

厚生労働科学研究費補助金  
平成22年度 分担研究報告書

研究課題名：「アジアの研究機関との連携におけるラボラトリーネットワーク」

分担研究：「*Vibrio cholerae* の新規 DNA フィンガープリンティング法、VCR-PCR 法と既知の PCR-based DNA フィンガープリンティング法との識別能等の比較」報告書

分担研究者 大澤 朗 神戸大学 農学研究科

研究要旨：

コレラ毒素産生性 *Vibrio cholerae* の血清型 01 および 0139 はゲノム DNA が高度に保存されていることから、通常有用な PFGE やリボタイピング法では菌株間の clonality (系統性) の異同の識別は難しいとされている。しかし、最近我々が行ったゲノムサブトラクション法による遺伝子解析で、*V. cholerae* の小染色体上に存在する Integron Island 領域に菌株レベルでの多型性が存在することが示唆された。そこで、我々は本領域に多数散在する *V. cholerae* repeat (VCR) に注目し、一昨年度この散在性を利用した新規 DNA フィンガープリンティング法、VCR-PCR の開発に成功した。本年度はマレーシアのマラヤ大学と連携して、本法と既知の3つの PCR を用いた DNA フィンガープリンティング法 (RAPD-PCR, ERIC-PCR, REP-PCR) による成績の信頼性、整合性等について比較した。その結果、VCR-PCR 行って系統解析した上で、同じ系統に属する菌株群について ERIC-PCR を行うことで疫学的に整合性のあるコレラ菌の DNA フィンガープリンティングを行えることが示された。

A. 研究目的

*Vibrio cholerae* によって引き起こされるコレラは、世界中で、特に、衛生状態の整っていない国々で流行している。この病原菌は、主に熱帯地域の水生環境中に常在している。しかしながら総ての *V. cholerae* がコレラを引き起こすわけではなく、腸管毒の1種であるコレラ毒素を産生する菌株、特に血清型 01 と 0139 に属する菌株がコレラの原因とされている。このため、*V. cholerae* の疫学解析に有用な DNA

フィンガープリンティング法をもって *V. cholerae* 01, 0139 の感染源、感染経路を正確かつ迅速に特定することは、本疾病の世界的な蔓延、特に我が国への侵入を阻止する上で大変重要である。しかしながら、コレラ毒素産生性 *V. cholerae* の血清型 01 および 0139 はゲノム DNA が高度に保存されていることから、多くの病原細菌の疫学解析に有用なパルスフィールド電気泳動 (PFGE) 法、リボタイピング法、あるいはハウスキーピング遺伝

子等の配列に注目した MLVA や MLST では菌株間の clonality (系統性) の異同の識別は難しいとされている。PFGE, MLVA, MLST はさらに、PCR に基づいた手法と比べると、高価な機器を使用する必要があり、アジアのラボラトリーのルーティン検査として容易に導入できない難点がある。

このような背景から、AP-PCR、BOX-PCR、Random Amplified Polymorphic DNA (RAPD)-PCR、Enterobacterial Repetitive Intergenic Consensus sequence (ERIC)-PCR、Repetitive Extragenic Palindromic (REP)-PCR のような PCR を基にした DNA フィンガープリンティング法による *V. cholerae* の疫学的解析がこれまでに多数行われており、この中でも、RAPD-PCR, ERIC-PCR, REP-PCR が有効であると報告されている。

我々は一昨年度の研究によって *V. cholerae* の分離地域や年代的な違いや血清学的違いを分類できる新たな DNA フィンガープリンティング法、VCR-PCR 法を開発した。この手法は、*V. cholerae* のインテグロンアイランド内における繰り返し配列の配列を標的としたものである。しかしながら、これは、2000 年以前に分離された毒素産生性の *V. cholerae* 01, 0139 株のみにしか試験を行っていない。

そこで本年度の研究では、2000 年代後半にマレーシアにおいて患者あるいは環境より分離された非毒素産生、毒素産生の *V. cholerae* 株を対象として、VCR-PCR と既に他の研究者らによ

って有用と報告されている上記 3 つの PCR を用いた DNA フィンガープリンティング法の疫学解析における有用性を比較検討した。

## B. 研究方法

1) 供試菌株：マレーシアのマラヤ大学 Biochemical Science and Molecular Microbiology の研究室が保有している人患者糞便、水、藻、海産物由来の *V. cholerae* 01 を 23 株、*V. cholerae* 0139 を 1 株、*V. cholerae* non-01/non-0139 を 19 株、総計 43 株を使用した。(Table 1)

2) テンプレート DNA 試料の準備：各株について、50  $\mu$ l の純水に 1 コロニーを懸濁し、99 $^{\circ}$ C で煮沸することで大雑把な DNA を採取した。短い遠心分離後、5  $\mu$ l の上清を PCR のために使用した。

3) RAPD-PCR：任意に合成した 10 個程度の塩基配列をプライマーに用い、ゲノム DNA の塩基配列上に個体差が存在した場合、増幅される複数の PCR 産物の泳動パターンに違いが見られることで DNA フィンガープリンティングする方法である。本研究では Chakraborty ら [2000] によって報告されている方法 (1281 というプライマーを使用) を若干改変した RAPD-PCR を行った。PCR ミクスチャーの組成は、1 $\times$  PCR buffer, 3 mM MgCl<sub>2</sub>, 200  $\mu$ M each dNTP, 1.0  $\mu$ M primer, 2 U Taq DNA polymerase (Promega, USA) で、計 25  $\mu$ l で行った。PCR のプログラムは、初めの変性過程として 94 $^{\circ}$ C で 5

