

## MLVA (VNTR) 情報の集計・可視化システムMLVA-mateの開発

南須原 亮<sup>a</sup>, 灘岡 陽子<sup>a</sup>, 草深 明子<sup>a</sup>, 石川 貴敏<sup>a</sup>, 安中 めぐみ<sup>b</sup>, 三宅 啓文<sup>b</sup>, 横山 敬子<sup>b</sup>, 貞升 健志<sup>c</sup>

Multiple Locus Variable-number tandem repeat Analysis (MLVA) 法やVariable-Number Tandem Repeat (VNTR) 法は菌株の遺伝子型を数値の配列として表す方法である。この手法においては、菌株の同一性の判断に数値配列の一致を利用する。MLVA法やVNTR法で得られるデータについて専用ソフトウェアを使用することなく、一般的に入手可能なソフトウェアのみでデータを集計し、解析結果を可視化する独自システムを開発した。本システムを利用することにより、MLVA法やVNTR法における一致例の検出や、集積データにおけるクラスターの傾向をminimum spanning tree (MST) 法により解析可能であることが示された。

**キーワード** : MLVA, VNTR, 腸管出血性大腸菌, 結核菌, minimum spanning tree

### 緒 言

従来、腸管出血性大腸菌の分子疫学解析には、PFGE (Pulsed Field Gel Electrophoresis) 法が用いられてきたが、検査日数がかかることや、散在的集団発生事例 (diffuse outbreak) などの場合に菌株の授受なしでは自治体間での比較が難しいこと等の欠点があった。2017年、関東で発生した系列総菜店で購入した総菜による腸管出血性大腸菌 O157 事例を契機に、2018年、腸管出血性大腸菌の遺伝子型別検査を反復配列多型解析法 (Multiple Locus Variable-number tandem repeat Analysis : MLVA) に統一し、情報共有するための手順が厚生労働省より示された (平成 30 年 6 月 29 日付事務連絡「腸管出血性大腸菌による広域的な感染症・食中毒に関する調査について」)。

一方、2011年に改正された結核に関する特定感染症予防指針 (平成 19 年厚生労働省告示第 72 号) に「分子疫学調査の積極的実施」が盛り込まれて以降、反復配列多型 (Variable-Number Tandem Repeat : VNTR) 分析が結核対策に活用されている。この分析方法を菌の異同判定だけでなく、実地疫学調査の結果と組み合わせることで、未知の感染経路の発見にも役立つことが期待されている。

このような状況の中、VNTR 法や MLVA 法のデータの増大に伴い、データの一元管理や一致度の分析、さらにはデータ全体の関連をとらえるために可視化を簡便に行える環境が必要とされるようになってきた。データは単なる数値の配列であるため、どのような特徴を持っているのかを読み取ることは難しいが、minimum spanning tree (以下 MST という) と呼ばれるグラフを作成することによって、全体像の把握が容易になる。

これらを統一的に管理・分析できるソフトウェアとして BioNumerics<sup>TM</sup> (Applied Maths 社) <sup>1)</sup> が知られているが、高価であり、利用可能な研究室は限られている。また、イ

ンターネット上で利用可能なフリーの解析・描画アプリケーション・サービス <sup>2)</sup> もあるが、データ管理の問題やわかりやすいマニュアルがない点などの難点がある。そこで、フリーで入手可能なソフトウェアのみで、検査現場において容易に分析および可視化するシステム MLVA-mate (MLVA mathematical and analytical tool for evolution) を開発したので報告する。

### MLVA-mateシステムの概要

#### 1. 構成

本システムはMicrosoft Excel, Webブラウザ, Data-Driven Documents (D3.js) <sup>3)</sup>, グラフ生成スクリプトから構成される。構成の概略を図1に示した。D3.jsはインターネット上で公開されているBSDライセンスの情報可視化ライブラリであり、無償で使うことができる。

利用者はPCに本システムをダウンロードして使用する。ネットワークドライブ上に設置すれば、複数PCからの使用も可能である。

プログラムは、Excelに付属するVisual Basic for Applications (VBA) と、D3.jsを利用するためのJavaScriptで作成した。動作環境はMicrosoft Windows 8.1, Excel 2010, Internet Explorer 11を推奨する。

