietnam	Clínica	O1
ERS013188	1992	América	Argentina	Clínica	O1
ERS013189	1992	América	Argentina	Clínica	O1
ERS013190	1980	América	USA	Clínica	O1
ERS013191	1984	América	USA	Clínica	O1
ERS013192	1985	América	USA	Clínica	O1
ERS013193	1986	América	USA	Clínica	O1
ERS013194	1986	América	USA	Clínica	O1
ERS013195	1991	América	México	Clínica	O1
ERS013196	1989	Ásia	Vietnam	Clínica	O1
ERS013197	1990	Europa	Suécia	Clínica	O1
ERS013198	2007	África	Djibuti	Clínica	O1
ERS013199	2007	Ásia	Bangladesh	Clínica	O1
ERS013200	1993	América	Argentina	Clínica	O1
ERS013201	1993	América	Argentina	Clínica	O1
ERS013202	1994	Ásia	Bangladesh	Clínica	O1
ERS013203	1987	Ásia	Bangladesh	Clínica	O1
ERS013204	1987	Ásia	Bangladesh	Clínica	O1
ERS013205	1987	Ásia	Bangladesh	Clínica	O1
ERS013206	2007	África	Djibuti	Clínica	O1
ERS013207	2007	África	Djibuti	Clínica	O1
ERS013208	2007	África	Quênia	Clínica	O1
ERS013209	2005	África	Quênia	Clínica	O1
ERS013210	2007	África	Quênia	Clínica	O1
ERS013211	2005	África	Quênia	Clínica	O1
ERS013212	2005	África	Quênia	Clínica	O1
ERS013213	2007	África	Quênia	Clínica	O1
ERS013214	2005	África	Quênia	Clínica	O1

ERS013215	2005	África	Quênia	Clínica	O1
ERS013216	2007	África	Quênia	Clínica	O1
ERS013217	2007	África	Quênia	Clínica	O1
ERS013218	2007	África	Quênia	Clínica	O1
ERS013219	2007	África	Quênia	Clínica	O1
ERS013220	2009	África	Quênia	Clínica	O1
ERS013221	2009	África	Quênia	Clínica	O1
ERS013224	2009	África	Quênia	Clínica	O1
ERS013225	2009	África	Quênia	Clínica	O1
ERS013226	2009	África	Quênia	Clínica	O1
ERS013232	2007	Ásia	Índia	Clínica	O1
ERS013233	2007	Ásia	Índia	Clínica	O1
ERS013234	2006	Ásia	Índia	Clínica	O1
ERS013235	2009	África	Tanzânia	Clínica	O1
ERS013236	2007	Ásia	Índia	Clínica	O1
ERS013237	2007	Ásia	Índia	Clínica	O1
ERS013238	2001	Ásia	Bangladesh	Clínica	O1
ERS013239	2005	Ásia	Índia	Clínica	O1
ERS013240	2007	Ásia	Índia	Clínica	O1
ERS013241	2005	África	Moçambique	Clínica	O1
ERS013242	2005	África	Moçambique	Clínica	O1
ERS013243	1980	Ásia	Índia	Clínica	O1
ERS013244	1973	desconhecido	desconhecido	Clínica	O1
ERS013245	1989	África	Angola	Clínica	O1
ERS013246	1957	Ásia	Indonésia	Clínica	O1
ERS013247	1979	Ásia	Bangladesh	Clínica	O1
ERS013248	1977	Ásia	Índia	Clínica	O1
ERS013249	1971	Ásia	Bangladesh	Clínica	O1
ERS013250	1979	Ásia	Bangladesh	Clínica	O1
ERS013251	1991	América	Peru	Clínica	O1
ERS013252	1995	Ásia	Vietnam	Clínica	O1
ERS013253	2002	Ásia	Vietnam	Clínica	O1
ERS013254	2004	Ásia	Índia	Clínica	O1
ERS013255	2006	Ásia	Índia	Clínica	O1
ERS013256	1991	Ásia	Bangladesh	Clínica	O1
ERS013257	2007	Ásia	Índia	Clínica	O1
ERS013258	2006	Ásia	Índia	Clínica	O1
ERS013259	2001	Ásia	Bangladesh	Clínica	O1
ERS013260	1999	Ásia	Bangladesh	Clínica	O1
ERS013261	2001	Ásia	Bangladesh	Clínica	O1
ERS013262	1994	Ásia	Bangladesh	Clínica	O1
ERS013263	2001	Ásia	Bangladesh	Clínica	O1
ERS013264	2007	Ásia	Vietnam	Clínica	O1
ERS013265	2005	África	Moçambique	Clínica	O1
ERS013266	2007	Ásia	Índia	Clínica	O1
ERS013267	2007	Ásia	Índia	Clínica	O1
ERS013268	2007	Ásia	Índia	Clínica	O1
ERS013269	2007	Ásia	Índia	Clínica	O1
ERS013270	2004	Ásia	Índia	Clínica	O1

ERS013271	2004	Ásia	Vietnam	Clínica	O1
ERS013273	2003	Ásia	Vietnam	Clínica	O1
ERS013274	1991	América	Peru	Clínica	O1
ERS013275	1991	América	Peru	Clínica	O1
ERS013276	1991	América	Peru	Clínica	O1
ERS013277	1985	América	USA	Clínica	O1
ERS013278	1994	Ásia	Bangladesh	Clínica	O1
ERS013279	2006	Ásia	Bangladesh	Clínica	O1
ERS013280	1993	América	Argentina	Clínica	O1
ERS013281	2007	Ásia	Bangladesh	Clínica	O1
ERS013282	1991	Ásia	Bangladesh	Clínica	O1
ERS016137	2000	Ásia	Bangladesh	Clínica	O1
INDRE91	1991	América	México	Clínica	O1
LMA1143	1992	América	Brasil	Ambiental	Não-O1
LMA1169	1993	América	Brasil	Ambiental	Não-O1
LMA1171	1993	América	Brasil	Ambiental	Não-O1
LMA1173	1993	América	Brasil	Ambiental	Não-O1
LMA3984	2007	América	Brasil	Ambiental	O1
LMA454	1995	América	Brasil	Ambiental	Não-O1
M66-2	1937	Ásia	Indonésia	Clínica	O1
MAK757	1937	Ásia	Indonésia	Clínica	O1
MJ1236	1994	Ásia	Bangladesh	Clínica	O1
MO10	1992	Ásia	Índia	Clínica	O139
MZO-2	2001	Ásia	Bangladesh	Clínica	O14
MZO-3	2001	Ásia	Bangladesh	Clínica	O37
N16961	1975	Ásia	Bangladesh	Clínica	O1
NCTC8457	1910	Ásia	Arábia Saudita	Clínica	O1
O395	1965	Ásia	Índia	Clínica	O1
R-18246	1992	América	Brasil	Clínica	Não-O1/Não-O139
R-18273	1992	América	Brasil	Ambiental	Não-O1/Não-O139
R-18303	2001	América	Brasil	Ambiental	Não-O1/Não-O139
R-18304	1993	América	Brasil	Clínica	O1
R-18308	1973	Ásia	Índia	Clínica	O1
R-18316	1993	América	Brasil	Clínica	O1
R-18317	1994	América	Brasil	Clínica	O1
R-18327	1999	América	Brasil	Clínica	O1
R-18332	1991	América	Brasil	Clínica	O1
R-18338	1992	América	Brasil	Clínica	Não-O1/Não-O139
R-18348	1999	América	Brasil	Clínica	Não-O1/Não-O139
R-18356	1991	América	Brasil	Clínica	O1
RC27	1991	Ásia	Indonésia	Clínica	O1
RC385	1998	América	USA	Ambiental	O135
RC9	1985	África	Quênia	Clínica	O1
TM11079-80	1980	América	Brasil	Ambiental	O1
TMA21	1982	América	Brasil	Ambiental	Não-O1/Não-O139
V107i	1980	África	Gana	Clínica	O1
V33i	2006	África	Gana	Clínica	O1
V34i	2006	África	Gana	Clínica	O1
V35i	2006	África	Gana	Clínica	O1

V42i	2006	África	Gana	Clínica	O1
V45i	2006	África	Gana	Clínica	O1
V47i	2006	África	Gana	Clínica	O1
V51	1987	América	USA	Clínica	O141
V51i	2006	África	Gana	Clínica	O1
V52	1968	África	Sudão	Clínica	O37
V52i	2006	África	Gana	Clínica	O1
V53i	2006	África	Gana	Clínica	O1
V78i	2006	África	Gana	Clínica	O1
V84i	2006	África	Gana	Clínica	O1
V90i	2006	África	Gana	Clínica	O1
V97i	2006	África	Gana	Clínica	O1
V98i	2006	África	Gana	Clínica	O1
VC04	ND	Ásia	Índia	desconhecido	O1
VC08	1992	Ásia	Índia	Clínica	O139
VC1001	2010	África	Nigéria	Clínica	O1
VC1004	2010	África	Nigéria	Clínica	O1
VC1005	2010	África	Nigéria	Ambiental	Não-O1/Não-O139
VC1006	2010	África	Nigéria	Clínica	Não-O1
VC1007	2010	África	Nigéria	Clínica	O1
VC1009	2010	África	Nigéria	Clínica	Não-O1
VC101	1979	África	Gana	Clínica	O1
VC102	1979	África	Gana	Clínica	O1
VC104	1972	África	Algéria	desconhecido	O1
VC106	1980	África	Gana	Clínica	O1
VC107	1980	África	Gana	Clínica	O1
VC111	1972	África	Nigéria	Clínica	O139
VC115	ND	África	Gana	Clínica	O1
VC120	1973	Ásia	Índia	desconhecido	O1
VC135	1993	América	Brasil	Ambiental	O1
VC137	1992	América	Brasil	Clínica	O1
VC14	1970	Ásia	Israel	desconhecido	O1
VC141	1991	América	Peru	Clínica	Não-O1
VC150	1992	Ásia	Nepal	Clínica	O139
VC151	1986	América	México	Ambiental	Não-O1
VC153	1991	América	Brasil	Ambiental	O1
VC156	1993	América	Brasil	Clínica	O1
VC163	1970	África	Gana	Clínica	O1
VC172	1993	Ásia	Bangladesh	Clínica	O139
VC174	ND	Ásia	Índia	Clínica	O1
VC179	1993	Ásia	Bangladesh	Clínica	O139
VC182	1983	Austrália	Austrália	Ambiental	O1
VC192	1980	África	Gana	Clínica	O1
VC193	1981	África	Gana	Clínica	O1
VC194	1981	África	Gana	Clínica	O1
VC198	1991	América	Brasil	Ambiental	O1
VC20	1993	África	Brasil	Clínica	O1
VC200	1986	Ásia	Paquistão	Clínica	O1
VC201	1986	Ásia	Paquistão	Clínica	O1

VC21	1970	Ásia	China	desconhecido	O1
VC214	desconhecido	Ásia	Paquistão	Clínica	O1
VC216	desconhecido	Ásia	Bangladesh	Clínica	O1
VC22	1970	Ásia	Myanmar	desconhecido	O1
VC225	1976	África	Gana	Clínica	O1
VC27	1971	Ásia	Índia	Clínica	O1
VC28	1993	América	Brasil	Clínica	O1
VC307	1986	Austrália	Austrália	Ambiental	O1
VC311	1986	Austrália	Austrália	Clínica	Não-O1/Não-O139
VC32	1977	Ásia	Kuwait	Clínica	O1
VC33	1979	África	Tanzânia	Clínica	O1
VC34	1970	África	Gana	Clínica	O1
VC341	1971	Europa	França	Clínica	O1
VC346	1991	América	Peru	Clínica	O1
VC347	1991	América	Brasil	Clínica	O1
VC348	1991	América	Brasil	Clínica	O1
VC350	1991	América	Brasil	Clínica	O1
VC37	1980	África	Gana	Clínica	O1
VC38	1977	Europa	Inglaterra	Ambiental	O1
VC39	1977	Europa	Inglaterra	Ambiental	O1
VC40	1977	Europa	Inglaterra	Ambiental	O1
VC409	1994	América	Brasil	Clínica	O1
VC434	1983	Ásia	Singapura	Clínica	O1
VC46	1966	África	Algéria	Ambiental	Não-O1/Não-O139
VC470	1999	América	Brasil	Clínica	O1
VC473	1994	América	Peru	Ambiental	O1
VC492	1993	Europa	França	Clínica	O1
VC500	1993	África	Ruanda	Clínica	O1
VC504	1994	Ásia	Nepal	Clínica	O1
VC508	1994	América	Guiana Francesa	Clínica	O1
VC512	1995	América	Brasil	Clínica	O1
VC514	1995	América	Brasil	Clínica	O1
VC527	1998	América	Peru	Clínica	O1
VC566	1991	América	Brasil	Clínica	O1
VC568	1991	América	Peru	Clínica	O1
VC570	1991	América	Colômbia	Clínica	O1
VC592	1999	América	Brasil	Clínica	O1
VC593	ND	África	Sudão	Clínica	O37
VC609	ND	América	Peru	Ambiental	Não-O1
VC610	ND	América	Peru	Ambiental	O1
VC620	ND/	América	Peru	Ambiental	Não-O1
VC622	ND	América	Peru	Ambiental	Não-O1
VC625	1998	América	Peru	Clínica	O1
VC626	1998	América	Peru	Clínica	O1
VC627	1998	América	Peru	Clínica	O1
VC630	ND	América	Peru	Clínica	O1
VC696	1979	América	Guatemala	Clínica	Não-O1
VC727	1992	América	Brasil	Clínica	O1
VC735	1992	América	Brasil	Clínica	O1

VC75	1971	Ásia	Indonésia	desconhecido	O1
VC754	ND	América	Argentina	Ambiental	Não-O1
VC755	2002	América	Argentina	Ambiental	O1
VC758	ND	América	Argentina	Ambiental	O1
VC77	1971	Ásia	Indonésia	desconhecido	O1
VC79	1971	África	Nigéria	Clínica	O1
VC815	1994	América	Brasil	Clínica	O1
VC818	1993	América	Brasil	Clínica	O1
VC820	1995	América	Brasil	Clínica	O1
VC821	1993	América	Brasil	Clínica	O1
VC822	1993	América	Brasil	Clínica	O1
VC826	1992	América	Brasil	Ambiental	O1
VC83	1971	Ásia	Myanmar	desconhecido	O1
VC832	2009	África	Nigéria	Clínica	O1
VC833	2010	África	Nigéria	Clínica	O1
VC834	2010	África	Nigéria	Clínica	O1
VC835	2010	África	Nigéria	Clínica	O1
VC836	2010	África	Nigéria	Clínica	O1
VC841	2009	África	Nigéria	Clínica	O1
VC869	ND	África	Nigéria	desconhecido	O1
VC875	ND	África	Tanzânia	desconhecido	O1
VC898	1993	América	Brasil	Ambiental	O1
VC909	1991	América	Brasil	Ambiental	O1
VC91	1971	Ásia	Rússia	Clínica	O1
VC910	1991	América	Brasil	Ambiental	O1
VC918	1991	América	Peru	Clínica	O1
VC919	1991	América	Peru	Clínica	O1
VC92	1979	África	Gana	Clínica	O1
VC920	1990s	América	Peru	desconhecido	O1
VC921	1990s	América	Peru	desconhecido	O1
VC922	1990s	América	Peru	desconhecido	O1
VC923	1990s	América	Peru	desconhecido	O1
VC924	1990s	América	Peru	desconhecido	O1
VC925	1990s	América	Peru	desconhecido	O1
VC927	1990s	América	Peru	desconhecido	O1
VC930	1990s	América	Peru	desconhecido	O1
VC931	1990s	América	Peru	desconhecido	O1
VC932	1990s	América	Peru	desconhecido	O1
VC933	1990s	América	Peru	desconhecido	O1
VC934	1990s	América	Peru	desconhecido	O1
VC938	1990s	América	Peru	desconhecido	O1
VC939	1990s	América	Peru	Clínica	O1
VC94	1979	África	Gana	Clínica	O1
VC944	1990s	América	Peru	desconhecido	O1
VC946	1990s	América	Peru	desconhecido	O1
VC95	1978	África	Gana	Clínica	O1
VC950	1990s	América	Peru	desconhecido	Não-O1
VC957	1990s	América	Peru	desconhecido	Não-O1
VC96	1978	África	Gana	Clínica	O1

VC960	1990s	América	Peru	desconhecido	Não-O1
VC966	1990s	América	Peru	desconhecido	Não-O1
VC967	1990s	América	Peru	desconhecido	Não-O1
VC969	1990s	América	Peru	desconhecido	Não-O1
VC97	1979	África	Gana	Clínica	O1
VC974	1990s	América	Peru	desconhecido	Não-O1
VC978	1990s	América	Peru	desconhecido	Não-O1
VC98	1979	África	Gana	Clínica	O1
VC984	1997	América	Peru	Clínica	O1
VC985	1997	América	Peru	Clínica	O1
VC987	1997	América	Peru	Clínica	O1
VC988	1997	América	Peru	Clínica	O1
VC989	1997	América	Peru	Clínica	O1
VC991	2010	África	Nigéria	Clínica	O1
VC996	2010	África	Nigéria	Clínica	Não-O1
VC997	2010	África	Nigéria	Clínica	O1
VC998	2010	África	Nigéria	Clínica	Não-O1/Não-O139
VC999	2010	África	Nigéria	Clínica	O1
VL426	1979	Europa	Inglaterra	Ambiental	Não-O1/Não-O139
Vm523-80	1980	América	USA	Clínica	O115

ND: informação não disponível. Vm523-80 é um *Vibrio mimicus*.

9.2 Anexo II: Iniciadores usados no estudo

Iniciadores	Sequência (5'-3')
Genes MLSA	
recA-01-F	TGARAARCARTTYGGTAAAGG
recA-02-R	TCRCCNTRTAGCTRACC
pyrH-04-F	ATGASNACBAAYCCWAAACC
pyrH-02-R	GTRAABGCNGMYARRTCCA
mdh-1	ATGAAAGTCGCTGTTATT
mdh-2	GTATCTAACATGCCATCC
Genes relacionados a biotipo	
Sorogroup O1	
O1-F	GTTTCACTGAACAGATGGG
O1-R	GGTCATCTGTAAGTACAAC
Genes relacionados a virulência	
VSP-II	
VC0511F1	CTTGCTGCGTACTTAGCA
VC0511R1	AGTAGCATCGCTCTCGTA
VC0513F1	CTGAGGTGTTATATGTTTCG
VC0513R1	TCAAATTTCTGACAGTTCC
VPI-I	
TCPA1	CACGATAAGAAAACCGGTCAAGAG
TCPA2	ACCAAATGCAACGCCGAATGGAGC
CTXΦ	
CT8	GCAGTCAGGTGGTCTTATTGC
CT10	TCCAGATATGCAATCCTCAG
Genes relacionados a resistência	
Relacionado a Integron	
INT1 F	AAAACCGCCACTGCGCCGTTA
INT1 R	GAAGACGGCTGCACTGAACG
INT2 F	GCGTTTTATGTCTAACAGTCC
INT2 R	AAGTAGCATCAGTCCATCC
Integrase SXT	
SXTF	TCGGGTATCGCCCAAGGGCA
SXTR	GCGAAGATCATGCATAGACC
Resistência a quinolona	
gyrA_vF	AATGTGCTGGGCAACGACTGG
gyrA_vR	GTGCGCGATTTTCGACATACG
gyrB_vF	GGAAATGACTCGCCGTAAGG
gyrB_vR	GTTGTGATAACGCAGTTTATCTGGG
parC_vF	GTCTGAGTTGGGTCTCTCGGC
parC_vR	AGAATCTCGGCAAACCTTGACAG
parE_vF	ATGCGTGCCAGCAAGAAAGTG
parE_vR	TTATCGCTGTCAGGGTCAATCC

9.3 Anexo III: Genes do superintegron de genomas de *V. cholerae* usados para a caracterização de linhagens

Gene	Anotação no genoma	Isolado	Gene	Anotação no genoma	Isolado
1	Integron integrase IntI4	N16961	349	hypothetical protein	O395
2	hypothetical protein	N16961	350	hypothetical protein	O395
3	hypothetical protein	N16961	351	Death on curing protein, Doc toxin	O395
4	conserved hypothetical protein	N16961	352	hypothetical protein	O395
5	hypothetical protein	N16961	353	hypothetical protein	O395
6	hypothetical protein	N16961	354	hypothetical protein	O395
7	hypothetical protein	N16961	355	Death on curing protein, Doc toxin	O395
8	Alcohol dehydrogenase, class IV	N16961	356	hypothetical protein	O395
9	Acetyltransferase	N16961	357	FIG032766: hypothetical protein	O395
10	Putative oxidoreductase	N16961	358	FIG001353: Acetyltransferase	O395
11	hypothetical protein	N16961	359	hypothetical protein	O395
12	hypothetical protein	N16961	360	conserved hypothetical protein	O395
13	hypothetical protein	N16961	361	hypothetical protein	O395
14	hypothetical protein	N16961	362	Glyoxalase family protein	O395
15	Death on curing protein, Doc toxin	N16961	363	hypothetical protein	O395
16	hypothetical protein	N16961	364	Histone acetyltransferase HPA2	O395
17	hypothetical protein	N16961	365	hypothetical protein	O395
18	hypothetical protein	N16961	366	hypothetical protein	O395
19	Deoxyguanosinetriphosphate triphosphohydrolase (EC 3.1.5.1)	N16961	367	hypothetical protein	O395
20	5-methylcytosine-specific restriction enzyme A (EC 3.1.21.-)	N16961	368	Phenazine biosynthesis protein PhzF	O395
21	hypothetical protein	N16961	369	conserved hypothetical protein	O395
22	ParE toxin protein	N16961	370	hypothetical protein	O395
23	ParD protein (antitoxin to ParE)	N16961	371	HigA protein (antitoxin to HigB)	O395
24	MutT/nudix family protein	N16961	372	HigB toxin protein	O395
25	hypothetical protein	N16961	373	hypothetical protein	O395
26	acetyltransferase, putative	N16961	374	hypothetical protein	O395
27	lipoprotein Blc	N16961	375	toxin resistance protein	O395
28	FIG001353: Acetyltransferase	N16961	376	ParD protein (antitoxin to ParE)	O395
29	FIG032766: hypothetical protein	N16961	377	ParE toxin protein	O395
30	hypothetical protein	N16961	378	hypothetical protein	O395
31	hypothetical protein	N16961	379	Integron integrase IntI4	O395
32	hypothetical protein	N16961	380	Large Subunit Ribosomal RNA	M66-2
33	YafQ toxin protein	N16961	381	5S RNA	M66-2
34	DNA-damage-inducible protein J	N16961	382	tRNA-Asp-GTC	M66-2
35	conserved hypothetical protein	N16961	383	CDP-diacylglycerol--serine O-phosphatidyltransferase (EC 2.7.8.8)	M66-2
36	hypothetical protein	N16961	384	Predicted acyltransferase	M66-2
37	biphenyl-2,3-diol 1,2-dioxygenase III-related protein	N16961	385	UDP-N-acetylenolpyruvoylglucosamine reductase (EC 1.1.1.158)	M66-2
38	Alcohol dehydrogenase, class IV	N16961	386	Biotin-protein ligase (EC 6.3.4.15) / Biotin operon repressor	M66-2
39	conserved hypothetical protein	N16961	387	Pantothenate kinase (EC 2.7.1.33)	M66-2
40	conserved hypothetical protein	N16961	388	Translation elongation factor Tu	M66-2
41	COG1764: Predicted redox protein, regulator of disulfide bond formation	N16961	389	Preprotein translocase subunit SecE (TC 3.A.5.1.1)	M66-2
42	hypothetical protein	N16961	390	Transcription antitermination protein NusG	M66-2
43	Helix-turn-helix protein, CopG family	N16961	391	LSU ribosomal protein L11p (L12e)	M66-2
44	hypothetical protein	N16961	392	LSU ribosomal protein L1p (L10Ae)	M66-2

45	hypothetical protein	N16961	393	LSU ribosomal protein L10p (P0)	M66-2
46	Muramoyltetrapeptide carboxypeptidase (EC 3.4.17.13)	N16961	394	LSU ribosomal protein L7/L12 (P1/P2)	M66-2
47	Glyoxalase family protein	N16961	395	DNA-directed RNA polymerase beta subunit (EC 2.7.7.6)	M66-2
48	hypothetical protein	N16961	396	DNA-directed RNA polymerase beta' subunit (EC 2.7.7.6)	M66-2
49	Death on curing protein, Doc toxin	N16961	397	Regulator of sigma D	M66-2
50	biphenyl-2,3-diol 1,2-dioxygenase III-related protein	N16961	398	NADH pyrophosphatase (EC 3.6.1.22)	M66-2
51	hypothetical protein	N16961	399	Uroporphyrinogen III decarboxylase (EC 4.1.1.37)	M66-2
52	Alcohol dehydrogenase, class IV	N16961	400	Transcriptional regulator, TetR family	M66-2
53	hypothetical protein	N16961	401	hypothetical protein	M66-2
54	H-REV 107-related protein	N16961	402	Cell wall endopeptidase, family M23/M37	M66-2
55	Glyoxalase family protein	N16961	403	2,3-bisphosphoglycerate-independent phosphoglycerate mutase (EC 5.4.2.1)	M66-2
56	YafQ toxin protein	N16961	404	Permease of the drug/metabolite transporter (DMT) superfamily	M66-2
57	DNA-damage-inducible protein J	N16961	405	Transporter, putative	M66-2
58	lipoprotein Blc	N16961	406	Phosphatidylserine decarboxylase (EC 4.1.1.65)	M66-2
59	hypothetical protein	N16961	407	Ribosome small subunit-stimulated GTPase EngC	M66-2
60	hypothetical protein	N16961	408	3'-to-5' oligoribonuclease (orn)	M66-2
61	Putative threonine efflux protein	N16961	409	Iron-sulfur cluster-binding protein	M66-2
62	hypothetical protein	N16961	410	ATPase YjeE, predicted to have essential role in cell wall biosynthesis	M66-2
63	hypothetical protein	N16961	411	N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28)	M66-2
64	hypothetical protein	N16961	412	DNA mismatch repair protein MutL	M66-2
65	hypothetical protein	N16961	413	tRNA delta(2)-isopentenylpyrophosphate transferase (EC 2.5.1.8)	M66-2
66	hypothetical protein	N16961	414	RNA-binding protein Hfq	M66-2
67	Plasmid stabilization element ParE, putative	N16961	415	GTP-binding protein HflX	M66-2
68	Predicted transcriptional regulator	N16961	416	HflK protein	M66-2
69	hypothetical protein	N16961	417	HflC protein	M66-2
70	hypothetical protein	N16961	418	Putative inner membrane protein YjeT (clustered with HflC)	M66-2
71	hypothetical protein	N16961	419	Protein slyX	M66-2
72	hypothetical protein	N16961	420	FOG: WD40 repeat	M66-2
73	hypothetical protein	N16961	421	FKBP-type peptidyl-prolyl cis-trans isomerase FkpA precursor (EC 5.2.1.8)	M66-2
74	hypothetical protein	N16961	422	YheO-like PAS domain	M66-2
75	hypothetical protein	N16961	423	tRNA 5-methylaminomethyl-2-thiouridine synthase TusD	M66-2
76	hypothetical protein	N16961	424	tRNA 5-methylaminomethyl-2-thiouridine synthase TusC	M66-2
77	transposase OrfAB, subunit B	N16961	425	tRNA 5-methylaminomethyl-2-thiouridine synthase TusB	M66-2
78	Transposase	N16961	426	SSU ribosomal protein S12p (S23e)	M66-2
79	hypothetical protein	N16961	427	SSU ribosomal protein S7p (S5e)	M66-2
80	hypothetical protein	N16961	428	Translation elongation factor G	M66-2
81	conserved hypothetical protein	N16961	429	Translation elongation factor Tu	M66-2
82	hypothetical protein	N16961	430	Bacterioferritin-associated ferredoxin	M66-2
83	hypothetical protein	N16961	431	Bacterioferritin	M66-2
84	hypothetical protein	N16961	432	SSU ribosomal protein S6p	M66-2
85	hypothetical protein	N16961	433	Primosomal replication protein N	M66-2
86	hypothetical protein	N16961	434	SSU ribosomal protein S18p	M66-2
87	hypothetical protein	N16961	435	LSU ribosomal protein L9p	M66-2
88	hypothetical protein	N16961	436	membrane protein	M66-2
89	acetyltransferase (putative)	N16961	437	Replicative DNA helicase (EC 3.6.1.-)	M66-2
90	hypothetical protein	N16961	438	Alanine racemase, biosynthetic (EC 5.1.1.1)	M66-2
91	ParE toxin protein	N16961	439	Uncharacterized conserved protein	M66-2
92	ParD protein (antitoxin to ParE)	N16961	440	Glucose-6-phosphate isomerase (EC 5.3.1.9)	M66-2
93	toxin resistance protein	N16961	441	membrane protein	M66-2
94	hypothetical protein	N16961	442	Chemotaxis protein CheX	M66-2
95	hypothetical protein	N16961	443	Zinc uptake regulation protein ZUR	M66-2
96	HigB toxin protein	N16961	444	tRNA dihydrouridine synthase A (EC 1.-.-.-)	M66-2