Table 1. *Vibrio cholerae* Malaysian strains used in this study

Strain no.	serogroup	biotype	serotype	Cholera toxin production	Year isolation	Source
VC21	O1	El Tor	Inaba	+	2004	Patient
VC35	O1	El Tor	Inaba	-	2004	Patient
VC37	O1	El Tor	Inaba	+	2004	Patient
VC38	O1	El Tor	Inaba	+	2004	Patient
VC40	O1	El Tor	Inaba	+	2004	Patient
VC52	O1	El Tor	Ogawa	+	2008	Patient
VC63	O1	El Tor	Ogawa	+	2008	Patient
VC64	O1	El Tor	Ogawa	+	2008	Patient
VC65	O1	El Tor	Ogawa	+	2008	Patient
VC66	O1	El Tor	Ogawa	+	2008	Patient
VC4	O1	El Tor	Ogawa	+	2008	Patient
VC6	O1	El Tor	Ogawa	+	2008	Patient
VC9	O1	El Tor	Ogawa	+	2008	Patient
VC10	O1	El Tor	Ogawa	+	2008	Patient
VC123	O1	El Tor	Ogawa	+	2008	Patient
VC1	O1	El Tor	Ogawa	+	2008	Water
VC3	O1	El Tor	Ogawa	+	2008	Water
VC3P	O1	El Tor	Ogawa	+	2008	Water
VC11	O1	El Tor	Ogawa	+	2008	Algae
VC70	O1	El Tor	Ogawa	+	2008	Water
VC87	O1	El Tor	Ogawa	+	2008	Water
VC88	O1	El Tor	Ogawa	+	2008	Water
VC90	O1	El Tor	Ogawa	+	2008	Water
VC4370	O139	-	-	+	2008	Water
VC17	Non-O1/non-O139	-	-	-	2008	Water
VC77	Non-O1/non-O139	-	-	-	2008	Water
VC81	Non-O1/non-O139	-	-	-	2008	Seafood
VC82	Non-O1/non-O139	-	-	-	2008	Seafood
VC83	Non-O1/non-O139	-	-	-	2008	Seafood
VC84	Non-O1/non-O139	-	-	-	2008	Seafood
VC85	Non-O1/non-O139	-	-	-	2008	Seafood
VC310	Non-O1/non-O139	-	-	-	2008	Seafood
VC374	Non-O1/non-O139	-	-	-	2008	Water
VC933	Non-O1/non-O139	-	-	-	2008	Seafood
VC3477	Non-O1/non-O139	-	-	-	2008	Water
VC4933	Non-O1/non-O139	-	-	-	2008	Water
VCGB	Non-O1/non-O139	-	-	-	2008	Water
VCSW	Non-O1/non-O139	-	-	-	2008	Algae
VCM1	Non-O1/non-O139	-	-	-	2008	Water
VCM2	Non-O1/non-O139	-	-	-	2008	Water
VCB2	Non-O1/non-O139	-	-	-	2008	Water
VCB4	Non-O1/non-O139	-	-	-	2008	Water
VCPSW	Non-O1/non-O139	-	-	-	2008	Water

分、変性として94℃で1分、アニーリングとして36℃で1分、伸長として72℃で2分のサイクルを45サイクル、最後の伸長過程として70℃で10分とした。

4) ERIC-PCR: 染色体DNAに散在する126bpのEnterobacterial Repetitive Intergenic Consensus sequenceという繰り返し配列を標的として、ゲノムDNAの塩基配列上に個体差が存在した場合、増幅される複数のPCR産物の泳動パターンに違いが見られることでDNAフィンガープリンティングする方法である。本研究ではVersalovicら[1991]によって報告されている方法(プライマーはERIC1RとERIC2を用いた)にてERIC-PCRを行った。PCRミクスチャーの組成は、1× PCR buffer, 1.5 mM MgCl<sub>2</sub>, 125 μM each dNTP, 1.0 μM each primer, 2.5 U Taq DNA polymerase (Promega, 128 USA)で計25 μlで行った。PCRのプログラムは、初めの変性過程として95℃で5分、変性として92℃で45秒、アニーリングとして52℃で1分、伸長として70℃で10分のサイクルを35サイクル、最後の伸長過程として70℃で20分とした。

5) REP-PCR: ゲノム上に多数存在する35~40bpのrepetitive extragenic palindromic sequence (遺伝子外回文繰り返し配列)を標的として、ゲノムDNAの塩基配列上に個体差が存在した場合、増幅される複数のPCR産物の泳動パターンに違いが見られることでDNAフィンガープリンティングする

方法である。本研究では、Naviaら[1999]の報告に書かれているREP-PCRを行った。PCRミクスチャーの組成は、2.5 mM MgCl<sub>2</sub>, 50 μM of each dNTPs, 0.6 μM of primer, 1.0 U Taq DNA polymerase (Promega, USA)とし、計25 μlで行った。PCRのプログラムは、初めの変性過程として94℃で4分、変性として94℃で1分、アニーリングとして42℃で1分、伸長として68℃で8分のサイクルを35サイクル、最後の伸長過程として72℃で8分とした。

6) VCR-PCR: 一昨年度の研究報告に記載されている、Tokunagaら[2010]が報告した方法を若干改変したVCR-PCR法を行った。PCRミクスチャーの組成は、2.5 mM MgCl<sub>2</sub>, 145 50 μM of each dNTPs, 0.5 μM of primer, 1.5 U Taq DNA polymerase (Promega, 147 USA)とし、計25 μlで行った。PCRのプログラムは、変性として94℃で1分、アニーリングとして65℃で30秒、伸長として72℃で3分のサイクルを35サイクルとした。

7) PCR産物の電気泳動: RAPD-PCR, ERIC-PCR, REP-PCR, VCR-PCRの単位複製配列は、1.2%のアガロースゲルを用い、90Vで6時間電気泳動した。分子サイズの基準のために1kbマーカを用いた。そのゲルは、エチジウムブロマイドで染色し、Gel Doc XRで観察した。

8) フィンガープリンティングパターンの解析: RAPD-PCR, ERIC-PCR, REP-PCR, VCR-PCRのバンドパターンをBio Numerics 6.0を用いて解析し

た。全ての PCR フィンガープリンティング法は、任意設定を割り当てた。0.15 の位置公差を使用した UPGMA に基づいた Dice coefficient, F. Cluster 解析によって、このプロファイル間での定量的に異同を判別した。9) 4つの PCR フィンガープリンティング法の性能比較：RAPD-PCR, ERIC-PCR, REP-PCR, VCR-PCR の分類力 (Typeability; 株レベルで特徴的なバンドパターンを示す確率)、再現性 (Reproducibility; 3カ所のラボで同じバンドパターンが得られる確率)、安定性 (Stability; 同一のラボで3回試行を行い、全サンプルが3回とも同じバンドパターンを示す確率) を比較した。

