

ウイルスを主とした広域事例調査手法の検討

研究分担者 野田 衛 (国立医薬品食品衛生研究所)

研究要旨

食品媒介ウイルスによる散在型集団食中毒事例(diffuse outbreak)を早期に探知するためには、各事例の患者等から検出されたウイルスの塩基配列を比較して、株間の異同性を調べるのが有用である。また、ウイルスによる食中毒が疑われる場合、調査において、食中毒か否かの判断や調理従事者からの二次汚染の有無の判断などを行うためには、患者、食品あるいは調理従事者から検出されたウイルスの遺伝子型や検出株間の相同性を調べるのが重要である。しかしながら、これらの解析を行政レベルで実施している自治体は多くないのが現状である。そこで、本研究では迅速な遺伝子型別や系統樹解析を行うために、エクセルファイルをデータソースとするローカルBlast検索システムおよびClustalWを基本プログラムとした自動遺伝子型別システムを開発してきた。今年度は、過去2年に開発したプログラムについて、以下の点を改良した。①Blast検索において、E-valueが最大なもの(E-valueが同値のものがあれば%identicalが最大なもの)を戻す機能の追加。②Blast検索において、ミスマッチが0のものをすべて戻す機能の追加。③ClustalWにおいて、Queryに指定した株ごとにReferenceに指定した株に対してClustalWを実行するプログラムを追加。最終バージョンは今後NESFD等をとおして公開し、必要な自治体に対してプログラムを配布する予定である。本システムは食中毒調査等における迅速な遺伝子型別や系統樹解析の実施に寄与するものと思われる。

研究協力者

北海道立衛生研究所	吉澄 志磨	東京都健康安全研究センター	小田 真悠子
青森県環境保健センター	筒井 理華	東京都健康安全研究センター	木本 佳那
青森県環境保健センター	菩提寺 誉子	東京都健康安全研究センター	秋場 哲哉
青森県環境保健センター	坂 恭平	新潟県保健環境科学研究所	田村 務
岩手県環境保健研究センター	佐藤 直人	新潟県保健環境科学研究所	新井 礼子
岩手県環境保健研究センター	梶田 弘子	新潟市衛生環境研究所	山本 一成
仙台市衛生研究所	関根 雅夫	新潟市衛生環境研究所	田岡 悠央
福島県衛生研究所	北川 和寛	富山県衛生研究所	名古屋 真弓
福島県衛生研究所	富田 望	富山県衛生研究所	稲崎 倫子
福島県衛生研究所	金成 篤子	福井県衛生環境研究センター	佐藤 かおり
群馬県衛生環境研究所	塚越 博之	山梨県衛生環境研究所	大沼 正行
千葉県衛生研究所	堀田 千恵美	長野県環境保全研究所	中沢 春幸
千葉県衛生研究所	秋田 真美子	長野県環境保全研究所	粕尾 しず子
千葉市環境保健研究所	西川 和佳子	長野県環境保全研究所	水澤 哲也
千葉市環境保健研究所	坂本 美砂子	三重県保健環境研究所	楠原 一
千葉市環境保健研究所	山崎 恵美	滋賀県衛生科学センター	石川 和彦
東京都健康安全研究センター	宗村 佳子	大阪市立環境科学研究所	入谷 展弘

大阪市立環境科学研究所	山元 誠司
堺市衛生研究所	三好 龍也
島根県保健環境科学研究所	辰巳 智香
広島市衛生研究所	藤井 慶樹
福岡県保健環境研究所	小林 孝行
福岡県保健環境研究所	芦塚 由紀
福岡市保健環境研究所	財津 修一
福岡市保健環境研究所	古川 英臣
福岡市保健環境研究所	松藤 貴久
大分県衛生環境研究センター	加藤 聖紀
沖縄県衛生環境研究所	加藤 峰史
国立医薬品食品衛生研究所	上間 匡
国立医薬品食品衛生研究所	三元 昌美