97	HigA protein (antitoxin to HigB)	N16961	445	hypothetical protein	M66-2
98	hypothetical protein	N16961	446	Putative inner membrane protein	M66-2
99	hypothetical protein	N16961	447	ABC-type Fe ³⁺ -hydroxamate transport system, periplasmic component	M66-2
100	Phenazine biosynthesis protein PhzF	N16961	448	Thymidylate kinase	M66-2
101	hypothetical protein	N16961	449	Sulfite reductase [NADPH] flavoprotein alpha-component (EC 1.8.1.2)	M66-2
102	hypothetical protein	N16961	450	Sulfite reductase [NADPH] hemoprotein beta-component (EC 1.8.1.2)	M66-2
103	hypothetical protein	N16961	451	Phosphoadenylyl-sulfate reductase [thioredoxin] (EC 1.8.4.8)	M66-2
104	acetyltransferase, putative	N16961	452	hypothetical protein	M66-2
105	Death on curing protein, Doc toxin	N16961	453	Small Subunit Ribosomal RNA; ssuRNA; SSU rRNA	M66-2
106	Glyoxalase family protein	N16961	454	tRNA-Ile-GAT	M66-2
107	hypothetical protein	N16961	455	tRNA-Ala-TGC	M66-2
108	conserved hypothetical protein	N16961	456	Large Subunit Ribosomal RNA; lsuRNA; LSU rRNA	M66-2
109	hypothetical protein	N16961	457	5S RNA	M66-2
110	lipoprotein Blc	N16961	458	tRNA-Asp-GTC	M66-2
111	hypothetical protein	N16961	459	Na ⁺ /H ⁺ antiporter, putative	M66-2
112	hypothetical protein	N16961	460	5-methyltetrahydrofolate--homocysteine methyltransferase (EC 2.1.1.13)	M66-2
113	hypothetical protein	N16961	461	Aspartokinase (EC 2.7.2.4)	M66-2
114	hypothetical protein	N16961	462	Serine--pyruvate aminotransferase (EC 2.6.1.51) / L-alanine:glyoxylate aminotransferase (EC 2.6.1.44)	M66-2
115	hypothetical protein	N16961	463	Lipid A core-O-antigen ligase	M66-2
116	hypothetical protein	N16961	464	Excinuclease ABC subunit A	M66-2
117	hypothetical protein	N16961	465	UTP--glucose-1-phosphate uridylyltransferase (EC 2.7.7.9)	M66-2
118	Death on curing protein, Doc toxin	N16961	466	Transcriptional regulator, LuxR family	M66-2
119	Glyoxalase family protein	N16961	467	Single-stranded DNA-binding protein	M66-2
120	conserved hypothetical protein	N16961	468	Glyoxalase family protein	MJ-1236
121	Death on curing protein, Doc toxin	N16961	469	acetyltransferase, putative	MJ-1236
122	acetyltransferase, putative	N16961	470	hypothetical protein	MJ-1236
123	hypothetical protein	N16961	471	DNA-damage-inducible protein J	MJ-1236
124	hypothetical protein	N16961	472	YafQ toxin protein	MJ-1236
125	hypothetical protein	N16961	473	hypothetical protein	MJ-1236
126	hypothetical protein	N16961	474	hypothetical protein	MJ-1236
127	Prevent host death protein, Phd antitoxin	N16961	475	hypothetical protein	MJ-1236
128	Death on curing protein, Doc toxin	N16961	476	hypothetical protein	MJ-1236
129	conserved hypothetical protein	N16961	477	hypothetical protein	MJ-1236
130	hypothetical protein	N16961	478	Maleylacetoacetate isomerase (EC 5.2.1.2) / Glutathione S-transferase	MJ-1236
131	hypothetical protein	N16961	479	No significant database matches	MJ-1236
132	hypothetical protein	N16961	480	acetyltransferase (putative)	MJ-1236
133	hypothetical protein	N16961	481	IS1004 transposase	MJ-1236
134	Death on curing protein, Doc toxin	N16961	482	acetyltransferase (putative)	MJ-1236
135	hypothetical protein	N16961	483	hypothetical protein	MJ-1236
136	Lactoylglutathione lyase and related lyases	N16961	484	hypothetical protein	MJ-1236
137	hypothetical protein	N16961	485	hypothetical protein	MJ-1236
138	hypothetical protein	N16961	486	hypothetical protein	MJ-1236
139	hypothetical protein	N16961	487	hypothetical protein	MJ-1236
140	hypothetical protein	N16961	488	hypothetical protein	MJ-1236
141	acetyltransferase, putative	N16961	489	hypothetical protein	MJ-1236
142	hypothetical protein	N16961	490	hypothetical protein	MJ-1236
143	conserved hypothetical protein	N16961	491	hypothetical protein	MJ-1236
144	Muramoyltetrapeptide carboxypeptidase (EC 3.4.17.13)	N16961	492	hypothetical protein	MJ-1236
145	hypothetical protein	N16961	493	hypothetical protein	MJ-1236
146	putative acetyltransferase	N16961	494	hypothetical protein	MJ-1236
147	lipoprotein Blc	N16961	495	hypothetical protein	MJ-1236
148	StbE replicon stabilization toxin	N16961	496	hypothetical protein	MJ-1236

149	StbD replicon stabilization protein (antitoxin to StbE)	N16961	497	hypothetical protein	MJ-1236
150	Death on curing protein, Doc toxin	N16961	498	hypothetical protein	MJ-1236
151	haemagglutinin associated protein	N16961	499	hypothetical protein	MJ-1236
152	hypothetical protein	N16961	500	hypothetical protein	MJ-1236
153	hypothetical protein	N16961	501	hypothetical protein	MJ-1236
154	conserved hypothetical protein	N16961	502	hypothetical protein	MJ-1236
155	Sulfate-binding protein Sbp	N16961	503	hypothetical protein	MJ-1236
156	ThiJ/Pfpl family protein	N16961	504	hypothetical protein	MJ-1236
157	hypothetical protein	N16961	505	Lipase, GDYG family	MJ-1236
158	hypothetical protein	N16961	506	Death on curing protein, Doc toxin	MJ-1236
159	hypothetical protein	N16961	507	Prevent host death protein, Phd antitoxin # A	MJ-1236
160	hypothetical protein	N16961	508	FIG001353: Acetyltransferase	MJ-1236
161	hypothetical protein	N16961	509	FIG032766: hypothetical protein	MJ-1236
162	Death on curing protein, Doc toxin	N16961	510	MazG-related protein	MJ-1236
163	conserved hypothetical protein	N16961	511	hypothetical protein	MJ-1236
164	hypothetical protein	N16961	512	hypothetical protein	MJ-1236
165	hypothetical protein	N16961	513	FIG032766: hypothetical protein	MJ-1236
166	biphenyl-2,3-diol 1,2-dioxygenase III-related protein	N16961	514	FIG001353: Acetyltransferase	MJ-1236
167	hypothetical protein	N16961	515	hypothetical protein	MJ-1236
168	hypothetical protein	N16961	516	conserved hypothetical protein	MJ-1236
169	Cytotoxic translational repressor of toxin-antitoxin stability system	N16961	517	hypothetical protein	MJ-1236
170	hypothetical protein	N16961	518	Histone acetyltransferase HPA2	MJ-1236
171	Acetyltransferase	N16961	519	StbE replicon stabilization toxin	MJ-1236
172	IS5 transposase	N16961	520	StbD replicon stabilization protein (antitoxin to StbE)	MJ-1236
173	Acetyltransferase	N16961	521	Glyoxalase family protein	MJ-1236
174	putative acetyltransferase	N16961	522	doc protein	MJ-1236
175	doc protein	N16961	523	putative acetyltransferase	MJ-1236
176	Glyoxalase family protein	N16961	524	Acetyltransferase	MJ-1236
177	StbD replicon stabilization protein (antitoxin to StbE)	N16961	525	IS5 transposase	MJ-1236
178	StbE replicon stabilization toxin	N16961	526	Acetyltransferase	MJ-1236
179	Histone acetyltransferase HPA2	N16961	527	hypothetical protein	MJ-1236
180	Death on curing protein, Doc toxin	N16961	528	Cytotoxic translational repressor of toxin-antitoxin stability system	MJ-1236
181	hypothetical protein	N16961	529	hypothetical protein	MJ-1236
182	hypothetical protein	N16961	530	hypothetical protein	MJ-1236
183	FIG001353: Acetyltransferase	N16961	531	biphenyl-2,3-diol 1,2-dioxygenase III-related protein	MJ-1236
184	FIG032766: hypothetical protein	N16961	532	hypothetical protein	MJ-1236
185	hypothetical protein	N16961	533	hypothetical protein	MJ-1236
186	hypothetical protein	N16961	534	conserved hypothetical protein	MJ-1236
187	MazG-related protein	N16961	535	hypothetical protein	MJ-1236
188	FIG032766: hypothetical protein	N16961	536	hypothetical protein	MJ-1236
189	FIG001353: Acetyltransferase	N16961	537	hypothetical protein	MJ-1236
190	Prevent host death protein, Phd antitoxin # A	N16961	538	hypothetical protein	MJ-1236
191	Death on curing protein, Doc toxin	N16961	539	hypothetical protein	MJ-1236
192	Lipase, GDYG family	N16961	540	hypothetical protein	MJ-1236
193	acetyltransferase (putative)	N16961	541	ThiJ/Pfpl family protein	MJ-1236
194	IS1004 transposase	N16961	542	Sulfate-binding protein Sbp	MJ-1236
195	acetyltransferase (putative)	N16961	543	conserved hypothetical protein	MJ-1236
196	No significant database matches	N16961	544	hypothetical protein	MJ-1236
197	Maleylacetoacetate isomerase (EC 5.2.1.2) / Glutathione S-transferase	N16961	545	hypothetical protein	MJ-1236
198	hypothetical protein	N16961	546	haemagglutinin associated protein	MJ-1236
199	hypothetical protein	N16961	547	hypothetical protein	MJ-1236
200	hypothetical protein	N16961	548	StbD replicon stabilization protein (antitoxin to StbE)	MJ-1236

201	hypothetical protein	N16961	549	StbE replicon stabilization toxin	MJ-1236
202	hypothetical protein	N16961	550	lipoprotein Blc	MJ-1236
203	YafQ toxin protein	N16961	551	putative acetyltransferase	MJ-1236
204	DNA-damage-inducible protein J	N16961	552	hypothetical protein	MJ-1236
205	hypothetical protein	N16961	553	Muramoyltetrapeptide carboxypeptidase (EC 3.4.17.13)	MJ-1236
206	acetyltransferase, putative	N16961	554	conserved hypothetical protein	MJ-1236
207	Glyoxalase family protein	N16961	555	hypothetical protein	MJ-1236
208	Glyoxalase family protein	O395	556	HigA protein (antitoxin to HigB)	MJ-1236
209	acetyltransferase, putative	O395	557	HigB toxin protein	MJ-1236
210	hypothetical protein	O395	558	hypothetical protein	MJ-1236
211	DNA-damage-inducible protein J	O395	559	toxin resistance protein	MJ-1236
212	YafQ toxin protein	O395	560	ParD protein (antitoxin to ParE)	MJ-1236
213	hypothetical protein	O395	561	ParE toxin protein	MJ-1236
214	hypothetical protein	O395	562	hypothetical protein	MJ-1236
215	hypothetical protein	O395	563	acetyltransferase (putative)	MJ-1236
216	hypothetical protein	O395	564	hypothetical protein	MJ-1236
217	hypothetical protein	O395	565	hypothetical protein	MJ-1236
218	Maleylacetoacetate isomerase (EC 5.2.1.2) / Glutathione S-transferase	O395	566	hypothetical protein	MJ-1236
219	No significant database matches	O395	567	hypothetical protein	MJ-1236
220	acetyltransferase (putative)	O395	568	hypothetical protein	MJ-1236
221	IS1004 transposase	O395	569	hypothetical protein	MJ-1236
222	acetyltransferase (putative)	O395	570	hypothetical protein	MJ-1236
223	Lipase, GDXX family	O395	571	hypothetical protein	MJ-1236
224	Death on curing protein, Doc toxin	O395	572	hypothetical protein	MJ-1236
225	Prevent host death protein, Phd antitoxin # A	O395	573	hypothetical protein	MJ-1236
226	FIG001353: Acetyltransferase	O395	574	Transposase	MJ-1236
227	FIG032766: hypothetical protein	O395	575	transposase OrfAB, subunit B	MJ-1236
228	MazG-related protein	O395	576	hypothetical protein	MJ-1236
229	hypothetical protein	O395	577	hypothetical protein	MJ-1236
230	hypothetical protein	O395	578	hypothetical protein	MJ-1236
231	FIG032766: hypothetical protein	O395	579	hypothetical protein	MJ-1236
232	FIG001353: Acetyltransferase	O395	580	hypothetical protein	MJ-1236
233	hypothetical protein	O395	581	hypothetical protein	MJ-1236
234	hypothetical protein	O395	582	hypothetical protein	MJ-1236
235	hypothetical protein	O395	583	hypothetical protein	MJ-1236
236	hypothetical protein	O395	584	Predicted transcriptional regulator	MJ-1236
237	hypothetical protein	O395	585	Plasmid stabilization element ParE, putative	MJ-1236
238	hypothetical protein	O395	586	hypothetical protein	MJ-1236
239	hypothetical protein	O395	587	hypothetical protein	MJ-1236
240	hypothetical protein	O395	588	hypothetical protein	MJ-1236
241	hypothetical protein	O395	589	hypothetical protein	MJ-1236
242	hypothetical protein	O395	590	Putative threonine efflux protein	MJ-1236
243	hypothetical protein	O395	591	hypothetical protein	MJ-1236
244	hypothetical protein	O395	592	hypothetical protein	MJ-1236
245	hypothetical protein	O395	593	lipoprotein Blc	MJ-1236
246	hypothetical protein	O395	594	DNA-damage-inducible protein J	MJ-1236
247	hypothetical protein	O395	595	YafQ toxin protein	MJ-1236
248	hypothetical protein	O395	596	Glyoxalase family protein	MJ-1236
249	hypothetical protein	O395	597	H-REV 107-related protein	MJ-1236
250	hypothetical protein	O395	598	hypothetical protein	MJ-1236
251	Similarity	O395	599	Alcohol dehydrogenase, class IV	MJ-1236
252	hypothetical protein	O395	600	hypothetical protein	MJ-1236

253	hypothetical protein	O395	601	biphenyl-2,3-diol 1,2-dioxygenase III-related protein	MJ-1236
254	hypothetical protein	O395	602	hypothetical protein	MJ-1236
255	putative acetyltransferase	O395	603	hypothetical protein	MJ-1236
256	StbE replicon stabilization toxin	O395	604	Glyoxalase family protein	MJ-1236
257	StbD replicon stabilization protein (antitoxin to StbE)	O395	605	Muramoyltetrapeptide carboxypeptidase (EC 3.4.17.13)	MJ-1236
258	Glyoxalase family protein	O395	606	hypothetical protein	MJ-1236
259	doc protein	O395	607	hypothetical protein	MJ-1236
260	putative acetyltransferase	O395	608	Helix-turn-helix protein, CopG family	MJ-1236
261	Acetyltransferase	O395	609	hypothetical protein	MJ-1236
262	IS5 transposase	O395	610	COG1764: Predicted redox protein, regulator of disulfide bond formation	MJ-1236
263	Acetyltransferase	O395	611	conserved hypothetical protein	MJ-1236
264	hypothetical protein	O395	612	conserved hypothetical protein	MJ-1236
265	Cytotoxic translational repressor of toxin-antitoxin stability system	O395	613	Alcohol dehydrogenase, class IV	MJ-1236
266	hypothetical protein	O395	614	biphenyl-2,3-diol 1,2-dioxygenase III-related protein	MJ-1236
267	hypothetical protein	O395	615	hypothetical protein	MJ-1236
268	hypothetical protein	O395	616	conserved hypothetical protein	MJ-1236
269	conserved hypothetical protein	O395	617	DNA-damage-inducible protein J	MJ-1236
270	hypothetical protein	O395	618	YafQ toxin protein	MJ-1236
271	hypothetical protein	O395	619	hypothetical protein	MJ-1236
272	hypothetical protein	O395	620	hypothetical protein	MJ-1236
273	hypothetical protein	O395	621	FIG032766: hypothetical protein	MJ-1236
274	ThiJ/Pfpl family protein	O395	622	FIG001353: Acetyltransferase	MJ-1236
275	Sulfate-binding protein Sbp	O395	623	lipoprotein Blc	MJ-1236
276	conserved hypothetical protein	O395	624	acetyltransferase, putative	MJ-1236
277	hypothetical protein	O395	625	hypothetical protein	MJ-1236
278	hypothetical protein	O395	626	MutT/nudix family protein	MJ-1236
279	Death on curing protein, Doc toxin	O395	627	ParD protein (antitoxin to ParE)	MJ-1236
280	haemagglutinin associated protein	O395	628	ParE toxin protein	MJ-1236
281	hypothetical protein	O395	629	hypothetical protein	MJ-1236
282	StbD replicon stabilization protein (antitoxin to StbE)	O395	630	5-methylcytosine-specific restriction enzyme A (EC 3.1.21.-)	MJ-1236
283	StbE replicon stabilization toxin	O395	631	Deoxyguanosinetriphosphate triphosphohydrolase (EC 3.1.5.1)	MJ-1236
284	lipoprotein Blc	O395	632	hypothetical protein	MJ-1236
285	putative acetyltransferase	O395	633	hypothetical protein	MJ-1236
286	hypothetical protein	O395	634	hypothetical protein	MJ-1236
287	Muramoyltetrapeptide carboxypeptidase (EC 3.4.17.13)	O395	635	hypothetical protein	MJ-1236
288	hypothetical protein	O395	636	hypothetical protein	MJ-1236
289	hypothetical protein	O395	637	hypothetical protein	MJ-1236
290	acetyltransferase, putative	O395	638	hypothetical protein	MJ-1236
291	hypothetical protein	O395	639	hypothetical protein	MJ-1236
292	hypothetical protein	O395	640	Putative oxidoreductase	MJ-1236
293	hypothetical protein	O395	641	Acetyltransferase	MJ-1236
294	Lactoylglutathione lyase and related lyases	O395	642	Alcohol dehydrogenase, class IV	MJ-1236
295	hypothetical protein	O395	643	hypothetical protein	MJ-1236
296	Death on curing protein, Doc toxin	O395	644	hypothetical protein	MJ-1236
297	hypothetical protein	O395	645	hypothetical protein	MJ-1236
298	hypothetical protein	O395	646	conserved hypothetical protein	MJ-1236
299	hypothetical protein	O395	647	Integron integrase IntI4	MJ-1236
300	hypothetical protein	O395	648	Integron integrase IntI4	LMA3984-4
301	conserved hypothetical protein	O395	649	Threonyl-tRNA synthetase	LMA3984-4
302	COG2315: Uncharacterized protein conserved in bacteria	O395	650	Translation initiation factor 3	LMA3984-4
303	hypothetical protein	O395	651	LSU ribosomal protein L35p	LMA3984-4
304	Death on curing protein, Doc toxin	O395	652	LSU ribosomal protein L20p	LMA3984-4

305	Prevent host death protein, Phd antitoxin # A	O395	653	Integron integrase IntI4	LMA3984-4
306	hypothetical protein	O395	654	hypothetical protein	LMA3984-4
307	acetyltransferase (putative)	O395	655	hypothetical protein	LMA3984-4
308	hypothetical protein	O395	656	hypothetical protein	LMA3984-4
309	hypothetical protein	O395	657	hypothetical protein	LMA3984-4
310	hypothetical protein	O395	658	conserved hypothetical protein	LMA3984-4
311	Isochorismatase (EC 3.3.2.1)	O395	659	MutT/nudix family protein	LMA3984-4
312	Death on curing protein, Doc toxin	O395	660	Predicted redox protein	LMA3984-4
313	Prevent host death protein, Phd antitoxin	O395	661	unknown	LMA3984-4
314	hypothetical protein	O395	662	hypothetical protein	LMA3984-4
315	hypothetical protein	O395	663	Predicted transcriptional regulator	LMA3984-4
316	hypothetical protein	O395	664	hypothetical protein	LMA3984-4
317	hypothetical protein	O395	665	hypothetical protein	LMA3984-4
318	lipoprotein Blc	O395	666	hypothetical protein	LMA3984-4
319	Alcohol dehydrogenase, class IV	O395	667	acetyltransferase (putative)	LMA3984-4
320	hypothetical protein	O395	668	unknown	LMA3984-4
321	biphenyl-2,3-diol 1,2-dioxygenase III-related protein	O395	669	hypothetical protein	LMA3984-4
322	hypothetical protein	O395	670	hypothetical protein	LMA3984-4
323	Dabb	O395	671	conserved hypothetical protein	LMA3984-4
324	lipoprotein Blc	O395	672	hypothetical protein	LMA3984-4
325	lipoprotein Blc	O395	673	Lactoylglutathione lyase	LMA3984-4
326	Glyoxalase family protein	O395	674	Sulfate-binding protein Sbp	LMA3984-4
327	conserved hypothetical protein	O395	675	hypothetical protein	LMA3984-4
328	acetyltransferase (putative)	O395	676	conserved hypothetical protein	LMA3984-4
329	hypothetical protein	O395	677	Cytotoxic translational repressor of toxin- antitoxin stability system	LMA3984-4
330	Muramoyltetrapeptide carboxypeptidase (EC 3.4.17.13)	O395	678	hypothetical protein	LMA3984-4
331	hypothetical protein	O395	679	Acetyltransferase	LMA3984-4
332	hypothetical protein	O395	680	StbD replicon stabilization protein (antitoxin to StbE)	LMA3984-4
333	hypothetical protein	O395	681	MazG-like protein	LMA3984-4
334	hypothetical protein	O395	682	hypothetical protein	LMA3984-4
335	hypothetical protein	O395	683	Acetyltransferase	LMA3984-4
336	hypothetical protein	O395	684	Prevent host death protein, Phd antitoxin	LMA3984-4
337	hypothetical protein	O395	685	Death on curing protein, Doc toxin	LMA3984-4
338	hypothetical protein	O395	686	Lipase, GDVG family	LMA3984-4
339	Predicted transcriptional regulator	O395	687	Mobile element protein	LMA3984-4
340	Plasmid stabilization element ParE, putative	O395	688	acetyltransferase (putative)	LMA3984-4
341	hypothetical protein	O395	689	hypothetical protein	LMA3984-4
342	hypothetical protein	O395	690	unknown	LMA3984-4
343	hypothetical protein	O395	691	hypothetical protein	LMA3984-4
344	hypothetical protein	O395	692	hypothetical protein	LMA3984-4
345	Putative threonine efflux protein	O395	693	hypothetical protein	LMA3984-4
346	hypothetical protein	O395	694	DNA-damage-inducible protein J	LMA3984-4
347	hypothetical protein	O395	695	acetyltransferase, putative	LMA3984-4
348	hypothetical protein	O395	696	Glyoxalase family protein	LMA3984-4

9.4 Anexo IV: Estudo do efeito de nucleotídeos e gaps randomicamente inseridos na predição e anotação de genes

A inserção foi em uma proporção de 1%, 2%, 5% e 10% do total do tamanho do genoma de *Escherichia coli* str. K-12 (Poster apresentado no RECOMB 2012).



Accuracy of the NGS sequencing and assembling on the gene prediction and annotation

Michel A. Marin* and Ana Carolina P. Vicente

Laboratory of Molecular Genetics of Microorganisms and Computational and Systems Biology Post Graduate Program, Oswaldo Cruz Institute (IOC) - Oswaldo Cruz Foundation (FIOCRUZ), Rio de Janeiro, Brazil

*Corresponding author. E-mail: mfabanto@gmail.com

Introduction

New sequencing technologies, commonly referred as next-generation sequencing (NGS), including pirosequencing (454 FLX and FLX Titanium), Illumina's clonal arrays (Illumina GAIIx and HiSeq2000), PacBio and IonTorrent, are enabling a remarkable increase in the number and diversity of sequenced genomes. Besides of the different read lengths and amounts of sequence data produced, each technology has a characteristic profile of average error rate, comprising between 0.001% and 15% of the total number of reads, including insertions, deletions and mismatches.

Methodology

A set of *V. cholerae* genomes previously sequenced by Ion Torrent technology, was assembled and the scaffolded contigs were uploaded to Rapid Annotations based on Subsystem Technology (RAST). The automated gene calling and annotation showed that some genomes would have an increase number of predict protein encoding genes (PEGs) when compared to *V. cholerae* gene content determined by Sanger sequencing. When the predicted genes assigned to each genome were checked, we observed that most of genes, of some genomes, had short lengths (fig. 1).

