

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同一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、本研究では迅速な同一性解析や遺伝子型別を行うために、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た。平成26年度は、①塩基配列データをFASTA形式とエクセル形式の相互に変換するプログラム、②検索したい株(問い合わせ株)と検索の対象となる株(参照株)のそれぞれをエクセルファイルで指定することにより、Blast検索を実行し、問い合わせ株に対し類似性が最も高い株の類似度や遺伝子型等の情報を還元するプログラム、③塩基配列データをもつ2つのエクセルファイルを統合するプログラムを開発した。平成27年度は、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。平成28年度は、過去2年で開発したプログラムについて、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の追加、の改良を行った。3年間にわたり本システムを利用し、研究協力地方衛生研究所と塩基配列データの共有を試行的に実施したところ、迅速な情報還元ができた。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。本システムはウイルス性食中毒調査等の精度向上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	仙台市衛生研究所	関根 雅夫
青森県環境保健センター	筒井 理華	福島県衛生研究所	北川 和寛
青森県環境保健センター	菩提寺 誉子	福島県衛生研究所	富田 望
青森県環境保健センター	武差 愛美	福島県衛生研究所	鈴木 理恵
青森県環境保健センター	坂 恭平	福島県衛生研究所	金成 篤子
岩手県環境保健研究センター	佐藤 直人	群馬県衛生環境研究所	丹羽 祥一
岩手県環境保健研究センター	梶田 弘子	群馬県衛生環境研究所	塚越 博之
岩手県環境保健研究センター	高橋 雅輝	千葉県衛生研究所	堀田 千恵美
		千葉県衛生研究所	秋田 真美子

千葉市環境保健研究所 田中 俊光
 千葉市環境保健研究所 土井 妙子
 千葉市環境保健研究所 水村 綾乃
 千葉市環境保健研究所 西川 和佳子
 千葉市環境保健研究所 坂本 美砂子
 千葉市環境保健研究所 山崎 恵美
 東京都健康安全研究センター 森 功次
 東京都健康安全研究センター 宗村 佳子
 東京都健康安全研究センター 永野 美由紀
 東京都健康安全研究センター 小田 真悠子
 東京都健康安全研究センター 木本 佳那
 東京都健康安全研究センター 秋場 哲哉
 新潟県保健環境科学研究所 田村 務
 新潟県保健環境科学研究所 広川 智香
 新潟県保健環境科学研究所 新井 礼子
 新潟市衛生環境研究所 山本 一成
 新潟市衛生環境研究所 南波 裕太
 新潟市衛生環境研究所 田岡 悠央
 富山県衛生研究所 滝澤 剛則
 富山県衛生研究所 名古屋 真弓
 富山県衛生研究所 稲崎 倫子
 福井県衛生環境研究センター 小和田 和誠
 福井県衛生環境研究センター 佐藤 かおり
 山梨県衛生環境研究所 大沼 正行
 長野県環境保全研究所 中沢 春幸
 長野県環境保全研究所 粕尾 しず子
 長野県環境保全研究所 水澤 哲也
 三重県保健環境研究所 楠原 一
 滋賀県衛生科学センター 米谷 僚子
 滋賀県衛生科学センター 石川 和彦
 大阪府立公衆衛生研究所 左近 直美
 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 飯塚 節子
 島根県保健環境科学研究所 三田 哲朗
 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 吉富 秀亮
 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 宮代 守
 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 喜屋 武向子
 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

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ことが有用である。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ことが重要である。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、本研究では迅速な相同性解析や遺伝子型別を行うために、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することを目的とした。

B. 研究方法

(倫理面への配慮)

倫理面で配慮すべき事項はない。

1 システムの開発

3年間にわたり、以下のプログラムを開発することとした。

(1) 平成26年度

- ① 塩基配列データをFASTA形式とエクセル形式の相互に変換するプログラムの開発。
- ② 検索したい株(問い合わせ株)と検索の対象となる株(参照株)のそれぞれをエクセルファイルで指定することにより、Blast検索を実行し、問い合わせ株に対し類似性が最も高い株の

類似度や遺伝子型等の情報を還元するプログラムの開発。

- ③ 塩基配列データをもつ2つのエクセルファイルを統合するプログラムの開発

(2) 平成27年度

- ① Blast検索と同様にエクセルファイルをデータソースとし、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する自動遺伝子型別システムの開発

(3) 平成28年度

過去2年で開発したプログラムについて、以下の改良を行う。

- ① Blast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
- ② ClustalW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の追加

2 システムの有用性の検証

研究協力地方衛生研究所と食中毒事例等から検出されたノロウイルスおよびサポウイルスの塩基配列情報を共有する中で、開発したプログラムを用いて解析を行い、その有用性や問題点を検証した。

C. 研究結果

1 システムの開発

ほぼ計画どおりのプログラムを開発することができた。

作成したプログラムの操作マニュアルを資料として添付した。

2 システムの有用性の検証

3年間にわたり本システムを利用し、研究協力地方衛生研究所と塩基配列データの共有を試行的に実施した。研究協力地方衛生研究所から連絡を受けた塩基配列デ

ータに対して、その還元データ(当該報告に先立って各地方衛生研究所から報告を受けたノロウイルス等の塩基配列のデータベースに対する相同性(Blast)検索および系統樹解析)の作成作業は概ね10分以内で終了することができた。また、次世代シーケンサーで解読された塩基配列データについて本システムのBLASTシステムを利用し遺伝子型別を実施した。これらのシステムの運用をとおり、課題を抽出し、システムの改善につなげた。

D. 考察

ウイルス性食中毒の調査精度の向上に寄与するために、エクセルファイルをデータソースとする、ローカルデータに対する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た。本システムの主な利用目的を表1に取りまとめた。

本システムの利用や本システムで利用するエクセルファイルを自治体間で共有することで、以下のようなことが可能となり、食中毒調査の精度向上が期待される。

- ① 広域食中毒事例の迅速な探知：塩基配列が一致するウイルスが他の事例で検出されているか迅速に把握できる。
- ② 迅速な遺伝子型別：複雑な遺伝子型別検査を自動で、かつ迅速に実施できる。
- ③ ウイルス以外の遺伝子検査に対応：シーケンズデータがあれば、菌種鑑別、魚種鑑別等、食中毒調査に必要な遺伝子解析も可能。
- ④ 疫学情報と遺伝子情報の一元管理と共有：喫食調査等の疫学情報と遺伝子情報をエクセルファイルで一元管理でき、かつ共通の様式で自治体間で共有できる。
- ⑤ 次世代シーケンサーの解析データに対応：膨大なシーケンズデータでも参照株があればBlast検索で型別できる。
- ⑥ シーケンズ検査の普及：シーケンズ検査に対応できない自治体でも、PCR検査が可能であれば、シーケン

ス解析を業者等に委託し、そのデータを本システムで解析することで上記1-5に対応できる。

ノロウイルスの遺伝子型別については、現在、RIVMが提供するNoronetの「Norovirus Genotyping Tool」(<http://www.rivm.nl/mpf/norovirus/typingtool>)を使用することにより、FASTAファイルを用いて簡便に遺伝子型別を実施することが可能である。しかし、同システムはノロウイルスの遺伝子型別、キメラ型（遺伝子組み換え型）の同定に特化したシステムであり、他のウイルスには利用できない。また、同システムでは複数の株を解析する場合、1株ごとに系統樹解析が行われるため、各事例から検出された株間の相同性については調べることができない欠点がある。今回開発したシステムはNorovirus Genotyping Toolを補完する意味でも有用であると思われる。

一方、本システムの利用や塩基配列情報の共有等に関する課題としては、以下の点があげられる。

- ① プログラムを使用できる作業環境が限定される。
- ② 市販ソフトであるエクセル(*.xlsx)を基本的に利用するので、ソフトのバージョンアップ等で、今後作動しなくなる可能性がある。
- ③ 食中毒事例における汚染経路の調査や共通の原因食品の関与の有無の調査などにおいて、自治体間での情報共有が必要な場合、それらに関する情報の取り扱いに関する取り決めやその方法が確立されていない。
- ④ ClustalWやBLASTでの遺伝子型別について、遺伝子型別されないためのカットオフ値が十分に検討されたものではないため、参照株との遺伝的距離が長い株等については検証が必要である。

これらの課題については、今後検討する必要がある。

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株間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、エクセルファイルをデータソースとする、ローカルデータを対象とした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を開発した。

F. 健康危険情報

該当なし

G. 研究発表

1. 論文発表

入谷展弘, 山元誠司, 改田 厚, 阿部仁一郎, 久保英幸, 西尾孝之, 伯井紀隆, 大平真由, 安井典子, 榊田晴美, 細井舞子, 松本珠実, 坂本徳裕, 廣川秀徹, 半羽宏之, 野田 衛: 2014年9~11月に発生したノロウイルスによる胃腸炎集団事例について—大阪市, 病原微生物検出情報, 36, 26-27 (2015)

楠原 一, 赤地重宏, 小林隆司, 西中隆道, 小林真美, 山口江里, 岩出義人, 田沼正路, 野田 衛: ノロウイルスGII. 17型の流行とその特徴について—三重県, 病原微生物検出情報, 36, 91-92 (2015)

入谷展弘, 山元誠司, 改田 厚, 阿部仁一郎, 上林大起, 久保英幸, 野田 衛: 2014-2015シーズンに流行したノロウイルスGI. 17について, 食品衛生研究, 65(10), 7-15 (2015)