(順不同)

A. 研究目的

ノロウイルス、A型肝炎ウイルス、E型肝炎ウイルス等の食品媒介性ウイルスによる感染症は毎年発生し、国民に多くの健康被害をもたらしている。多くの食品媒介ウイルスは食品を介さないヒト-ヒト感染も起こすことから、ウイルスによる食中毒が疑われる集団事例が発生した場合、行政対応上、食中毒か感染症かの判断が求められる。また、調理従事者等からウイルスが検出された場合、汚染源としての調理従事者等の関与の有無についても判断する必要がある。これらの行政対応上の判断を行うためには、聞き取りによる疫学調査が重要であることは言うまでもないが、患者、食品あるいは調理従事者等から検出されたウイルスの遺伝子型や検出株間の相同性等に関する科学的根拠も重要な役割を担うと考えられる。しかしながら、実際の食中毒調査において、上記のような行政判断が求められる時点において、シークエンス検査およびその結果に基づくそれらの解析を実施している自治体は多くないのが現状である。この要因としては、人員数、業務量等いくつか理由が考えられるが、得られたシークエンスデータのアライメント、系統樹解析などを専用ソフトを用いて手作業で行う解析作業の煩雑さは重要な要因であると思われる。

昨年度までの本研究において、ウイルス

による散在型集団食中毒事例の早期探知や迅速な遺伝子型別を行うために、ローカルデータベースを対象にしたBLAST検索システムおよび遺伝子のアライメントや近隣結合法による系統樹解析のプログラムとして広く利用されているClustalWを基本プログラムとした自動遺伝子型別システムを開発した。本年度は、それらのプログラムにおける問題点を抽出して、それらを改良し、より利用しやすいものを構築することを目的とした。

B. 研究方法およびC. 研究結果

(倫理面への配慮)

倫理面で配慮すべき事項はない。

1 昨年まで開発したプログラムの問題点

① ローカルBlast検索

Blast検索においては、同プログラムにおけるアルゴリズムに従いScore(Bits)の高い順に表示されるシステムになっている。一昨年開発したプログラムでは、Scoreが最も高い株をエクセルファイルに戻り値とする仕様となっていた。試験運用において、検索対象であるReference株の中に塩基配列が一致する株が存在しているにも関わらず、エクセルに還元される株は一部塩基配列が異なるものが還元される場合が認められた。この原因を調べたところ、Blast検索においては、Scoreの演算に比較する塩基配列の長さが影響することが判明した(たとえば、比較する塩基長が100塩基で1塩基に違いがある場合と、塩基長が50塩基で違いがない場合は、前者のScoreが高くなるというようなこと)。また、比較するデータにギャップが生じた(どちらかの株に比較対象となる塩基配列のデータが欠損している)場合、比較可能なデータ間で塩基配列が100%一致した場合でも、%Identitesは100%とはならないことも判明した。塩基配列データは、PCR増幅部位の間のすべてのデータがそろっていることが望ましいが、実際にはすべてのデータがそろっていない場合もあり得る。また、シークエンサーの解読により、

本来あるべき塩基が読まれておらず、1塩基データが欠損しているという場合も少なくない。一方、広域事例の探知には塩基配列が一致する株を迅速に把握することが重要である。これらのことから、ローカルBlast検索において、ミスマッチが0（ギャップは無視される）のものは、条件設定した件数内において、すべての株名を表示する機能とその件数を表示する機能を追加することにした。

また、従来、Score(=Bits)が最も高いものを戻り値とする方法に加えて、E-valueが最も高いものを戻り値とする（仮にE-valueが同じものが複数ある場合は、%Identitiesが最も高いものを戻り値とする）メニューを追加することとした。なお、ミスマッチ=0の株名と件数の表示は、従来のScoreのTop1を戻す場合とE-valueがTop1を戻す場合のいずれにおいても、実行される仕様とした。