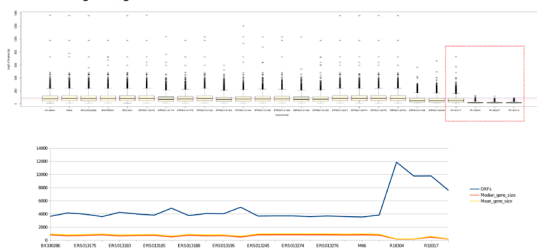


Figure 1. Top: Variability of gene length identified and annotated by RAST of the *V. cholerae* genomes. Box-and-whiskers graph represent interquartile ranges (upper, median and lower quartile) and the minimum and maximum values are show. The red and blue lines represent the median and mean values and the red dot box represents genomes with lowest length of genes. Bottom: Number of ORFs identified by RAST by the median and mean of gene length.

In order to test the effect of erroneous or Ns characters introduced by either the sequencing error rate or derivative of the assembly process, we simulated different error rate levels in the complete genome of *Escherichia coli* str. K-12 substr. W3110 (NC_007779.1). We build an R script to replaced randomly the original nucleotide sequence by a set of characters (A, T, G, C, N, -) reproducing the potential errors generated by NGS technologies and on the contigs or scaffold building. The random characters were incorporated in a proportion of 1%, 2%, 5% and 10% of total genome size. The obtained chimeras and the original genomes were then annotated using RAST.

Results

The results showed that, keeping the genome size, the number of coding sequences in the chimeras increased in a ranged from 4474 to 8356 according to the proportion of the error introduced (Fig. 2). The same result was found in subsystem features of the RAST annotation (Fig. 3).

Table. Main attributes found by RAST annotation

Attribute	eco	eco1	eco2	eco5	eco10
Size	4646332	4646332	4646332	4646332	4646332
Number of contigs (with PEGs)	1	1	1	1	1
Number of Subsystems	590	593	586	585	577
Number of Coding Sequences	4474	5046	5476	6834	8356
Number of RNAs	108	108	107	103	96

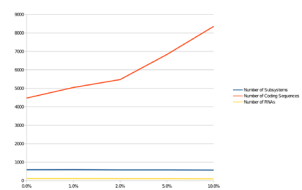


Figure 2. Number of ORFs identified by RAST with 0%, 1%, 2%, 5% and 10% of characters inserted in the original *Escherichia coli* genome sequence.

Interestingly, comparing the predicted genes involved in metabolic pathways between canonical *E. coli* (eco) and *E. coli* with 10% error rate (eco10), we found that 100 PEGs were present solely in eco, while 47 PEGs were specific of eco10. Overall, our results showed that the number of predicted coding sequences strongly correlated with the average error rate.

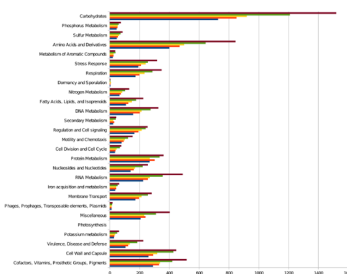


Figure 3. Functional analysis based on Rapid Annotation using Subsystem Technology (RAST) indicated same subsystem proportion in *E. coli* genomes with 0%, 1%, 2%, 5% and 10% of characters inserted.

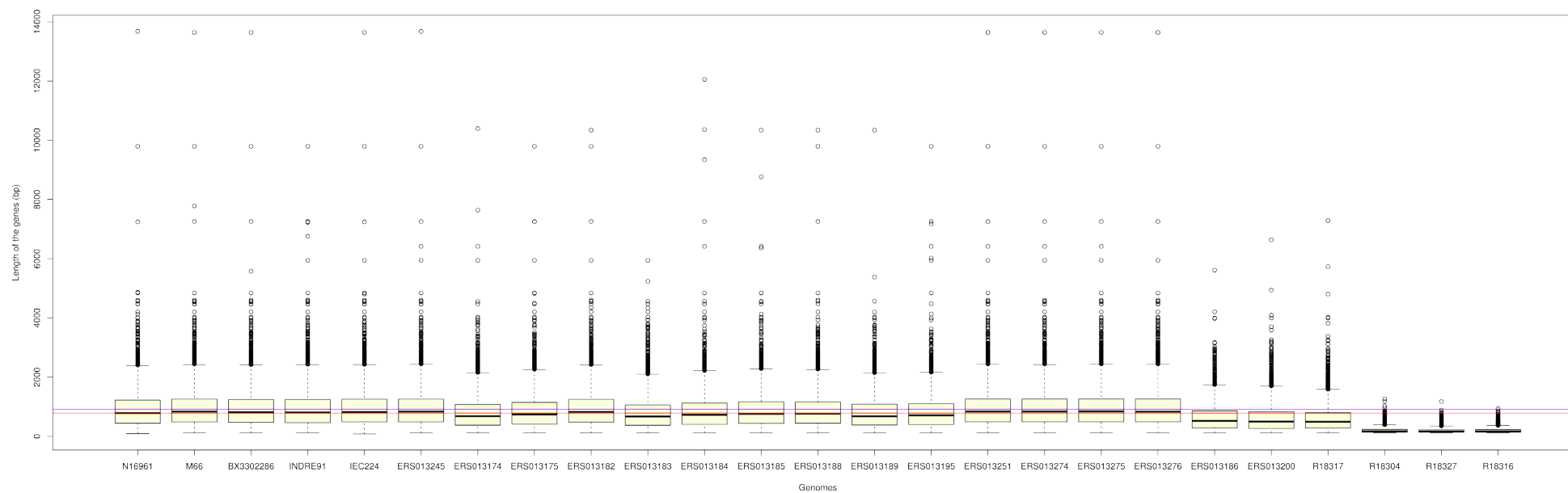
Conclusions

The gene prediction and annotation could be affected by presence of erroneous characters introduced by new sequencing technologies or for assembling and scaffolded genome process. The Comparative analyzes using draft genomes to deduce gene content, pangenome calculation, gene lost/acquisition among others could be misleading.

Supported by: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), FCTIS/FIOCRUZ.

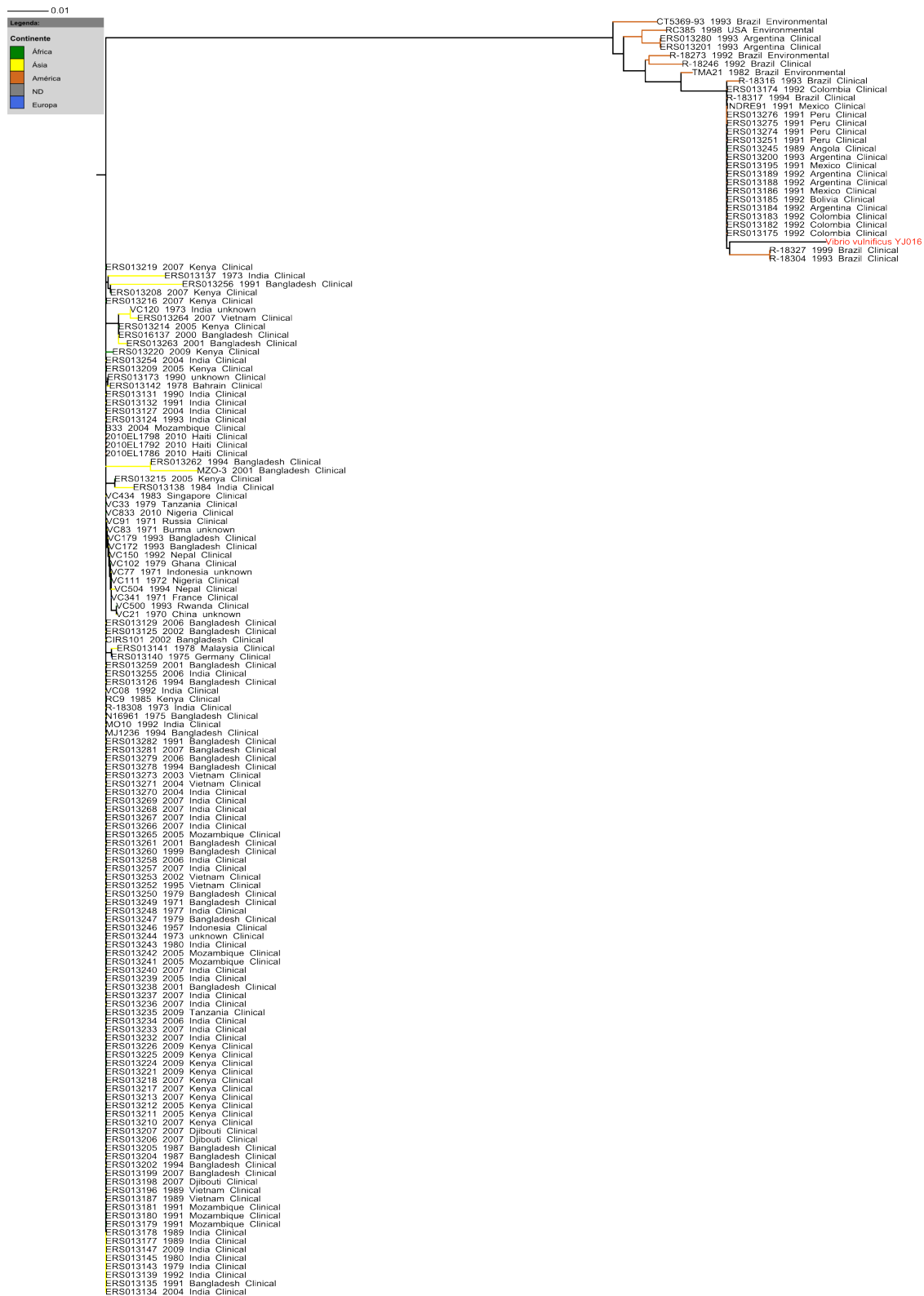
9.5 Anexo V: Comprimento dos genes codificadores dos genomas de *V. cholerae*

A média (linha azul) e a mediana (linha vermelha) são mostradas. Na análise de ortólogos foram considerados somente os genomas que apresentavam um comprimento dos genes codificadores maior que a média.



9.6 Anexo VI: Árvore de máxima verossimilhança de homólogos da integrase VC0516 de uma população de *V. cholerae*

A cepa mais estreitamente relacionada identificada pelo genome-BLAST *Vibrio vulnificus* YJ016 (Número de acesso GenBank BA000037) foi incluída e é mostrada em vermelho.



9.7 Anexo VII: Publicações geradas durante tese

Correspondence

Epidemiol. Infect. (2011).
doi:10.1017/S0950268811001944

Variants of *Vibrio cholerae* O1 El Tor from Zambia showed new genotypes of *ctxB*

To the Editor

In the article ‘Genetic characterization of *Vibrio cholerae* O1 strains isolated in Zambia during 1996–2004 possessing the unique VSP-II region of El Tor variant’ Bhuiyan and colleagues [1], reported a Zambian variant of El Tor biotype harbouring El Tor-type *rstR* but producing classical *ctxB*. They also showed that these recent variants lack VC0493 and VC0498 in the VSP-II island. The authors characterized the *ctxB* gene by mismatch amplification mutation PCR assay (MAMA)-PCR analysis [2]. Although this method is useful for epidemiological surveillance, it is specific for the *ctxB* allele and misses polymorphisms outside the target position. Based on the classification of Olsvik *et al.* [3], the authors concluded that the Zambian *V. cholerae* O1 strains isolated in 1996–1997 were identical to the *ctxB* genotype 3 of El Tor reference strain N16961, while the strains from 2003–2004 that harboured *ctxB* genotype 1 corresponded to the classical biotype.

In 2010, Safa and co-workers [4] characterized nine *ctxB* genotypes and showed polymorphism throughout the *ctxB* coding sequence as well as in five additional positions, not previously considered by Olsvik *et al.* [3]. We analysed the *ctxB* sequences from Bhuiyan *et al.* [1] (GenBank: EU932878-EU932884) and found that the Zambian sequences represent three distinct *ctxB* alleles and not two as originally stated by the authors. According to Safa’s characterization, EU932883 corresponds to genotype 1 and is identical to the classical type. The sequences EU932878-80 and EU932881-82, EU932884 are similar to genotype 6 with mutations at Thr⁶⁸Ile and Tyr³⁹His, respectively.

We designated the new variants genotypes 10 and 11 (Table 1).

Bhuiyan’s paper established a relationship between the *ctxB* genotypes and the two major clusters of the *V. cholerae* strains from Zambia. However, it was not possible to determine here the distribution of the *ctxB* genotypes 1, 10 and 11 among Zambian strains, due to lack of information, regarding the year of isolation and the strain identification of the sequences deposited at GenBank. The exception was the 237/02 strain (1996) (EU932878), classified as genotype 10 (an El Tor prototype according to the authors), belonging to cluster A. Concerning the 2003–2004 isolates, it is possible to state that they are unique due to the characteristics of their VSP-II region and to the new *ctxB* allele.

We agree with the authors on the implication of the heterogeneity in the B subunit of cholera toxin for vaccine development and diagnosis, as well as concerning the impact of this variation in cholera severity [5]. The two new emerging *ctxB* genotypes from Zambia are labels that could be considered in the epidemiological surveillance of cholera epidemics.

Declaration of Interest

None.

References

1. **Bhuiyan NA, *et al.*** Genetic characterization of *Vibrio cholerae* O1 strains isolated in Zambia during 1996–2004 possessing the unique VSP-II region of El Tor variant. *Epidemiology and Infection*. Published online: 31 May 2011. doi:10.1017/S0950268811000926.
2. **Morita M, *et al.*** Development and validation of a mismatch amplification mutation PCR assay to monitor the dissemination of an emerging variant of *Vibrio cholerae* O1 biotype El Tor. *Microbiology and Immunology* 2008; **52**: 314–347.
3. **Olsvik O, *et al.*** Use of automated sequencing of polymerase chain reaction-generated amplicons to identify

Table 1. *Distribution and comparison of ctxB genotypes in V. cholerae strains**

Subtype	ctxB genotype	Nucleotide position (amino-acid position)								Ref.
		58 (20)	72 (24)	83 (28)	101 (34)	115 (39)	138 (46)	165 (55)	203 (68)	
O1	1 (Classical, 569B)	C (His)	A (Gln)	A (Asp)	A (His)	C (His)	T (Phe)	A (Lys)	C (Thr)	[3, 6]
	2 (El Tor, Australia)	C (His)	A (Gln)	A (Asp)	A (His)	C (His)	G (Leu)	A (Lys)	C (Thr)	
	3 (El Tor, N16961)	C (His)	A (Gln)	A (Asp)	A (His)	T (Tyr)	T (Phe)	A (Lys)	T (Ile)	
	7 (Altered El Tor)	A (Asn)	A (Gln)	A (Asp)	A (His)	C (His)	T (Phe)	A (Lys)	C (Thr)	
	10	C (His)	A (Gln)	A (Asp)	C (Pro)	T (Tyr)	T (Phe)	A (Lys)	T (Ile)	[1]
	11	C (His)	A (Gln)	A (Asp)	C (Pro)	C (His)	T (Phe)	A (Lys)	C (Thr)	(this study)
O139	3	C (His)	A (Gln)	A (Asp)	A (His)	T (Tyr)	T (Phe)	A (Lys)	T (Ile)	[3, 7]
	4	C (His)	A (Gln)	A (Asp)	A (His)	T (Tyr)	T (Phe)	A (Lys)	C (Thr)	
	5	C (His)	A (Gln)	C (Ala)	A (His)	C (His)	T (Phe)	A (Lys)	C (Thr)	
	6	C (His)	A (Gln)	A (Asp)	C (Pro)	T (Tyr)	T (Phe)	A (Lys)	C (Thr)	
O27	8	C (His)	C (His)	C (Ala)	A (His)	C (His)	T (Phe)	A (Lys)	C (Thr)	[8]
O37	9	C (His)	A (Gln)	A (Asp)	A (His)	C (Leu)	G (Leu)	C (Asn)	C (Thr)	[9]

* Adapted from Safa *et al.* [4].

- three types of cholera toxin subunit B in *Vibrio cholerae* O1 strains. *Journal of Clinical Microbiology* 1993; **31**: 22–25.
4. Safa A, Nair GB, Kong RYC. Evolution of new variants of *Vibrio cholerae* O1. *Trends in Microbiology* 2010; **18**: 46–54.
 5. Ryan ET. The cholera pandemic, still with us after half a century: time to rethink. *PLoS Neglected Tropical Diseases* 2011; **5**: e1003.
 6. Goel AK, *et al.* A new variant of *Vibrio cholerae* O1 El Tor causing cholera in India. *Journal of Infection* 2008; **57**: 280–281.
 7. Bhuiyan NA, *et al.* Changing genotypes of cholera toxin (CT) of *Vibrio cholerae* O139 in Bangladesh and description of three new CT genotypes. *FEMS Immunology and Medical Microbiology* 2009; **57**: 136–141.
 8. Li M, *et al.* Evidence for the emergence of non-O1 and non-O139 *Vibrio cholerae* strains with pathogenic potential by exchange of O-antigen biosynthesis regions. *Infection and Immunity* 2002; **70**: 2441–2453.
 9. Yamamoto K, *et al.* Amino acids of the cholera toxin from *Vibrio cholerae* O37 strain S7 which differ from those of strain O1. *Gene* 1995; **163**: 155–156.

M. A. MARIN, A. C. P. VICENTE

Laboratory of Molecular Genetics of Microorganisms,
Oswaldo Cruz Institute (IOC) – Oswaldo Cruz Foundation
(FIOCRUZ), Rio de Janeiro, Brazil

Author for correspondence:

Dr A. C. P. Vicente, Instituto Oswaldo Cruz (IOC)/
FIOCRUZ,
Av. Brasil, 4365, PO Box 926 CEP 21045-900,
Rio de Janeiro, Brazil
(Email: anapaulo@ioc.fiocruz.br)

Epidemiol. Infect. (2011).
doi:10.1017/S0950268811002019

Variants of *Vibrio cholerae* O1 El Tor from Zambia showed new genotypes of ctxB: a reply to Marin & Vicente

We appreciate and thank Marin & Vicente for a quick analysis of *ctxB* sequences that we deposited in GenBank under the accession numbers EU932878-EU932884 [1].

In fact the sequences of the *ctxB* gene of Zambian strains were deposited as additional information to confirm the result of MAMA-PCR as well as to type the strains following Olsvik *et al.* overlooking the actual result of the genotype [2]. After the report by Safa and co-workers [3] we are now sequencing a large number of variant strains of *V. cholerae* O1 for the *ctxB* gene that were isolated from cholera-endemic areas globally in order to understand the actual number and location of new genotypes of *ctxB*. In our published paper, we concentrated our analysis on targeting the pattern of nine virulence regions of *V. cholerae* O1 especially the regions for VSP-I and VSP-II [1]. However, Marin & Vicente, in their letter, have clearly identified the new genotypes 10 and 11 in Zambian strains. They have correctly identified three distinct *ctxB* alleles among Zambian strains, not the two types that we stated in our published paper [1]. We did not analyse the clonal lineages in terms of genotype of the strain. However, as



Vibrio cholerae O1 lineages driving cholera outbreaks during seventh cholera pandemic in Ghana

Cristiane C. Thompson^a, Fernanda S. Freitas^a, Michel A. Marin^a, Erica L. Fonseca^a, Iruka N. Okeke^b, Ana Carolina P. Vicente^{a,*}

^aLaboratory of Molecular Genetics of Microorganisms, Oswaldo Cruz Institute (IOC) – Oswaldo Cruz Foundation (FIOCRUZ), Rio de Janeiro, Brazil

^bDepartment of Biology, Haverford College, 370 Lancaster Avenue, Haverford, PA 19041, USA

ARTICLE INFO

Article history:

Received 19 April 2011

Received in revised form 5 August 2011

Accepted 22 August 2011

Available online 30 August 2011

Keywords:

Vibrio cholerae O1 Amazonia

Cholera and Africa

Ghana

Multilocus Sequence Analysis

Class 2 integron

Vibrio cholerae resistance

Altered El Tor

ABSTRACT

In recent years, the frequency of cholera epidemics across Africa has increased significantly with thousands of people dying each year. However, there still exists a lack of information concerning the *Vibrio cholerae* O1 lineages driving early and contemporary epidemics since the seventh cholera pandemic started in the continent. This compromises the understanding of the forces determining the epidemiology of cholera in Africa and its control. This study aimed to analyze a collection of *V. cholerae* O1 strains from the beginning of the seventh cholera pandemic in Ghana and to compare them with recent isolates to understand the evolution of the cholera epidemic in Ghana. *V. cholerae* O1 strains were characterized by means of Multilocus Sequence Analysis (MLSA), genes from the virulence core genome (VCG), and genes related to the cholera phenotype. Our results revealed two major clusters of Ghanaian *V. cholerae* O1 strains, El Tor and Amazonia/Ghana. Concerning the virulence genes, all strains harbored the set of VCG and most were positive for VSP-II genomic island. The *ctxB* gene of the contemporary strains was characterized as Altered El Tor. The strains from 1970 to 1980 were susceptible to all antibiotics tested, except for the Amazonia/Ghana cluster that was resistant to aminoglycosides and carried the class 2 integron with the *sat2-aadA1* arrangement. This study showed that distinct *V. cholerae* O1 were the determinants of cholera outbreaks in Ghana. Thus, in endemic regions, such as Africa, cholera can be caused by various *V. cholerae* O1 genotypes.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Vibrio cholerae, the causative agent of cholera, is a major pathogen which is responsible for outbreaks of life-threatening diarrheal disease worldwide, particularly in developing countries. Only in 2009, the ancient disease Cholera killed nearly 4500 people in Africa (Mintz and Guerrant, 2009). The management of cholera outbreaks has changed little over time. Drinking water and sewage treatment, and oral/intravenous rehydration are the measures applied to prevent and treat the disease. *V. cholerae* is phenotypically characterized by the O somatic antigen. While over 200 serogroups are currently recognized, only the O1 and O139 serogroups have been responsible for epidemics and pandemics of cholera (Dharmasena et al., 2009). The pandemic *V. cholerae* O1 lin-

eages are classified into two major biotypes, classical and El Tor (Kaper et al., 1995). Seven cholera pandemics have been recorded, the first one in 1817. The sixth pandemic, and probably the earlier ones as well, were caused by the classical *V. cholerae* O1 biotype. The current seventh pandemic originated in Indonesia in 1961, is the most extensive in geographic spread and duration, and is caused by the El Tor biotype, which almost completely replaced the classical biotype (Kaper et al., 1995). Several clinical *V. cholerae* O1 strains cannot be classified into any of these two biotypes. Recently, the term 'atypical El Tor' has been applied to these unusual strains, some of which present a mix of both Classical and El Tor traits and which at least carry the Classical *ctxB* allele (Safa et al., 2010). Among the atypical El Tor strains, some lineages have been characterized, i.e., Matlab and Mozambique variants, Altered El Tor and Hybrid El Tor (Safa et al., 2010).

Cholera has contributed to a considerable disease burden in Africa, particularly in the early years of the 21st century (Mintz and Guerrant, 2009). The seventh pandemic entered Africa in 1970 by the Western countries of Guinea, Sierra Leone, Liberia, Ghana and Nigeria and spread inland along rivers and trade routes where it remains an ongoing source of mortality (Gaffga et al., 2007). In 2006, the coastal West African countries reported large

* Corresponding author. Address: Instituto Oswaldo Cruz (IOC)/FIOCRUZ, Av. Brasil 4365, P.O. Box 926, CEP 21045-900, Rio de Janeiro, Brazil. Tel.: +55 21 38658168; fax: +55 21 22604282.

E-mail addresses: thompson@ioc.fiocruz.br (C.C. Thompson), freitas@ioc.fiocruz.br (F.S. Freitas), mfabanto@ioc.fiocruz.br (M.A. Marin), ericafon@ioc.fiocruz.br (É.L. Fonseca), iokeke@haverford.edu (I.N. Okeke), anapaulo@ioc.fiocruz.br (A.C.P. Vicente).

numbers of cases and fatality rates of 1–6.2%. Between January and June 2006, in particular, 1869 cases and 79 deaths (4.2% fatality rate) were reported in Ghana (Opintan et al., 2008). Studies focusing on the molecular epidemiology of *V. cholerae* O1 in Africa are rare, despite of the severity of epidemics and the importance of this information in the understanding and control of this permanent public health threat in the continent. One exception was a study which conducted a surveillance of cholera in Beira/Mozambique that characterized some *V. cholerae* O1 strains as genetic hybrids denominated the Matlab variants (Ansaruzzaman et al., 2004).

In order to determine the population structure of cholerae *V. cholerae* O1 from Ghana, we analyzed a collection of clinical strains isolated between 1970 and 1981, representing the beginning of the seventh cholera epidemic in Africa, as well as strains from the 2006 outbreaks. These strains were characterized by Multilocus Sequence Analysis (MLSA), using housekeeping genes, and by macrorestriction profiles, for inter and intra-lineages analysis, respectively. The following genes involved in the cholerae phenotype were analyzed: the major virulence determinants (cholera toxin (*ctxAB*) and toxin-coregulated pilus (*tcpA*)), the *Vibrio* seventh pandemic island-II (VSP-II), and genes from the virulence core genome (VCG) (Gu et al., 2009). The VCG genes are involved in the hemolytic activity (*hlyB*), in bile acid resistance (*gshB*, *hepA*, *recO* and *tolC*), RTX toxin (*rtxA*) and haemagglutinin/protease (*hapA*) (Gu et al., 2009).

2. Materials and methods

A total of 32 clinical *V. cholerae* strains from cholera outbreaks were used in this study (Table 1). The strains are from the *Vibrio* culture collection of the Laboratory of Molecular Genetics of Microorganisms/FIOCRUZ. For MLSA the amplification and sequencing of the *pyrH*, *recA* and *rpoA* genes were performed as described previously (Thompson et al., 2005, 2008) (Table 2). The sequences were aligned using ClustalW (Thompson et al., 1994). Phylogenetic analyses were conducted using MEGA version 4.0 (Tamura et al., 2007), based on the genetic distance Neighbor-Joining method (Saitou and Nei, 1987) using concatenated sequences. Distance estimations were obtained by Kimura two parameter model. The reliability of the tree topology was checked by 2000 bootstrap replications. DNA macrorestriction profile was obtained by using *NotI* enzyme and Pulsed Field Gel Electrophoresis (PFGE) as previously described (Fonseca et al., 2006). The PFGE dendrogram was constructed using the BioNumerics software (Applied Maths, Belgium). The similarity between the strains was determined using the Dice coefficient, and cluster analysis was performed using the Unweighted Pair-Group Method with Arithmetic Mean (UPGMA).