#### 2. 入力データ

本システムに入力するデータを図2に示した。

菌株データのカテゴリ (属性) は、感染地域や感染経路、薬剤耐性等疫学情報の特徴をカテゴリ化したものとする。これらの情報は、遺伝子型別の表やグラフを表示する際に、必要に応じて使うことができる。

遺伝子型データは、既に遺伝子型別の名称が付けられている場合に使用する。例えば、結核菌では ST11, 腸管出

<sup>a</sup> 東京都健康安全研究センター企画調整部健康危機管理情報課  
169-0073 東京都新宿区百人町3-24-1

<sup>b</sup> 東京都健康安全研究センター微生物部病原細菌研究科

<sup>c</sup> 東京都健康安全研究センター微生物部

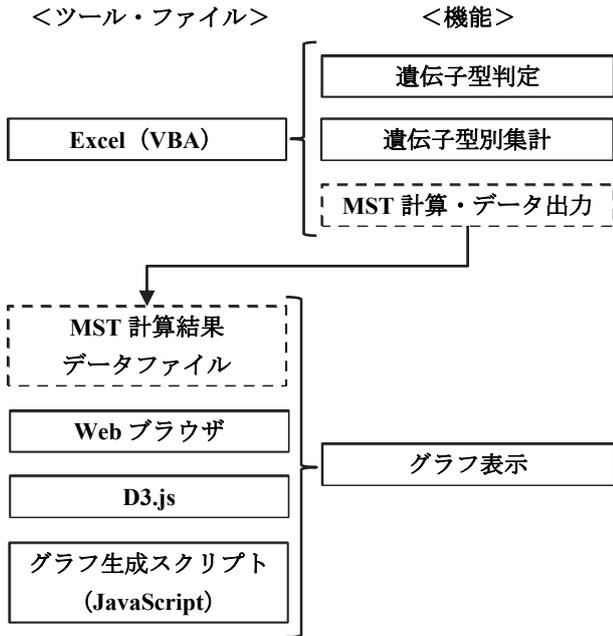


図 1. システムの構成



図 2. 入力データ

血性大腸菌では 18m0095 のように、広域事例などで同定された菌などが考えられる。

MLVA (VNTR) 法の解析領域数は、解析対象の細菌や血清型によって異なるため、任意の領域数に設定することができる。また近縁遺伝子型として何領域の差まで許容するかについても、設定可能である。

3. データ処理

MST とは、ノード (頂点) とエッジ (辺) から構成されるグラフで、全ての頂点が連結していて、辺の長さの総和が最小となり、かつ閉路を含まないグラフである。各遺伝子型を頂点とし、各遺伝子間の変異の度合いを辺の長さに対応させて作成した MST は、対象としている遺伝子型全ての進化の関係を一望できる表示法の一つである。

Excel により、データ集計と MST の作成に必要な計算を行い、計算結果をファイルとして出力する。なお、MST を計算する際に必要となる辺の長さは、その辺が連結する頂点に相当する遺伝子型の各領域を比較し、双方の反復数が不一致となる領域数とした。

MSTの描画はWebブラウザが備えている、Scalable Vector Graphics形式の画像描画機能を利用して行う。Excel による計算結果を元に、情報可視化ライブラリD3.js、描画方法を定義したグラフ生成スクリプトを組み合わせ、MSTを表示する。グラフ描画アルゴリズムとして、頂点に分子間力モデル、辺にバネモデルを用いて、力学的エネルギーの最小化を図る力学モデルを用いた。

4. 機能

1) 遺伝子型の判定

菌株データと遺伝子型データを入力して解析ボタンを押すと、反復数の照合が行われ、各菌株データには一致した遺伝子型名が表示される。一致した反復数が見つからない菌株データの場合には、新たな遺伝子型名を自動付与する (図3)。