以上の変更の概要を図1と図2に示した。また、Blast検索で得られた結果を別シートにリンク貼り付けで作成した一覧表の例を図3に示した。

② ClustalWの条件設定

ClustalWにおける遺伝子の距離計算は、ライメントされた塩基配列データにおいて、比較するすべて株のデータがそろっている部位のデータのみが利用されている。したがって、分析に用いた株の中に塩基配列データが短いものが含まれている場合、最も短い塩基配列のデータの範囲内において比較が行われるため、データの解析精度が低下することになる。その場合、参照株との距離計算が不正確なものになり、場合によっては系統樹を作成した場合、本来とは異なる分岐の仕方を示す場合が認められる。

そこで、問い合わせ株(=Query)に指定された株ごとに参照株(=Rerence)に指定された株に対してClustalWを実行し、遺伝子型別と遺伝子距離計算を行うシステムを追加することとした(図4)。これにより、Reference株との遺伝子的距離を、Queryに

指定する株ごとに正確に求めることが可能となった。ただし、このメニューにおいては、Queryに指定された株ごとにClustalWが実行されるので、システムの実行に時間が長くなることになる。

また、ClustalWでQueryに指定する株の最低長を指定するパラメータを追加した。この指定された塩基長を満たさない株がQueryのデータに存在した場合、その株は解析対象から除外されることになる。

D. 考察

一般に、調理従事者が関連した食中毒事例においては、患者から検出されたノロウイルスと調理従事者から検出されたノロウイルスの塩基配列が一致することが、調理従事者の関与を示唆する根拠となる。また、広域事例においては、ある事例の患者からの検出株と他の事例の患者からの検出株の塩基配列が一致することが、両事例の関連性を示す根拠となる。すなわち、食中毒調査においては、塩基配列が一致する株を簡便に把握することは、きわめて重要な点である。今回、改良したBlast検索では、ミスマッチ=0の株をすべて還元するメニューを新たに追加した。このことにより、Referenceに指定した株の中に塩基配列が一致した株が何株含まれているのか、また、その株名はなになのかを速やかに把握することが可能となった。ただし、一致する株名と件数はBlast検索のパラメータ設定における、「EntryHit」で上限を設定することとなっている。網羅的ゲノム解析など膨大なデータを扱う場合は、この設定値を高くするとプログラムの実行により時間が必要となる場合も想定されるので、注意が必要である。また、いうまでもないが、塩基配列データの中には、データの読み間違いがある場合には、ミスマッチ=0とはならないので、Blast検索のログファイルである「*.bln」をテキストデータとして開き、実際の塩基配列の比較画面でデータの確認を行うことが望まれる。

ClustalWにおいては、結果で述べたように、解析に用いた株の中に塩基配列のデー

タが短いものが含まれている場合、参照株との距離計算が不正確なものになる危険性があり、場合によっては系統樹を作成した場合、本来とは異なる分岐の仕方を示す場合が認められていたことから、Queryに指定された株ごとにReference株に指定された株に対してClustalWを実行し、遺伝子型別と遺伝子距離計算を行うシステムを追加した。現在、RIVMが提供しているNorovirus Genotyping Tool (<http://www.rivm.nl/mpf/norovirus/typingtool>)は基本的にこの方法を採用している。通常はこの機能を使用することは多くはないが、得られた系統樹の正確性を確認したい場合や株（データ）ごとに正確な遺伝的距離を計算したい場合などに利用するとよいと思われる。

3年間をかけてエクセルファイルをデータソースとして利用するローカルBlast検索とClustalWを基本プログラムとする遺伝子型別システムを開発した。本システムは食中毒調査等における迅速な遺伝子型別や系統樹解析の実施に寄与するものと思われる。