### C. 研究結果

1) RAPD-PCR では、250bp~2000bp の 14~24 バンドにより、43 株は、38 タイプに分類された (Fig. 1)。3 回の試行のうち 2 回で同じバンドを示していた 4 株を除き、全株において再現性があった。相同性 75% により、4 クラスタ (RAPD 1-4) に分類される (Fig. 2a)。RAPD 1 は、環境由来の non-01/non-0139 が 2 株、01 株が 1 株である。RAPD 2 は、もっとも大きいクラスタであり、01 株が 20 株、non-01/non-0139 が 10 株、0139 が 1 株である。RAPD 3 と RAPD 4 はそれぞれ 2 株ずつであった。Non-01/non-0139 株の 4 株、01 の 1 株は、いかなるグループにもクラスタ分類されなかった。4 株と VC6 (01 血清型) は全く異

なっていた。

2) ERIC-PCR では、12~14 バンドにより、43 株を 40 プロファイルに分類した。バンドパターンは再現性が有った。75% の相同性から 4 クラスタに分類された (ERIC 1-4) (Fig. 2b)。ERIC 1 は、non-01/non-0139 が 2 株であった。ERIC 3 は、01 株が 12 株、non-01/non-0139 株が 3 株であるのに対し、ERIC 2 は、01 株が 11 株、non-01/non-0139 株が 3 株であった。ERIC 4 は、non-01/non-0139 が 8 株、0139 が 1 株であった。

3) REP-PCR では、12~20 バンドにより、43 株を 35 プロファイルに分類した。75% の相同性から 5 クラスタに分類された (REP 1-5) (Fig. 2c)。REP 1, REP 4 は、各々 non-01/non-0139 6 株で構成されている。REP 2 は、non-01/non-0139 株が 1 株、0139 株が 1 株であった。REP 3 は、non-01/non-0139 株が 2 株であった。REP 5 は、01 が 21 株であった。

4) VCR-PCR では、43 株を 31 プロファイルに分類した。75% の相同性から各血清型により 3 つにクラスタ分類された (VCR 1-3) (Fig. 2d)。VCR1, VCR3 には、non-01/non-0139 で構成されていた。01 は全株とも VCR 2 に分類された。

各手法の分類力、再現性、安定性を Table 2 にまとめた。RAPD-PCR と REP-PCR は高い菌株識別力 (typeability) を持っているが、成績の安定性 (stability) や再現性 (reproducibility) が低いために、*V.*