All samples were screened for genes coding for the major determinants of cholera disease, (*ctxAB* and *tcpA*), genes from the virulence core genome of the epidemic *V. cholerae* O1 (*gshB*, *hapA*, *hepA*, *hlyB*, *recO*, *rtxA* and *tolC*) and VSP-II (*Vibrio* seventh

Table 1
Phenotypic and genotypic traits of clinical *V. cholerae* O1 strains used in this study.

<i>V. cholerae</i> O1 strains	Isolation	Antimicrobial resistance profile	SXT/class 2 integron	CVG/ <i>ctxB</i> / <i>tcpA</i> gene	VSP-II	Reference*
O395	India, 1965	ND	–/–	+/+	–/–	(1)
LMG 21698T	Asia	ND	ND	ND	ND	(1)
MAK757	Celebes Island, 1937	ND	–/–	+/+	–/–	(1)
N16961	Bangladesh, 1975	ND	–/–	+/+	+/+	(1)
TM 11079-80	Brazil, 1980	ND	–/–	+/-	–/–	(1)
V33	Ghana, 2006	SXT, AMP, NAL	+/-	+/-ND	ND	(2,3)
V34	Ghana, 2006	SXT, NAL	+/+	+/-ND	ND	(2,3)
V35	Ghana, 2006	SXT, NAL	+/+	+/-ND	ND	(2,3)
V42	Ghana, 2006	SXT	+/+	+/-ND	ND	(2,3)
V47	Ghana, 2006	SXT	+/+	+/-ND	ND	(2,3)
V51	Ghana, 2006	SXT	+/+	+/-ND	ND	(2,3)
V52	Ghana, 2006	SXT	+/+	+/-ND	ND	(2,3)
V53	Ghana, 2006	SXT, AMP, NAL	+/+	+/-ND	ND	(2,3)
V78	Ghana, 2006	SXT, AMP, NAL, CAZ, CRO, CHL, CXM	+/+	+/-ND	ND	(2,3)
V84	Ghana, 2006	SXT, AMP, TET, CXM	+/+	+/-ND	ND	(2,3)
V90	Ghana, 2006	SXT, AMK, CAZ, CRO, CTXCHL, GEN	+/+	+/-ND	ND	(2,3)
V97	Ghana, 2006	SXT	-/+	+/-ND	ND	(2,3)
V98	Ghana, 2006	SXT, AMP	+/+	+/-ND	ND	(2,3)
VC33	Tanzania, 1979	Susceptible	–/–	+/+	+/+	(3)
VC34	Ghana, 1970	Susceptible	–/–	+/+	+/+	(3)
VC37	Ghana, 1980	Susceptible	–/–	+/-	–/–	(3)
VC92	Ghana, 1980	STP, SPT	-/+	+/-	–/–	(3)
VC94	Ghana, 1979	STP, SPT	-/+	+/-	–/–	(3)
VC95	Ghana, 1978	STP, SPT	-/+	+/-	–/–	(3)
VC96	Ghana, 1978	STP, SXT	–/–	+/+	+/+	(3)
VC97	Ghana, 1979	STP, SPT	-/+	+/-	–/–	(3)
VC98	Ghana, 1979	STP, SPT	-/+	+/-	–/–	(3)
VC102	Ghana, 1979	STP	–/–	+/-	+/+	(3)
VC104	Algeria, 1972	STP	–/–	+/-	+/+	(3)
VC106	Ghana, 1980	Susceptible	–/–	+/+	+/+	(3)
VC107	Ghana, 1980	Susceptible	–/–	+/+	+/+	(3)
VC121	India, 1973	Susceptible	–/–	+/+	+/+	(3)
VC192	Ghana, 1980	Susceptible	–/–	+/+	+/+	(3)
VC193	Ghana, 1981	Susceptible	–/–	+/-	+/+	(3)
VC225	Ghana, 1976	Susceptible	–/–	+/+	+/+	(3)
VC355	Amazonia (Brazil), 1992	Susceptible	–/–	+/-	ND	(3)
VC356	Amazonia (Brazil), 1991	STP, SPT	-/+	+/-	ND	(3,4)

+, Positive; –, negative; ND, not determined Antibiotics tested; SXT, trimethoprim/sulfamethoxazole; AMK, amikacin; AMP, ampicillin; CAZ, ceftazidime; CRO, ceftriaxone; CTX, cefotaxime; CHL, chloramphenicol; GEN, gentamicin; TET, tetracycline; NAL, nalidixic acid; CXM, cefuroxime; STP, streptomycin; SPT, spectinomycin; TRM, trimethoprim. CVG: Core Virulence Genome (*gshB*, *hepA*, *hepA*, *hlyB*, *recO*, *rtxA* and *tolC*). VSP-II: *Vibrio* seventh pandemic island-II (VC0511–VC0513).

* (1) GenBank; (2) Opintan et al. (2008); (3) This study; (4) Sá et al. (2010).

Table 2
Primers used in this study.

Primer name	Primer sequence	Target region	Reference
pyrH-04-F	ATGASNACBAAYCCWAAACC	<i>pyrH</i>	Ansaruzzaman et al. (2004) and Thompson et al. (2008)
pyrH-02-R	GTRAABGCNGMYARRTCCA		
recA-01-F	TGARAARCARITYGGTAAAGG	<i>recA</i>	Ansaruzzaman et al. (2004) and Thompson et al. (2008)
recA-02-R	TCRCNTRTAGCTRTACC		
rpoA-01-F	ATGCAGGGTTCTGTDACAG	<i>rpoA</i>	Ansaruzzaman et al. (2004) and Thompson et al. (2008)
rpoA-03-R	GHGGCCARTTTTCHARRCCG		
CT8	GCAGTCAGGTGGTCTTATTGC	<i>ctxAB</i>	This study
CT10	TCCAGATATGCAATCCTCAG		
TCPA1	CACGATAAGAAAACCGTCAAGAG	<i>tcpA</i>	Keasler and Hall (1993)
TCPA2	ACCAAATGCAACGCCAATGGAGC		
GSHBF	ATCTGGCACAAAGGTCTGC	<i>gshB</i>	This study
GSHBR	CTCGGTATTGTGATGGACCC		
HAPF	TGAATACGGCAGTAACGG	<i>hapA</i>	This study
HAPR	GCCACAAGAAGCATTGAC		
HEPAF	AGCAAATGCACATTACGCAC	<i>hepA</i>	This study
HEPAR	CTTTAGCGAGGCCGATAACC		
HLYBF	CAAGCCTTCGCAATAAC	<i>hlyB</i>	This study
HLYBR	CCACTTTTTCCCTTCACC		
RECOF	CCGCTTGCACTGTTCTTTG	<i>recO</i>	This study
RECOR	TGCAACGCTGTTTTGTTCTG		
RTXA1	TCTTTACCATCACCACCCC	<i>rtxA</i>	This study
RTXA2	ACCACCTTCACTTATACGCC		
TOLCF	AGACTCTCAATACGCTGCC	<i>tolC</i>	This study
TOLCR	GACCGCACATCTTTAAAC		
INT1 F	AAAACCCGCACTGCGCCGTTA	Class 1 integrase gene	Fonseca et al. (2005)
INT1 R	GAAGACGGCTGCACTGAACG		
INT2 F	GCGTTTTATGTCTAACAGTCC	Class 2 integrase gene	Fonseca et al. (2005)
INT2 R	AAGTAGCATCAGTCCATCC		
INF	GGCATCCAAGCAGCAAGC	Class 1 integron variable region	Lévesque et al. (1995)
INB	AAGCAGACTTGACCTGAT		
INF2	TGGGTGAGATAATGTGCATC	Class 2 integron variable region	Sá et al. (2010)
INB2	TCCAGAGAGGATATGGAAGG		
SXTF	TCCGGTATCGCCCAAGGGCA	<i>SXT</i>	This study
SXTR	GCGAAGATCATGCATAGACC		
VC0511F1	CTTGCTGCGTACTTAGCA	<i>VC0511</i>	Nusrin et al. (2009)
VC0511R1	AGTAGCATCGCTCTCGTA		
VC0513F1	CTGAGGTGTTATATGTTCCG	<i>VC0513</i>	Nusrin et al. (2009)
VC0513R1	TCAAATTCCTGACAGTTCC		

pandemic island-II gene cluster) by PCR and sequencing (Table 2). Antimicrobial susceptibility testing was performed according to the disc diffusion method described by the CLSI guidelines using cephalothin, cefoxitin, nalidixic acid, ampicillin, sulfamethoxazole–trimethoprim, sulfonamide, chloramphenicol, gentamicin, tetracycline, ceftriaxone, streptomycin, and spectinomycin. The strains were screened for the presence of integrons and their variable regions. PCR reactions were performed using primers targeting the *intI1*, *intI2* and *intI3* genes (Fonseca et al., 2005) and the *SXT* (sulfamethoxazole–trimethoprim element) (Table 2).

The gene sequences have been deposited at DDBJ/EMBL/GenBank under the Accession Nos. HM003832 to HM003893.

3. Results

3.1. Genetic relationship of *V. cholerae* O1 strains by MLSA and PFGE

The MLSA results revealed two clusters of clinical *V. cholerae* O1 from Ghana. One, containing the majority of the strains, belonged to the pandemic *V. cholerae* O1 El Tor cluster. Strains from this cluster were isolated in Ghana between 1970 and 2006 (Fig. 1 and Table 1). The other cluster was composed of Ghanaian strains VC92, VC94, VC95, VC97 and VC98 isolated from 1978 to 1980. These strains grouped with 100% MLSA similarity with clinical Brazilian *V. cholerae* O1 Amazonia strains VC356 and VC355 isolated in 1991 and 1992, respectively, and with TM 11079-80 isolated in 1980 (Fig. 1 and Table 1). This cluster, which we called *V. cholerae* O1 Amazonia/Ghana cluster, was clearly separated from the

V. cholerae O1 El Tor cluster with at least 97% similarity in MLSA, and supported by high bootstrap values (99%). The *recA* gene showed the highest resolution for differentiating these two clusters (3% gene sequence divergence). The PFGE dendrogram (Fig. 2) showed two major clusters, the El Tor and Amazonia/Ghana clusters. The El Tor strains VC121 and VC20 (representative from the seventh cholera pandemic) and VC34 and VC37 from Ghana showed > 89.5% similarity. The Ghanaian strains VC92, VC94, VC95, VC97 and VC98 were similar to the Brazilian *V. cholerae* O1 Amazonia strains (VC355 and VC356) (at least 91.5% similarity) despite being temporally distant. The similarity between the two major clusters is 76.2%.

These results showed the presence of a lineage shared by the two countries however, PFGE sensitivity reveals recent genetic diversification within the strains from Ghana/Amazonia and El Tor clusters.

3.2. Characterization of elements associated with antibiotic resistance

All the Ghana strains belonging to the Amazonia/Ghana cluster were positive for class 2 integron (Table 1) as well as the *V. cholerae* O1 Amazonia VC356 from the Brazilian cholera epidemic (Sá et al., 2010). Sequence analysis of the class 2 integron gene cassette region revealed the presence of the *sat1/aadA1* array (streptothricin/streptomycin and spectinomycin resistance).

All the strains isolated between 1970 and 1981, the beginning of the pandemic in Ghana, were susceptible to all antimicrobial agents tested (Table 1). Strains VC92, VC94, VC95, VC97 and VC98, belonging to the Amazonia/Ghana cluster, were the

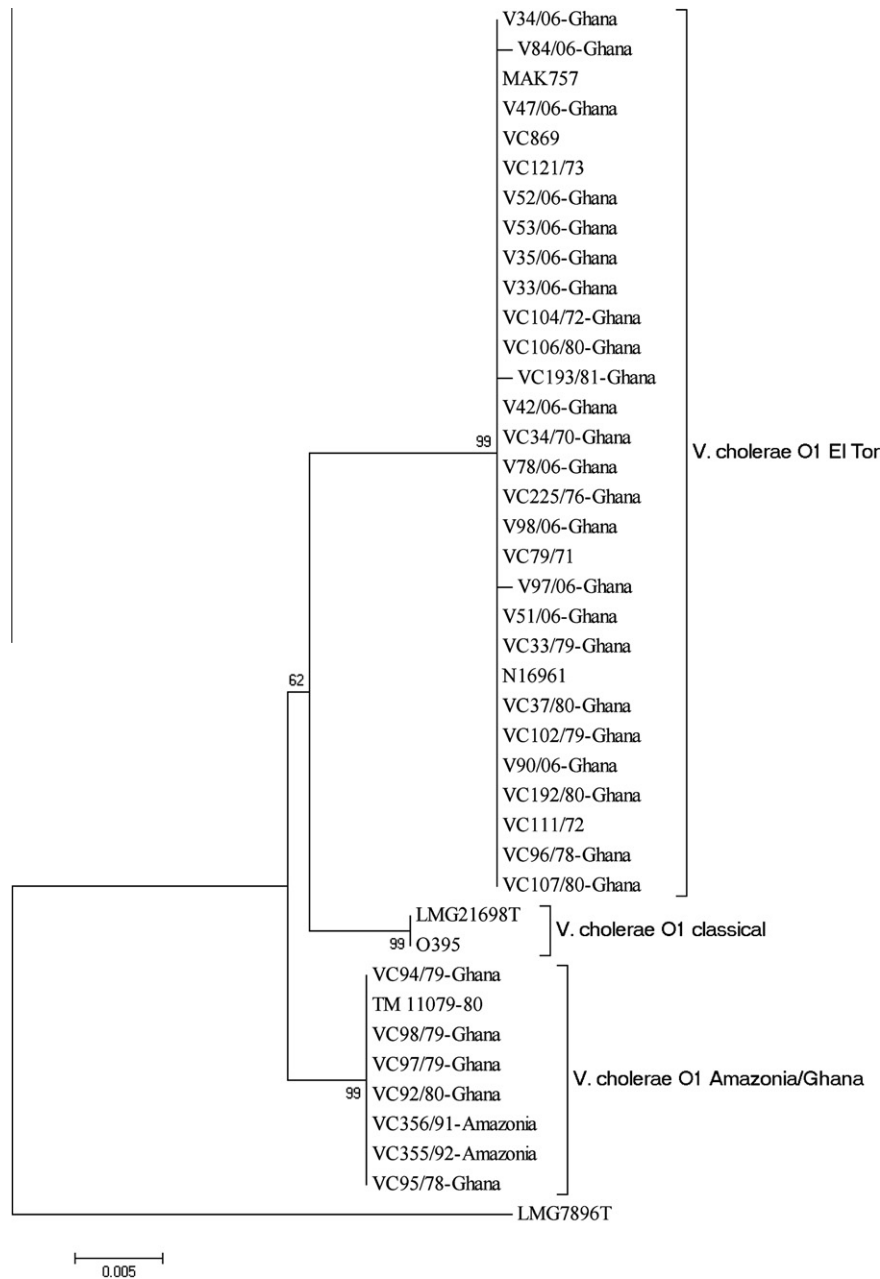


Fig. 1. Phylogenetic tree based on the Neighbor-joining method using concatenated sequences of the *pyrH*, *recA* and *rpoA* housekeeping genes. Distance estimation obtained by the Kimura two-parameter model. Bootstrap percentages after 2000 replications are shown. LMG 21698^T is the type strain of *V. cholerae* species. *V. mimicus* LMG 7896^T was included as an outgroup. Bar, estimated sequence divergence. The year and origin of strain isolation are placed after the slash and hyphen, respectively.

exception and were resistant to streptomycin and spectinomycin, probably due to the presence of the *aadA1* (aminoglycoside adenylyltransferase) gene inserted in the class 2 integrons. All strains (1970–1981) were negative for the SXT element (SXT encodes resistances to several antibiotics, including sulfamethoxazole and trimethoprim (Wozniak et al., 2009)).

3.3. Detection and characterization of *ctxAB* and *tcpA* genes, VCG genes and the VSP-II region

The two major virulence determinants, *ctxAB* and *tcpA* genes, were not present in the strains belonging to the Amazonia/Ghana cluster. On the other hand, most *V. cholerae* O1 from the El Tor cluster carried these genes (Table 1). Interestingly, the *ctxB* gene carried by the 2006 epidemic El Tor isolates from Ghana are

identical to that of the Classical biotype, characterizing an Altered El Tor variant. The *gshB*, *hapA*, *hepA*, *hlyB*, *recO*, *rtxA* and *tolC* genes that are part of the virulence core genome were present in all *V. cholerae* O1 strains from the El Tor and Amazonia/Ghana lineages (Table 1). The strains from the El Tor cluster were positive for the presence of ORFs VC0511 and VC0513 from VSP-II genomic island, except VC37. These ORFs were absent in the Amazonia/Ghana strains.

4. Discussion

The world is under the seventh cholera pandemic since the 1960s when the El Tor lineage emerged. Cholera has continuously emerged in places where it has been absent for a century as recently occurred in Haiti in October, 2010. Variants of the pandemic

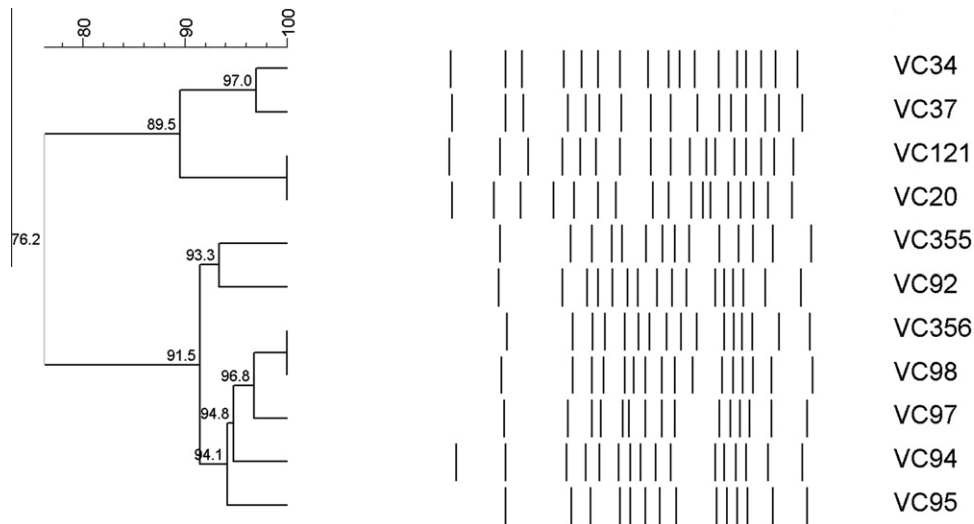


Fig. 2. Dendrogram constructed from the PFGE profiles of representative *V. cholerae* strains O1. The similarity between the strains was determined using the Dice coefficient, and cluster analysis was performed using the UPGMA.

V. cholerae O1 El Tor have been characterized causing cholera outbreaks. Therefore, consistent surveillance programs characterizing *V. cholerae* strains from the distinct outbreaks worldwide are necessary to understand the epidemiology of the disease. In this study, the genetic characterization of contemporary and old *V. cholerae* strains causing cholera in Ghana was performed for the first time.

Our study revealed that, in the beginning of the seventh cholera pandemic in Ghana, at least two *V. cholerae* O1 lineages, Amazonia/Ghana and El Tor, were co-circulating causing cholera. A similar scenario had been demonstrated in the Brazilian cholera epidemic when the *V. cholerae* O1 Amazonia lineage was first characterized (Coelho et al., 1995). This lineage was isolated from cholera outbreaks in northern Brazil/Amazonia (1991 and 1992), and was proved to be distinct from the prevailing El Tor seventh pandemic lineage (Thompson et al., In Press). So far, *V. cholerae* O1 Amazonia was thought to be restricted to villages of the Amazon Basin and was probably unable to compete with the invading El Tor lineage (Coelho et al., 1995). However, our results clearly showed that *V. cholerae* O1 Amazonia was already circulating at the beginning of the seventh pandemic in Ghana (1978–1980). Amazonia strains might have escaped detection owing to their low prevalence, and/or because they share a common important phenotypic marker, the O1 antigen, used to characterize *V. cholerae* isolates during cholera outbreaks (Chun et al., 2009). The occurrence of the Amazonia/Ghana lineage in distinct times and places, Brazil/1991–1992 and Ghana/1978–1980, indicates its ability in spreading, fitness, and epidemic potential.

Chin et al. (2011) corroborates the Lam et al. (2010) hypothesis that the *V. cholerae* O1 strains from the Latin American cholera epidemic were introduced from Africa, as opposed to the previous hypothesis suggesting their origin in Asia. The authors argue that the mass migration from Africa to South America in the 1970s brought *V. cholerae* into the region. Our findings, showing the presence of strains belonging to the Amazonia/Ghana cluster in Ghana (1978–1980), support the Africa to South America route in the epidemiology of the seventh cholera pandemic.

The presence of class 2 integron in strains of Amazonia/Ghana lineage but not in the El Tor strains, at the beginning of seventh cholera pandemic in Ghana, contrast with results from Opintan et al. (2008). The authors showed that most of *V. cholerae* O1 El Tor strains from the 2006 cholera outbreak in Ghana carried class 2 integron, which is a rare trait in *V. cholerae* O1. All together these

findings suggest that this element was horizontally transferred from the Amazonia/Ghana lineage to El Tor, in the course of the epidemics in the country. The current class 2 integron carries the canonical *dfrA1*, *sat* and *aadA1* array while the class 2 integron of the Amazonia/Ghana strains, from 1970 to 1980, lacked *dfrA1*. The evolution of this variable region can be due to the acquisition of *dfrA1*, by *in trans* activity mediated by *intVchA*, from the chromosomal integron (Ahmed et al., 2006).

Recently, several works demonstrated the emergence of El Tor strains encoding the classical cholera toxin causing epidemics in Asian and African countries (Safa et al., 2010). In some of these outbreaks, due to these Altered El Tor strains, a much higher proportion of patients presenting severe dehydration was noticed. So far, in Africa, these Altered El Tor strains have been described in Mozambique and Zambia (Safa et al., 2010) as successful clinical clones and as having replaced the prototypical O1 El Tor (Grim et al., 2010). We demonstrated the occurrence of the Altered *V. cholerae* O1 strains in the Ghana outbreak for the first time.

Opintan et al. (2008) showed that Ghana strains from 2006 were resistant to multiple antibiotics, including trimethoprim/sulfamethoxazole, that is associated with the presence of the SXT element, presented in the strains. These results, together with our findings related to the susceptibility of the isolates from the beginning of the 7th pandemic, showed that *V. cholerae* O1 antibiotic resistance has been evolving during the cholera pandemic in Ghana. Concerning the emergence of the SXT element, it was showed that Altered *V. cholerae* O1 strains are carrying this element, as was also recently observed in India (Goel and Jiang, 2010).

Studies on the molecular epidemiology of *V. cholerae* O1 are of extreme importance to better understand the dynamics of this bacterium in epidemics and pandemics of cholera, and for the implementation of control measures. The work applying genomic analysis to determine the origin of the Haitian cholera outbreak strain is a landmark in this field (Chin et al., 2011). Our work demonstrated that in endemic regions, such as Africa, several genotypes could be driving cholera outbreaks.

Acknowledgements

The authors thank CAPES, FAPERJ, FIOCRUZ and the Society-in-Science (the Branco-Weiss Fellowship) for the financial support. We thank Koko Otsuki and Bing Dao Zhang for excellent technical

assistance, Carlos André Salles for *Vibrio* strains and fruitful discussions, and the PDTIS/FIOCRUZ DNA sequencing facility. We are grateful to Japheth Opintan, Mercy Newman and Owusu Agyemang Nsiah-Poodoh for DNA from Ghana 2006 isolates.