三好龍也, 内野清子, 岡山文香, 芝田有理, 吉田永祥, 小林和夫, 左近直美, 土生川洋, 田中智之, 野田 衛: 臨床検体および下水検体を用いた堺市内のA型肝炎の流行解析, 病原微生物検出情報, 36, 6-7 (2015)

入谷展弘, 山元誠司, 改田厚, 阿部仁一郎, 久保英幸, 平井有紀, 上林大起, 野田 衛, 西尾孝之: 2014-2015シーズンに大阪市内で認められたノロウイルス流行, 大阪市立環境科学研究所報告 調査・研究年報, 77, 13-16 (2015)

2. 学会発表

入谷展弘, 山元誠司, 改田厚, 阿部仁一郎, 久保英幸, 野田衛: 2013/14シーズンに大阪市内において集団胃腸炎事例から検出さ

れたノロウイルスGII.6株の分子疫学的解析, 第62回日本ウイルス学会学術集会, 横浜市, 11/10 (2014)

山本美和子, 伊藤文明, 野田衛: 広島市で検出されたA型肝炎ウイルスの分子疫学的解析, 第62回日本ウイルス学会学術集会, 横浜市, 11/10 (2014)

入谷展弘, 山元誠司, 改田 厚, 阿部仁一郎, 久保英幸, 上林大起, 野田 衛: 大阪市におけるノロウイルス GII.17 の流行状況, 第63回日本ウイルス学会学術集会, 福岡市, 11/22 (2015)

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。

表1 開発したシステムの主な利用方法

機能	使用目的	用いるエクセルファイル	
		参照株	問い合わせ株
FASTA-Excel 変換	FASTA ファイルをエクセルファイルに変換 エクセルファイルで保管されている塩基配列データを FASTA ファイルに変換		
Blast	広域事例の探知 疫学的関連性	他の事例由来株	検出株
	遺伝子型別 (特に NGS 解析)	遺伝子型別参照株	検出株
ClustalW	遺伝子型別 遺伝的距離計算	遺伝子型別参照株	検出株
	広域事例の探知 疫学的関連性	他の事例由来株	検出株
Merge	データの追加	保存株	新規検出株

V-Nus Net Tool 操作マニュアル

Ver.1.0 (2017年2月20日)

V-Nus Net Tool について

V-Nus Net Tool は、Excel ファイルを利用して、BLAST を用いたローカルなデータベースに対する相同性検索、ClustalW を用いた遺伝子型別等を自動で実行するためのシステムです。

本システムはノロウイルス等の遺伝子型別や A 型肝炎ウイルス等による広域食中毒事例(diffuse outbreak)の探知等のウイルス性食中毒の調査の精度向上を目的として開発されたものですが、塩基配列データに基づく解析であれば、その他のウイルスの他、食中毒菌や寄生虫の鑑別あるいはフグ等の魚種鑑別等にも利用できると思われる。

また、Excel ファイルを利用するので、データの管理や分析も容易に行うことができます。

食中毒調査等の遺伝子解析に広く活用いただければ幸いです。

内容

1 V-Nus Net Tool の利用にあたって	4
2 作業に必要な環境	4
3 V-Nus Net Tool のインストール	4
(1) プログラムのインストール	
(2) データファイルの保存場所	
4 V-Nus Net Tool の構成	5
5 V-Nus Net で使用する Excel ファイルについて	5
6 FASTA ファイルから Excel ファイルへの変換	8
7 V-Nus Net Tool の実行	11
(1) BLAST(相同性検索、遺伝子型別等)	12
(2) ClustalW(遺伝子型別、系統樹解析等)	19
(3) Merge(データの結合)	25
(4) パラメータの変更	27
8 BLAST プログラムの更新	28

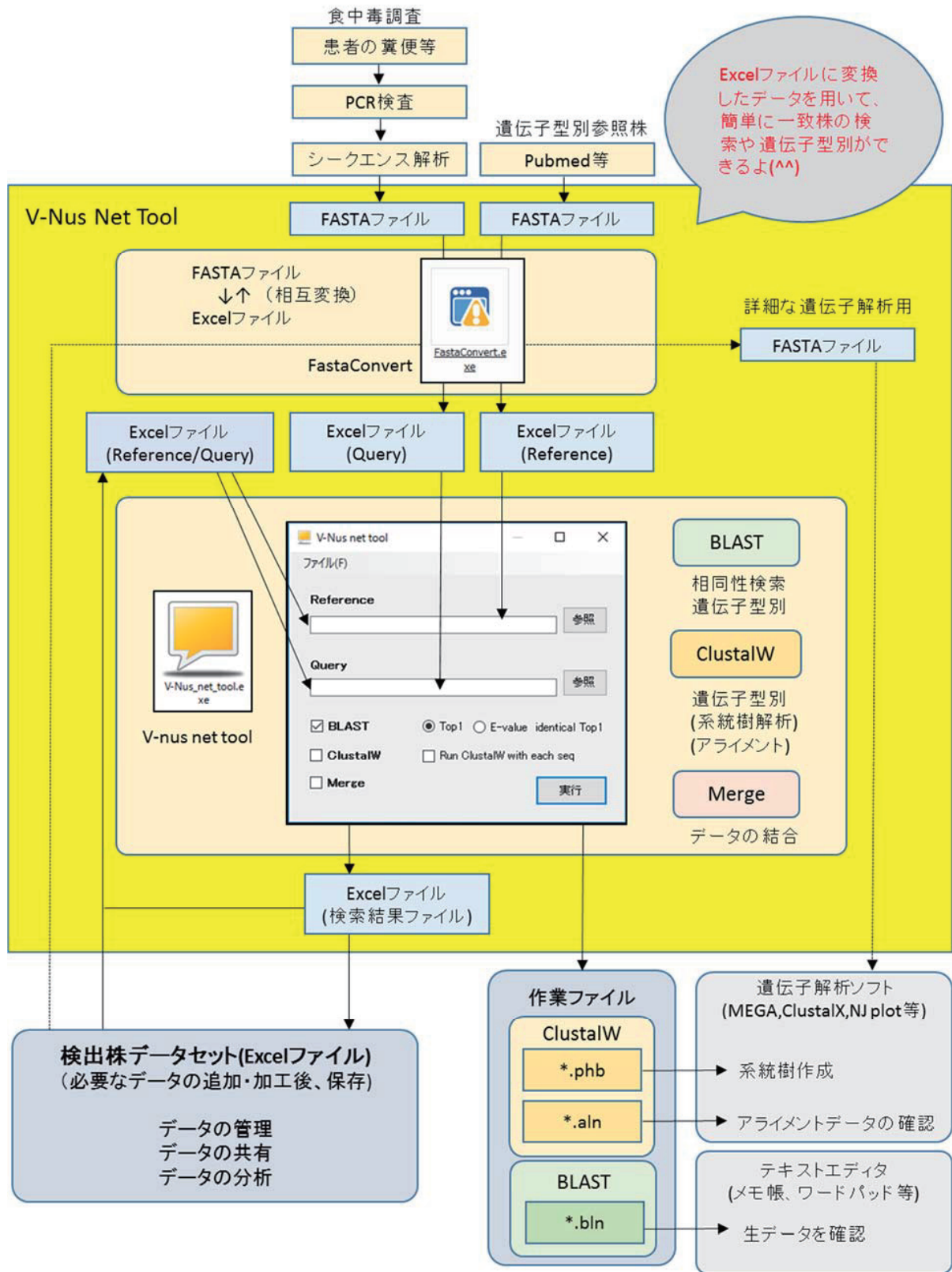


図 V-Nus Net Tool を利用した遺伝子検査の流れの概要

1 V-Nus Net Tool の利用にあたって

配布されたプログラムは、配布にあたって承諾いただいた利用方法を遵守の上、ご利用ください。

- 原則として食中毒調査関連業務等の行政検査に利用してください。
- それぞれの自治体において利用する場合は、複数の部署において複数のパソコンで利用していただき構いませんが、利用している部署およびパソコン並びに利用状況を把握しておいてください。今後、バージョンアップへの対応等のため、利用状況について照会させていただく場合があります。
- 他の自治体や外部機関へのプログラムの再配布はお控えください。

2 作業に必要な環境

本プログラムの作動には以下のハード的およびソフト的な作業環境が必要です。

- ① パソコンは 64bit マシンであること。
- ② OS は 64bit 版の Windows7、Windows8、Windows10 などであること。
- ③ Windows の 64bit 版の Excel2007、Excel2010、Excel2013 など(拡張子「*.xlsx」のファイルが利用できる)がインストールされていること。

(注) 上記を満たした作業環境においても、必ずしも作動を保証するものではありません。また、作動しない場合の対応はできません。

3 V-Nus Net Tool のインストール

(1) プログラムファイルのインストール

プログラムのインストールは、各自治体におけるセキュリティ対策等に関する規則を遵守の上、行ってください。プログラムファイルが保存されている「work」フォルダは、そのまま、Cドライブの直下にコピーしてください。

Cドライブ>work>blast_tool>(各種フォルダやファイル)

(2) データファイルの保存場所

「data」フォルダ(ノロウイルス、サポウイルス、A型肝炎ウイルスの遺伝子型別参照株データを含む Excel ファイルが含まれている)は、任意の場所にコピーしてください。