課題としては、以下の点があげられる。

- ① 本システムは、64bitのWindows (Windows7, Windows8, Windows10)をOSとするパソコン上で、エクセル(64bit) (拡張子は基本的に、「xlsx」)で作動する。今後、パソコンやソフトのバージョンアップにより、作動しなくなる可能性がある。
- ② より汎用性を持たせるためには、CSV形式のデータへの対応が望ましい。
- ③ データの有効利用や塩基配列情報の共有化を推進し、行政対応での利用を一般化するためには、Webベースでのシステムの構築が望ましい。

また、これまで地方衛生研究所の協力のもと、試行的にデータの共有と還元を実施してきたが、シーケンス検査の一般化や広域事例の早期探知等の食中毒調査におけるシーケンスデータの利用を行政レベルで実施可能とするためには、食中毒事

例あるいは食中毒疑い事例のシーケンスデータの共有を行政レベルで実施できるシステムの構築が必要である。

E. 結論

ウイルスによる散在型集団食中毒事例の早期探知のための相同性解析や食中毒検査における迅速な遺伝子型別検査を実施可能とするために、ローカルデータベースを対象としたBLAST検索システムおよびClustalWを基本プログラムとした自動遺伝子型別システムの問題点を把握し、その改良を行った。

F. 健康危険情報

なし。

G. 研究発表

1. 論文発表
なし。
2. 学会発表
なし。

H. 知的財産権の出願・登録状況

1. 特許取得
なし。
2. 実用新案登録
なし。
3. その他
なし。



システム	昨年までの機能	変更後の機能
Blast	Top1(Score)を表示	(1) Top1を表示 (2) E-valueが最も良い、E-valueが同値が複数あれば%identicalが最もよいものを表示 (3) (1)(2)いずれでも、ミスマッチ0の株をすべて表示
ClustalW	Queryに登録された株をまとめて解析	(1) Queryに登録された株をまとめて解析 (2) Queryに登録された株を個別に解析

図1 システムの改良点とメニュー画面

改良後のフォーマット

SeqID	株名	シーケンス	ウイルス	類似株名(Blast)	遺伝子型	Score	Expect	Identities	Strand	片数	一致株名(Blast)	一致片数
1	1612-441-O-160902	TTTGGGAGGGGGGAT ATCGCAATGTGGGTC CCAAATTTGTGAATG AAGATGGGCTGGAAT GAGGGGDTGCATG TAATGATGGTGGAGG GGGGTGGTGGGAGC AAAGTAAACAATGAGG TCATGGCTTTGAAC CCGTTGGGAGGGGAT TCTTGGGAGGGGAT	NoV	1612-272-F-16099	株2	584	7.00E-170	318/319 (99.69%)	Plus/Plus	6	1612-272-F-16099, 1701-402-F-20172024, 1612-101-F-180332, 1611-241-F-D10N31, 1611-241-F-C0N24, 1611-241-F-C7N23, 1611-241-F-C5N15, 1610-241-F-C3N6	6
2	1612-441-O-161002	TTTGGGAGGGGGGAT ATCGCAATGTGGGTC CCAAATTTGTGAATG AAGATGGGCTGGAAT GAGGGGDTGCATG TAATGATGGTGGAGG GGGGTGGTGGGAGC AAAGTAAACAATGAGG TCATGGCTTTGAAC CCGTTGGGAGGGGAT TCTTGGGAGGGGAT	NoV	1612-272-F-16099	株2	584	7.00E-170	320/320 (99.69%)	Plus/Plus	6	1612-272-F-16102, 1612-272-F-16099, 1612-342-F-2166212F, 1611-241-F-C0N29, 1610-342-O-2105009F, 1609-342-O-1100385F, 1611-471-F-16FP679	6
4	1701-441-O-170202	TTTGGGAGGGGGGAT ATCGCAATGTGGGTC CCAAATTTGTGAATG AAGATGGGCTGGAAT GAGGGGDTGCATG TAATGATGGTGGAGG GGGGTGGTGGGAGC AAAGTAAACAATGAGG TCATGGCTTTGAAC CCGTTGGGAGGGGAT TCTTGGGAGGGGAT	NoV	1612-272-F-16102	株2	590	2.00E-171	319/319 (100.00%)	Plus/Plus	6	1612-272-F-16102, 1612-272-F-16099, 1612-342-F-2166212F, 1611-241-F-C0N29, 1610-342-O-2105009F, 1609-342-O-1100385F, 1611-471-F-16FP679	6