References

- Ahmed, A.M., Kawaguchi, F., Shimamoto, T., 2006. Class 2 integrons in *Vibrio cholerae*. *J. Med. Microbiol.* 55, 643–644.
- Ansaruzzaman, M., Bhuiyan, N.A., Nair, B.G., Sack, D.A., Lucas, M. Mozambique Cholera vaccine Demonstration Project Coordination Group, 2004. Cholera in Mozambique, variant of *Vibrio cholerae*. *Emerg. Infect. Dis.* 10, 2057–2059.
- Chin, C.S., Sorenson, J., Harris, J.B., Robins, W.P., Charles, R.C., Jean-Charles, R.R., Bullard, J., Webster, D.R., Kasarskis, A., Peluso, P., Paxinos, E.E., Yamaichi, Y., Calderwood, S.B., Mekalanos, J.J., Schadt, E.E., Waldor, M.K., 2011. The origin of the Haitian cholera outbreak strain. *N. Engl. J. Med.* 364, 33–42.
- Chun, J., Grim, C.J., Hasan, N.A., Lee, J.H., Choi, S.Y., Haley, B.J., Taviani, E., Jeon, Y.S., Kim, D.W., Lee, J.H., Brettin, T.S., Bruce, D.C., Challacombe, J.F., Detter, J.C., Han, C.S., Munk, A.C., Chertkov, O., Meincke, L., Saunders, E., Walters, R.A., Huq, A., Nair, G.B., Colwell, R.R., 2009. Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. *Proc. Natl. Acad. Sci. USA* 106, 15442–15447.
- Coelho, A., Andrade, J.R., Vicente, A.C., Salles, C.A., 1995. New variant of *Vibrio cholerae* O1 from clinical isolates in Amazonia. *J. Clin. Microbiol.* 33, 114–118.
- Dharmasena, M.N., Krebs, S.J., Taylor, R.K., 2009. Characterization of a novel protective monoclonal antibody that recognizes an epitope common to *Vibrio cholerae* Ogawa and Inaba serotypes. *Microbiology* 155, 2353–2364.
- Fonseca, E.L., Vieira, V.V., Cipriano, R., Vicente, A.C., 2005. Class 1 integrons in *Pseudomonas aeruginosa* isolates from clinical settings in Amazon region, Brazil. *FEMS Immunol. Med. Microbiol.* 44, 303–309.
- Fonseca, E.L., Vieira, V.V., Cipriano, R., Vicente, A.C., 2006. Emergence of *dhfr*XVb and *bla*_{CARB-4} gene cassettes in class 1 integrons from clinical *Pseudomonas aeruginosa* isolated in Amazon region. *Mem. Inst. Oswaldo Cruz* 101, 81–84.
- Gaffga, N.H., Tauxe, R.V., Mintz, E.D., 2007. Cholera: a new homeland in Africa? *Am. J. Trop. Med. Hyg.* 77, 705–713.
- Goel, A.K., Jiang, S.C., 2010. Genetic determinants of virulence, antibiogram and altered biotype among the *Vibrio cholerae* O1 isolates from different cholera outbreaks in India. *Infect. Genet. Evol.* 10, 815–819.
- Grim, C.J., Hasan, N.A., Taviani, E., Haley, B., Chun, J., Brettin, T.S., Bruce, D.C., Detter, J.C., Han, C.S., Chertkov, O., Challacombe, J., Huq, A., Nair, G.B., Colwell, R.R., 2010. Genome sequence of hybrid *Vibrio cholerae* O1 MJ-1236, B-33, and CIRS101 and comparative genomics with *V. Cholerae*. *J. Bacteriol.* 192, 3524–3533.
- Gu, J., Wang, Y., Lilburn, T., 2009. A comparative genomics, network-based approach to understanding virulence in *Vibrio cholerae*. *J. Bacteriol.* 191, 6262–6272.
- Kaper, J.B., Morris Jr., J.G., Levine, M.M., 1995. Cholera. *Clin. Microbiol. Rev.* 8, 48–86.
- Keasler, S.P., Hall, R.H., 1993. Detecting and biotyping *Vibrio cholerae* O1 with multiplex polymerase chain reaction. *Lancet* 341, 1661.
- Lam, C., Octavia, S., Reeves, P., Wang, L., Lan, R., 2010. Evolution of seventh cholera pandemic and origin of 1991 epidemic, Latin America. *Emerg. Infect. Dis.* 16, 1130–1132.
- Lévesque, C., Piché, L., Larose, C., Roy, P.H., 1995. PCR mapping of integrons reveals several novel combinations of resistance genes. *Antimicrob. Agents Chemother.* 39, 185–191.
- Mintz, E.D., Guerrant, R.L., 2009. A lion in our village—the unconscionable tragedy of cholera in Africa. *N. Engl. J. Med.* 360, 1060–1063.
- Nusrin, S., Gil, A.I., Bhuiyan, N.A., Safa, A., Asakura, M., 2009. Peruvian *Vibrio cholerae* O1 El Tor strains possess a distinct region in the *Vibrio* seventh pandemic island-II that differentiates them from the prototype seventh pandemic El Tor strains. *J. Med. Microbiol.* 58, 342–354.
- Opintan, J.A., Newman, M.J., Nsiah-Poodoh, O.A., Okeke, I.N., 2008. *Vibrio cholerae* O1 from Accra, Ghana carrying a class 2 integron and the SXT element. *J. Antimicrob. Chemother.* 62, 929–933.
- Sá, L.L., Fonseca, E.L., Pellegrini, M., Freitas, F., Loureiro, E.C., Vicente, A.C., 2010. Occurrence and composition of class 1 and class 2 integrons in clinical and environmental O1 and non-O1/non-O139 *Vibrio cholerae* strains from the Brazilian Amazon. *Mem. Inst. Oswaldo Cruz* 105, 229–232.
- Safa, A., Nair, G.B., Kong, R.Y., 2010. Evolution of new variants of *Vibrio cholerae* O1. *Trends Microbiol.* 18, 46–54.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599.
- Thompson, C.C., Thompson, F.L., Vicente, A.C.P., 2008. Identification of *Vibrio cholerae* and *Vibrio mimicus* by Multi locus sequence analysis (MLSA). *Int. J. Syst. Evol. Microbiol.* 58, 617–621.
- Thompson, C.C., Marin, M.A., Dias, G.M., Dutilh, B.E., Edwards, R.A., Iida, T., Thompson, F.L., Vicente, A.C.P. Genome sequence of the human pathogen *Vibrio cholerae* Amazonia. *J. Bacteriol.*, in press.
- Thompson, F.L., Gevers, D., Thompson, C.C., Dawyndt, P., Naser, S., Hoste, B., Munn, C.B., Swings, J., 2005. Phylogeny and molecular identification of vibrios on the basis of multilocus sequence analysis. *Appl. Environ. Microbiol.* 71, 5107–5115.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- Wozniak, R.A.F., Fouts, D.E., Spagnoletti, M., Colombo, M.M., Ceccarelli, D., Garriss, G., et al., 2009. Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. *PLoS Genet.* 5 (12), e1000786, doi:10.1371/journal.pgen.1000786.

Cholera Outbreaks in Nigeria Are Associated with Multidrug Resistant Atypical El Tor and Non-O1/Non-O139 *Vibrio cholerae*

Michel A. Marin¹, Cristiane C. Thompson¹, Fernanda S. Freitas¹, Erica L. Fonseca¹, A. Oladipo Aboderin², Sambo B. Zailani³, Naa Kwarley E. Quartey⁴, Iruka N. Okeke⁴, Ana Carolina P. Vicente^{1*}

1 Laboratory of Molecular Genetics of Microorganisms, Oswaldo Cruz Institute (IOC), Oswaldo Cruz Foundation (FIOCRUZ), Rio de Janeiro, Brazil, **2** Department of Medical Microbiology & Parasitology, College of Health Sciences, Obafemi Awolowo University, Ile-Ife, Nigeria, **3** Department of Medical Microbiology and Parasitology, University of Maiduguri, Maiduguri, Nigeria, **4** Department of Biology, Haverford College, Haverford, Pennsylvania, United States of America

Abstract

Background: The current millennium has seen a steep rise in the number, size and case-fatality of cholera outbreaks in many African countries. Over 40,000 cases of cholera were reported from Nigeria in 2010. Variants of *Vibrio cholerae* O1 El Tor biotype have emerged but very little is known about strains causing cholera outbreaks in West Africa, which is crucial for the implementation of interventions to control epidemic cholera.

Methodology/Principal Findings: *V. cholerae* isolates from outbreaks of acute watery diarrhea in Nigeria from December, 2009 to October, 2010 were identified by standard culture methods. Fifteen O1 and five non-O1/non-O139 strains were analyzed; PCR and sequencing targeted regions associated with virulence, resistance and biotype were performed. We also studied genetic interrelatedness among the strains by multilocus sequence analysis and pulsed-field gel electrophoresis. The antibiotic susceptibility was tested by the disk diffusion method and E-test. We found that multidrug resistant atypical El Tor strains, with reduced susceptibility to ciprofloxacin and chloramphenicol, characterized by the presence of the SXT element, and *gyrA*^{Ser83Ile}/*parC*^{Ser85Leu} alleles as well CTX phage and TCP cluster characterized by *rstR*^{ElTor}, *ctxB-7* and *tcpA*^{CIRS} alleles, respectively, were largely responsible for cholera outbreaks in 2009 and 2010. We also identified and characterized a *V. cholerae* non-O1/non-O139 lineage from cholera-like diarrhea cases in Nigeria.

Conclusions/Significance: The recent Nigeria outbreaks have been determined by multidrug resistant atypical El Tor and non-O1/non-O139 *V. cholerae* strains, and it seems that the typical El Tor, from the beginning of seventh cholera pandemic, is no longer epidemic/endemic in this country. This scenario is similar to the East Africa, Asia and Caribbean countries. The detection of a highly virulent, antimicrobial resistant lineage in Nigeria is worrisome and points to a need for vaccine-based control of the disease. This study has also revealed the putative importance of non-O1/non-O139 *V. cholerae* in diarrheal disease in Nigeria.

Citation: Marin MA, Thompson CC, Freitas FS, Fonseca EL, Aboderin AO, et al. (2013) Cholera Outbreaks in Nigeria Are Associated with Multidrug Resistant Atypical El Tor and Non-O1/Non-O139 *Vibrio cholerae*. PLoS Negl Trop Dis 7(2): e2049. doi:10.1371/journal.pntd.0002049

Editor: Joseph M. Vinetz, University of California San Diego School of Medicine, United States of America

Received: October 11, 2012; **Accepted:** December 19, 2012; **Published:** February 14, 2013

Copyright: © 2013 Marin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: MAM, CCT and ELF obtained scholarships from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), respectively. INO received financial support from Society-in-Science, ETHZ, Switzerland. This study was also supported by Oswaldo Cruz Institute Grant. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mfabanto@ioc.fiocruz.br

Introduction

Vibrio cholerae is a human pathogen that causes cholera, a severe acute watery diarrhea. There are more than 200 *V. cholerae* serogroups, however only O1 and O139 are responsible for most epidemics and pandemics of cholera. Serogroup O139 is restricted to some parts of Asia but serogroup O1, classified as El Tor and classical biotypes, was found worldwide [1,2]. The classical biotype was responsible for severe clinical manifestation but the conventionally less virulent El Tor is better adapted to the environment [3].

The major virulence factors are the cholera toxin (CT) encoded by the *ctxAB* operon on the lysogenic bacteriophage CTXΦ [4],

and the toxin-coregulated pilus (TCP) encoded by the VPI-I island, an essential factor for the colonization and also receptor for the CTXΦ [5]. Classical and El Tor biotypes can be distinguished by several genetic markers, such as *rstR*, *tcpA*, *ctxB* genes, the *Vibrio* seventh pandemic island-I (VSP-I) and VSP-II, and by the number of genes in the repeat toxin region (RTX) [2,6].

Recently, *V. cholerae* strains have been identified showing markers of both classical and El Tor biotypes. These strains were assigned as hybrid biotype, altered or El Tor variants presenting a typical El Tor phenotype but with non-El Tor *ctxB* allele [2]. In contrast to *V. cholerae* from the beginning of 7th cholera pandemic, these atypical biotype strains are frequently multidrug resistant. Changes in the antibiotic resistance profiles of *V. cholerae* are

Author Summary

Cholera is acute watery diarrhoea, severely dehydrating, caused by *Vibrio cholerae*, a bacterium ubiquitous in aquatic environments. Cholera is a global threat, particularly, in areas where sanitary conditions, such as drinking water and sewage, are not available. Seven cholera pandemics, all originating in Asia, occurred. The ongoing pandemic, the 7th, has been caused by *V. cholerae* El Tor biotype. Recently, El Tor has undergone genetic changes and the strains being referred to as “atypical” El Tor are rapidly replacing the original El Tor in many areas. The atypical El Tor is characterized by multi-antibiotic resistance and changes in the major virulence determinants. Cholera caused by atypical strains may be more clinically severe. In Africa, cholera outbreaks are occurring with increasing frequency and severity, as demonstrated by the recent major outbreaks in Nigeria, Angola, Mozambique and Zimbabwe. Here, we performed a comprehensive characterization of *V. cholerae* isolated from different recent outbreaks in Nigeria. Our results show that cholera outbreaks in Nigeria are driven by atypical El Tor strains, as worldwide.

associated with mutations in housekeeping genes as well as acquisition of Integrative Conjugative Elements (ICEs) and other mobile elements [7,8].

Africa is endemic for cholera and frequently affected by outbreaks and epidemics, but there are few molecular epidemiology studies characterizing the determinants of these episodes. Nigeria is in one of the three major current cholera foci in the world [9]. In 2009, outbreaks began in Nigeria and other countries at the Lake Chad basin [8] with the first reports coming from Maiduguri, a city in the far north-east of the country. Subsequently, outbreaks were reported from distant locales in Northern and Western Nigeria, and in 2010 a severe outbreak, which started in the Northern Nigeria spreading through the country, was projected as the worst outbreak in Nigeria since 1991. This outbreak was marked with highest case-fatality [10], what could be in part due to changes in *V. cholerae* infectivity even though the organism remains largely unknown. It can be hypothesized that an index strain has been disseminated cross-country by human travel.

Here, we performed a comprehensive characterization of representative *V. cholerae* strains from sequential outbreaks in Nigeria (Maiduguri/2009, Bauchi/2010 and Ile Ife/2010) by means of biotyping, multilocus sequence analysis (MLSA), pulsed-field gel electrophoresis (PFGE), as well as detection and sequencing of virulence related genes and genetic determinants of antimicrobial resistance. Our results show that recent cholera outbreaks in Nigeria are driven by atypical El Tor strains and we reported the presence of a non-O1/non-O139 lineage. The atypical El Tor strains showed one unique combination of virulence factor alleles and antimicrobial resistance to sulphonamides, trimethoprim/sulfamethoxazole, streptomycin, nalidixic acid and reduced susceptibility to ciprofloxacin and chloramphenicol.

Materials and Methods

Bacterial strains

Twenty *V. cholerae* strains (15 O1 and five non-O1/non-O139 strains) isolated from outbreaks in Nigeria and reference strains were analyzed (Table 1). We included in our analysis three isolates

that originated from Nigeria in the beginning of the seventh pandemic (1970's), 16 isolates from three recent cholera outbreaks and one strain isolated from a water sample. *V. cholerae* isolates were identified and confirmed by standard culture methods at isolation during the epidemics and again prior to molecular characterization, and their serogroups and biotypes were identified by biochemical and molecular methods [11].

PCR amplification and sequencing

PCR were performed targeting the genetic elements and regions associated with *V. cholerae* virulence, resistance and biotype. Virulence: VSP-II (VC0511, VC0513), VPI-I (*tcpA*), CTXΦ (*ctxB*), heat stable enterotoxin (NAG-ST), type three secretion system (T3SS), type VI secretion system (T6SS), enterotoxigenic hemolysin (*hlyA*) and RTX toxin (*rtxA*). Resistance: *IntI1*, *IntI2* integrases and variable regions from class 1 and 2 integrons and the 3' conserved sequence from class 1 integron, SXT element-integrase gene and associated SXT resistance genes (*floR*, *sul2*, *dfrA1* and *strAB*) and genes related to quinolone resistance (*gyrA*, *gyrB*, *parC* and *parE*). Biotype-specific repeat sequence transcriptional regulator (*rstR*), *ctxB* alleles and *rfb* gene specific for O1 serogroup 1 were also amplified. The nucleotide sequences of primers employed are listed in Table 2.

MLSA

Amplification and sequencing of the *pyrH*, *recA* and *mdh* housekeeping genes were performed as described previously (Table 2). Sequence alignments and phylogenetic analyses were conducted using MEGA5 [12], based on the minimum evolution method using concatenated sequences. Distance estimations were obtained by Kimura two-parameter model. Two thousand bootstrap replicates were performed.

Pulsed-Field Gel Electrophoresis

DNA macrorestriction profile was obtained using *NotI* enzyme according to procedures described elsewhere [13]. The PFGE dendrogram was constructed using BioNumerics software (Applied Maths, Belgium). The similarity was determined by Dice coefficient and cluster analysis was carried out with the unweighted-pair group method using average linkages (UPGMA).

Disk diffusion susceptibility testing and Etest

Antibiotic susceptibility testing was performed by the disk diffusion method, with Muller-Hinton Agar (Oxoid, Basingstoke, UK), according to Clinical and Laboratory Standards Institute (CLSI, 2010) standards. Strains were tested for resistance to: Nalidixic acid, Ampicillin, Cephalothin, Cefpirom, Cefoxitin, Cefuroxime sodium, Ciprofloxacin, Chloramphenicol, Erythromycin, Spectinomycin, Compound sulphonamides, Streptomycin, Sulphamethoxazole, Sulphamethoxazole-Trimethopim, Trimethopim and Tetracycline. Minimum Inhibitory Concentrations (MICs) for Ciprofloxacin and Chloramphenicol were determined with the E-test (bioMerieux, Marcy l'Etoile, France).

Results

V. cholerae genotyping

We determined the genetic relationships among *V. cholerae* strains consisting of Nigeria strains from 2009/2010 and 1971/1972, one environmental isolate (2010) and well-characterized reference strains. A MLSA tree based on concatenated sequence of *recA* (666 bp), *pyrH* (456 bp) and *mdh* (591 bp) gene fragments were constructed.

Table 1. Genetic and phenotypic characteristics of *V. cholerae* from Nigeria.

Strain	Serogroup (biotype)	Place of isolation	Year of isolation	Virulence related genes and biotypes						Related to antibiotic resistance						
				VSP-II	ctxB*	tcpA	rstR	IntI1	IntI2	SXT	gyrA	parC	MIC ^{CIP} (µg/ml)	MIC ^{CM} (µg/ml)		
Reference [‡]				VC0511	VC0513											
N16961	O1 (El Tor)	Bangladesh	1975	+	+	3	ET	ET	ET	-	-	-	Ser83	Ser85	ND	ND
CIRS101	O1 (El Tor)	Dhaka, Bangladesh	2002	-	+	1	ET ^{CRIS}	ET	ET	-	-	+	Ser83lle	Ser85	ND	ND
MO10	O139	Madras, India	1992	-	+	3	ET	ET	ET	-	-	+	Ser83	Ser85	ND	ND
O395	O1 (Classical)	India	1965	-	-	1	CL	CL	CL	-	-	-	Ser83	Ser85	ND	ND
RC27	O1 (Classical)	Indonesia	1991	-	-	1	CL	CL	CL	-	-	-	Ser83	Ser85	ND	ND
Nigeria strains																
VC79	O1	Nigeria	1971	+	+	-	-	ET	ET	-	-	-	Ser83	Ser85	0.008	ND
VC111	O1	Nigeria	1972	+	+	3	ET	ET	ET	-	-	-	Ser83	Ser85	0.008	ND
VC869	O1	Nigeria	1971	+	+	3	ET	ET	ET	-	-	-	Ser83	Ser85	0.008	ND
VC832	O1	Maiduguri	2009	-	+	7	ET ^{CRIS}	ET	ET	-	-	+	Ser83lle	Ser85Leu	0.380	8.0
VC841	O1	Maiduguri	2009	-	+	7	ET ^{CRIS}	ET	ET	-	-	+	Ser83lle	Ser85Leu	0.250	8.0
VC833	O1	Ile Ife	2010	-	+	7	ET ^{CRIS}	ET	ET	-	-	+	Ser83lle	Ser85Leu	0.250	6.0
VC834	O1	Ile Ife	2010	-	+	7	ET ^{CRIS}	ET	ET	-	-	+	Ser83lle	Ser85Leu	0.250	6.0
VC835	O1	Ile Ife	2010	-	+	7	ET ^{CRIS}	ET	ET	-	-	+	Ser83lle	Ser85Leu	0.380	6.0
VC991	O1	Ile Ife	2010	-	+	7	ET ^{CRIS}	ET	ET	-	-	+	Ser83lle	Ser85Leu	0.380	6.0
VC997	O1	Ile Ife	2010	-	+	7	ET ^{CRIS}	ET	ET	-	-	+	Ser83lle	Ser85Leu	0.380	6.0
VC1001	O1	Ile Ife	2010	-	+	7	ET ^{CRIS}	ET	ET	-	-	+	Ser83lle	Ser85Leu	0.380	6.0
VC1007	O1	Ile Ife	2010	-	+	7	ET ^{CRIS}	ET	ET	-	-	+	Ser83lle	Ser85Leu	0.380	8.0
VC836	O1	Bauchi	2010	-	+	7	ET ^{CRIS}	ET	ET	-	-	+	Ser83lle	Ser85Leu	0.380	8.0
VC996	Non-O1	Bauchi	2010	-	-	-	-	-	-	-	-	-	Ser83	Ser85	0.023	0.5
VC998	Non-O1	Bauchi	2010	-	-	-	-	-	-	-	-	-	Ser83	Ser85	0.008	0.5
VC999	O1	Bauchi	2010	-	+	7	ET ^{CRIS}	ET	ET	-	-	+	Ser83lle	Ser85Leu	0.380	6.0
VC1004	O1	Bauchi	2010	-	+	7	ET ^{CRIS}	ET	ET	-	-	+	Ser83lle	Ser85Leu	0.500	6.0
VC1006	Non-O1	Bauchi	2010	-	-	-	-	-	-	-	-	-	Ser83	Ser85	0.012	0.5
VC1009	Non-O1	Bauchi	2010	-	-	-	-	-	-	-	-	-	Ser83	Ser85	0.008	0.5
VC1005 [§]	Non-O1	Ile Ife	2010	-	-	-	-	-	-	-	-	-	Ser83lle	Ser85Leu	0.250	ND

CL, classical; El Tor; CIP, ciprofloxacin; CM, Chloramphenicol; ND, not determined;

* ctxB genotype [16];

‡ The genetic characteristics were determined by *in silico* analysis;

§ VC1005 is an environmental isolate.

doi:10.1371/journal.pntd.0002049.t001

Table 2. PCR primers used in this study.

Primer	Sequence (5'-3')	Reference
Housekeeping gene		
recA-01-F	TGARAARCARTTYGGTAAAGG	[43]
recA-02-R	TCRCCNTRTAGCTRACC	
pyrH-04-F	ATGASNACBAAYCCWAAACC	[43]
pyrH-02-R	GTRAABGCNGMYARRTCCA	
mdh-1	ATGAAAGTCGCTGTATT	[44]
mdh-2	GTATCTAACATGCCATCC	
Biotype related genes		
Serogroup O1		
O1-F	GTTTCACTGAACAGATGGG	
O1-R	GGTCATCTGTAAGTACAAC	[11]
rstR		
rstRclaf	TTTGCTACTTCTTGGTT	
rstRETF	TGAGCATAAGCTCTTGATT	
rstAR	CCGTGAAAGTCATCAACG	[45]
Virulence gene primers		
VSP-II		
VC0511F1	CTTGCTGCGTACTTAGCA	
VC0511R1	AGTAGCATCGCTCTCGTA	[14]
VC0513F1	CTGAGGTGTTATATGTTTCG	
VC0513R1	TCAAATTTCTGACAGTTCC	[14]
VPI-I		
TCPA1	CACGATAAGAAAACCGTCAAGAG	
TCPA2	ACCAAATGCAACGCCAATGGAGC	[46]
CTX Φ		
CT8	GCAGTCAGGTGGTCTTATTGC	
CT10	TCCAGATATGCAATCCTCAG	[35]
NAG-ST		
ST-1	GAGAAACCTATTATTGCA	
ST-2	GCAAGCTGGATTGCAAC	[47]
T3SS		
vcsV2-F	GGCTCACCAGCTGTTATGGT	
vcsV2-R	CGTATTGCACAAGTAGCCGC	This study
T6SS		
vasH-F	TGTTGATGGGCGAGAGTCAC	
vasH-R	ACGTGTGTGGCAGATACCAG	This study
HlyA		
VC0489F	AGATCAACTACGATCAAGCC	
VC0489R	AGAGGTGCTATGCTTTCTAC	[44]
RTX		
rtxA1	GCGATTCTCAAAGAGATGC	
rtxA2	CACTATTCCGATAACCAC	[44]
Resistance gene primers		
Integron related primers		
INT1 F	AAAACCGCCACTGCGCCGTTA	
INT1 R	GAAGACGGCTGCACTGAACG	[13]
INT2 F	GCGTTTTATGTCTAACAGTCC	
INT2 R	AAGTAGCATCAGTCCATCC	[24]
SXT integrase		
SXTF	TCGGGTATCGCCCAAGGGCA	

Table 2. Cont.

Primer	Sequence (5'-3')	Reference
SXTR	GCGAAGATCATGCATAGACC	[48]
ICE's related genes		
dfrA1 F	CGTGACAGGTTTGCGAATC	
dfrA1 R	ATGGAGTGCCAAAGGTGAAC	This study
strB F	CGTTGCTCCTCTTCTCCATC	
sul2 R	CGTCAACATAACCTCGGACA	This study
FloR F	GTGATTTTTGGTCCGCTCTC	
FloR R	TCGGTAGGATGAAGGTGAGG	This study
Quinolone resistance		
gyrA_vF	AATGTGCTGGCAACGACTGG	
gyrA_vR	GTGCGGATTTTCGACATACG	[49]
gyrB_vF	GGAATGACTCGCCGTAAGG	
gyrB_vR	GTTGTATAACGCAGTTTATCTGGG	[49]
parC_vF	GTCTGAGTTGGTCTCTCGGC	
parC_vR	AGAATCTCGGCAACTTTGACAG	[49]
parE_vF	ATGCGTGCCAGCAAGAAAGTG	
parE_vR	TTATCGCTGCAGGGTCAATCC	[49]
oqxA_F	CTCGGCGCATGATGCT	
oqxA_R	CCACTCTCACGGGAGACGA	[50]
qepA_F	GCAGGTCCAGCAGCGGGTAG	
qepA_R	CTTCGTCGCCGAGTATCGTG	[51]
qnrA_F	TTCAGCAAGAGGATTTCTCA	
qnrA_R	GGCAGCACTATTACTCCAA	[52]
qnrB_F	CCTGAGCGGCACTGAATTTAT	
qnrB_R	GTTTGTCTGCTGCCAGTCGA	[52]
qnrS_F	CAATCATAATATCGGCACC	
qnrS_R	TCAGGATAAACAACAATACCC	[52]
qnrVC_F	AATTTTAAGCGCTCAAACCTCCG	
qnrVC_R	TCCTGTTGCCACGAGCATATTTT	[53]
aac(6')-Ib-cr_F	TTGCGATGCTCTATGAGTGGCTA	
aac(6')-Ib-cr_R	CTCGAATGCCTGGCGTGT	[54]

doi:10.1371/journal.pntd.0002049.t002

The MLSA tree showed three major clusters: (i) O1 El Tor, O139 (MO10 strain) and the O1 clinical Nigeria strains, including the ones from the 1970s (Figure 1), (ii) O1 classical (O395 and RC27) and (iii) non-O1/non-O139 Nigeria strains. The average pairwise difference for the *V. cholerae* concatenated sequences was 1.5%. The divergence between the El Tor/O139 cluster and the classical cluster was 1.2% while the El Tor/O139 cluster and non-O1 cluster was 1.9%. This grouping was also revealed by macrorestriction analysis using representative strains of the MLSA analysis (Figure 2. VC04 and VC200 strains were used as representative classical strains of the 6th pandemic and El Tor VC121 of the beginning of the 7th pandemic from India. The PFGE profiles showed that the Nigeria strains from the beginning of the 7th pandemic (VC869, VC111, VC79) and the recent strains (VC1004, VC832 and VC835) are genetically related and shared a profile with El Tor VC121. The non-O1 strains (VC1006 and VC998) shared a unique profile distinct from the O1 strains.