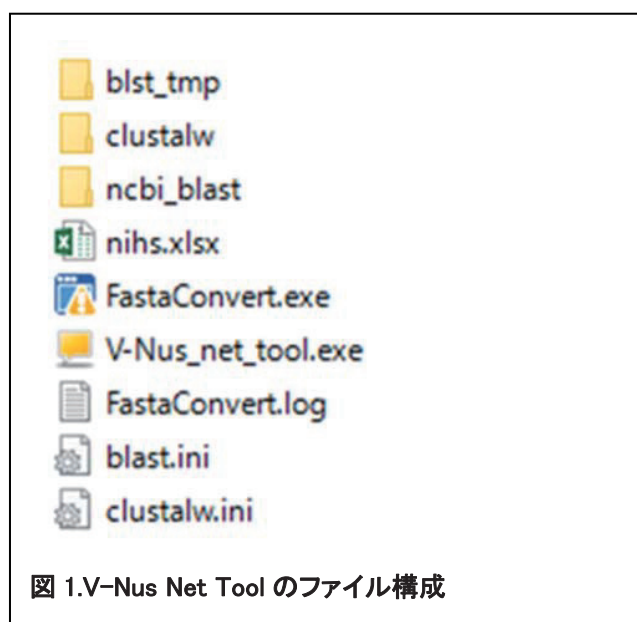
フォルダ名は、「data」から任意のフォルダ名に変更したり、サブフォルダを作成しても問題ありません。

4 V-Nus Net Tool の構成

V-Nus Net Tool は、以下の外部モジュール(=ツール)と実行を制御するプログラムで構成されています。

- ① NCBI Blast+ ver30:Blast 検索の実行プログラム
- ② ClustalW 2.0.12:ClustalW の実行プログラム
- ③ Ruby 64bit ver2.1:①②の制御プログラム
- ④ C# :①②の制御プログラム

V-Nus Net Tool の実行プログラム本体のファイル構成は図 1 のとおりです。



実際の作業において、利用者が主に使用するフォルダやファイルは以下のものです。

- ① FastaConvert.exe
FASTA ファイルから Excel ファイルへの変換あるいは Excel ファイルから FASTA ファイルへの変換を実行します。
- ② V-Nus_net_tool.exe
Blast 検索や ClustalW などを実行します。
- ③ nihs.xlsx
FASTA ファイルから変換される Excel ファイルのテンプレートとなるファイルです。プログラムに関連しない列名は変更できます。
- ④ blst_temp
BLAST 検索や ClustalW を実行した際に生成される各種の作業ファイルが保存されるフォルダです。系統樹作成に利用する「*.phb」ファイルは本フォルダ内に作成されます。

5 V-Nus NetTool で使用する Excel ファイルについて

V-Nus Net Tool において、塩基配列データはすべて Excel ファイルを利用します。以下の点に注意してください。

① Excel ファイルは「*.xlsx」の形式で使用する

Excel ファイルは原則として Excel2007 以降の「*.xlsx」の形式で利用してください。「*.xls」のファイル形式でも利用可能ですが、「*.xls」のファイル形式と拡張子が一致しません。ファイルが破損しているか、安全ではない可能性があります。発行元が信頼できない場合は、このファイルを開かないでください。ファイルを開きますか? というメッセージがでる場合があります。また、V-Nus Net Tool 自体は基本的に作動しますが、「sheet1」以外のシートのデータが反映されないなど、機能が制限されるようです。

② Excel ファイルのフォーマット

FastaConvert.exe による FASTA ファイルから作成される Excel ファイルは、nihs.xlsx と同様のフォーマット(☒2)になります(nihs.xlsx に塩基配列データが収納され、別名で保存される)。

本システムでは Excel ファイルの「sheet1」のシートで作業が行われます。このシートには作業に必須の列(1 行目を赤で塗りつぶしてある A~S の列)と任意に列名を変更できる列があります。A~S の列の 2 行目に記載してある列名は変更しないでください。また、1 行目の「#」も変更しないでください。

列名が変更可能な列(T~)は、任意の列名に変更できます(列 W 以降も利用できます)。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	#																					
	#SeqID	株名	シークエンス	ウイルス名	領域	遺伝子型	アクセシオンNo	類似株名(Blist)	遺伝子型(B)	Score	Expect	Identities	Strand	件数	一致株名(Blist)	一致件数	類似株名(CW)	遺伝子型(C)	類似株との距離	備考1	備考2	備考3
2																						

図 2. 作業シートのフォーマット

作業に必須な列名と格納されるデータの概要は表 1 のとおりです。

表 1. V-Nus Net Tool で使用する Excel ファイルのフォーマットとデータ内容の概要

列名の変更	分類	列名	内容	データ例
不 可	基本 情報	#SeqID	システムが利用する ID	1
		株名	FastaConvert.exe により FASTA ファイルで使用されている株名が自動入力されます	1611-NIHS-F-1234
		シークエンス	FastaConvert.exe により FASTA ファイルのシークエンスデータが自動入力されます	GTTTCT...
		ウイルス名	ウイルス名等: 基本的に手作業で入力します	NoV
		領域	シークエンスデータの領域等: 手作業で入力します	N/S
		遺伝子型	遺伝子型等: BLAST や ClustalW での解析に利用されます。手作業で入力します	GII.2
		アクセシオン No	DDBJ 等のアクセシオン No: BLAST により得られた類似株の「アクセシオン No」が自動入力されます	XXXXXXXX
		類似株名 (Blist)	BLAST により得られた類似株の「株名」が自動入力されます	1612-151-F-5467
		遺伝子型(B)	BLAST により得られた類似株の「遺伝子型」が自動入力されます	GII.2
		Score	BLAST により得られた類似株の「Score(=bits)」の値が自動入力されます	623
BLAST 検索	可能	Expect	BLAST により得られた類似株の「E-value」の値が自動入力されます	0
		Identities	BLAST により得られた類似株の「Identities」の値が自動入力されます	337/337(100.00%)
		Strand	BLAST により得られた類似株と問い合わせ株の「Strand(plusあるいはminus)」が自動入力されます	Plus/Plus
		件数	BLAST により、reference に指定した株の中の得られた類似株と同じ類似性(Score または E-value)をもつ株の件数が自動入力されます*	5
		一致株名 (Blist)	BLAST により、 mismatches=0 の株名が自動入力されます*	1612-151-F-5467
		一致件数	BLAST により、 mismatches=0 の株の件数が自動入力されます*	5
		類似株名 (CW)	ClustalW により得られた類似株の「株名」が自動入力されます*	GII.2_Melksham/89/UK
		遺伝子型 (C)	ClustalW により得られた類似株の「遺伝子型」が自動入力されます*	GII.2
		類似株との距離	ClustalW により得られた類似株との遺伝的距離が自動入力されます*	0.02809
		備考 1	任意の疫学情報のための列として利用	
備考 2	任意の疫学情報のための列として利用			
備考 3	任意の疫学情報のための列として利用			

*: 設定により条件が決まります(「パラメータの変更」参照)。

6 FASTA ファイルから Excel ファイルへの変換

シーケンサーによって決定された検出株(問い合わせ株=Query)の塩基配列データあるいは Pubmed (<https://www.ncbi.nlm.nih.gov/pubmed>)等から入手した遺伝子型別用参照株(参照株=reference)など、解析に必要なデータを FASTA ファイル(図 3)で準備します。

```

>株名 1
塩基配列データ 1
>株名 1
塩基配列データ 2
... (以下、同様)

(例)
>1611-NIHS-F-1
GTTTTCTGACTTGAGCACGTGGGAGGGCGATCGCAATCTGGCTCCCAATTTTGTGAATGAAGATGGC
GTCGAATGACGCCGCTCCATCTACTGATGGTGCAGCCGGCCTCGTGCCAGAAA...
>1611-NIHS-F-2
GTTCTCAGATCTAAGCACATGGGAGGGCGATCGCAATCTGGCTCCCAGTTTTGTGAATGAAGATGGC
GTCGAATGACGCCGCTCCATCTAATGATGGTGTGCTGCTGGTCTCGTACCAGAGG...
>1612-NIHS-F-3
ATTCTCTGACCTCAGCACATGGGAGGGCGATCGCAATCTTGCTCCCGAGGGTGTGAATGAAGATGG
CGTCGAATGACGCTGTTCCATCGAATGATGGTGTGCTGCCAACCTCGTACCAGAGG...
>1612-NIHS-F-4
CCTTGGCCTGTGGACAGGAGATCGCAATCTCCTGCCCGAATTCGTAAATGATGATGGCGTCTAAGGA
CGCCCCAACATCCCCTGATGGCGCCAGTGGCGCCGGCCAGCTGGTACCGGAGG...