菌株名	カテゴリー	遺伝子型名	D	E	F	G	H	I	J	K	L
44 MLVA-43	○○地域	Type-1	2	1	1	2	3	12	13	-2	-2
45 MLVA-44	○○地域	Type-1	2	1	1	2	3	12	13	-2	-2
46 MLVA-45	△△地域	Type-3	2	-2	1	4	-2	5	4	-2	-2
47 MLVA-46	△△地域	Type-4	2	-2	1	6	-2	11	5	-2	-2
48 MLVA-47	○○地域	Type-5	2	-2	1	4	-2	6	4	-2	-2
49 MLVA-48	○○地域	Type-6	2	-2	1	4	-2	7	4	10	-2
50 MLVA-49	××地域	unknown-0001	2	1	1	2	3	12	15	12	-2
51 MLVA-50	××地域	unknown-0002	2	-2	1	6	-2	11	5	-2	-2
52 MLVA-51	××地域	unknown-0003	2	1	1	2	3	7	35	-2	-2
53 MLVA-52	△△地域	unknown-0004	2	1	1	2	5	7	12	9	17
54 MLVA-53	△△地域	unknown-0005	2	-2	1	4	-2	5	4	13	-2
55 MLVA-54	○○地域	unknown-0006	2	-2	1	4	-2	6	4	-2	-2
56 MLVA-55	××地域	unknown-0007	4	1	5	2	-2	12	9	-2	3
57 MLVA-56	××地域	unknown-0008	2	1	1	2	3	8	21	10	-2
58 MLVA-57	××地域	unknown-0009	2	-2	1	4	-2	6	4	7	-2
59 MLVA-58	△△地域	unknown-0010	2	-2	1	4	-2	7	4	10	-2
60 MLVA-59	○○地域	Type-1	2	1	1	2	3	12	13	-2	-2
61 MLVA-60	○○地域	Type-1	2	1	1	2	3	12	13	-2	-2
62 MLVA-61	○○地域	Type-3	2	-2	1	4	-2	5	4	-2	-2
63 MLVA-62	○○地域	Type-4	2	-2	1	6	-2	11	5	-2	-2
64 MLVA-63	○○地域	Type-5	2	-2	1	4	-2	6	4	-2	-2

図 3. 遺伝子型判定の例

2) 遺伝子型別の集計

遺伝子型ごとの菌株数、属する菌株名、近縁遺伝子型名の集計表 (図4)、およびカテゴリー別に菌株数の集計表 (図5) が生成される。

遺伝子型名	菌株数	菌株名	近縁遺伝子型名
Type-6	7	MLVA-24, MLVA-30, MLVA-36, MLVA-42, MLVA-48, MLVA-64, MLVA-93	
Type-7	2	MLVA-85, MLVA-94	
Type-8	2	MLVA-86, MLVA-95	Type-3, unknown-0014
Type-9	2	MLVA-87, MLVA-96	
Type-10	2	MLVA-88, MLVA-97	
Type-11	2	MLVA-89, MLVA-98	
Type-12	2	MLVA-70, MLVA-99	
unknown-0001	1	MLVA-49	Type-1
unknown-0002	1	MLVA-50	Type-4
unknown-0003	1	MLVA-51	Type-1
unknown-0004	1	MLVA-52	
unknown-0005	1	MLVA-53	
unknown-0006	1	MLVA-54	
unknown-0007	1	MLVA-55	unknown-0021
unknown-0008	1	MLVA-56	
unknown-0009	1	MLVA-57	
unknown-0010	1	MLVA-58	
unknown-0011	2	MLVA-71, MLVA-100	
unknown-0012	1	MLVA-72	

図 4. 遺伝子型別集計表の例

遺伝子型名	〇〇地域	××地域	△△地域	カテゴリなし
Type-1	18	6	2	
Type-2		5	1	
Type-3	2	2	3	
Type-4	2	2	3	
Type-5	3		4	
Type-6	5		2	
Type-7	1		1	
Type-8		1	1	
Type-9		1	1	
Type-10		1	1	
Type-11			2	
Type-12	1		1	
unknown-0001		1		
unknown-0002		1		
unknown-0003		1		
unknown-0004			1	
unknown-0005			1	
unknown-0006	1			
unknown-0007		1		
unknown-0008		1		
unknown-0009		1		