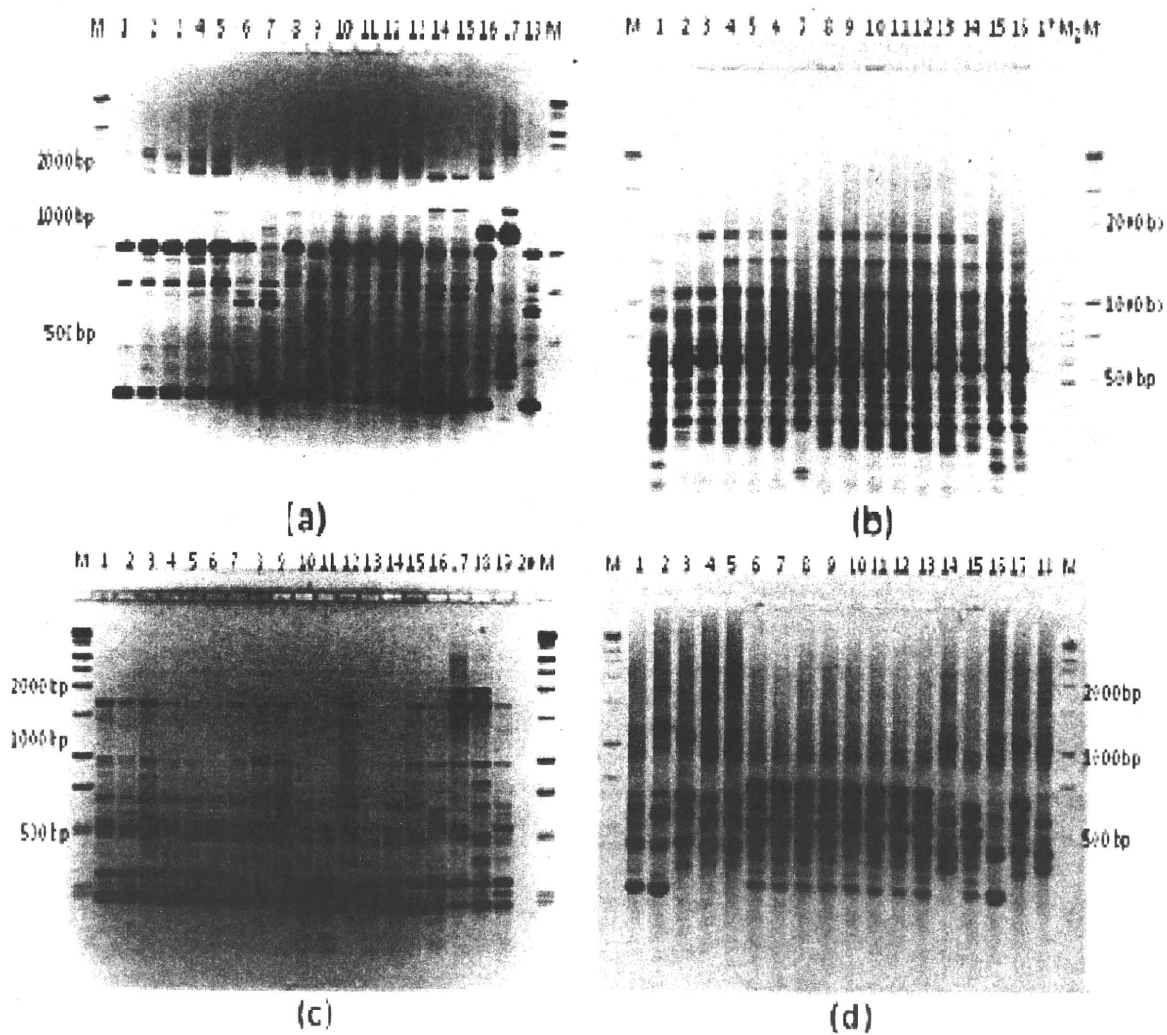


Fig. 1. RAPD-PCR (a), ERIC-PCR (b), REP-PCR (c), VCR-PCR (d) による電気泳動図

	Typeability* $T = N_t / N$	Reproducibility* $R = N_r / N$	Stability* $S = N_s / N$
RAPD-PCR	0.88	0.90 (39 strains)	0.6 (2/3)
ERIC-PCR	0.93	1.00 (43 strains)	1.00 (3/3)
REP-PCR	0.81	0.95 (41 strains)	0.6 (2/3)
VCR-PCR	0.72	1.00 (43 strains)	1.00 (3/3)

Fig. 2a.

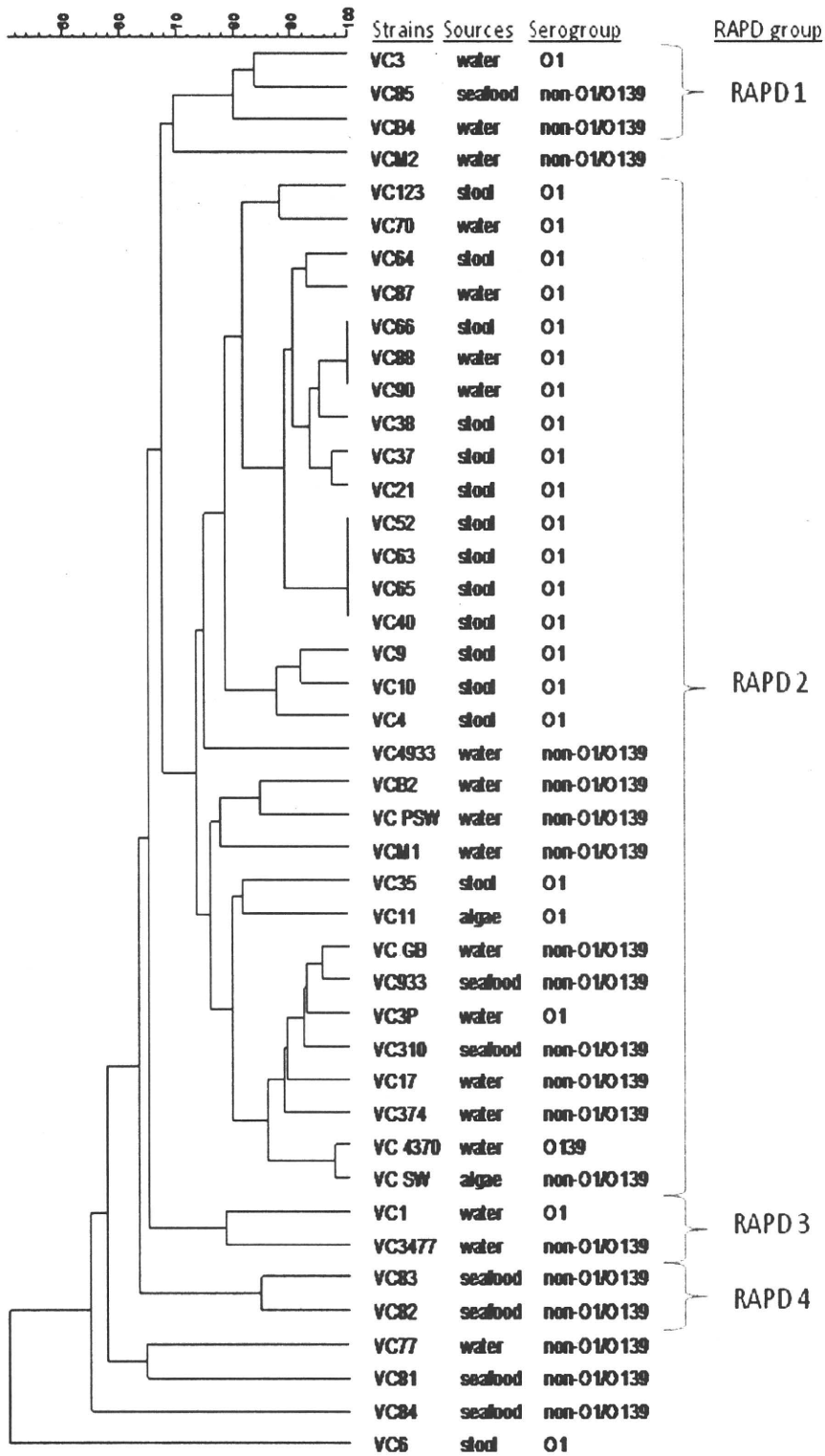


Fig. 2b.