```

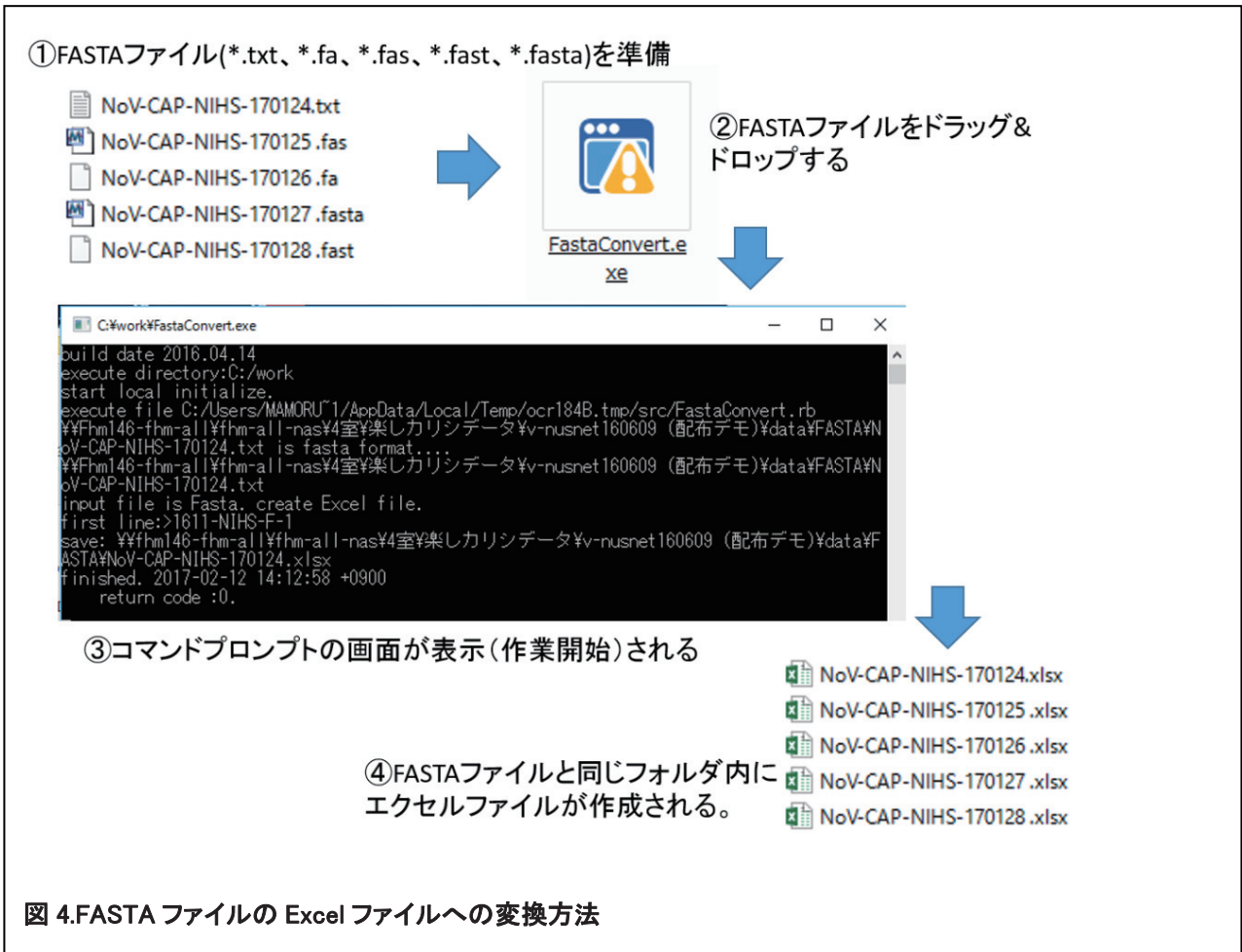
図 3. FASTA ファイル形式(例)

FASTA 形式のファイル(拡張子は*.txt、*.fa、*.fas、*.fast、*.fasta のいずれにも対応)を FastaConvert.exe にドラッグ&ドロップすると、コマンドプロンプト画面が表示され、変換作業が開始されます。黒のコマンドプロンプト画面が消えたら作業が終了し、変化元の FASTA ファイルと同じフォルダ内に変換された同名の Excel ファイルが作成されます(図 4)。

作成された Excel ファイル(図 5)は、株名が B 列に、塩基配列データが C 列に入力されたものとなります。

(参考)

本システムで利用している Excel ファイルのデータを各種の遺伝子解析ソフトで利用したい場合は、Excel ファイルを同じように FastaConvert.exe にドラッグ&ドロップすることで、同名の FASTA ファイルを作成することができます。その場合、株名(B 列)および塩基配列のデータ(C 列)以外の各列のデータは反映されません。



	A	B	C	D	E	F	G	H
1	#							
2	#SeqID	株名	シークエンス	ウイルス名	領域	遺伝子型	アクセシオン	類似株名(B)
3	1	1611-NIHS-F-1	GT TTTCTGACTTGAGCACGTTGGGAGGGCGATCGCAATCTGGCTCCC AATTTTGTGAATGA					
4	2	1611-NIHS-F-2	GT TCTCAGATCTAAGCACATGGGAGGGCGATCGCAATCTGGCTCCCAGTTTGTGAATGA					
5	3	1612-NIHS-F-3	ATTCTCTGACCTCAGCACATGGGAGGGCGATCGCAATCTTGCTCCCGAGGGTGTGAATGA					
6	4	1612-NIHS-F-4	CCTTGGCCTGTGGACAGGAGATCGCAATCTCCTGCCCGAATTCGTAAATGATGATGGCGT					

図 5. 作成されたエクセルファイル

解析したい株(問い合わせ株=Query)については、上記の作業を毎回行った後、BLAST による類似株検索や ClustalW による遺伝子型別等を行います。

遺伝子型別用の参照株(参照株=Reference)についても、基本的に同様の作業が必要です。

参照株については、FASTA ファイルをエクセルファイルに変換した後、「ウイルス名」、「領域」、「遺伝子型」および「アクセシオン番号(Genbank 等の登録番号)」について必要なデータを入力しておきます(図 6)。F 列の「遺

伝子型」に入力されたデータが BLAST 検索や ClustalW 等を実行した場合の戻り値となるので、入力が必要で
す。

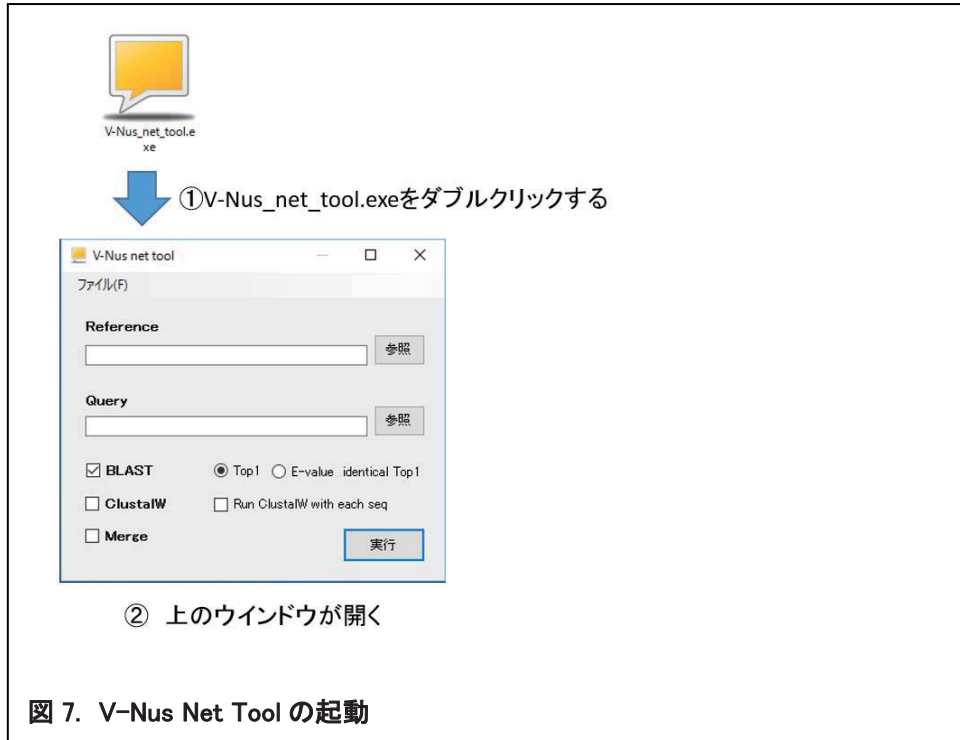
また、本システムをウイルス以外の細菌、寄生虫、魚種(フグ等)等の鑑別に利用したい場合は、参照株
(Reference)として、同様の Excel ファイルを準備します。

	A	B	C	D	E	F	G
1	#						
2	#SeqID	株名	シーケンス	ウイルス名	領域	遺伝子型	アクセションNo
3	1	GI.1_M87661 Norwalk/68/US	ATGATGATGGC	NoV	N/S	GI.1	M87661
4	2	GI.2_L07418 Southampton/91/UK	ATGATGATGGC	NoV	N/S	GI.2	L07418
5	3	GI.3_U04469 DesertShieldDSV395/90/US	ATGATGATGGC	Nov	N/S	GI.3	U04469
6	4	GI.4_AB042808 Chiba407/87/JP	ATGATGATGGC	Nov	N/S	GI.4	AB042808
7	5	GI.5_AJ277614 Musgrove/89/UK	ATGATGATGGC	Nov	N/S	GI.5	AJ277614
8	6	GI.6_AF093797 BS5/98/GE	ATGATGATGGC	Nov	N/S	GI.6	AF093797
9	7	GI.7_AJ277609 Winchester/94/UK	ATGATGATGGC	Nov	N/S	GI.7	AJ277609

図 6. 参照株(Reference)の例

7 V-Nus Net Tool の実行(図 7)

V-Nus Net Tool の各機能は、「V-Nus_net_tool.exe」のアイコンをダブルクリックし、作業ウィンドウを開き、各種のプログラムを実行します。



V-Nus Net Tool の機能は大きく 3 種類に分けられます。

① BLAST

インターネットで公開されている迅速な相同性検索システムである BLAST (Basic Local Alignment Search Tool)のプログラムを利用し、Query で指定した株について Reference で指定した株に対して検索を行い、相同性が高い株を調べます。

② ClustalW

インターネットで公開されている近隣結合法 (neighbor-joining method=NJ 法) に基づく系統樹解析を実行するプログラムである ClustalW を利用して、遺伝子型別等を行います。