図 5. 遺伝子型別カテゴリー表の例

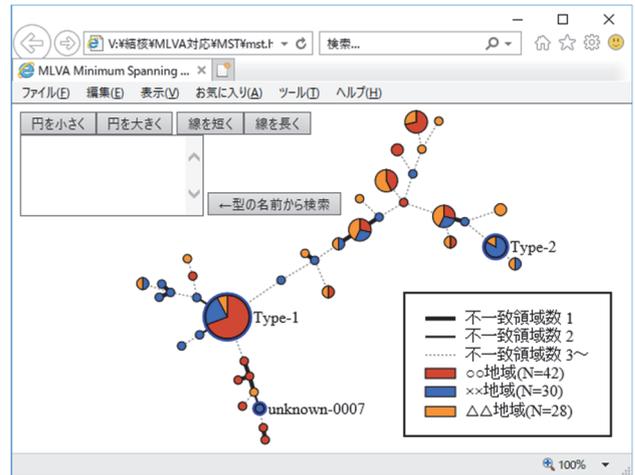


図 6. minimum spanning tree の例

### 3) MSTによる可視化

BioNumerics<sup>TM</sup>)を参考にしたMSTを表示する(図6)。各遺伝子型は円で表わされ、菌株数が多いほど円は大きくなり、不一致領域数が少ない程、近くに配置される。菌株データにカテゴリーが入力されていた場合は、カテゴリー別に色分けした円グラフとなる。

辺は遺伝子型間の不一致領域数に応じて線の種類や太さが変わる。

円をクリックすることによって当該遺伝子型名を表示させたり、マウスのホイール操作で図形の拡大・縮小が可能である。また、遺伝子型名を入力することにより遺伝子型の位置を検索することもできる。

### 5. 本システムの実用例

#### 1) 結核菌VNTR解析

東京都結核菌検査事業によって、2011年4月から2018年5月に当センターに搬入された菌株のうち、VNTR型別検査を実施した922株について、遺伝系統、ストレプトマイシン(SM)耐性の有無をカテゴリーとして本システムに入力しMSTを作成した。

(1) 遺伝系統 結核菌は北京型(新興型と祖先型)、非北京型に分類されることが知られているが、VNTR各領域の傾向から最大事後確率(maximum a posteriori: MAP)推定法によって、祖先型北京株がさらに4系統(ST11/26, STK, ST3, ST25/19)に分けられ、全体で6系統に分類で