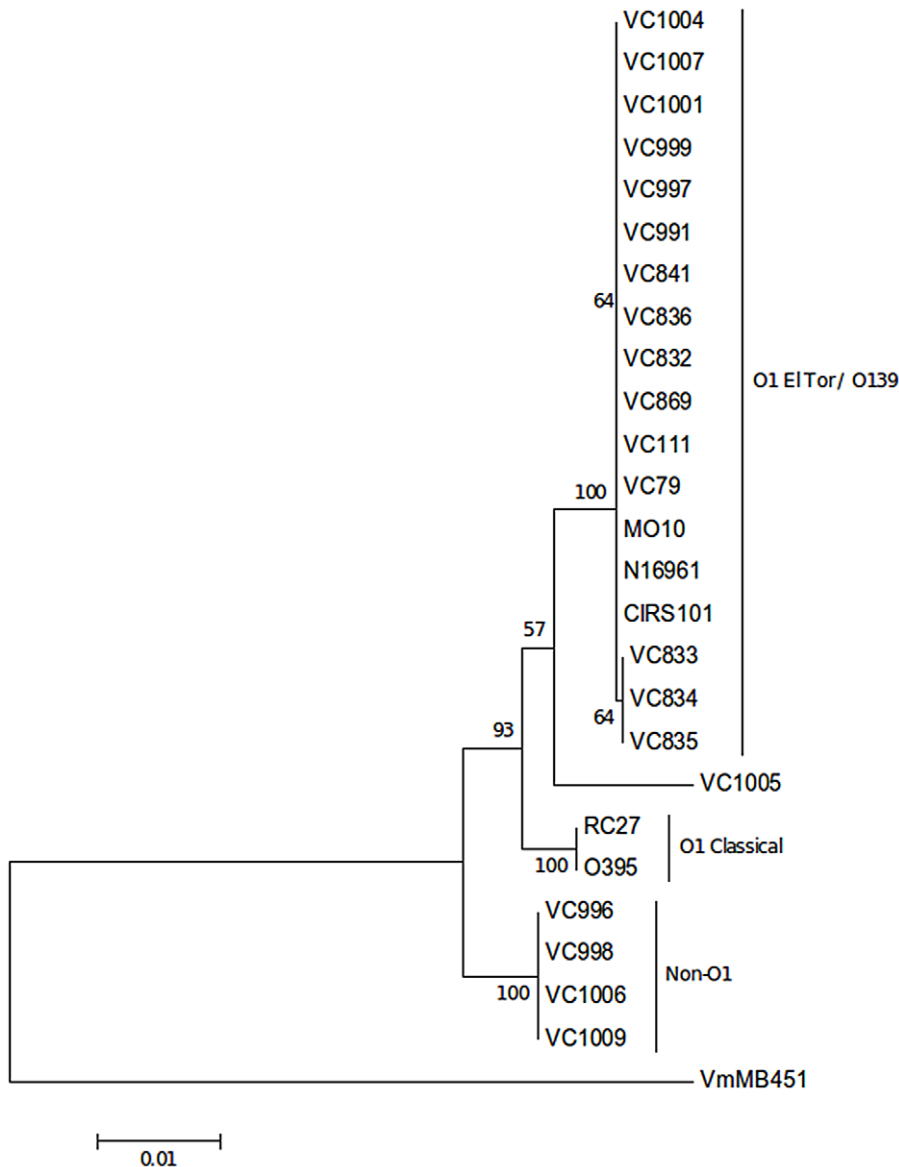


Figure 1. Genetic relatedness of Nigeria *V. cholerae* strains. The epidemic O1 Nigeria strains are within the El Tor/139 cluster and the Non-O1/ Non-O139 strains characterized a distinct cluster. This Minimum Evolution tree was constructed based on Kimura-two-parameter method using the concatenated nucleotide sequences of *recA*, *pyrH* and *mdh* gene fragments. Two thousand bootstrap replicates were performed; bootstrap values are given at the nodes. *Vibrio mimicus* MB451 strain was used as outgroup. doi:10.1371/journal.pntd.0002049.g001

Biotyping and virulence gene typing

PCR-verified serogroup analysis demonstrated that 15 of the 20 strains studied, belonged to serogroup O1 and five were non-O1 isolates (Table 1). *V. cholerae* O1 strains are classified into El Tor, classical and atypical El Tor [2] based on *tcpA* (toxin co-regulated pilin A), *ctxB* (cholera toxin B), *rstR* (repeat sequence transcriptional regulator), as well as variations in VSP-II (Vibrio seventh pandemic island II). The VSP-II gene cluster characterizes the El Tor strains, and comprises a region of ~24 ORFs. In this study, *V. cholerae* O1 strains were genotyped based on sequence information of *rstR*, *ctxB*, *tcpA*, VSP-II regions (locus VC0511 and VC0513).

We screened for the presence of two loci of the VSP-II, VC0511 and VC0513, which are considered the most variable loci in the VSP-II island [14,15]. The Nigeria strains from 1970's and the prototypical N16961, were positive for VC0511 and VC0513 loci,

while the more recently recovered 2009/10 O1 El Tor Nigeria isolates harbor only VC0513 (Table 1), indicating that these lineages carry a different version of VSP-II than earlier isolates. The current non-O1 Nigeria strains were negative for both, VC0511 and VC0513 loci.

The *ctxB* present in all Nigeria current O1 El Tor strains belong to the atypical genotype 7, characterized by the non conservative change His20Asn [16]. El Tor strains from the beginning of the 7th pandemic in Nigeria, as worldwide, were characterized by the canonical El Tor *ctxB* genotype 3 and keep their own repressor *rstR*^{ElTor} gene of the CTX prophage.

Concerning the *tcpA* alleles, there are *tcpA*^{ElTor}, *tcpA*^{classical} and *tcpA*^{NonO1} disseminated among *V. cholerae* strains. All current O1 Nigeria strains harbor *tcpA*^{CIRS} allele (Table 1), recently described in the CIRS101 strain, a *V. cholerae* O1 El Tor from Bangladesh,

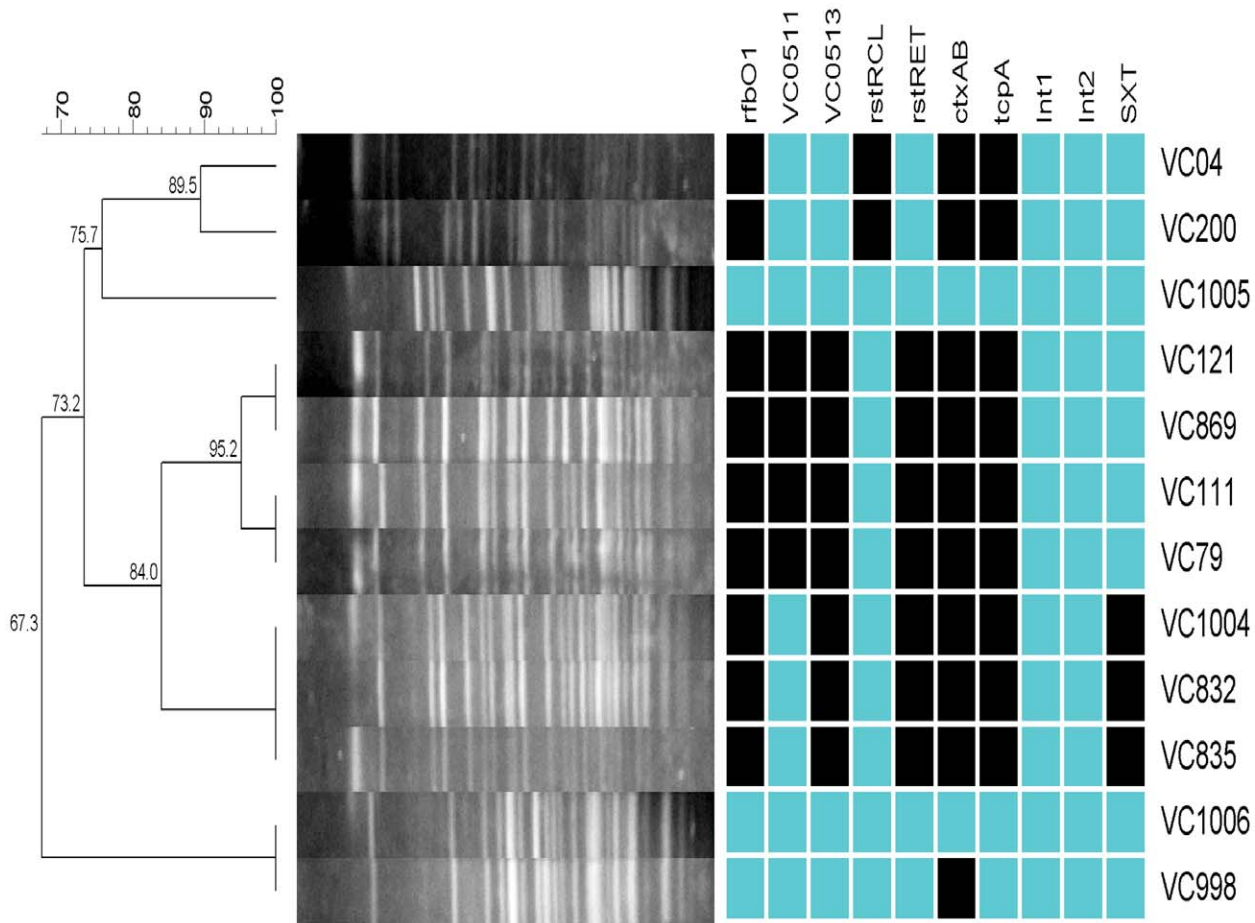


Figure 2. Dendrogram of PFGE patterns of representative *V. cholerae* O1 and Non-O1/Non-O139 strains. The similarity was determined using the Dice coefficient, and cluster analysis was performed using the UPGMA. The right panel shows the presence (dark box) and absence (light box) of genomic traits. VC04 and VC200 belong to the classical lineage; VC121 is an India strain from the beginning of the 7th cholera pandemic. doi:10.1371/journal.pntd.0002049.g002

2002 [17]. Based on these genotypes, the current epidemic *V. cholerae* O1 strains from Nigeria are classified as atypical El Tor biotype.

Although, the current non-O1/non-O139 Nigeria strains were negative for *rstR*, *ctxB*, *tcpA*, VSP-II regions (locus VC0511 and VC0513) as well NAG-ST and T3SS (*vasV2* gene), they are positive for T6SS (*vasH* gene), *hlyA* and *rtxA* genes, all virulence associated factors involved in diarrheogenic property [18–21]. Therefore, a non-O1/non-O139 lineage harboring a set of virulence determinants could be associated with a cholera-like diarrhea [22].

Characterization of elements associated with antibiotic resistance

Antibiotic multidrug resistance is becoming increasingly common among the atypical *V. cholerae* strains, mostly associated with acquisition of genes and/or modification in the antibiotic target genes [23]. According to our results, the current O1 Nigeria strains were resistant to streptomycin, trimethoprim and sulfonamides. In *V. cholerae*, these resistances, are frequently associated with class 1 and 2 integrons [24] and SXT element, which is a *V. cholerae*-derived ICE (integrating and conjugative element) [25]. Thus, we investigated the presence of these elements in the Nigeria strains. All the current O1 strains harbor an ICE element, determined by the

presence of the SXT integrase gene, contrasting with non-O1 and the O1 from the 1970’s (Table 1). No evidence was found for the presence of class 1 and 2 integrons (which we sought by PCR for *intI1* and *intI2*, respectively. Table 1). We performed PCR targeting ICE associated genes *floR*, *sul2*, *dfxA1*, and *strAB*, associated with chloramphenicol, sulfamethoxazole, trimethoprim and streptomycin resistance, respectively [7,26]. All these genes were identified, explaining the resistance profile, including reduced susceptibility to chloramphenicol, of the current O1 strains.

Majority of the 2009/10 Nigeria O1 strains showed reduced susceptibility to ciprofloxacin (Table 1) as well as resistance to nalidixic acid. Quinolone resistance has been attributed to mutations in *gyrA* (Ser83Ile) and in *parC* (Ser85Leu) [26,27] and these alleles were found in these strains. We additionally screened the 2009/10 isolates for horizontally transmitted quinolone-resistance genes *qnrA*, *qnrB*, *qnrS*, *qnrVC*, *qepA*, *oqxAB* and *aac(6’)-Ib-cr*. The isolates were all negative for all of these targets. None of these resistance phenotypes or genotypes detected in the 2009/10 isolates were found in strains isolated in Nigeria during the 1970’s. Additionally, resistance phenotype and genes were not seen in the 2009/10 non-O1/non-O139 isolates, with the exception of the environmental non-O1/non-O139 VC1005 strain, which showed resistance to nalidixic acid attributable to mutations in *gyrA* and *parC* (Table 1).

Discussion

Since 1995 and over a period of more than a decade the canonical El Tor has been replaced in Kolkata, India by atypical El Tor. They were originally described in South Asia [2,28,29] but recent reports have shown their spread to all continents, reaching even Mexico [30] and Haiti [26]. Based only on *ctxB* genotyping, atypical El Tor have already been reported in some African countries, where cholera has become a serious public health threat in recent years [7,8,31–35].

Despite the great impact from cholera, few investigators focused on the molecular epidemiology of *V. cholerae* in West Africa. A recent genomic survey of 154 isolates worldwide did not include any isolates from West Africa [36]. Due to the plasticity of *V. cholerae* resulting in the constant emergence of variants, surveillance and characterization of outbreak strains, and their antibiotic resistance determinants, is essential to defining the complex scenario of cholera in this continent as well as worldwide. Recently, it was shown that different *V. cholerae* O1 lineages were responsible for cholera outbreaks in Ghana [35]. Here, we provide evidence that different *V. cholerae* lineages are driving cholera outbreaks in Nigeria.

Molecular typing of 20 *V. cholerae* strains from Nigeria, allowed us to identify atypical O1 strains as well as a non-O1/non-O139 lineage. To date, there are only two studies performing molecular characterization of *V. cholerae* strains causing cholera in Nigeria and very few microbiological outbreak investigations [8,17]. Both molecular studies focused on the allelic characterization of the two major virulence genes and one of them addressed also the quinolone resistance determinants. Talkington *et al*, who analyzed two 2008 isolates from Nigeria, reported two profiles: *ctxB-1/tcpA^{ElTor}/rstR^{ElTor}* and *ctxB-1/tcpA^{CIRS}/rstR^{ElTor}* [17]. Quilici and collaborators analyzing ten strains from 2009 showed a unique allelic profile: *ctxB-7/tcpA^{ElTor}/rstR^{ElTor}* [8]. They reported that these isolates were identical to nine isolates from neighboring Cameroon, which also borders the Lake Chad basin. These isolates are temporally and geographically proximal to the 2009 Maiduguri outbreak from which we obtained isolates VC832 and VC841 for our study. In our study, considering these set of genes, the 12 *V. cholerae* O1 strains, from 2009/2010 cholera outbreaks including the 2009 Maiduguri outbreak, presented a new combination of alleles: *ctxB-7/tcpA^{CIRS}/rstR^{ElTor}*. Interestingly, a strain carrying same combination of *ctxB-7* and *tcpA^{CIRS}* alleles, as the profile identified in this work, was characterized causing cholera in Cameroon and Haiti in 2010 [17]. Strains from the 2010 Bauchi outbreak in North Central Nigeria and the 2010 Ile-Ife outbreak, both evaluated in this study, were also predominantly of the *ctxB-7* and *tcpA^{CIRS}* genotype. Therefore it is likely that these four genotypes were co-circulating in the Chad basin at that time [8].

In the Quilici *et al* study [8], *V. cholerae* strains showed intermediate susceptibility to chloramphenicol, reduced susceptibility to ciprofloxacin and resistance to trimethoprim/sulfamethoxazole, sulfonamides and nalidixic acid, and the *gyrA^{Ser83Ile}/parC^{Ser85Leu}* alleles related to quinolone resistance. This same resistance phenotypic and

genotypic profile was found in isolates from all three outbreaks represented in this study (Table 1). Moreover, we found that the 2009/10 O1 Nigeria strains carried *sul2*, *dfrA1* and *floR* genes, conferring resistance to sulfonamides, trimethoprim, and intermediate susceptibility to chloramphenicol, respectively, and *strA/B*, which confers to streptomycin resistance. These genes have been associated with integrative and conjugative elements (ICEs) and the 2009/10 O1 Nigeria strains were positive for the SXT integrase, indicating that these strains harbor an ICE. ICEs have been identified in *V. cholerae* strains from Ghana [24], Kenya [36], Mozambique [37] and Angola [7]. In contrast to the 2009/10 isolates, the Nigeria strains from the beginning of the 7th cholera pandemic had the canonical El Tor allelic profile *ctxB-3/tcpA^{ElTor}/gyrA^{Ser83}/parC^{Ser85}* and lacked most of the resistance genes.

Interestingly, as reported elsewhere [38], we found that a lineage of *V. cholerae* non-O1/non-O139, lacking the major virulence determinants and the resistance phenotype described above, but harboring a set of virulence associated genes was co-circulating with atypical El Tor strains being recovered from patients with cholera-like disease in Bauchi, Nigeria in 2010. This outbreak followed the Maiduguri outbreak and were accompanied by a slightly improved public engagement response. It is probable that individuals with cholera-like disease are more likely to present to health facilities when they have received information about an on-going cholera outbreak. Our findings suggest that non-O1/non-O139 *V. cholerae* may be important causes of endemic diarrheal disease and outbreaks. Recovery of a similar strain from water tested in Ile-Ife supports this possibility.

Population structure of clinical *V. cholerae* strains can be influenced by epidemiological factors like transmission dynamics, clonal expansion during epidemics, human travel, and selective pressure from antimicrobial drugs. The data in our study showed that recent outbreaks in Nigeria are caused by multidrug resistant atypical El Tor O1 strains, which are reportedly highly virulent [8,39–42], a common worrisome trend in the current cholera outbreaks around the world. The data suggest that guidelines for managing and containing cholera outbreaks in Nigeria (which include, in addition to rehydration, using the antimicrobials trimethoprim, and more recently ciprofloxacin) need to be urgently revised. The data also show that *V. cholerae* non-O1/non-O139 strains were involved in outbreaks in Nigeria suggesting that interventions to control epidemic cholera such as improvements in water supplies and sanitation, as well as vaccination, may result in broad gains.

Acknowledgments

We thank Babatunde W Odetoyn for excellent technical assistance.

Author Contributions

Conceived and designed the experiments: MAM ACPV. Performed the experiments: MAM CCT FSF ELF AOA SBZ NKEQ. Analyzed the data: MAM CCT ELF ACPV INO. Contributed reagents/materials/analysis tools: ACPV AOA. Wrote the paper: MAM CCT ELF INO ACPV.

References

- Salim A, Lan R, Reeves PR (2005) Pathogenic Clones. *Emerging Infectious Diseases* 11: 1758–1760.
- Safa A, Nair GB, Kong RYC (2010) Evolution of new variants of *Vibrio cholerae* O1. *Trends in microbiology* 18: 46–54. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19942436>. Accessed 25 April 2011.
- Sack D a, Sack RB, Nair GB, Siddique a K (2004) Cholera. *Lancet* 363: 223–233. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19678971>.
- Waldor MK, Mekalanos JJ (1996) Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science (New York, NY)* 272: 1910–1914. Available: <http://www.ncbi.nlm.nih.gov/pubmed/8658163>.
- Karaolis DKR, Johnson JA, Bailey CC, Boedeker EC, Kaper JB, et al. (1998) A *Vibrio cholerae* pathogenicity island associated with epidemic and pandemic strains. *Proceedings of the National Academy of Sciences of the United States of America* 95: 3134. Available: <http://www.pnas.org/content/95/6/3134.short>. Accessed 8 April 2011.
- Kumar P, Thulaseedharan A, Chowdhury G, Ramamurthy T, Thomas S (2011) Characterization of novel alleles of toxin co-regulated pilus A gene (*tcpA*) from environmental isolates of *Vibrio cholerae*. *Current microbiology* 62: 758–763. Available: <http://www.ncbi.nlm.nih.gov/pubmed/20967447>. Accessed 19 July 2011.

7. Ceccarelli D, Spagnoletti M, Bacciu D, Cappuccinelli P, Colombo MM (2011) New V. cholerae atypical El Tor variant emerged during the 2006 epidemic outbreak in Angola. *BMC microbiology* 11: 130. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3131240&tool=pmcentrez&rendertype=abstract>. Accessed 18 July 2011.
8. Quilici ML, Massenot D, Gake B, Bwaki B, Olson DM (2010) Vibrio cholerae O1 variant with reduced susceptibility to ciprofloxacin, Western Africa. *Emerging infectious diseases* 16: 1804–1805. Available: <http://www.cdc.gov/eid/content/16/11/1804.htm>. Accessed 9 July 2012.
9. Piarroux R, Faucher B (2012) Cholera epidemics in 2010: respective roles of environment, strain changes, and human-driven dissemination. *Clinical microbiology and infection: the official publication of the European Society of Clinical Microbiology and Infectious Diseases* 18: 231–238. Available: <http://www.ncbi.nlm.nih.gov/pubmed/22288560>.
10. Adagbada AO, Adesida SA, Nwaokorie FO, Niemogha M-T, Coker AO (2012) Cholera Epidemiology in Nigeria: an overview. *The Pan African medical journal* 12: 59. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3428179&tool=pmcentrez&rendertype=abstract>. Accessed 3 October 2012.
11. Hoshino K, Yamasaki S, Mukhopadhyay A K, Chakraborty S, Basu A, et al. (1998) Development and evaluation of a multiplex PCR assay for rapid detection of toxigenic Vibrio cholerae O1 and O139. *FEMS immunology and medical microbiology* 20: 201–207. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9566491>.
12. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular biology and evolution*. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21546353>. Accessed 18 July 2011.
13. Fonseca EL, Vieira V V, Cipriano R, Vicente ACP (2006) Emergence of dhfrXVb and blaCARB-4 gene cassettes in class 1 integrons from clinical Pseudomonas aeruginosa isolated in Amazon region. *Memórias do Instituto Oswaldo Cruz* 101: 81–84. Available: <http://www.ncbi.nlm.nih.gov/pubmed/16699713>. Accessed 2 May 2011.
14. Nusrin S, Gil AI, Bhuiyan NA, Safa A, Asakura M, et al. (2009) Peruvian Vibrio cholerae O1 El Tor strains possess a distinct region in the Vibrio seventh pandemic island-II that differentiates them from the prototype seventh pandemic El Tor strains. *Journal of medical microbiology* 58: 342–354. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19208885>. Accessed 13 December 2010.
15. Taviani E, Grim CJ, Choi J, Chun J, Haley B, et al. (2010) Discovery of novel Vibrio cholerae VSP-II genomic islands using comparative genomic analysis. *FEMS microbiology letters* 308: 130–137. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2925232&tool=pmcentrez&rendertype=abstract>. Accessed 27 April 2011.
16. Marin MA, Vicente ACP (2011) Variants of Vibrio cholerae O1 El Tor from Zambia showed new genotypes of ctxB. *Epidemiology and infection*: 1–2. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21939579>. Accessed 27 September 2011.
17. Talkington D, Bopp C, Tarr C, Parsons MB, Dahourou G, et al. (2011) Characterization of Toxigenic Vibrio cholerae from Haiti, 2010–2011. *Emerging infectious diseases* 17: 2122–2129. Available: <http://www.ncbi.nlm.nih.gov/pubmed/22099116>.
18. Chen Y, Johnson J A, Pusch GD, Morris JG, Stine OC (2007) The genome of non-O1 Vibrio cholerae NRT36S demonstrates the presence of pathogenic mechanisms that are distinct from those of O1 Vibrio cholerae. *Infection and immunity* 75: 2645–2647. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1865779&tool=pmcentrez&rendertype=abstract>. Accessed 8 March 2012.
19. Shin OS, Tam VC, Suzuki M, Ritchie JM, Bronson RT, et al. (2011) Type III Secretion Is Essential for the Rapidly Fatal Diarrheal Disease Caused by Non-O1, Non-O139 Vibrio cholerae. *mBio* 2: 1–11. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3111608&tool=pmcentrez&rendertype=abstract>. Accessed 26 July 2011.
20. Ma AT, Mekalanos JJ (2010) In vivo actin cross-linking induced by Vibrio cholerae type VI secretion system is associated with intestinal inflammation. *Proceedings of the National Academy of Sciences of the United States of America* 107: 4365–4370. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2840160&tool=pmcentrez&rendertype=abstract>. Accessed 13 November 2012.
21. Chatterjee S, Ghosh K, Raychoudhuri A, Chowdhury G, Bhattacharya MK, et al. (2009) Incidence, virulence factors, and clonality among clinical strains of non-O1, non-O139 Vibrio cholerae isolates from hospitalized diarrheal patients in Kolkata, India. *Journal of clinical microbiology* 47: 1087–1095. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2668327&tool=pmcentrez&rendertype=abstract>. Accessed 26 November 2012.
22. Dalsgaard A, Albert MJ, Taylor DN, Shimada T, Meza R, et al. (1995) Characterization of Vibrio cholerae non-O1 serogroups obtained from an outbreak of diarrhea in Lima, Peru. *Journal of clinical microbiology* 33: 2715–2722. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=228562&tool=pmcentrez&rendertype=abstract>. Accessed 17 December 2012.
23. Okeke IN, Aboderin O A, Byarugaba DK, Ojo KK, Opintan J A (2007) Growing problem of multidrug-resistant enteric pathogens in Africa. *Emerging infectious diseases* 13: 1640–1646. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3375797&tool=pmcentrez&rendertype=abstract>.
24. Opintan J A, Newman MJ, Nsiah-Poodoh OA, Okeke IN (2008) Vibrio cholerae O1 from Accra, Ghana carrying a class 2 integron and the SXT element. *The Journal of antimicrobial chemotherapy* 62: 929–933. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2566517&tool=pmcentrez&rendertype=abstract>. Accessed 16 April 2012.
25. Waldor MK, Tschäpe H, Mekalanos JJ (1996) A new type of conjugative transposon encodes resistance to sulfamethoxazole, trimethoprim, and streptomycin in Vibrio cholerae O139. *Journal of bacteriology* 178: 4157–4165. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=178173&tool=pmcentrez&rendertype=abstract>.
26. Sjölund-Karlsson M, Reimer A, Folster JP, Walker M, Dahourou GA, et al. (2011) Drug resistance mechanisms in Vibrio cholerae O1 outbreak strain, Haiti, 2010. *Emerg Infect Dis* 17: 2151–2154. Available: http://wwwnc2.cdc.gov/eid/pdfs/11-0720-ahead_of_print.pdf. Accessed 12 December 2011.
27. Kim H Bin, Wang M, Ahmed S, Park CH, LaRocque RC, et al. (2010) Transferable quinolone resistance in Vibrio cholerae. *Antimicrobial agents and chemotherapy* 54: 799–803. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2812163&tool=pmcentrez&rendertype=abstract>. Accessed 8 July 2011.
28. Ang GY, Yu CY, Balqis K, Elina HT, Azura H, et al. (2010) Molecular evidence of cholera outbreak caused by a toxigenic Vibrio cholerae O1 El Tor variant strain in Kelantan, Malaysia. *Journal of clinical microbiology* 48: 3963–3969. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3020861&tool=pmcentrez&rendertype=abstract>. Accessed 28 July 2011.
29. Nair GB, Qadri F, Holmgren J, Svennerholm A-M, Safa A, et al. (2006) Cholera due to altered El Tor strains of Vibrio cholerae O1 in Bangladesh. *Journal of clinical microbiology* 44: 4211–4213. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1698305&tool=pmcentrez&rendertype=abstract>. Accessed 18 July 2011.
30. Alam M, Nusrin S, Islam A, Bhuiyan N A, Rahim N, et al. (2010) Cholera between 1991 and 1997 in Mexico was associated with infection by classical, El Tor, and El Tor variants of Vibrio cholerae. *Journal of clinical microbiology* 48: 3666–3674. Available: <http://www.ncbi.nlm.nih.gov/pubmed/20668130>. Accessed 14 October 2010.
31. Mintz ED, Guerrant RL (2009) A lion in our village—the unconscionable tragedy of cholera in Africa. *The New England journal of medicine* 360: 1060–1063. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19279337>. Accessed 3 October 2012.
32. Ansaruzzaman M, Bhuiyan NA, Nair BG, Sack DA, Lucas M, et al. (2004) Cholera in Mozambique, variant of Vibrio cholerae. *Emerging infectious diseases* 10: 2057–2059. Available: <http://www.ncbi.nlm.nih.gov/pubmed/16010751>. Accessed 18 July 2011.
33. Islam MS, Mahmud ZH, Ansaruzzaman M, Faruque SM, Talukder K A, et al. (2011) Phenotypic, Genotypic and Antibiotic Sensitivity Patterns of Strains Isolated from the Cholera Epidemic in Zimbabwe. *Journal of clinical microbiology*: 1–13. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21471347>. Accessed 19 April 2011.
34. Bhuiyan NA, Nusrin S, Ansaruzzaman M, Islam A, Sultana M, et al. (2011) Genetic characterization of Vibrio cholerae O1 strains isolated in Zambia during 1996–2004 possessing the unique VSP-II region of El Tor variant. *Epidemiology and infection*: 1–9. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21676349>. Accessed 18 July 2011.
35. Thompson CC, Freitas FS, Marin M A, Fonseca EL, Okeke IN, et al. (2011) Vibrio cholerae O1 lineages driving cholera outbreaks during seventh cholera pandemic in Ghana. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases* 11: 1951–1956. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21896336>. Accessed 16 September 2011.
36. Mutreja A, Kim DW, Thomson NR, Connor TR, Lee JH, et al. (2011) Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature*. Available: <http://www.nature.com/doi/10.1038/nature10392>. Accessed 24 August 2011.
37. Taviani E, Grim CJ, Chun J, Huq A, Colwell RR (2009) Genomic analysis of a novel integrative conjugative element in Vibrio cholerae. *FEBS letters* 583: 3630–3636. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19850044>. Accessed 27 April 2011.
38. Hasan N A, Choi SY, Eppinger M, Clark PW, Chen A, et al. (2012) Genomic diversity of 2010 Haitian cholera outbreak strains. *Proceedings of the National Academy of Sciences of the United States of America*: 1–8. Available: <http://www.pnas.org/cgi/doi/10.1073/pnas.1207359109>. Accessed 26 June 2012.
39. Ghosh-Banerjee J, Senoh M, Takahashi T, Hamabata T, Barman S, et al. (2010) Cholera toxin production by the El Tor variant of Vibrio cholerae O1 compared to prototype El Tor and classical biotypes. *Journal of clinical microbiology* 48: 4283–4286. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3020813&tool=pmcentrez&rendertype=abstract>. Accessed 8 July 2011.
40. Ceccarelli D, Spagnoletti M, Cappuccinelli P, Burrus V, Colombo MM (2011) Origin of Vibrio cholerae in Haiti. *The Lancet infectious diseases* 11: 262. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21453867>. Accessed 18 July 2011.