③ Merge

Reference で指定した Excel ファイルのデータに Query で指定した Excel ファイルのデータを追加します。

(1) BLAST**1) 使用目的**

BLAST は以下のような例に使用します。

目的		用いる Excel ファイル(例)	
		Reference	Query
①	患者と調理従事者、食品、ふき取りなどから検出されたノロウイルスの塩基配列が一致するか調べたい。	患者から検出された株	調理従事者、食品、ふき取り等から検出された株
②	当該事例から検出された株が他の事例から検出された株と一致するものがあるか調べたい(広域事例かどうか調べたい)	他の事例から検出された株	当該事例から検出された株
③	網羅的ゲノム解析等で得られた大量の塩基配列データについて遺伝子型別を行いたい。	遺伝子型別用参照株	型別したい株

2) 使用方法(図 8)

- ① Reference 欄に、検索対象となるデータ(例:患者から検出された株のデータ、他の事例から検出された株のデータ、遺伝子型別の参照株等)をドラッグ&ドロップする(あるいは、参照をクリックして、該当するファイルを指定する。「以下同様」)
- ② Query 欄に検索したいデータ(調理従事者から検出された株のデータ、当該事例から検出された株のデータ、遺伝子型別したい株のデータ等)を指定する。
- ③ 実行をクリックする。
- ④ 検索結果を含む Excel ファイルの保存場所、ファイル名、ファイル形式の指定を行い、保存をクリックする。
- ⑤ 作業ウインドウが表示される。
- ⑥ 終了後、終了を示すウインドウが表示される。
- ⑦ 検索結果データを含む Excel ファイル(検索結果ファイル)と作業ファイル(*.bln)#が、指定されたフォルダに作成される(#は、Blast のパラメータの設定が debug=true の場合のみ作成される)。

検索結果ファイルの基本情報のうち、「ウイルス名」、「領域」、「遺伝子型」の各データは Query ファイルの情報を受け継ぐため、Query ファイルにデータがない場合は、検索結果ファイルも空欄のままです。そのため、「ウイルス名」、「領域」、「遺伝子型」のデータは、BLAST 実施後手作業で入力する必要があります。

特に、「**遺伝子型**」については、**検索結果ファイルを Reference ファイルとして利用する場合には、「遺伝子型(B)のデータを代入するなどして、必ず入力しておいてください。**

① 検索の対象となるファイルをドラッグ&ドロップする

② 検索したいファイルをドラッグ&ドロップする

③ 実行をクリックする

④ 検索結果を含むエクセルファイルの保存場所、ファイル名、ファイル形式の指定を行い、保存をクリックする。

⑤ 作業ウィンドウが表示される

⑥ 作業終了を示すウィンドウが表示される

⑦ 検索結果データを含むエクセルファイルと作業ファイル(.bln)*が、指定されたフォルダに作成される (*は、debug=trueの場合のみ作成される)

図 8. BLAST の実施方法

3) BLAST で得られるデータについて

本システムでは、Query で指定した株について Reference で指定した株に対して BLAST 検索を行い、以下の項目について結果を戻します。本システムでは、BLAST の結果、

- ① Score(=bits)が最も高いもの (Top1)を類似株とする方法
- ② E-value が最も高いもの(最も高い E-value を持つ株が複数ある場合は、%identities が最も高いものの Top1)を類似株とする方法

の 2 種類が選択できます(通常は、Top1 のままで使用してください)。

列名	戻り値
類似株名 (Blst)	Reference に含まれる株の中で、最も Score(=bits)が高いものの Top1 となった株*1(①)の「株名」が戻り値となります。
遺伝子型(B)	①の株の遺伝子型が戻り値となります。
Score	①と当該株との類似性を示す Score(=bits)が戻り値となります。
Expect	①と当該株との類似性を示す E-value が戻り値となります。
Identities	①と当該株との Identities*2 が戻り値となります。
Strand	①と当該株の Strand(Reference 株を Plus 鎖とした場合)が戻り値となります。
件数	①と同じ類似性を示す株の件数が戻り値となります*3。
一致株名(Blst)	ミスマッチ*2=0 の株名が戻り値となります*4。
一致件数	ミスマッチ=0 の株の件数が戻り値となります*5。

*1: E-value identical Top1 で実行した場合は、E-value が最も高いもの(最も高い E-value を持つ株が複数ある場合は、Identities が最も高いものの Top1)が戻り値になります。

*2: Identied とミスマッチは以下の例のように計算されます。

Reference : ATGCATGCAT	の場合	Identities=9/10 (90%)
Query : ATGC-TGCAT		ミスマッチ=0

(重要)

*1 と*2 において、BLAST のパラメータ(「(4) パラメータの設定」参照)に指定した既定値(デフォルト)を満たす値を持つ株が Reference に存在しない場合、戻り値は「該当なし」となります。つまり、これらの値を緩く設定すると相対性が低いものでもヒットすることになり、逆に値を厳しく設定するとヒットしにくくなります。「遺伝子型(B)」には、それらの条件を満たした上で、最も類似性が高い株の遺伝子型が戻り値となります。そのため、「遺伝子型(B)」で得られた遺伝子型が、Query で指定した株の遺伝子型になるとは必ずしも限りません。

*3: 10 株以上の同じ類似性(同じ Score 値など)を示す株が 10 株以上あった場合は「10+」と表示されます。

*4: BLAST パラメータの設定した上限値(例 EntryHit=50)を上回る場合は、上限値(この例の場合は 50)の株名のみが表示されます。

*5: BLAST パラメータの設定した上限値(例 EntryHit=50)を上回る場合は、上限値(この例の場合は 50)が戻り値になります。

検索結果ファイルのデータの由来については、[図 9](#) に示したように、基本情報は Query に指定した株のデータが引き継がれます。

図 10 に作成された Excel ファイルの例を示します。

また、「BLAST 検索」シートには、表 2 に示したように、遺伝子型別結果と塩基配列が一致した(ミスマッチ=0)の株名およびその件数をまとめた一覧表が作成されます(Excel のリンク形式で「sheet1」のデータの必要部分を表示したもの)。必要に応じて、「nihs.xls」ファイルにリンク形式で、必要なフォーマットの表等のテンプレートを作成しておく、検索結果ファイルにも反映されるので便利です。

表 2. BLAST 検索により一致した株の一覧

株名	遺伝子型	一致した株	件数
1611-NIHS-F-1	GII.2	1611-272-F-16087	1
1611-NIHS-F-2	GII.17	1611-272-F-16088, 1605-181-F-22804N04, 1604-181-O-22801N17	3
1612-NIHS-F-3	GII.6	1612-272-F-16097	1
1612-NIHS-F-4	GI.6	1612-272-F-16101	1
0	0	0	0

分類	列名	Reference ファイル	Query ファイル	検索結果ファイル
基本情報	#SeqID	12	1	1
	株名	1611-272-F-16088	Isolate-1	Isolate-1
	シークエンス	ATGATGAT...	TTGATGAT...	TTGATGAT...
	ウイルス名	NoV		
	領域	N/S		
	遺伝子型	GII.17		
	アクセッション No	(AB01)		(AB01)
	類似株名 (Blst)			1611-272-F-16088
	遺伝子型(B)			GII.17
	Score			625
BLAST 検索	Expect			0
	Identities			338/338(100.00%)
	Strand			Plus/Plus
	件数			1
	一致株名 (Blst)			1611-272-F-1 1605-181-F-22 1604-181-O-28
	一致件数			3
	類似株名 (CW)			
	遺伝子型 (C)			
	類似株との距離			
	ClustalW			
		BLAST 検索結果		
		Top1		
		Score(=bits)が最大(Top1)		
		E-value が最大 (複数ある場合は、%Identities が最大)		
		ミスマッチ=0 の株名		
		ミスマッチ=0 の件数		

図 9.BLAST 検索の結果、作成される Excel ファイルのデータの内容

Referenceに他の事例から検出された株のデータセットを指定した場合のBlast検索例

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	#SeqID	株名	シロアケセンシオン	類似株名(Blast)	遺伝子型(B)Score	Expect	Identities	Strand	件数	一致株名(Blist)	一致件数					
2	1	1611-NIHS-F-1	C	1611-272-F-16087	GII.2	625	0.338/338(100.00%)	Plus/Plus	1	1611-272-F-16087	1					
3	2	1611-NIHS-F-2	T	1611-272-F-16088	GII.17	625	0.338/338(100.00%)	Plus/Plus	1	1611-272-F-16088, 1605-181-F-22804N04, 1604-181-O-22801N17	3					
4	3	1612-NIHS-F-3	T	1612-272-F-16097	GII.6	625	0.338/338(100.00%)	Plus/Plus	1	1612-272-F-16097	1					
5	4	1612-NIHS-F-4	C	1612-272-F-16101	GII.6	632	0.342/342(100.00%)	Plus/Plus	1	1612-272-F-16101	1					

Referenceに遺伝子型参照株を指定した場合のBlast検索例

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	#SeqID	株名	シロアケセンシオンNo.	類似株名(Blist)	遺伝子型(B)Score	Expect	Identities	Strand	件数	一致株名(Blist)	一致件数					
2	1	1611-NIHS-F-1	GTTT(X81879)	GII.2_X81879Melksham/89/UK	GII.2	477	4.00E-137	274/282(97.16%)	Plus/Plus	1	0					
3	2	1611-NIHS-F-2	GTTT(AB983218)	GII.17[P17]_AB983218_Kawasaki323/GII.17	GII.17	603	6.00E-175	334/338(98.82%)	Plus/Plus	1	0					
4	3	1612-NIHS-F-3	ATTCT(AJ277620)	GII.6_AJ277620_Seacroft/90/UK	GII.6	368	2.00E-104	254/282(90.07%)	Plus/Plus	1	0					
5	4	1612-NIHS-F-4	CCTT(AB081723)	GII.6_AB081723WUG1/00/JP	GII.6	486	6.00E-140	283/293(96.59%)	Plus/Plus	1	0					