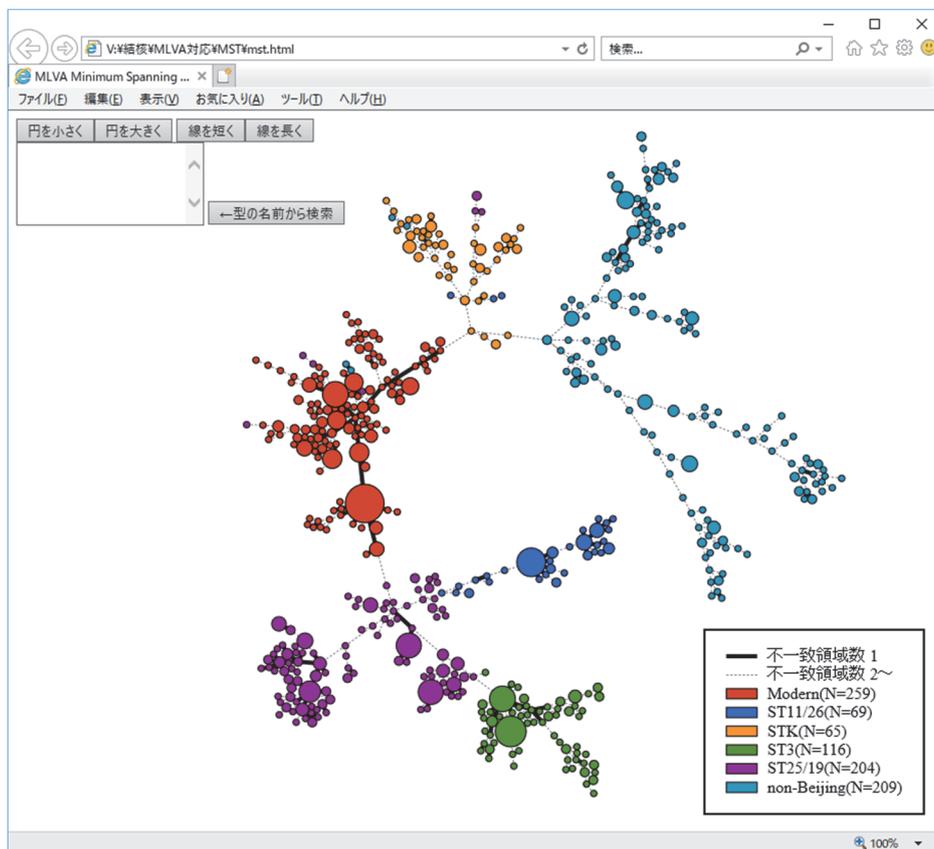


図 7. 結核菌 MAP 推定法による遺伝系統を示した minimum spanning tree

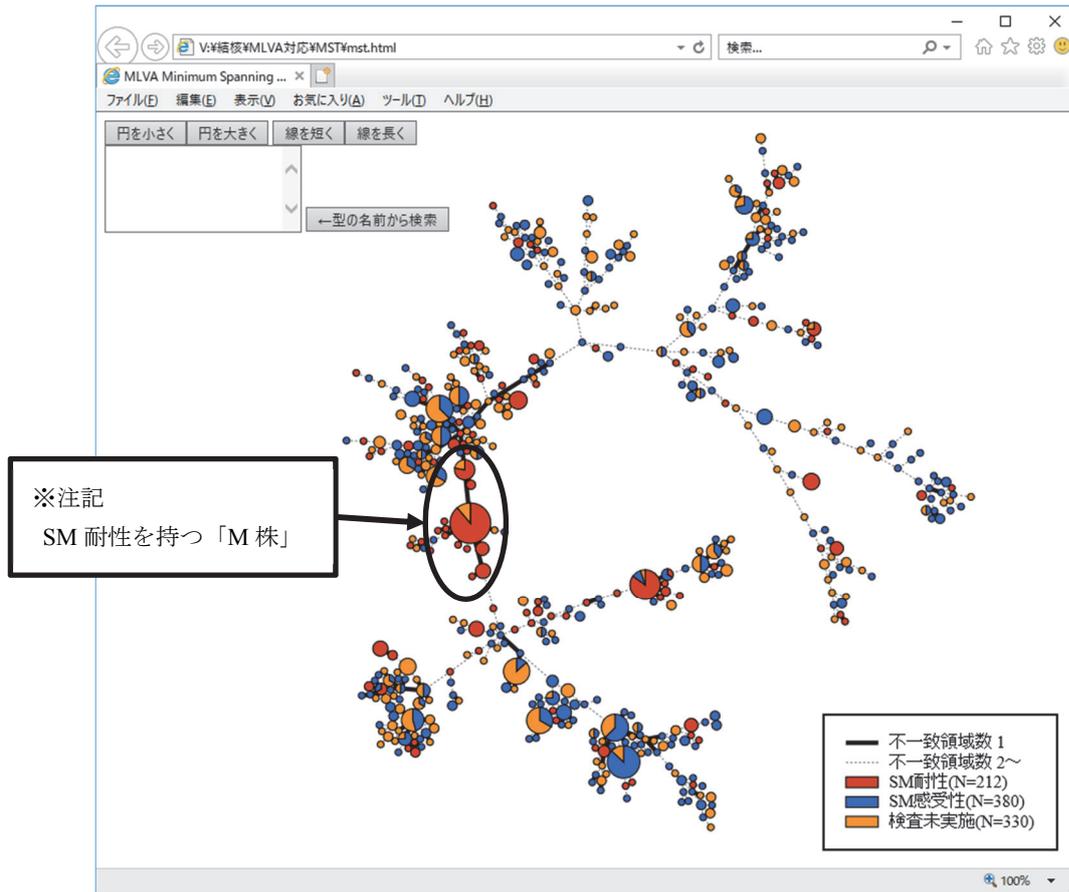


図 8. 結核菌 SM 耐性を示した minimum spanning tree

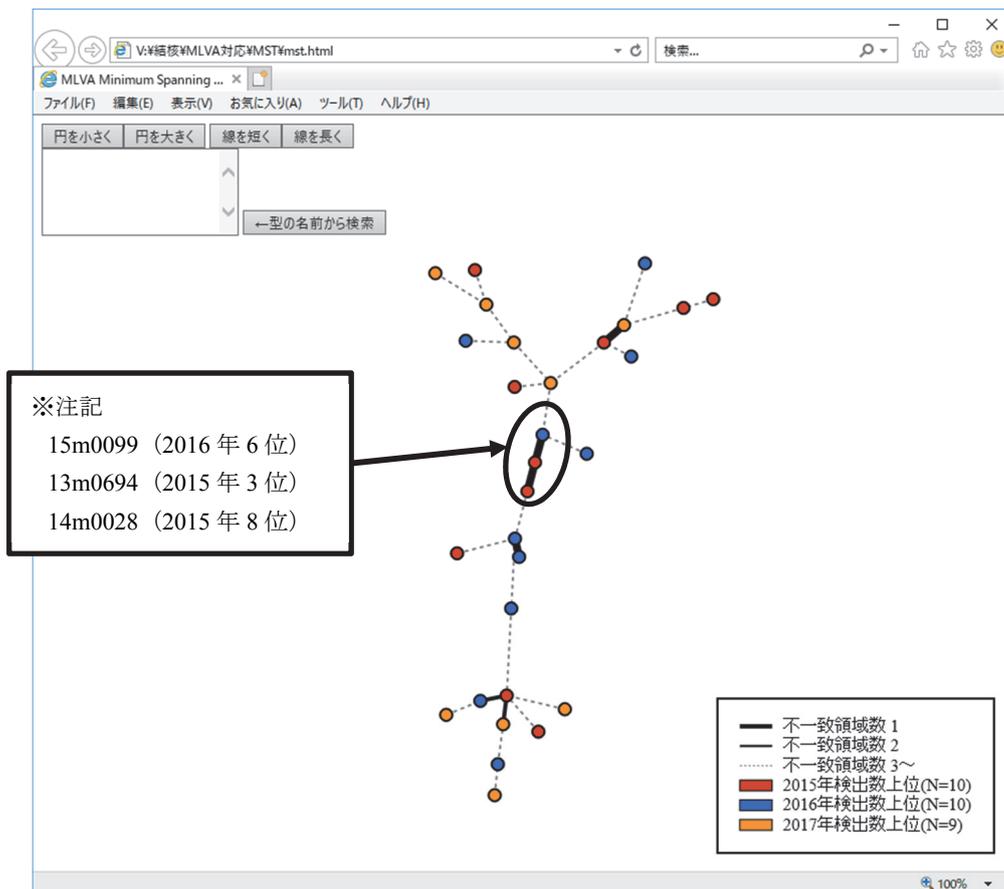


図 9. 腸管出血性大腸菌 2015~2017 年検出数上位を示した minimum spanning tree

きることが報告されている<sup>4)</sup>。当センターの検査結果を図7に示した。ほぼ遺伝系統ごとに区別できた。

(2) **SM耐性結核菌** 結核菌VNTRデータにSM耐性の有無をカテゴリーとして加え、SM耐性、SM感受性、薬剤感受性検査未実施の3種類に分類した。図8に結果を示す。SM耐性を持ち、菌株数が最も多い型を確認したところ、全国で検出されているいわゆる「M株<sup>5)</sup>」であった。

## 2) 腸管出血性大腸菌MLVA解析

国立感染症研究所による2015年から2017年の集団事例関連株、広域株等のうちMLVA型別検出数が多かった上位9～10株<sup>6-8)</sup>を年別に分類してMSTを作成した(図9)。その結果、2015年検出数3位のMLVA型13m0694と2016年6位の15m0099は不一致領域数が1であった。このように、ある年に多数分離されたMLVA型の近縁型が、別の年でも多数分離される場合があることが判明した。