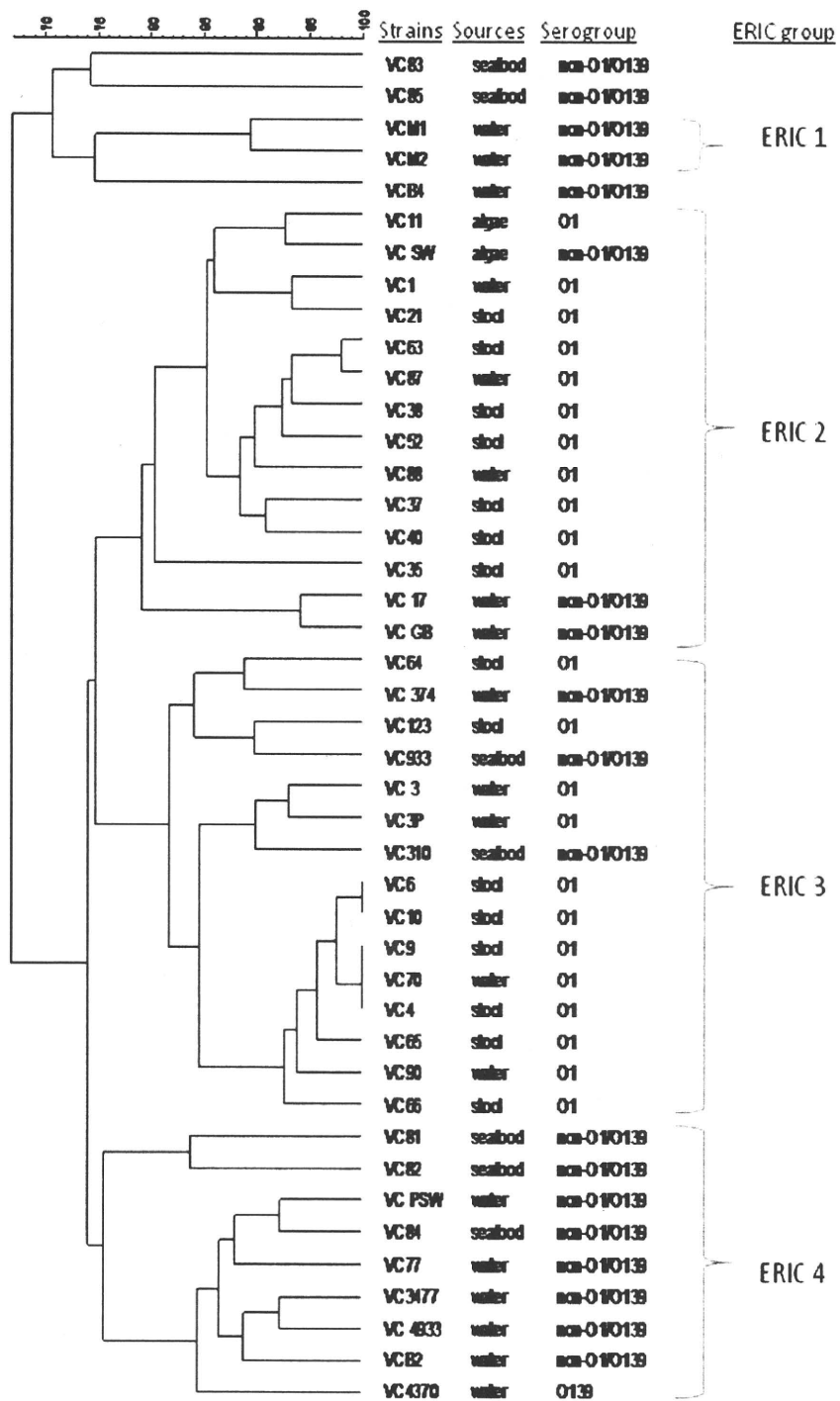




Fig 2c.