41. Ryan ET (2011) The Cholera Pandemic, Still with Us after Half a Century: Time to Rethink. *PLoS Neglected Tropical Diseases* 5: e1003. Available: <http://dx.plos.org/10.1371/journal.pntd.0001003>. Accessed 26 January 2011.
42. Siddique a K, Nair GB, Alam M, Sack D a, Huq a, et al. (2010) El Tor cholera with severe disease: a new threat to Asia and beyond. *Epidemiology and infection* 138: 347–352. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19678971>. Accessed 26 June 2012.
43. Thompson CC, Thompson FL, Vicente ACP (2008) Identification of *Vibrio cholerae* and *Vibrio mimicus* by multilocus sequence analysis (MLSA). *International journal of systematic and evolutionary microbiology* 58: 617. Available: <http://ijsb.sgmjournals.org/cgi/content/abstract/58/3/617>. Accessed 25 April 2011.
44. O'Shea YA, Reen FJ, Quirke AM, Boyd EF (2004) Evolutionary genetic analysis of the emergence of epidemic *Vibrio cholerae* isolates on the basis of comparative nucleotide sequence analysis and multilocus virulence gene profiles. *Journal of clinical microbiology* 42: 4657–4671. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=522369&tool=pmcentrez&rendertype=abstract>. Accessed 12 October 2010.
45. Lee JH, Choi SY, Jeon Y-S, Lee HR, Kim EJ, et al. (2009) Classification of hybrid and altered *Vibrio cholerae* strains by CTX prophage and RS1 element structure. *Journal of microbiology (Seoul, Korea)* 47: 783–788. Available: <http://www.ncbi.nlm.nih.gov/pubmed/20127474>. Accessed 25 April 2011.
46. Keasler SP, Hall RH (1993) Detecting and biotyping *Vibrio cholerae* O1 with multiplex polymerase chain reaction. *Lancet* 341: 1661. Available: <http://www.ncbi.nlm.nih.gov/pubmed/8100020>. Accessed 5 August 2011.
47. Vicente AC, Coelho AM, Salles CA (1997) Detection of *Vibrio cholerae* and *V. mimicus* heat-stable toxin gene sequence by PCR. *Journal of medical microbiology* 46: 398–402. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9152035>. Accessed 17 December 2012.
48. Bhanumathi R, Sabcena F, Isac SR, Shukla BN, Singh D V (2003) Molecular characterization of *Vibrio cholerae* O139 bengal isolated from water and the aquatic plant *Eichhornia crassipes* in the River Ganga, Varanasi, India. *Applied and environmental microbiology* 69: 2389–2394. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=154771&tool=pmcentrez&rendertype=abstract>. Accessed 3 March 2011.
49. Baranwal S, Dey K, Ramamurthy T, Nair GB, Kundu M (2002) Role of active efflux in association with target gene mutations in fluoroquinolone resistance in clinical isolates of *Vibrio cholerae*. *Antimicrobial agents and chemotherapy* 46: 2676–2678. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=127385&tool=pmcentrez&rendertype=abstract>. Accessed 28 July 2011.
50. Kim H Bin, Wang M, Park CH, Kim E-C, Jacoby GA, et al. (2009) *oqxAB* encoding a multidrug efflux pump in human clinical isolates of *Enterobacteriaceae*. *Antimicrobial agents and chemotherapy* 53: 3582–3584. Available: <http://aac.asm.org/cgi/content/long/53/8/3582>. Accessed 3 October 2012.
51. Liu J-H, Deng Y-T, Zeng Z-L, Gao J-H, Chen L, et al. (2008) Coprevalence of plasmid-mediated quinolone resistance determinants *QepA*, *Qnr*, and *AAC(6')-Ib-cr* among 16S rRNA methylase *RmtB*-producing *Escherichia coli* isolates from pigs. *Antimicrobial agents and chemotherapy* 52: 2992–2993. Available: <http://www.ncbi.nlm.nih.gov/pubmed/2493129>. Accessed 3 October 2012.
52. Wu J-J, Ko W-C, Tsai S-H, Yan J-J (2007) Prevalence of plasmid-mediated quinolone resistance determinants *QnrA*, *QnrB*, and *QnrS* among clinical isolates of *Enterobacter cloacae* in a Taiwanese hospital. *Antimicrobial agents and chemotherapy* 51: 1223–1227. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1855486&tool=pmcentrez&rendertype=abstract>. Accessed 3 October 2012.
53. Kim H Bin, Wang M, Ahmed S, Park CH, LaRocque RC, et al. (2010) Transferable quinolone resistance in *Vibrio cholerae*. *Antimicrobial agents and chemotherapy* 54: 799–803. Available: <http://aac.asm.org/cgi/content/long/54/2/799>. Accessed 3 October 2012.
54. Park CH, Robicsck A, Jacoby GA, Sahn D, Hooper DC (2006) Prevalence in the United States of *aac(6')-Ib-cr* encoding a ciprofloxacin-modifying enzyme. *Antimicrobial agents and chemotherapy* 50: 3953–3955. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1635235&tool=pmcentrez&rendertype=abstract>. Accessed 3 October 2012.



RESEARCH ARTICLE

Architecture of the superintegron in *Vibrio cholerae*: identification of core and unique genes [v1; ref status: approved 1, <http://f1000r.es/w6>]

Michel A Marin, Ana Carolina P Vicente

Laboratory of Molecular Genetics of Microorganisms, Oswaldo Cruz Institute (IOC), Rio de Janeiro, 4365, PO Box 926 CEP 21045-900, Brazil

v1 First published: 27 Feb 2013, 2:63 (doi: 10.3410/f1000research.2-63.v1)
 Latest published: 27 Feb 2013, 2:63 (doi: 10.3410/f1000research.2-63.v1)

Abstract


Background: *Vibrio cholerae*, the etiologic agent of cholera, is indigenous to aquatic environments. The *V. cholerae* genome consists of two chromosomes; the smallest of these harbors a large gene capture and excision system called the superintegron (SI), of ~120 kbp. The flexible nature of the SI that results from gene cassette capture, deletion and rearrangement is thought to make it a hotspot of *V. cholerae* diversity, but beyond the basic structure it is not clear if there is a core genome in the SI and if so how it is structured. The aim of this study was to explore the core genome structure and the differences in gene content among strains of *V. cholerae*.

Methods: From the complete genomes of seven *V. cholerae* and one *Vibrio mimicus* representative strains, we recovered the SI sequences based on the locations of the structural gene *IntI4* and the *V. cholerae* repeats. Analysis of the pangenome, including cluster analysis of functional genes, pangenome profile analysis, genetic variation analysis of functional genes, strain evolution analysis and function enrichment analysis of gene clusters, was performed using a pangenome analysis pipeline in addition to the R scripts, splitsTree4 and genoPlotR.

Results and conclusions: Here, we reveal the genetic architecture of the *V. cholerae* SI. It contains eight core genes when *V. mimicus* is included and 21 core genes when only *V. cholerae* strains are considered; many of them are present in several copies. The *V. cholerae* SI has an open pangenome, which means that *V. cholerae* may be able to import new gene cassettes to SI. The set of dispensable SI genes is influenced by the niche and type species. The core genes are distributed along the SI, apparently without a position effect.

Article Status Summary

Referee Responses

Referees	1
v1 published 27 Feb 2013	 report

1 **Thandavarayan Ramamurthy**, National Institute of Cholera and Enteric Diseases India

Latest Comments

No Comments Yet

Corresponding author(s): Michel A Marin (mfabanto@ioc.fiocruz.br)

How to cite this article: Marin MA, Vicente ACP (2013) Architecture of the superintegron in *Vibrio cholerae*: identification of core and unique genes [v1; ref status: approved 1, <http://f1000r.es/w6>] *F1000Research* 2013, 2:63 (doi: 10.3410/f1000research.2-63.v1)

Copyright: This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

Grant information: This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) doctoral fellowship to MAM. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: No competing interests were disclosed.

First published: 27 Feb 2013, 2:63 (doi: 10.3410/f1000research.2-63.v1)

First Approved: 27 Feb 2013, 2:63 (doi: 10.3410/f1000research.2-63.v1)

Introduction

Vibrio cholerae is a diverse, environmental, gram-negative bacterial species that can be pathogenic and can cause cholera, a severe diarrheal disease that occurs most frequently in epidemic form^{1,2}. The *V. cholerae* genome consists of two chromosomes. The largest chromosome of 2.96 Mbp encodes most essential genes. The 1.07 Mbp small chromosome contains few essential genes and the superintegron (SI), a large gene capture and excision system of ~120 kbp² (Figure 1). The SI is characterized by a site-specific integrase gene (*IntI4*) closely associated with a cognate recombination site *attI* and a promoter Pc followed by a large array of gene cassettes. Within the SI, the gene cassettes generally consist of a promoterless open reading frame (ORF) flanked by two recombination sites termed *V. cholerae* repeats (VCRs)³. Cassettes can be excised from any position in the array through VCR × VCR recombination mediated by the integrase. The resulting circular intermediate can then be integrated, preferentially through *attI* × VCR recombination by the integrase, bringing the cassette under control of Pc^{4,5}. Since gene cassettes are usually promoterless, only the first few cassettes are expressed by Pc and the rest of the array can be seen as a reservoir of standing genetic variation⁵.

The functions of the majority of the SI genes are unknown; however, a few genes have been characterized and it has been suggested that they are involved in adaptive functions such as toxin-antitoxin (TA) loci. TA loci consist of two genes in an operon encoding a ‘toxin’ and an ‘antitoxin’. The expression of the toxins reduces cell growth and prevents colony formation, thus exerting a bacteriostatic rather than bacteriocidal condition. However, cell viability can be rescued by later overproduction of the cognate antitoxins⁶.

The pangenome describes the complete repertoire of genes in a bacterial species, which includes the “core genome” containing genes present in all strains, a “dispensable genome” containing genes present in two or more strains, and “unique genes”

specific to single strains⁷. Previous phylogeographic analysis, considering *V. cholerae* strains and its sister species *Vibrio metecus*⁸, showed that, in contrast to the core genome, the SI displays strong geographical differentiation, and cassettes from the *V. cholerae* group cluster with those of *V. metecus* from the same place rather than with cassettes from geographically distinct *V. cholerae*. It suggested that SI structure is influenced by geographic boundaries and in response to environmental conditions. The flexible nature of the SI that results from gene cassette capture, deletion and rearrangement is thought to make it a hotspot of *V. cholerae* diversity, but beyond the basic structure it is not clear if there is a core genome in the SI and if so how it is structured. The aim of this work was to explore the core genome structure and the differential gene content among strains of *V. cholerae*.

Methodology

Based on the complete genomes of seven *V. cholerae* and one *V. mimicus* representative strains (Table 1), we searched repeats above 10 nucleotides and used one VCR sequence (AAC AAA CGC CTC AAG AGG GAC TGT CAA CGC GTG GCG TTT CCA GTC CCA TTG AGC CGT GGT GGT TTC GGT TGT TGT GTT TGA GTT TAG TGT TAT GCG TTG TCA GCC CCT TAG GCG GGC G) to search for sequences with more of 45% nucleotide identity. The SI sequences were recovered using the locations of the structural gene *IntI4* and VCRs identified with the UGENE software⁹. Cluster analysis of functional genes was performed using the pangenome analysis pipeline¹⁰, which searches for homologs or orthologs among multiple genomes using the MultiParanoid (MP) method (based on a 90% nucleotide identity threshold). For each pair of genes in the same cluster, the local matched region is no less than 25% of the longer protein coding sequence and the global matched region is no less than 50% of the longer protein coding sequence. The minimum score value and E-value in BLAST are 50 and 1e-8¹⁰. The gene content was converted to a presence/absence (0/1) matrix and then the

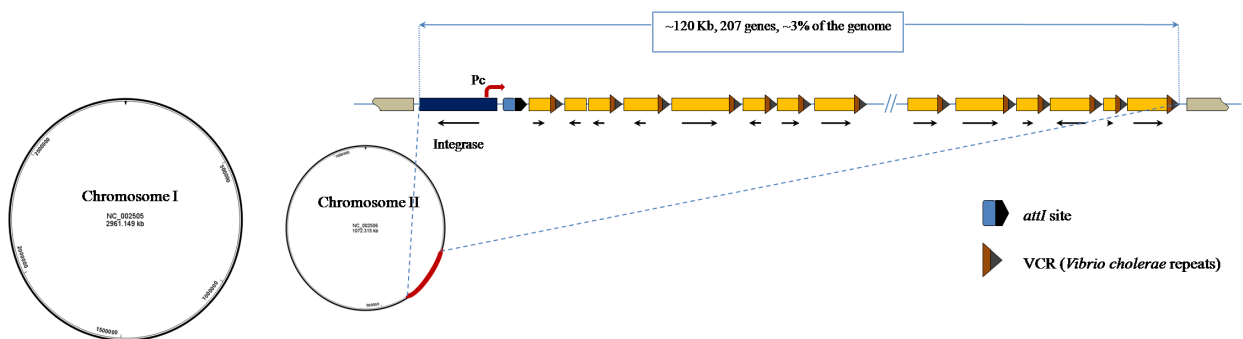


Figure 1 Schematic organization of the *Vibrio cholerae* genome and the superintegron (SI). The functional platform of the SI consists of an integrase gene, a cassette promoter (Pc), and a primary recombination site (*attI*). The system maintains an array of several cassettes, which generally consist of a promoterless ORF flanked by two recombination sites termed VCR (*V. cholerae* repeats).

Table 1 Superintegron regions extracted from *V. cholerae* and *V. mimicus* genomes.

Organism	Serogroup/ Biotype	Geographical origin	Source of isolation	Year of isolation	Start	End*	Size (bp)	G+C(%)	ORFs	Locus <i>IntI4</i>	Accession in NCBI
<i>V. cholerae</i> N16961	O1 El Tor	Bangladesh	Clinical	1975	309750	435418	125669	42.20	166	VCA0291	NC_002506
<i>V. cholerae</i> 2010EL1786	O1 El Tor	Haiti	Clinical	2010	36195	135658	99464	42.08	138	Vch1786_I10037	NC_016446
<i>V. cholerae</i> MJ-1236	O1 El Tor	Matlab, Bangladesh	Clinical	1994	931735	1050596	118862	41.46	135	VCD_000984	NC_012667
<i>V. cholerae</i> O395	O1 Classical	India	Clinical	1965	799827	916350	116524	41.35	175	VCO395_0938	NC_009456
<i>V. cholerae</i> LMA3984	O1	Para, Brazil	Environmental	2007	294428	332847	38420	42.70	47	VCLMA_B0259	NC_017269
<i>V. cholerae</i> M66-2	O1	Indonesia	Clinical	1937	310949	409433	98485	42.15	133	VCMM66_A0290	NC_012580
<i>V. cholerae</i> IEC224	O1	Para, Brazil	Clinical	1990s	309717	435237	125521	42.21	167	O3Y_14823	NC_016945
<i>V. mimicus</i> MB-451	ND	Matlab, Bangladesh	Clinical	ND	744870	872905	128036	41.39	115	VII_000636	NZ_ADAAF01000002

*Nucleotide position on the chromosome. ND, not determined.

core, dispensable and unique genes were identified by in-house R scripts. The phylogenetic tree based in gene content and split network for gene content were constructed with SplitsTree⁴¹ using the GeneContentDistance method¹². The SI structure and comparison of seven *V. cholerae* and, their sister species, *V. mimicus* were performed using genoPlotR¹³.

Results and discussion

SI regions were extracted from the seven *V. cholerae* and one *V. mimicus* genomes (Table 1). The 1285 genes recovered were clustered and a total of 408 clusters were detected (Figure 2A; Table S1). The pangenome of the SI of *Vibrio* strains evaluated was 408 genes, of which eight correspond to core genes, 196 are distributed or dispensable genes and 204 are unique genes. Six of the eight core genes are present in many copies (Table 2). The pangenome profile analysis shows that the cluster numbers of core genome are almost the same, when the SI considered reaches nine, while the pangenome is still increasing (Figure 2A). We infer that the *V. cholerae* SI has an open pangenome, which means that *V. cholerae* may have the ability to import new SI gene cassettes, which affect its plasticity and diversity. On the other hand, the set of SIs, from clinical and environmental lineages, used in this study are apparently representative of this species because allowed to establish that the core genome is close to being completed.

Function enrichment analysis of gene clusters were performed according to description of gene annotation (File S1) supplied

to the pangenome analysis pipeline¹⁰. From the 408 clusters, 329 were unclassified by the function enrichment analysis. Following the categorization of Cluster of Orthologous Groups (COG), the characterized clusters were rich in the following categories: translation, ribosomal structure and biogenesis, transcription, replication, recombination and repair, cell cycle control, cell division, chromosome partitioning, defense mechanisms, cell wall/membrane/envelope biogenesis and posttranslational modification, protein turnover, chaperones, amino acid transport and metabolism, nucleotide transport and metabolism, lipid transport and metabolism, secondary metabolites biosynthesis, transport and catabolism (Figure 2B).

In the SI, random excisions occur throughout the cassette array to form nonreplicative circular intermediates containing one or several cassettes; integration events preferentially occur at the *attI* site⁵ and are subjected to selection. It is expected that SI core genes would be arranged and stay together; however, we found the core genes are distributed along the SI (Figure 3), apparently without any position effect.

We identified 204 unique genes, 94 belonging to *V. mimicus* MB451, nine to LMA3984, 45 to O395, nine to 2010EL1786, 14 to MJ1236, seven to IEC224, 20 to M66, and six to N16961 (Figure 3; Table S1). Considering only the *V. cholerae* SI, there are 21 core genes, most of them present in many copies and rich in the transcription, replication, recombination and repair, translation, ribosomal structure and biogenesis categories.

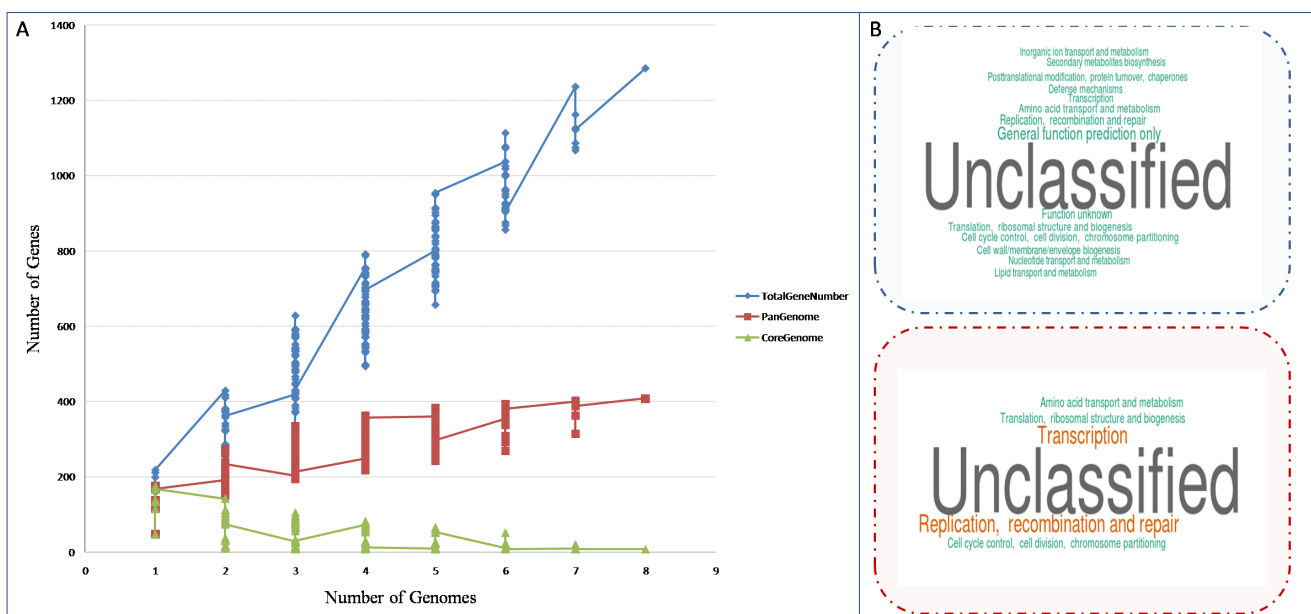


Figure 2 (A) Pangenome plot of the SI region considering seven *V. cholerae* and one *V. mimicus* genomes. 1285 total genes, 408 pangenome clusters and eight core clusters were identified. **(B) Word clouds of cluster function enrichment comparison** according to clusters of orthologous groups (COG) for whole and core clusters identified are shown at the top and bottom, respectively. Clusters that are not assigned in the COG classification were excluded from the figure.

Table 2 Core genes of the *V. cholerae* SI. The table shows the clusters, conservation level between genomes, the functional categories, gene description and the corresponding locus tag in the reference N16961 genome.

Cluster ID	Conservation level	COG*	Description	Locus_tag in N16961
1	8	–	hypothetical protein	VCA0407,VCA0353,VCA0336,VCA0297,VCA0302
2	8	COG0456R	acetyltransferase	VCA0470
3	8	–	lipoprotein	VCA0425,VCA0414
4	8	–	hypothetical protein	VCA0381,VCA0435,VCA0357,VCA0306
5	8	–	hypothetical protein	VCA0434,VCA0411
7	8	COG4974L	site-specific recombinase IntI4	VCA0291
8	8	–	relB protein	VCA0349,VCA0504
9	8	COG1670J	acetyltransferase	VCA0505,VCA0436,VCA0417,VCA0316
24	7	COG0110R	acetyltransferase	VCA0473
25	7	COG3668R	plasmid stabilization element ParE	VCA0359
27	7	COG2944K	virulence gene repressor RsaL	VCA0469
31	7	–	hypothetical protein	VCA0497
32	7	COG1694R	mazG-related protein	VCA0485
33	7	–	cytotoxic translational repressor of toxin-antitoxin stability system	VCA0468
34	7	COG0346E	glyoxalase/bleomycin resistance protein	VCA0506,VCA0347
35	7	–	hypothetical protein	VCA0486
37	7	COG2161D	antitoxin of toxin-antitoxin stability system	VCA0477
40	7	COG0456R	GCN5-related N-acetyltransferase	VCA0382
41	7	COG1943L	IS1004 transposase	VCA0493
43	7	COG3668R	plasmid stabilization system protein	VCA0489
44	7	COG3636K	hypothetical protein	VCA0498

*COG: Cluster of Orthologous Groups; "–" depicts no COG assignment.

Pandey and Gerdes¹⁴ identified 13 TA loci within the SI of the N16961 strain. Here we identified six TA genes as part of core SI genes (Table 2), of which the *relB* genes (VCA0349 and VCA0504) were present in all *V. cholerae* strains (including *V. mimicus*) SIs. The *parE* (VCA0359), *relB* (VCA0477) and *relE* (VCA0489) genes were present in all *V. cholerae* SIs. Moreover, we also identified two *higBA* loci (VCA0469 and VCA468), which encode mRNA cleaving enzymes and can stabilize plasmids⁶, as well as SI genes. The previous authors¹⁴ also identified *higBA-I* TA loci (VCA0392 and VCA0391); in our results, these two TA loci are present in all clinical *V. cholerae* strains (Table S1). These results suggest that *V. cholerae* TA loci function as essential stress response elements that help cells survive⁶, as well as act to stabilize the massive arrays of SI cassettes, as reported previously¹⁵.

A previous study suggested that SI structure is influenced by geographic boundaries in response to environmental conditions⁸. Here, we found that the clinical nature of the *V. cholerae*

and *V. mimicus* strains evaluated were not grouped together by the analyses performed. Therefore, the ability of *V. cholerae* to cause disease must be explained by other virulence factors found outside the SI region.