図 10. BLAST の結果、作成された検索結果ファイル(例)

③ BLAST の作業ファイルについて

BLAST パラメータの設定で、debug=true と設定した場合、検索結果ファイルと同じフォルダ内に BLAST 検索の作業ファイル(*.bln)が作成されます。このファイルをテキストエディタソフト(メモ帳、ワードパット等)で開くと、[図 11](#)に示したように、BLAST で作成された生データをみる事ができるので、解析結果の確認等に利用できます。

```

BLASTN 2.2.31+
(中略)
Database: blst_tmp¥20170212-160411s.fna
          97 sequences; 31,824 total letters
Query= 0001_1611-NIHS-F-1

Length=338

Sequences producing significant alignments:

0055_1611-272-F-16087          625    0.0
0083_1611-151-0-954/031/Ki/MU  619    3e-180
(中略)
>0055_1611-272-F-16087
Length=338

Score = 625 bits (338), Expect = 0.0
Identities = 338/338 (100%), Gaps = 0/338 (0%)
Strand=Plus/Plus

Query 1   GTTTTCTGACTTGAGCACGTGGGAGGGCGATCGCAATCTGGCTCCCAATTTTGTGAATGA 60
          |||
Sbjct 1   GTTTTCTGACTTGAGCACGTGGGAGGGCGATCGCAATCTGGCTCCCAATTTTGTGAATGA 60
    
```

図 11. BLAST 検索により作成された「*.bln」ファイルの例(一部のみ表示)

(2) ClustalW**1) 使用目的**

ClustalW は以下のような例に使用します。

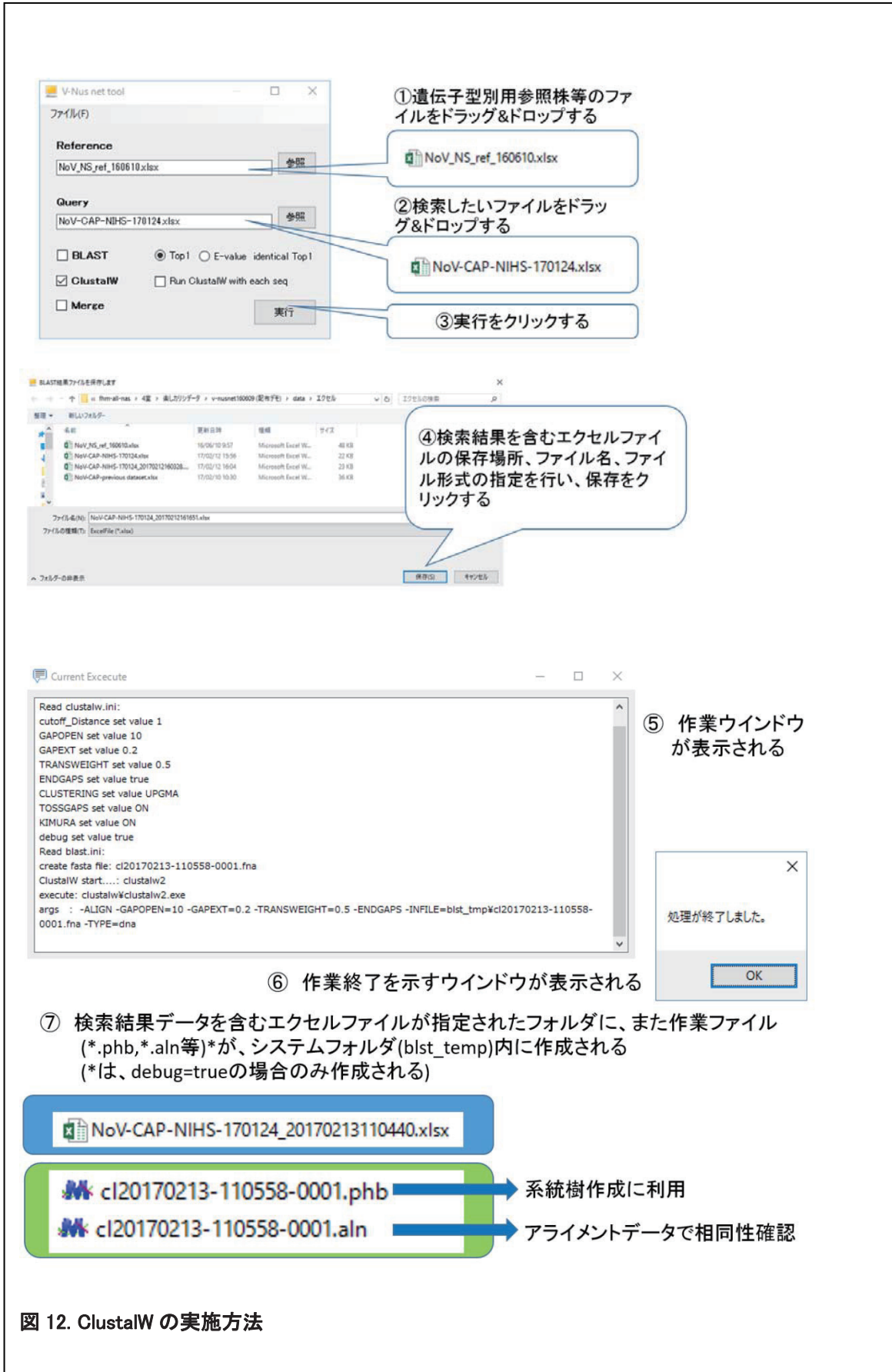
目的		用いる Excel ファイル(例)	
		Reference	Query
①	検出されたノロウイルス等の遺伝子型を決定したい	遺伝子型参照株	遺伝子型別したい株
②	系統樹を作成したい		
③	患者由来株と調理従事者由来株をアライメント(整列)して、比較したい	患者由来株	調理従事者等由来株
④	その他の微生物等(食中毒菌、寄生虫、フグ等の種鑑別等)の遺伝学的鑑別を行いたい	型別の参照株	型別したい株

2) 使用方法(図 12)

- ① Reference 欄に、遺伝子型別用参照株等の Excel ファイルをドラッグ&ドロップする(あるいは、参照をクリックして、該当するファイルを指定する。「以下同様」)
- ② Query 欄に遺伝子型別したい株等の Excel ファイルを指定する。
- ③ 実行をクリックする。
- ④ 検索結果を含むエクセルファイルの保存場所、ファイル名、ファイル形式の指定を行い、保存をクリックする。
- ⑤ 作業ウインドウが表示される。
- ⑥ 終了後、終了を示すウインドウが表示される。
- ⑦ 検索結果データを含むエクセルファイルが指定されたフォルダに作成される。また、作業ファイル(*.phb,*.aln 等)#が、システムフォルダ内の「blst_temp」フォルダ内に作成される(#は、ClustalW のパラメータの設定が debug=true の場合のみ作成される)。

検索結果ファイルの基本情報のうち、「ウイルス名」、「領域」、「遺伝子型」の各データは Query ファイルの情報を受け継ぐため、Query ファイルにデータがない場合は、検索結果ファイルも空欄のままです。そのため、「ウイルス名」、「領域」、「遺伝子型」のデータは、ClustalW 実施後手作業で入力する必要があります。

特に、「**遺伝子型**」については、**検索結果ファイルを Reference ファイルとして利用する場合には、「遺伝子型(C)のデータを代入するなどして、必ず入力しておいてください。**



3) ClustalW で得られるデータについて

本システムでは、以下の作業が自動で行われます。


- ① Query で指定した株と Reference で指定した株を合わせて、FASTA 形式のファイルを作成する。
- ② ClustalW でアライメントを行う。
- ③ BootStrap 検定後、N-J 法により距離計算を行う。
- ④ Query で指定した株のそれぞれに対して、Reference で指定した株のうち、最も遺伝的距離に近い株の「株名」と「遺伝子型」および両者の「遺伝的距離」が入力された検索結果ファイルを作成する。


この場合、ClustalW のパラメータ設定で、「cutoff_Distance」(初期値=1)より小さい遺伝子的距離を持つ株が Reference に指定した株の中に存在しない場合は「該当なし」と表示される。

(重要)

上記のように、V-Nus Net Tool における ClustalW による遺伝子型別は、Reference で指定した株のうち、最も遺伝的距離に近い株の「遺伝子型」が「遺伝子型(C)」の戻り値となり、パラメータの設定(「(4) パラメータの設定」参照)で、「cutoff_Distance」(初期値=1)より小さい遺伝子的距離を持つ株が Reference に指定した株の中に存在しない場合は「該当なし」と表示されます。すなわち、この値を高く設定すると、本来型別不能となる株が遺伝子型別されたり、逆に低く設定すると遺伝子型別される株が「該当なし(型別不能)」と判定されることが生じやすくなります。