## 今後の課題

本システムは、TAIMS(東京都高度情報化推進システム)ネットワーク上のファイルサーバに保存しており、TAIMS端末でも実行可能である。本システムの課題は以下の通りである。

①ExcelやWebブラウザの機能に依存しているため、プログラムのバージョンによっては正常に動作しない可能性がある。

②MSTの計算量はデータ数の二乗に比例するため、集計対象の遺伝子型が増加すると出力に時間がかかる。

③グラフは辺の数が多いと、位置計算の結果が期待した値にならず辺同士が交差して表示されてしまう場合がある。ただし、マウスで円をドラッグすることにより位置調整が可能であるため、数か所程度の問題であれば、手動にて回避できる。

④グラフの辺の長さは、不一致となる領域数が少ないほど短くなるように計算されるが、周囲に頂点が多いと、頂点間の反発力が強く作用し、辺が長くなる場合がある。

⑤比較する領域がどちらも未定義の場合、一致しているとみなしているが、不一致と判定するようなアルゴリズムを検討する必要がある。

今後、地方衛生研究所ネットワークWebサイト(URLは<https://www.chieiken.gr.jp/>)で本システムを提供する予定である。

## ま と め

MLVA法やVNTR法による大量データの一元管理、一致度の分析およびMSTを計算・描画するシステムMLVA-mateを開発した。全てフリーで入手できるソフトを利用している。TAIMS端末で実行可能であり、十分実用に耐えうる。

## 文 献

- 1) Applied Maths: BioNumerics, <http://www.applied-maths.com/bionumerics> (2018年8月10日現在, なお本URLは変更または抹消の可能性がある)
- 2) Gonçalves, B., Carriço, J.A., Francisco, A.P., *et al.*: PHYLOViZ Online, <https://online.phyloviz.net/index> (2018年8月10日現在, なお本URLは変更または抹消の可能性がある)
- 3) Bostock, M.: D3.js - Data-Driven Documents, <https://d3js.org/> (2018年8月10日現在, なお本URLは変更または抹消の可能性がある)
- 4) Seto, J., Wada, T., Iwamoto, T., *et al.*: *Infect Genet Evol*, **35**, 82-88, 2015.
- 5) Wada, T., Iwamoto, T., Tamaru, A., *et al.*: *PLoS ONE*, **10**(3), e0118495, 2015.
- 6) 泉谷秀昌, 石原朋子, 李 謙一, 他: 病原微生物検出情報, **37**, 93-95, 2016.
- 7) 泉谷秀昌, 石原朋子, 李 謙一, 他: 病原微生物検出情報, **38**, 100-101, 2017.
- 8) 泉谷秀昌, 李 謙一, 伊豫田 淳, 他: 病原微生物検出情報, **39**, 81-82, 2018.

**Development of an Aggregation and Visualization System (MLVA-mate) for MLVA and VNTR Sequence Typing**

Ryo NASUHARA<sup>a</sup>, Yoko NADAOKA<sup>a</sup>, Akiko KUSABUKA<sup>a</sup>, Takatoshi ISHIKAWA<sup>a</sup>,  
Megumi ANNAKA<sup>a</sup>, Hirofumi MIYAKE<sup>a</sup>, Keiko YOKOYAMA<sup>a</sup> and Kenji SADAMASU<sup>a</sup>

Multiple Locus Variable-number Tandem Repeat Analysis (MLVA) and Variable-Number Tandem Repeat (VNTR) are molecular typing methods for numerically expressing gene sequences of bacterial species. In these methods, bacterial species with matching sequence numbers are interpreted as the same strain. We have developed a proprietary system that could aggregate and visualize data from MLVA or VNTR sequence typing without using dedicated software. Researchers can process the data using versatile commonly used software. Our system allows not only the detection of genomic matches but also the analysis of cluster trends from accumulated data using Minimum Spanning Tree method.

**Keywords:** MLVA, VNTR, Enterohemorrhagic *Escherichia coli* (EHEC), *Mycobacterium tuberculosis*, Minimum Spanning Tree

---

<sup>a</sup> Tokyo Metropolitan Institute of Public Health  
3-24-1, Hyakunin-cho, Shinjuku-ku, Tokyo 169-0073, Japan