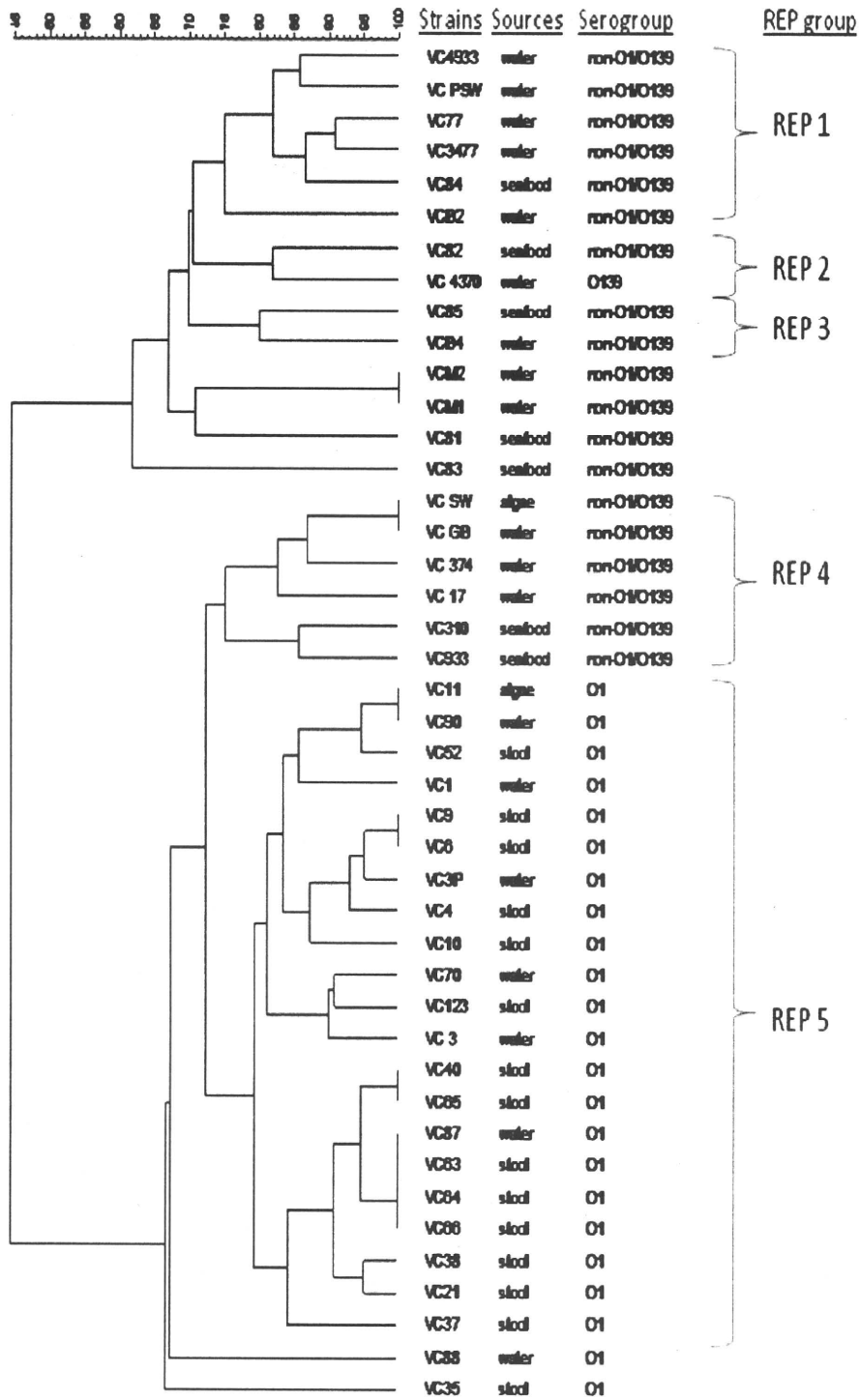
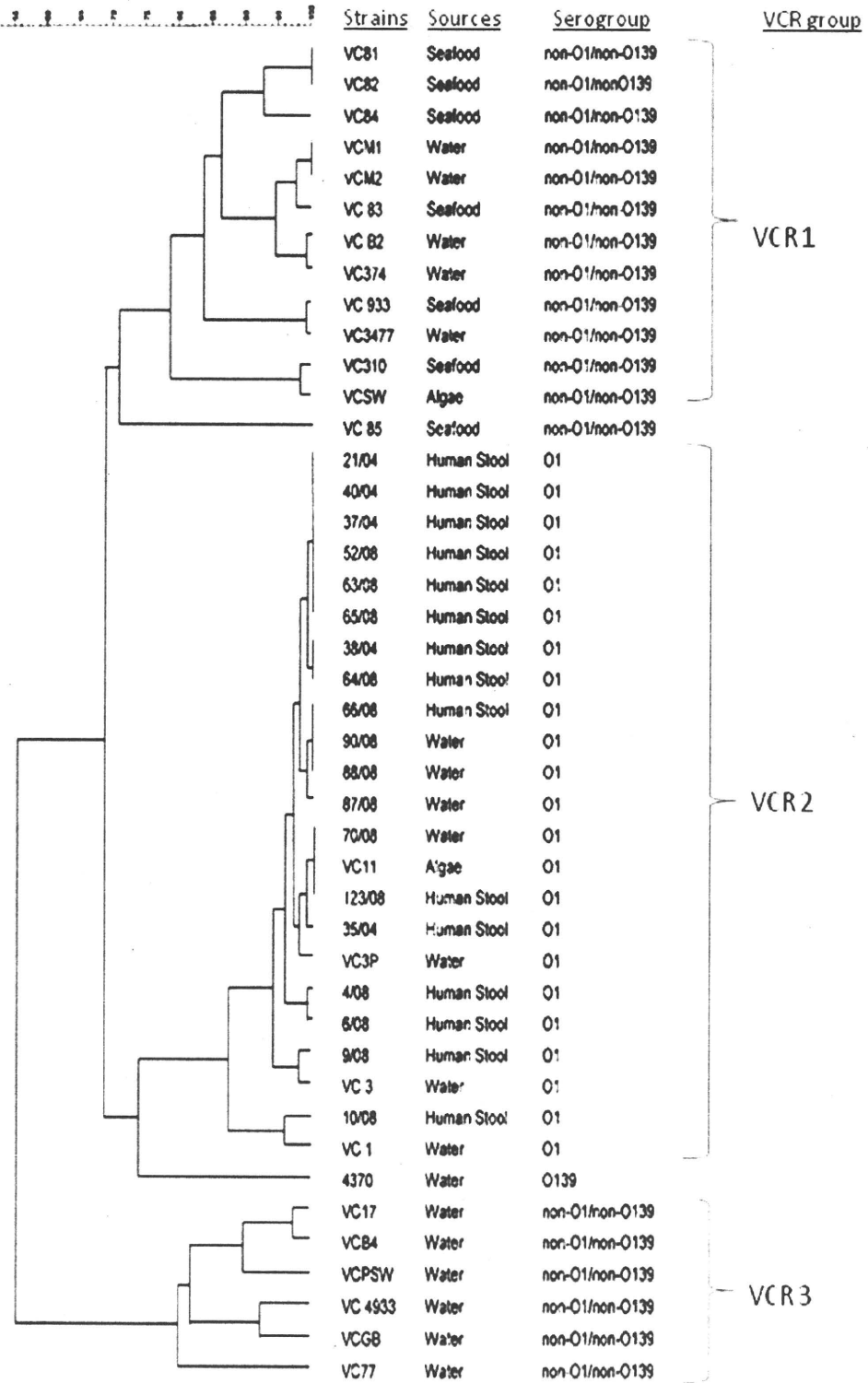


Fig 2d.



*cholerae* のフィンガープリンティング法としては不適であった。ERIC-PCR の菌株識別力、成績の安定性、再現性も高いが、異なる血清型が同一系統と判定されてしまうようなバンドパターンを示すため、疫学的系統解析は困難であった。一方、VCR-PCR は、菌株識別力は ERIC-PCR に劣るものの、成績の安定性、再現性が高く、さらに血清型と整合性のある系統分けが可能であった。

#### D. 考察

近年マレーシアでは、大規模なコレラの発生はないが、散発事例は、毎年報告があり、死者も発生する。それゆえに、菌株の迅速な系統解析はコレラの発生を回避あるいは最小限に食い止めるために大変重要である。4つのPCR法を評価し、適用することで、菌株の関係性を決定した。電気泳動の時間を含めてもPCRにかかる時間は14時間以内である。さらに01、0139とnon-01/non-0139を区別するためには、VCR-PCRが有用であった。Tokunagaら[2010]は、13カ国から集めた01、0139を使用して、8つのパターンに分類した。Eパターンは、ペルーやボリビアで分離された株であるのに対し、A、B、C、D、F、Gパターンは、1971年から1997年にアジアで分離された01であった。0139は、香港、デンマーク、バングラデシュ、インド、中国、日本、タイ、シンガポール、ネパールから1993年から2006年に分離された株であるが、Hパターンのみであった。こ

の報告に基づき、01と0139のVCRパターンから視覚的な比較が行われた。今回の研究で使用した01の12株はEパターンを除き、同様のパターンを示した。0139のVCRパターンは、Hパターンと同様のものであった。この結果は、0139全株が分離地や分離年に関係無く、同一の遺伝子型を示すという結果は、Tokunagaら[2010]の報告と一致していた。