つまり、正しい遺伝子型別を実施するためには、「cutoffDistance」の設定がきわめて重要となります。既定値として入力された値は暫定的なもので、ノロウイルスについて利用する場合でも、正しく判定できないことが起こる可能性があります。また、対象とする微生物や生物の種類によって鑑別の基準となる遺伝的距離は必ずしも同じではないため、それぞれに適切な「cutoffDistance」を設定する必要があります。そのため、本システムで得られた型別結果について、参照株との遺伝的距離が通常より多い場合などは、手作業で分析するなどして型別結果の信頼性について検証しておく必要があります。ノロウイルスの遺伝子型別については、Norovirus Genotyping Tool が利用できるため、その型別結果と比較することも有用です。

作成される Excel ファイルと Reference や Query に指定した Excel ファイルのデータの関係性は  13 のとおりです。

また、ClustalW の実行で作成された検索結果ファイル(例)を  14 に示します。

(参考)

Norovirus Genotyping Tool

ノロウイルスの遺伝子型別については、RIVM(オランダ国立公衆衛生環境研究所)が提供する Noronet の「Norovirus Genotyping Tool」(<http://www.rivm.nl/mpf/norovirus/typingtool>)を使用することにより、FASTA ファイルを用いて簡便に遺伝子型別を実施することができます。同システムはノロウイルスの遺伝子型別、キメラ型(遺伝子組み換え型)の同定に特化したシステムです。

分類	列名	Reference ファイル	Query ファイル	検索結果ファイル
基本情報	#SeqID	12	1	1
	株名	GI.1_M87661Norwalk/68/US	Isolate-2	Isolate-2
	シークエンス	ATGATGAT...	ATGATGAT...	ATGATGAT...
	ウイルス名	NoV		
	領域	N/S		
	遺伝子型	GI.1		
	アクセシオン No	M87661		
	類似株名 (Blst)			
	遺伝子型(B)			
	Score			
BLAST 検索	Expect			
	Identities			
	Strand			
	件数			
	一致株名(Blst)			
	一致件数			
	類似株名 (CW)			
	遺伝子型 (C)			
	類似株との距離			
	ClustalW			
				GI.1_M87661Norwalk/68/US
				GI.1
				0.03061

図 13. ClustalW の結果、作成される Excel ファイルのデータの内容

ClustalWによる遺伝子型別結果

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
1	#																			
2	#SeqID	株名	通SEIS件 類似株名(CW)																遺伝子型(C)	類似株との距離
3	1	1611-NIHS-F-1	GTTTCTGACTTG_GII.2_X81879Melksham/89/UK																GII.2	0.03021
4	2	1611-NIHS-F-2	GTTCTCAGATCTA_GII.1.7[P1.7]_AB983218_Kawasaki323/2014/JP																GII.1.7	0.0171
5	3	1612-NIHS-F-3	ATTCTCTGACCTC_GII.6_AB039776SaitamaU3/97/JP																GII.6	0.09847
6	4	1612-NIHS-F-4	CCTTGGCCTGTGEGI.6_AB081723WUG1/00/JP																GI.6	0.03458

ClustalW(Run ClustalW with each seq)による遺伝子型別結果

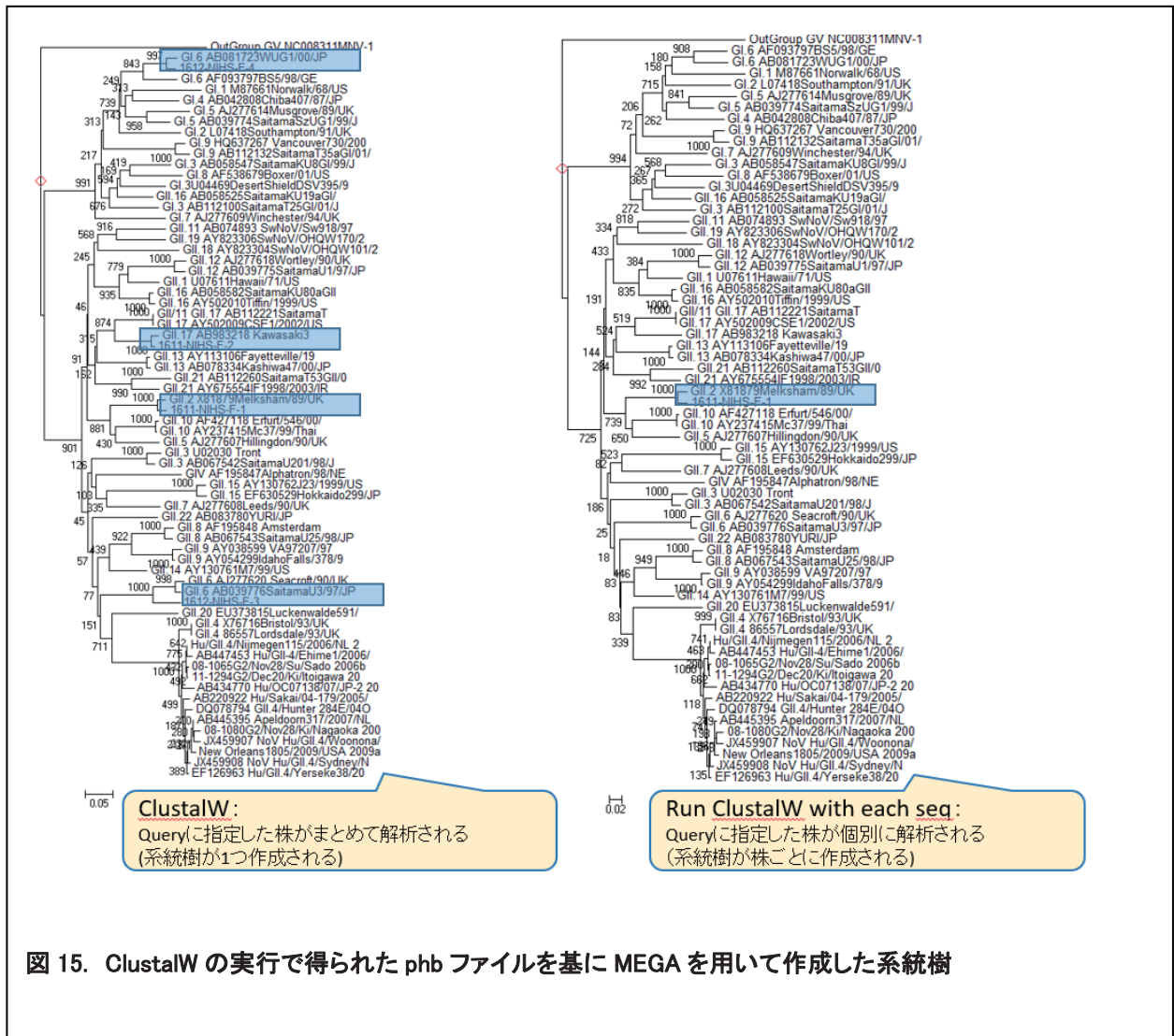
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
1	#																			
2	#SeqID	株名	通SEIS件 類似株名(CW)																遺伝子型(C)	類似株との距離
3	1	1611-NIHS-F-1	AAGATGGCG_GII.2_X81879Melksham/89/UK																GII.2	0.03061
4	2	1611-NIHS-F-2	AAGATGGCG_GII.1.7[P1.7]_AB983218_Kawasaki323/2014/JP																GII.1.7	0.01724
5	3	1612-NIHS-F-3	AAGATGGCG_GII.6_AB039776SaitamaU3/97/JP																GII.6	0.09837
6	4	1612-NIHS-F-4	TCTAAGGAC(GI.6_AB081723WUG1/00/JP																GI.6	0.03503

図 14. ClustalW を実行した結果、作成された検索結果ファイル

4) ClustalW と ClustalW run with each seq

ClustalW の実行メニューとしては、ClustalW と Run ClustalW with each seq の 2 種類があります。その違いは以下のとおりです(系統樹は [図 15](#) 参照)。通常は、ClustalW を使用します。

	ClustalW	Run ClustalW with each sequence
処理内容	Query と Reference に指定した株全体について処理を行う	Query に指定した株ごとに Reference に指定した株と合わせて処理を行う
処理回数	1 回(作業ファイルは 1 組)	Query に指定した株の数だけ、作業が繰り返される(作業ファイルは Query に指定した株ごとに作成される)
特徴	通常の解析に用いる	NJ 法による正確な遺伝子的距離を求めたい場合などに利用する



③ ClustalW の作業ファイルについて

ClustalW のパラメータの設定で、debug=true と設定した場合、ClustalW の作業ファイル(*.phb、*.aln、*.deb.log、*.clustalW.log、*.dnd、*.fna)がシステムフォルダ内の「blst_temp」内に作成されます。

このうち、主に使用するのは、以下の 2 つです。

作業ファイル	目的	使用するソフト*
*.phb	系統樹の描画	MEGA、NJ plot 等の系統樹描画ソフト
*.aln	アライメントされた塩基配列データを視覚的に確認し、株間の相同性や配列データの正確性を調べる	MEGA、ClustalX 等の遺伝子解析ソフト

*: 各ソフトはホームページからフリーでダウンロードできます(各自治体におけるセキュリティ対策等の規則を遵守してダウンロードして使用してください。)

MEGA: <http://www.megasoftware.net/>

NJ plot: <http://doua.prabi.fr/software/njplot>

ClustalX: <http://www.clustal.org/>

(本システムに利用している ClustalW でも実施可能ですが、コマンド方式なので、操作が煩雑です)