修正部分 (Top1またはE-valueが最も良い(E-valueが同値が複数あれば%identicalが最もよいもの)を表示)

追加部分 (ミスマッチ=0の株をすべて表示)

追加部分 (ミスマッチ=0の株数)

図2 システムの改良点(エクセルデータ)

	A	B	C	D
1	Blast検索により一致した株の一覧			
2	株名	遺伝子型	一致した株	件数
3	1612-441-O-160902	GII.2		0
4	1612-441-O-161002	GII.2	1612-272-F-16099,1701-402-F-20172024,1612-101-F-180GII.2,1611-241-F-C10N31,1611-241-F-C8N24,1611-241-F-C7N23,1611-241-F-C5N15,1610-241-F-C3N6	8
5	1701-441-O-170101	GII.2	1612-272-F-16099,1701-402-F-20172024,1612-101-F-180GII.2,1611-241-F-C10N31,1611-241-F-C8N24,1611-241-F-C7N23,1611-241-F-C5N15,1610-241-F-C3N6	8
6	1701-441-O-170202	GII.2	1612-272-F-16102,1612-272-F-16098,1612-342-F-2166212F,1611-241-F-C8N29,1610-342-O-2165009F,1609-342-O-1160385F,1611-471-F-16FP679	7

ある自治体の検出株

報告された検出株

図3 データの利用例:別シートでの表示
(エクセルのデータのリンク機能利用)

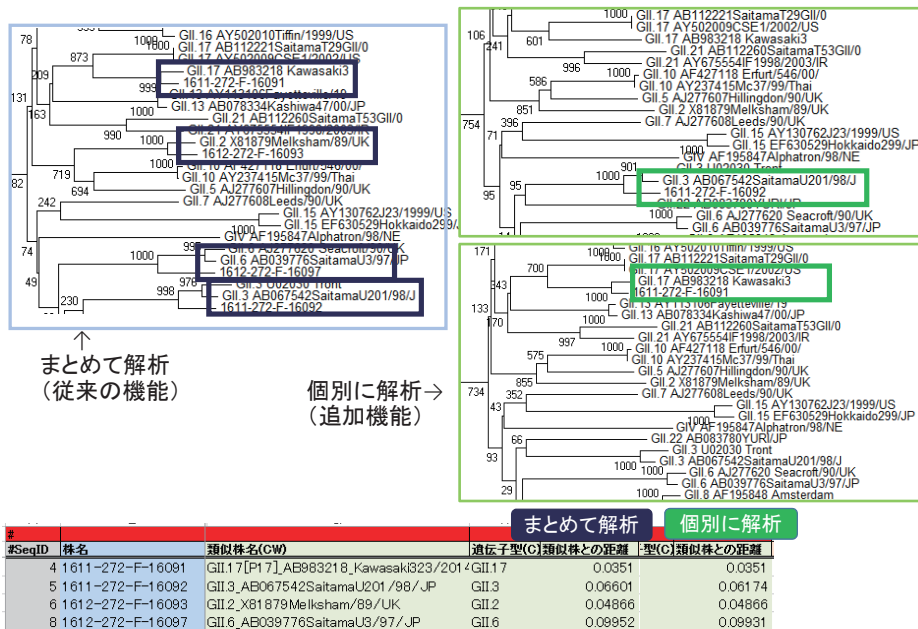


図4 システムの改良点 (ClustalW)