4つのPCRを用いたフィンガープリントの結果より、43株のマレーシア分離の*V. cholerae*が遺伝子的に多様であることが明らかとなった。Non-01/non-0139は、大きく2つのグループに分類される。A) 01との遺伝子的な相同性が80%以上のもの、B) 01との遺伝子的な相同性が80%以下のものの2つである。例えば、VCSW, VC374, VC310, VC933のnon-01/non-0139株は、RAPD-PCR, ERIC-PCR, REP-PCR, VCR-PCRの結果では01と高い相同性を示した。VCM1とVCM2においては、REP-PCRとVCR-PCRによりサブタイピングを行ったところ、100%の相同性を示した。しかしながら、この株はRAPD-PCRでは68%、ERIC-PCRでは、89%の相同性であった。RAPD-PCR, ERIC-PCR, REP-PCRの結果から作成したデンドログラムから、0139は、non-01/non-0139に近縁種であることが示された。しかしながら、VCR-PCRでは、0139は、01やnon-01/non-0139と区別され、01株により近い結果であった。それ故に、海水等から分離されたある0139株は、

01 El Tor 株の 01 抗原合成遺伝子やハウスキーピング遺伝子が non-01/non-0139 の環境分離株へ水平伝播した結果であるという可能性が考えられた。

前述のごとく RAPD-PCR と REP-PCR は高い菌株識別力を持っているが、成績の安定性や再現性が低いために、*V. cholerae* のフィンガープリンティング法としては不適であった。ERIC-PCR の菌株識別力、成績の安定性、再現性も高いが、異なる血清型が同一系統と判定されてしまうようなバンドパターンを示すため、疫学的系統解析は困難であった。一方、VCR-PCR は、菌株識別力は ERIC-PCR に劣るものの、成績の安定性、再現性が高く、さらに血清型と整合性のある系統分けが可能であった。このようなことから、*V. cholerae* の疫学的情報と整合性を保った菌株レベルでの DNA フィンガープリンティングは VCR-PCR と ERIC-PCR を併せて使用することが望ましい。

## E. 結論

*Vibrio cholerae* の DNA フィンガープリンティングは VCR-PCR 行って疫学的な付帯情報と整合性のある系統解析した上で、同じ系統に属する菌株群について ERIC-PCR を行い菌株レベルでの系統識別を行うことが適当である。

## F. 研究発表

### 1. 論文発表

Teh CS, Thong KL, Osawa R, Chua, KH

(in press): Comparative PCR-based fingerprinting of Malaysian *Vibrio cholerae*. J Gen Appl Microbiol

### 2. 学会発表

1) Osawa R: Novel PCR-based DNA fingerprinting, using genomic variability between repetitive sequences of toxigenic *Vibrio cholerae* 01 El Tor and 0139 strains. The international conference on food security during challenging times, University Putra in Malaysia (2010. 7).

2) 山口博史, 井口純, 森田昌知, 勢戸和子, 渡辺治雄, 大澤朗: コレラ毒素産生性 *Vibrio cholerae* エルトール 01, 0139 株の Integron island を標的とした PCR-RFLP 解析、第 83 回日本細菌学会総会、日本細菌学会誌 65 巻 1 号、p. 100 (2010. 3).

### G. 参考文献

1) Chakraborty S, Mukhopadhyay AK, Bhadra RK, Ghosh AN, Mitra R, Shimada T, Yamasaki S, Faruque SM, Takeda Y, Colwell RR, Nair GB (2000): Virulence genes in environmental strains of *Vibrio cholerae*. Appl Environ Microbiol 66: 4022- 4028.

2) Versalovic J, Koeuth T, Lupski JR (1991): Distribution of repetitive DNA sequences in eubacteria and applications to fingerprinting of bacterial genomes.

Nucleic Acids Res 19: 6823-6831.

3) Navia MM, Capitang L, Ruiz J, Vargas M, Urassa H, Schelleberg D (1999): Typing and characterization of mechanisms of resistance of *Shigella* spp. isolated from feces of children under 5 years of age from Ifakara, Tanzania. J Clin Microbiol 37: 3113-3117.

4) Tokunaga A, Yamaguchi H, Morita M, Arakawa E, Izumiya H, Watanabe H, Osawa R (2010): Novel PCR-based genotyping, using genomic variability between repetitive sequences of toxigenic *Vibrio cholerae* O1 El Tor and O139. Mol Cell Probes 24: 99-103.

**Title: annual report of PulseNet China**

Bacteriophages determine the epidemic strength of *Shigella flexneri*  
serotypes

Name of researcher: Jianguo Xu

**Affiliation:** State Key Laboratory for Infectious Diseases Prevention and Control,  
National Institute for Communicable Disease Control and Prevention, Chinese Center  
for Disease Control and Prevention.

**Objective:** The serotype-converting phages determine not only serotype specificity  
but also epidemic potential of *S. flexneri*.

**Achievements:**

**Introduction**

Bacillary dysentery caused by *Shigella* is one of the major infectious diseases in  
developing countries (1). Shigellosis in China is one of the top four notifiable infectious  
diseases, with close to half a million cases each year (<http://www.moh.gov.cn>) as  
reported by the Chinese Internet-based Reporting System of Notifiable Infectious  
Diseases which began operation in 2005. This figure is widely believed to be an

underestimate (2,3). The predominant *Shigella* species in developing countries such as China where the standards of sanitation and hygiene is low, is *S. flexneri* and the distribution of *S. flexneri* serotypes varies geographically and temporally. Significant *S. flexneri* serotype changes over time were observed in several developing countries including Indonesia, Bangladesh, Pakistan and China. Epidemiological studies show that some serotypes such as *S. flexneri* 2a are predominant and responsible for the majority of cases (4). This suggests that certain serotypes are more likely to cause epidemics than others. In other words, some serotypes may have higher epidemic strength than others. The phenomenon has not been well studied.