これらの遺伝子解析用ソフトの使用方法は、各自治体における専門家にお問合せください。

(3) Merge

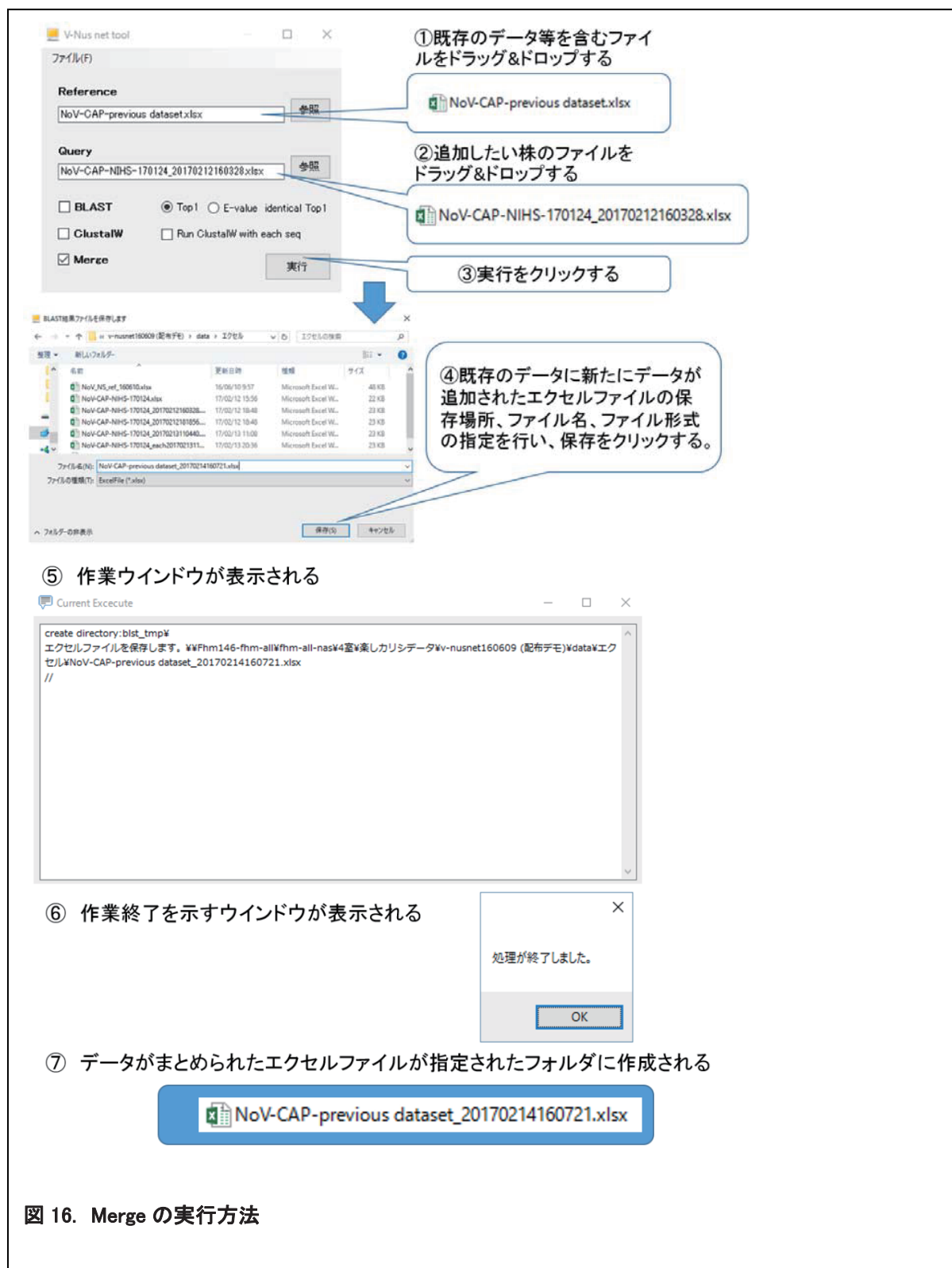
1) 使用目的

Merge は Query に指定した Excel ファイルのデータを reference に指定された Excel ファイルのデータに追加する場合に使用します。

目的	Reference	Query
新たに検出され、遺伝子型別された株を既存の検出株リストに追加したい	既存の検出株リストのファイル	追加したい検出株のファイル
遺伝子型別参照株に新たに参照株を追加したい	既存の遺伝子型別参照株のファイル	新たに参照株として追加したい株のファイル

2) 使用方法(図 16)

- ① Reference 欄に既存のデータ等を含むファイルをドラッグ&ドロップします(または、参照をクリックしてファイルを指定する)。
- ② Query 欄に追加したい株のファイルをドラッグ&ドロップします。
- ③ 実行をクリックします。
- ④ 既存のデータに新たにデータが追加されたエクセルファイルの保存場所、ファイル名、ファイル形式の指定を行い、保存をクリックします。
- ⑤ 作業ウィンドウが表示されます。
- ⑥ 作業終了を示すウィンドウが表示されます。



⑦ データがまとめられたエクセルファイルが指定されたフォルダに作成されます。

(4) パラメータの変更

BLAST 検索および ClustalW のパラメータを変更することができます。ツールのファイルメニューから「設定を開く」をクリックし、BLAST パラメータか ClustalW パラメータを選択します(図 17)。

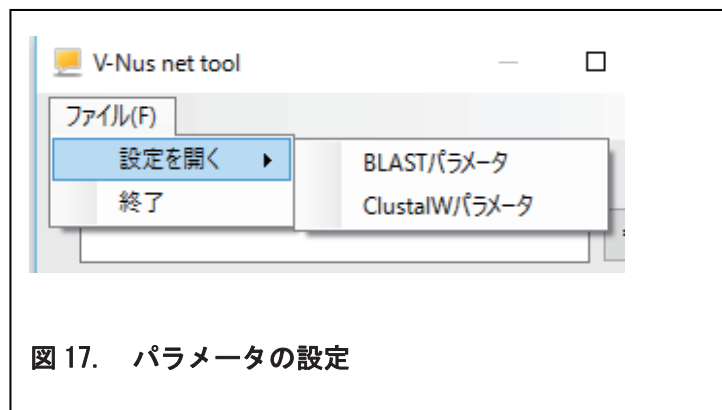


図 17. パラメータの設定

通常変更する場合があるパラメータは以下のとおりです。また、間違えて設定した場合、元の条件に戻す必要がある場合等に備え、ファイルの内容を控えておくことをお勧めします。

各システムの詳細については、下記のホームページ等を参考にしてください。

BLAST: <http://www.ddbj.nig.ac.jp/search/help/blasthelp-j.html> (DDBJ)

ClustalW: <http://www.ddbj.nig.ac.jp/search/help/clustalwhelp-j.html> (DDBJ)

BLAST のパラメータ

パラメータ	既定値	内容
Evalue	<0.0001	E-value は、「偶然に同じスコアになる配列の数の期待値」で、E-value が小さいほど、互いの配列の類似性が高いこととなります。Reference に指定した株の中に本条件を満たす株がない場合、「該当なし」が戻り値になります。
identify	>90	Reference と query の株の相同性(解析の対象となる塩基長に対する配列が一致した塩基長の割合)を示します。Reference に指定した株の中に本条件を満たす株がない場合、「該当なし」が戻り値になります。
Gap	<1	ギャップの塩基数のカットオフ値。
EntryHit	=50	1 配列当たりの最大 Hit 数。
MatchLengthMin	=200	マッチした塩基配列の最小の長さ。 この長さを満たさない場合は無視されます。
threads	=4	最大使用 CPU コア数(threads)
debug	=true (=false)	true: 作業ファイルを残します。 false: 作業ファイルを消去します。

ClustalWのパラメータ

[ClustalW -ALIGN]		
QueryLengthMin	=200	Query に指定した株の解析対象となる塩基配列長の最小値です。この値より短いデータは無視されます。
cutoff_Distance	=1	Reference に指定した株の中で、最も近縁であった株との遺伝的距離のカットオフ値です。 Reference に指定した株の中に本値より小さい値をもつ株がない場合は、戻り値が「該当なし」となります。
GAOPEN	=10	Gap 開始時の Gap penalty を指定します。デフォルトは 10 です。
GAPEXT	=0.2	Gap 継続時の Gap penalty を指定します。デフォルトは 0.2 です。
TRANSWEIGHT	=0.5	TRANSWEIGHT を指定します。デフォルトは 0.5 です。
ENDGAPS	=true	端の Gap について指定します。デフォルトは true です。
[ClustalW -BOOTSTRAP]		
clustering	=NJ (=UPGMA)	Clustering のタイプを指定します。デフォルトは NJ です。
TOSSGAPS	=ON	系統樹計算時に gap を無視するかを指定します。デフォルトは ON (gap を無視する) です。
KIMURA	=ON	系統樹計算時に使用する correction を指定します。デフォルトは Kimura です。
debug	=true (=false)	true: 作業ファイルを残します (デフォルト)。 False: 作業ファイルを消去します。

参考

##の行は無視されます (メモあるいはパラメータ変更の際に#を消去する等で利用)

8 BLAST プログラムの更新

NCBI Blast (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) の最新プログラムは以下の URL において公開されています。

<ftp://ftp.ncbi.nlm.nih.gov/blast/executables/LATEST>

この URL から

`ncbi-blast-[バージョン]+-x64-win64.tar.gz`

(2017年2月13日現在:`ncbi-blast-2.6.0+-x64-win64.tar.gz`)

のファイルをダウンロードし、解凍したディレクトリをシステムフォルダ内の[ncbi_blast]に上書き保存すると最新のNCBI Blast が利用できます。