There are 199 clusters involved with indel or mutation events (Table S2). As for the non-synonymous/synonymous substitution (dN/dS) ratio, we found that 30 clusters were suffering positive selection pressure (dN/dS > 1). At the same time, we could also select those variable clusters as the markers for different strains. Based on pangenome profiles and single nucleotide polymorphism (SNP) information, gene content and phylogenetic trees were constructed (Figure 4). The SNP information from SI was useful for separating *V. cholerae* from *V. mimicus*, but nevertheless lacked the resolution to distinguish between the different lineages of *V. cholerae*. However, using gene content information (Figure 4), a good resolution was reached that was coherent with the evolution of the species and the environmental or clinical nature of the strains. These results

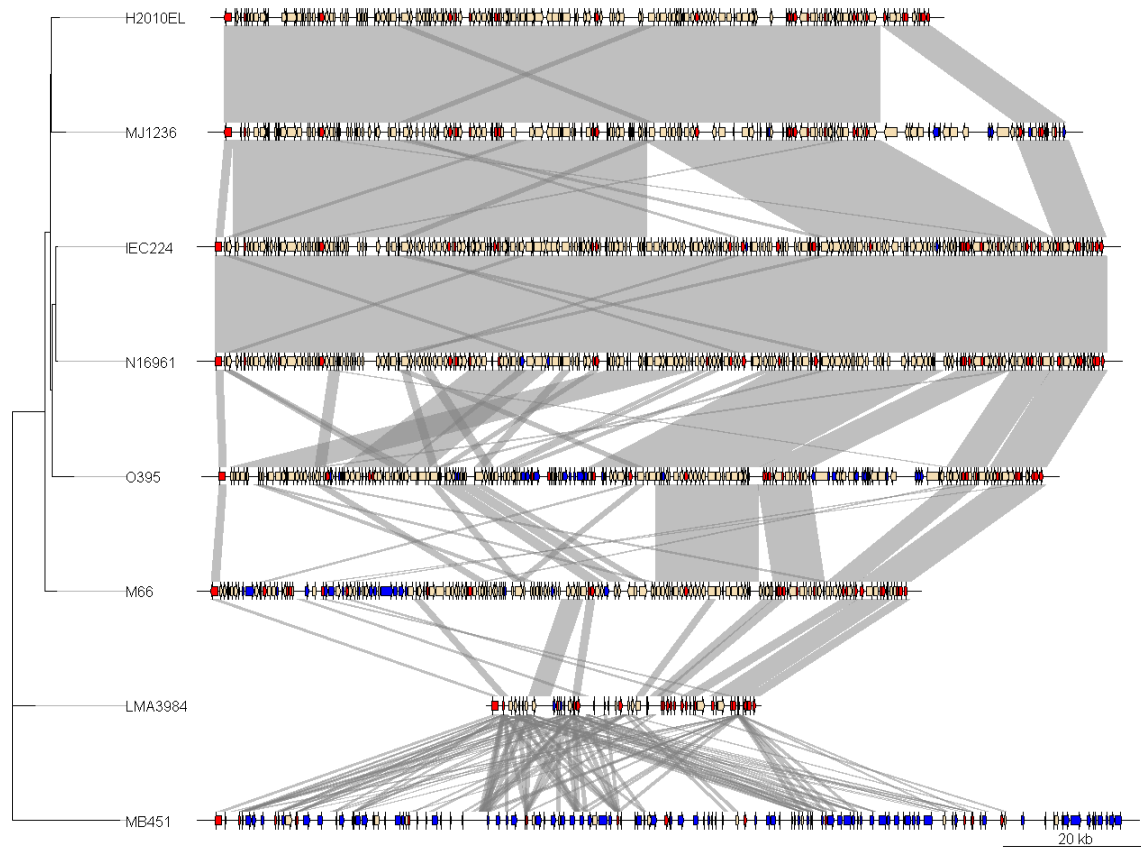


Figure 3 Superintegron (SI) structure and comparison of seven strains of *V. cholerae* and a strain of *V. mimicus*. The core, dispensable and unique genes are indicated by red, cream and blue arrows, respectively. Vertical blocks between sequences indicate regions with more than 1 kb of shared similarity shaded according to BLASTn. A phylogenetic tree based on gene content of the SI is shown on the left.

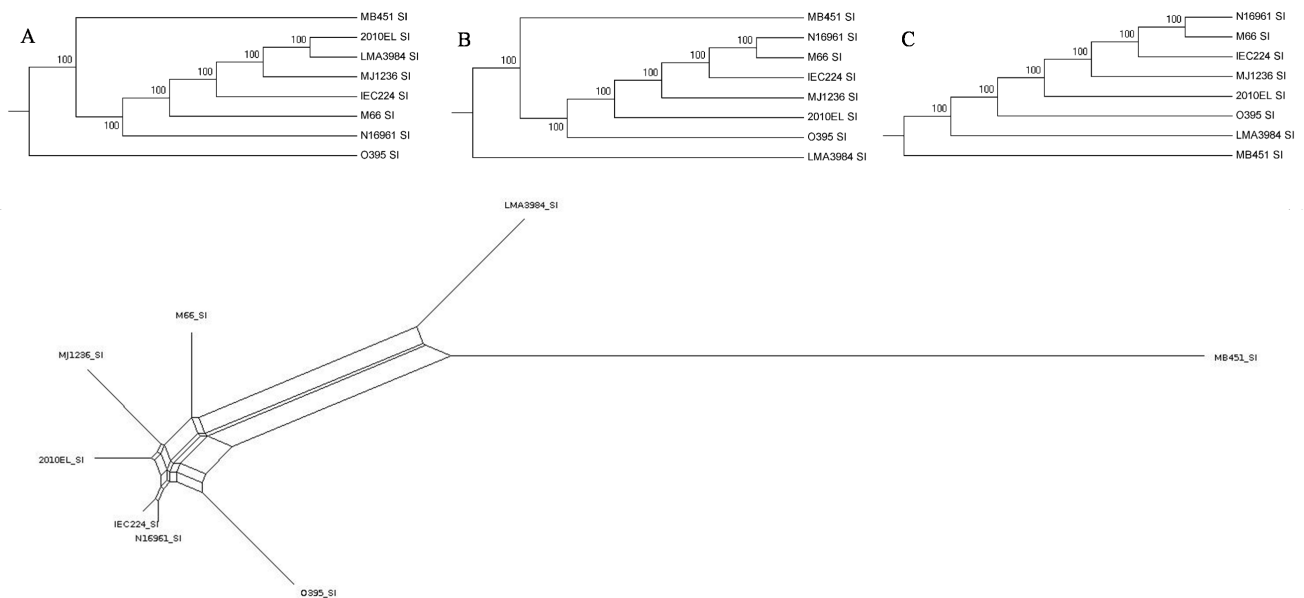


Figure 4 Top: Phylogenetic trees for the *V. cholerae* SI based on SNPs constructed by the Maximum Likelihood (A), Neighbor-Joining (B) and UPGMA (C) methods. The numbers indicate the bootstrap values. Bottom: Split network for gene content based on the 408 genes in seven *V. cholerae* and one *V. mimicus* genomes. The network was constructed with SplitsTree4 using the GeneContentDistance method¹².

indicate that the evolution of *V. cholerae* into different lineages is reflected in the diversity of the SI, which would be also influenced by horizontal gene transfer in these region, as proposed elsewhere^{8,16,17}.

File S1

1 Data File

<http://dx.doi.org/10.6084/m9.figshare.177719>

File S2

1 Data File

<http://dx.doi.org/10.6084/m9.figshare.177727>

Table S1

1 Data File

<http://dx.doi.org/10.6084/m9.figshare.177734>

Table S2

2 Data Files

<http://dx.doi.org/10.6084/m9.figshare.177736>

Supplementary tables

Table S1. Orthologs clusters identification among SIs from *V. cholerae* and *V. mimicus* genomes. These clusters were identified using the pangenome analysis pipeline¹⁰, strains without genes in the cluster are marked with “-”.

Table S2. Clusters involved with indel or mutation events. The 1st column is the Cluster ID, which is consistent with the ID in Table S1. The 2nd column is the cluster conservation of current cluster. The 3rd column is the variation position, which counts according to the alignment result of protein sequences in this cluster. For indel events, the position is an integer. For synonymous mutation and non synonymous mutation, the position is a floating number, in which the integer part marks the position of the amino acid in the alignment result of protein sequences,

Conclusions

In this study, we have revealed the genetic architecture of the *V. cholerae* SI, which contains eight core genes, many of them present in many copies. The *V. cholerae* SI has an open pangenome, which means that *V. cholerae* may have the ability to import new gene cassettes into the SI. The set of the dispensable SI gene cassettes is influenced by the niche and type species. The core genes are distributed along the SI, apparently without a position effect.

Author contributions

MAM and ACPV designed the study. MAM and ACPV analyzed the data. MAM and ACPV wrote the article. All authors have approved the final manuscript for publication.

Competing interests

No relevant competing interests were disclosed.

Grant information

This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) doctoral fellowship to MAM.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript

while the decimal part mark the position of codon. The 4th column shows the amino acid types on current position. The 5th column shows the nucleotide types on current position, indel is marked with “-”. The 6th column shows all gene nucleotide profile in current position (for indel, amino acid will be listed). The 7th column shows the variation type (indel, synonymous and non synonymous). The CDS.variation.analysis spreadsheet shows the summary result for CDS.variation.

Supplementary files

File S1. Gene data of SI region from seven *V. cholerae* and one *V. mimicus* genomes used in this study.

File S2. R scripts used in this study.

References

1. Blokesch M, Schoolnik GK: Serogroup conversion of *Vibrio cholerae* in aquatic reservoirs. *PLoS Pathog.* 2007; **3** (6): e81.
2. Heidelberg JF, Eisen JA, Nelson WC, *et al*: DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature.* 2000; **406** (6795): 477–83.
3. Mazel D: Integrons: agents of bacterial evolution. *Nat Rev Microbiol.* 2006; **4** (8): 608–20.
4. Rowe-Magnus DA, Guérout AM, Mazel D: Super-integrons. *Res Microbiol.* 1999; **150** (9–10): 641–51.
5. Cambray G, Guérout AM, Mazel D: Integrons. *Annu Rev Genet.* 2010; **44**: 141–66.
6. Christensen-Dalsgaard M, Gerdes K: Two *higBA* loci in the *Vibrio cholerae* superintegron encode mRNA cleaving enzymes and can stabilize plasmids. *Mol Microbiol.* 2006; **62** (2): 397–411.
7. Medini D, Donati C, Tettelin H, *et al*: The microbial pan-genome. *Curr Opin Genet Dev.* 2005; **15** (6): 589–94.
8. Boucher Y, Cordero OX, Takemura A, *et al*: Local mobile gene pools rapidly cross species boundaries to create endemicity within global *Vibrio cholerae* populations. *mBio. Am Soc Microbiol.* 2011; **2** (2): e00335–10.
9. Okonechnikov K, Golosova O, Fursov M: Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics.* 2012; **28** (8): 1166–7.
10. Zhao Y, Wu J, Yang J, *et al*: PGAP: pan-genomes analysis pipeline. *Bioinformatics.* 2012; **28** (3): 416–8.
11. Huson DH, Bryant D: Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.* 2006; **23** (2): 254–67.
12. Huson DH, Steel M: Phylogenetic trees based on gene content. *Bioinformatics.* 2004; **20** (13): 2044–9.
13. Guy L, Kultima JR, Andersson SGE: genoPlotR: comparative gene and genome visualization in R. *Bioinformatics.* 2010; **26** (18): 2334–5.
14. Pandey DP, Gerdes K: Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res.* 2005; **33** (3): 966–76.
15. Rowe-Magnus DA, Guérout A, Biskri L, *et al*: Comparative analysis of superintegrons: engineering extensive genetic diversity in the Vibrionaceae. *Genome Res.* 2003; **13** (3): 428–42.
16. Gao Y, Pang B, Wang HY, *et al*: Structural variation of the superintegron in the toxigenic *Vibrio cholerae* O1 El Tor. *Biomed Environ Sci.* 2011; **24** (6): 579–92.
17. Feng L, Reeves PR, Lan R, *et al*: A recalibrated molecular clock and independent origins for the cholera pandemic clones. *PLoS one.* 2008; **3** (12): e4053.

Current Referee Status:

Referee Responses for Version 1



Thandavarayan Ramamurthy, National Institute of Cholera and Enteric Diseases, Kolkata, India
Approved: 09 May 2013

Ref Report: 09 May 2013

In this manuscript the authors have explored the pattern of the superintegrons (SIs) in *V. cholerae* using the published DNA sequences of relevant strains. This study has shown the dynamic nature of *V. cholerae* O1 and genetic relatedness of SIs at the biotype level.

Comments:

- Abstract: Results and Conclusions: replace the word 'reveal' with "describe"
- Introduction: first paragraph: replace the word 'standing' with "standby"
- Page 6. The role of TA gene should be validated with non-toxigenic strains of *V. cholerae* (e.g., sequence comparison with non-O1, non-O139 strains of *V. cholerae*)

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Widespread distribution of the multidrug resistance ICEV_{ch}Hai1 in epidemic *Vibrio cholerae*.

MSc. Michel Abanto Marin¹, Dr. Adriana C. Cabral, MSc. Bruno Gabriel Nascimento Andrade, Dr. Erica Lourenço Fonseca, Dr. Ana Carolina P. Vicente.

¹ Corresponding author: Michel Abanto, Laboratory of Molecular Genetics of Microorganisms, Oswaldo Cruz Institute (IOC), Fiocruz, Rio de Janeiro, Brazil. Av. Brasil 4365 Manguinhos, Rio de Janeiro, Brazil, 21040-360; phone number: +55-21-38658168; Fax number: +55-21-22604282; email: mfabanto@ioc.fiocruz.br

Word counts:

Abstract: 51

Text: 1097

Running Title: Widespread distribution of ICEV_{ch}Hai1

Keywords: Multidrug resistance, Integrative and Conjugative Element, ICE, *Vibrio cholerae*, cholera, Haiti.

Abstract - 51

The emergence of *Vibrio cholerae* multidrug resistant overlaps with integrative and conjugative elements (ICEs) acquisition by this pathogen. ICEs are vectors disseminating antibiotic resistance. We identified ICEs in epidemic *V. cholerae* widespread in Africa and Asia sharing the same genetic content and conferring multidrug resistance as the 2010 Haiti epidemic strains.

Text – 1097

Vibrio cholerae is the etiological agent of cholera, a severe watery diarrhoeal disease that remains a serious health threat. In the last decades cholera epidemics are often caused by multidrug resistance strains (1). This resistance is mainly associated to mobile and mobilized elements, which affect not only the *V. cholerae* infection control, but also can be extended to overall bacterial infections (2).

ICEs are self-transmissible mobile elements, able to integrate into the host bacterial chromosome, excise and transfer to a new host genome (3). To date, more than 50 related ICEs were identified belonging to the SXT/R391 family. The acquisition of the multidrug resistant SXT/R391 elements by *V. cholerae* has shaped its pandemic spread (4). These ICEs have variable sizes due to deletion or insertion of sequences that confer specific characteristics to each of these elements. Recently, Ceccarelli et al. (5) proposed the term “siblings” for ICEs sharing the same gene content.

Here we identified that the specific ICE, harbored by 2010 Haitian cholera epidemic strains, carrying *floR* (chloramphenicol), *strAB* (streptomycin), *sul2* (sulfamethoxazole), and *dfrA1* (trimethoprim) antimicrobial resistance genes, is a sibling ICE widespread in *V.*

cholerae strains causing epidemics/outbreaks in three continents over a period of more than two decades.

The study

A total of 32 multidrug resistant *V. cholerae* O1 clinical strains from the Bacteria Culture Collection of Environment and Health (CBAS) at the Oswaldo Cruz Foundation, FIOCRUZ, were screened for the presence of genetic elements associated with antibiotic resistance, such as ICEs, class 1 and class 2 integrons. We identified strains from cholera outbreaks in Nigeria (n=12, 2010) (6) and Nepal (n=1, 1994) that carried the *int*_{SXT} gene (SXT integrase), indicating the presence of an SXT/R391 ICE element, however class 1 or class 2 integrons were not found in these strains. Mutations in the Quinolone Resistance Determining Regions (QRDR) of GyrA (Ser83Ile) and ParC (Ser85Leu), associated with resistance to nalidixic acid were observed only in the Nigerian strains. The strains from Nigeria and Nepal showed susceptibility to tetracycline, reduced susceptibility to chloramphenicol (MIC 6-8 mg/L), and resistance to both streptomycin (MIC > 1024 mg/L) and trimethoprim/sulfamethoxazole (MIC > 32 mg/L). MICs were determined by the E-test method according to the breakpoint established by the Clinical and Laboratory Standards Institute (CLSI, 2011).

A representative strain from Nigeria and the one from Nepal were selected for ICE characterization by sequencing using primers flanking the five regions considered hotspots (HS) for insertion of foreign DNA (7-9), as well as additional variable regions and the core genes (9). The gaps in the obtained sequences were closed by custom primer walking. These ICEs were named ICE*Vch*Nig1 and ICE*Vch*Nep1 (temporary identifier: GDSUB21770 and GDSUB21776, respectively). They contained the same set of 94 open reading frames (ORFs), including the core genes, the variable regions and the five HS regions. Comparative genomic

analysis of these ICEs revealed that all ORFs are similar and syntenic to ICEVchHai1, ICEVchInd5 and ICEVchBan5, which were previously identified and characterized in multidrug resistant strains causing outbreaks/epidemics in Haiti (10, 11), India (5) and Bangladesh (9), respectively. The ICEVchNig1 and ICEVchNep1 differed from ICEVchHai1 by two and eight single-nucleotide polymorphisms, respectively. According to our analyses, these ICEs shared the same gene content and are therefore siblings. We refer to these sibling ICEs as ICEVchHai1 due to the impact of the 2010 Haitian cholera epidemic.

Most of the variable sequences in ICE elements are present in five hotspot (HS) regions (9). Comparison of the ICEVchHai1 HS sequences using BLASTN (<http://blast.ncbi.nlm.nih.gov/>) searches in GenBank (October, 2012) revealed that HS4 and HS5 gene content is unique and characteristic of ICEVchHai1 siblings. The ORF content of HS4 (5394 bp) and HS5 (15040 bp) regions were characterized previously (9). These regions are possible candidates to be used as markers to identify specifically ICEVchHai1 siblings. These two markers, together with the *int*_{SXT} sequence (marker of SXT/R391 ICEs), were used for screening 202 *V. cholerae* genomes available in GenBank, Mutreja et al. (4), and Hendriksen et al. (12). We identified 62 *V. cholerae* genomes from the 7th cholera pandemic containing these two specific regions, as well as the entire set of ORFs present in ICEVchHai1: Kenya (2005-2009, n=17); India (1984-2009, n=15); Bangladesh (1991-2007, n=13); Nepal (1994 and 2010, n=9); Djibouti (2007, n=3); Haiti (2010, n=3); Nigeria (2010, n=1); Tanzania (2010, n=1) (Figure). The SXT/R391 family element were previously identified in some of these genomes (4). Here, we characterized that some of SXT/R391 from these strains are in fact ICEVchHai1 siblings (Figure). It is noteworthy that the ICEVchHai1 carry a variable region that tend to accumulate antibiotic resistance genes and is associated with a transposable element that could be transferred to other bacteria.

Considering that the sibling ICE*Vch*Hai1, characterized by HS4 and HS5, is present in strains spread in several continents, and that HS4 and HS5 sequences have not been found in other context, we searched for their presence in the CAMERA marine environmental metagenomes (<http://camera.calit2.net/>, October 2012), in order to gain insights into the possible origin and dispersion of these regions. A BLASTN search was performed with a similarity threshold based on an e-value of 10^{-20} . Although no hits were obtained for HS4 ORFs, eight hits related to HS5 ORFs were identified in two metagenomes (13): two hits corresponding to ORFs 4 and 5 from a mangrove in the Fernandina Island (Galapagos) and six hits corresponding to ORFs 1, 3 and 4 were found in Antarctic. We also detected hits for the *int*_{SXT} gene in these metagenomes. Taken together, these results show that HS5 and *int*_{SXT} are present in the environment as part of an ICE or in other genomic context. In the last case, there is the possibility of further association of these elements, the emergence of new ICEs and the spread of multidrug resistance elements to the microbiota.

One notable factor in the ongoing evolution of pandemic cholera was the acquisition of the SXT/R391 ICE family (4). Although the genetic relatedness of some of these ICEs were previously inferred (9, 10), our results showed an unexpected dispersion of siblings 2010 ICE*Vch*Hai1, carrying multidrug resistance genes in association with a transferable element, over three continents and their persistence for more than two decades. Beyond their multidrug resistance character these elements could have a role in the success of the current epidemic/pandemic *V. cholerae* O1 strains. The persistence and widespread distribution of this particular ICE could be due to its presence in the contemporary pandemic atypical *V. cholerae* strain or because this mobile element can easily be transferred and maintained by positive selection among *V. cholerae* populations as well as to be extended to other pathogens.

Acknowledgements:

This work was supported by Instituto Oswaldo Cruz grant (PROEP), FAPERJ, CAPES, Fiocruz fellowships.

Biographical Sketch:

Mr Abanto is a doctoral student at the Laboratório de Genética Molecular de Microrganismos, Instituto Oswaldo Cruz, FIOCRUZ, in Rio de Janeiro. His research interest is comparative genomics, metagenomics and *Vibrio cholerae* genetics.

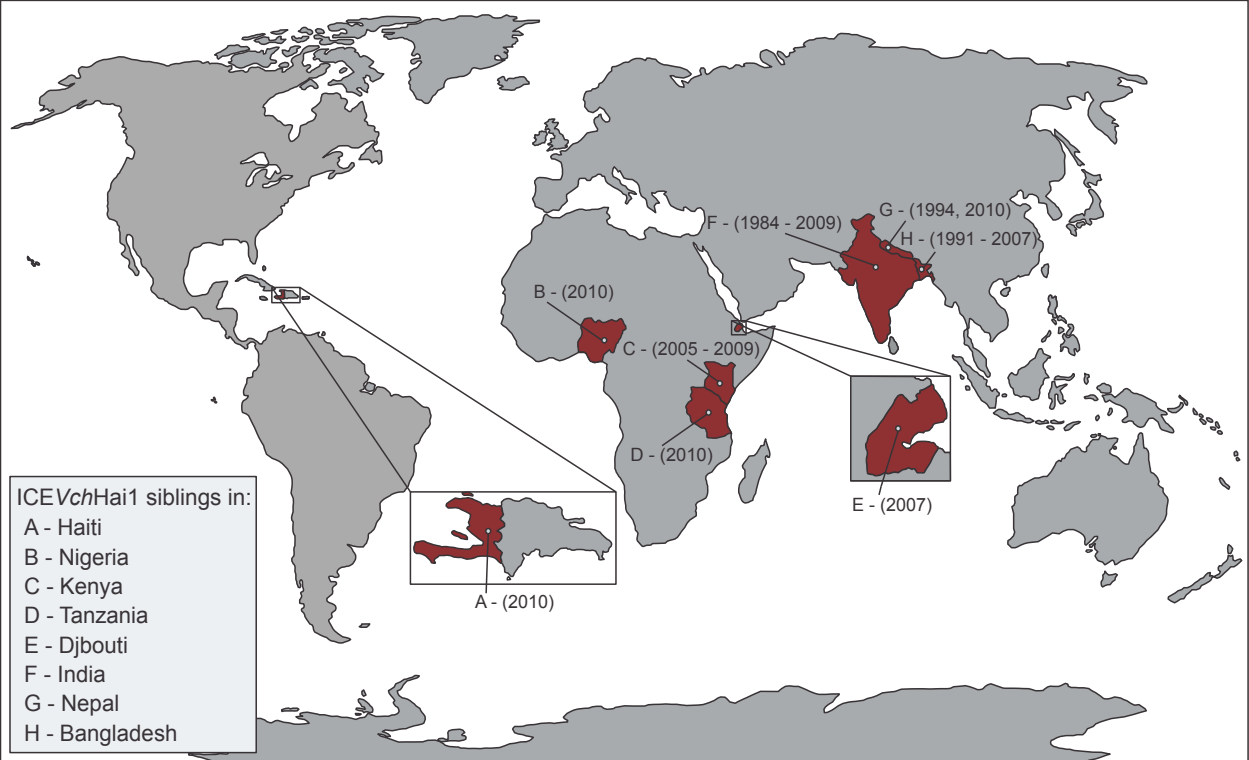
References:

1. Kitaoka M, Miyata ST, Unterweger D, Pukatzki S. Antibiotic resistance mechanisms of *Vibrio cholerae*. *J Med Microbiol*. 2011;60:397-407.
2. Okeke IN. Cholera vaccine will reduce antibiotic use. *Science* 2009;325:674.
3. Burrus V, Waldor MK. Shaping bacterial genomes with integrative and conjugative elements. *Res Microbiol*. 2004;155:376-86.
4. Mutreja A, Kim DW, Thomson NR, Connor TR, Lee JH, Kariuki S, et al. Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature*. 2011;477:462-5.
5. Ceccarelli D, Spagnoletti M, Bacciu D, Danin-Poleg Y, Mendiratta DK, Kashi Y, et al. ICEVchInd5 is prevalent in epidemic *Vibrio cholerae* O1 El Tor strains isolated in India. *Int J Med Microbiol*. 2011;301:318-24.
6. Marin MA, Thompson CC, Freitas FS, Fonseca EL, Aboderin a. O, Zailani SB, et al. Cholera Outbreaks in Nigeria Are Associated with Multidrug Resistant Atypical El Tor and Non-O1/Non-O139 *Vibrio cholerae*. *PLoS Negl Trop Dis*. 2013;7:e2049.

7. Beaber JW, Burrus V, Hochhut B, Waldor MK. Comparison of SXT and R391, two conjugative integrating elements: definition of a genetic backbone for the mobilization of resistance determinants. *Cell Mol Life Sci.* 2002;59:2065-70
8. Burrus V, Marrero J, Waldor MK. The current ICE age: biology and evolution of SXT-related integrating conjugative elements. *Plasmid.* 2006;55:173-83.
9. Wozniak RA, Fouts DE, Spagnoletti M, Colombo MM, Ceccarelli D, Garriss G, et al. Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. *PLoS Genet.* 2009;5:e1000786.
10. Sjolund-Karlsson M, Reimer A, Folster JP, Walker M, Dahourou GA, Batra DG, et al. Drug-resistance mechanisms in *Vibrio cholerae* O1 outbreak strain, Haiti, 2010. *Emerg Infect Dis.* 2011;17:2151-4.
11. Ceccarelli D, Spagnoletti M, Cappuccinelli P, Burrus V, Colombo MM. Origin of *Vibrio cholerae* in Haiti. *Lancet Infect Dis.* 2011;11:262.
12. Hendriksen RS, Price LB, Schupp JM, Gillece JD, Kaas RS, Engelthaler DM, et al. Population Genetics of *Vibrio cholerae* from Nepal in 2010: Evidence on the Origin of the Haitian Outbreak. *mBio* 2011;2:e00157-11.
13. Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K, et al. The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol.* 2007;5:e16.

Legend:

Figure: Spatio-temporal distribution of ICE*Vch*Hai1 siblings. The countries and years, where the strains carrying the ICE were isolated, are shown.



ICEVchHai1 siblings in:

- A - Haiti
- B - Nigeria
- C - Kenya
- D - Tanzania
- E - Djibouti
- F - India
- G - Nepal
- H - Bangladesh