There are 15 known *S. flexneri* serotypes (2). *S. flexneri* serotypes except serotype 6 are modifications of a common basic O antigen by serotype conversion phages (6). The basic O-antigen of *S. flexneri* is referred to as serotype Y and consists of repeating units of the tetrasaccharide, N-acetylglucosamine- rhamnose- rhamnose -rhamnose. The serotypes vary through the addition of glucosyl and/or O-acetyl groups to one or more of its four sugars in the tetrasaccharide unit. These modifications were carried out by phage encoded genes. There are six known serotype-converting phages namely Sfl, SflI,

Sf6, SfIV, SfV and SfX in *S. flexneri*. All except Sf6 encode three genes involved in a three-step process of the glucosylation of the O antigen. The first gene (*gtrA*) is responsible for translocation of lipid-linked glucose across the cytoplasmic membrane. The second gene (*gtrB*) encodes an enzyme catalyzing the transfer of the glucose residue from UDP-glucose to a lipid carrier. The third gene (*gtrtype*) encodes a serotype-specific glucosyltransferase which is largely responsible for the final step, i.e. attaching the glucosyl molecules onto the correct sugar residue of the O antigen repeating unit. Unlike glucosylation above, Sf6 encodes a single gene for O acetylation.

In a previous study we found that 9 *S. flexneri* serotypes including 1a, 2a, 2b, 3a, 4a, 5b, X, X variant and Y were present in China over a nine-year surveillance. We further showed that 95% of the *S. flexneri* infections in China were caused by a single epidemic clone namely sequence type (ST) 91. One of the factors leading to the persistence of the epidemic was found to be contributed by that ST19 switches its serotype frequently (14).

In this study we report that epidemic strength of a *S. flexneri* serotype reflected by prevalence in clinical isolation is determined by bacteriophages. The *S. flexneri* strains carrying high epidemic strength phages, such as SfX, SfII and Sfl, are prevalent and



epidemic. Nine phage genes associated with SfX epidemic potential were identified.

## **Results and discussion**

Epidemic serotypes and epidemic associated serotype-converting phages of *S. flexneri*

In a total of 1608 *S. flexneri* isolates collected from various parts of China during a

nine-year period from 2000 to 2008, the most frequently isolated serotypes were X

variant (597 isolates), 2a (334 isolates), X (268 isolates), 1a (233 isolates) and 2b (117

isolates). A very small proportion of isolates were serotypes 4a, 3a, 5b and Y with only

59 isolates in total. Since *S. flexneri* serotype specificity is determined by

serotype-converting phages (1), we converted the serotypes observed to

serotype-converting phages they carry and found that a total of six phages were present

among these isolates, namely Sfl, SflI, Sf6, SflV, SfV and SfX, with the percentages

ranging from 0.4% to 61.9%(2). The three most prevalent phages, SfX, SflI and Sfl,

were present in 996 (61.9%), 451(28%) and 233 (14.5%) isolates respectively,

accounting for 97% of the 1608 isolates. It was surprising to find that 865 (53.8%) of

the 1608 isolates carry phage SfX only, including 597 serotype X variant isolates and

268 serotype X isolates. Note that serotype X variant is determined by SfX and one

unknown modifying factor (13). An additional 131 isolates had SfX and another serotype-converting phage, namely SfX+SfV for serotype 5b, SfX+Sf6 for 3a and SfX+SfII for 2b. In total, 996 (61.9%) of 1608 isolates has SfX, alone or with others.

Therefore SfX is regarded as a phage with high epidemic strength. In contrast, Sf6 and SfV, were only carried by eight (0.5%) and six (0.4%) of the 1608 isolates studied, and regarded as having low epidemic strength.

**Infectivity and immunity of the serotype-converting phages with high epidemic strength**

To understand further the epidemic potential of the three predominant serotype-converting phages in *S. flexneri*, we performed phage induction and infection experiments. All isolates used were from serotypes expected as well as confirmed by PCR to carry the relevant phages. Phage SfX was successfully induced from three of the five isolates tested. In contrast, Sfl was induced from only one of the 10 isolates and SfII from none of 60 isolates attempted. It seems that SfX is more readily inducible than Sfl and SfII.

We tested the infectivity of the induced phages. SfX successfully infected all six

serotype Y isolates and one serotype 3b isolate attempted to serotype X and serotype 3a respectively while Sfl successfully infected all five serotype Y isolates attempted to serotype 1a. We also performed sequential infection with SfX and Sfl. A serotype Y strain 2003036 was first infected by phage SfX and was converted to X. The serotype X convertant was further infected by Sfl and was converted to a novel serotype which we refer to as serotype I:X. These convertants were serologically confirmed by slide agglutination using Shigella diagnostic sera and genetically confirmed by PCR and sequencing of the serotype-specific gtr genes. The resulted *S. flexneri* serotype I:X 2003036 agglutinated with both serotype X-specific monovalent anti-7,8 group antisera and serotype 1a-specific monovalent typing sera. The conventional serological identification results were further confirmed by Western blot assay using monovalent anti-7,8 group antisera and serotype 1a-specific monovalent typing sera. Both serotype-specific genes gtrX and gtrI were also detected by PCR.

In order to see how easily the novel serotype could be created, we performed Sfl infection of 10 clinical isolates of serotype X separately. Surprisingly, all of the 10 isolates were successfully converted to I:X. Considering that a high proportion (53.8%)

of the clinical isolates carry SfX, the probability of these isolates being infected by SfX in nature is expected to be high. However, it should be noted that SfX failed to infect any of the 2a, 5a, 4a, 4b and 1a isolates with one to four isolates tested, likely due to phage immunity from modified O antigen receptors (9).

#### Genomic characterization of the high epidemic strength phage, SfX

We then characterized SfX, induced from the *S. flexneri* X variant strain 2002017 (12).

Electron microscopy showed that the phage particle consists of an icosahedral head and a short tail as a typical type C phage (7), and is identical to the morphology previously reported (5). The complete phage genome sequence was constructed based on prophage genome sequence from strain *S. flexneri* X variant 2002017 (GenBank accession number CP001383-CP001388) by joining of the prophage ends which were determined by sequencing PCR products across the ends from purified phage DNA preparations.

The SfX genome is 37,355 bp, with an average G+C content of 47% which is slightly lower than the host genome G+C content (50%)(12). The phage genome contains 59 ORFs with protein-coding regions occupying 92.6% of the genome. Similar to other bacteriophages, the majority of the predicted ORFs (36) are transcribed to the right,