

BioNumerics v8.1 manual
wgMLST tools local plugin

注意

本日本語マニュアルは、AppliedMaths 社が作成した英語版「wgMLST_Local_manual.pdf」を簡易的に日本語に訳したものです。詳細はオリジナル英語マニュアルをご覧ください。

BioNumerics では以下の 3rd-party ソフトウェアツール及びライブラリを使っています。

- Python 3.8 release from the Python Software Foundation,
<https://www.python.org/>
- Xerces library for XML input and output from the Apache Software Foundation,
<https://xerces.apache.org/>
- NCBI toolkit version 2.11.0, <https://www.ncbi.nlm.nih.gov/BLAST/>
- SRA Toolkit, <https://ncbi.github.io/sra-tools/>
- Boost c++ libraries, <https://www.boost.org/>
- Samtools for interacting with SAM / BAM files, <https://www.htslib.org/download/>
- 7-Zip (7za.exe), <https://www.7-zip.org/>
- Zlib library, <https://zlib.net/>
- Pigz for parallel gzip compression, <https://zlib.net/pigz/>
- Cairo 2D graphics library version 1.12.14, <https://cairographics.org/>
- Crypto++ library version 5.5.2, <https://www.cryptopp.com/>
- OpenSSL library, <https://www.openssl.org/>
- libSVM library for Support Vector Machines,
<https://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- SQLite version 3.7.17, <https://www.sqlite.org/>
- pymzML Python module version 2.4.7, <https://github.com/pymzml/pymzML>
- NumPy Python library version 1.19.1, <https://www.numpy.org/>
- BioPython Python library version 1.78, <https://www.biopython.org/>
- pyodbc Python module version 4.0.30, <https://pypi.org/project/pyodbc/>
- jinja2 Python library version 2.11.2, <https://pypi.org/project/Jinja2/>
- MarkupSafe Python library version 1.1.1, <https://pypi.org/project/MarkupSafe/>
- regex Python library version 2.5.91, <https://pypi.org/project/regex/>
- Chromium Embedded Framework,
<https://bitbucket.org/chromiumembedded/cef/wiki/Home>
- SPAdes genome assembler version 3.15.3, <https://bioinf.spbau.ru/spades> *
- SKESA version 2.3.0, <https://github.com/ncbi/SKESA/releases>
- Unicycler version 0.5.0, <https://github.com/rrwick/Unicycler/releases> *
- Velvet for Windows, source code can be downloaded from
<https://www.bionumerics.com/download/open-source>

BioNumerics v8.1 manual
wgMLST tools local plugin

- Bowtie2 version 2.2.5 (<https://bowtie-bio.sourceforge.net/bowtie2/index.shtml>)*
- SNAP version 2.0.0, <https://www.microsoft.com/en-us/research/project/snap/>
- RAxML version 8.2.11, <https://github.com/stamatak/standard-RAxML/releases>
- FastTree version 2.1.10, <https://www.microbesonline.org/fasttree/>
- CFSAN SNP pipeline version 2.2.0,
<https://github.com/CFSAN-Biostatistics/snp-pipeline>*
- Prokka version 1.14.5, <https://github.com/tseemann/prokka> *
- sourmash version 4.1.0, <https://github.com/dib-lab/sourmash> **
- SeqSero2 for Windows, source code can be downloaded from
<https://www.bionumerics.com/download/open-source>
- Fastp version 0.22.0, <https://github.com/OpenGene/fastp>

*: On Calculation Engine only

** : 英語マニュアルの3ページをご覧ください。

Chapter 1 : WGS tools local plugin の目的


計算エンジンは BioNumerics デスクトップクライアントアプリケーションから、全ゲノムシーケンス (WGS) データ解析における、計算負荷の高いタスクを実行するための、高性能コンピューティング (HPC) 環境へのアクセスを提供するサーバーアプリケーションです。wgMLST 用の命名法サービスは、計算エンジンに統合されています。

計算エンジンは、強力なコンピュータクラスタ (オンプレ物理サーバーまたはクラウド) にインストールされます。デフォルトの計算エンジンインスタンスは、Applied Maths Cloud Calculation Engine で、米国の Amazon Web Services (AWS) サーバーでホスティングされ、世界中のどこからでもアクセスできます。

201 年 10 月、Biomerieux 社は BioNumerics ソフトウェアの段階的な開発及びサポート停止を発表しました。この段階的な措置は、2022 年 10 月 6 日にリリースされた BioNumerics v8.1.1 で、BioNumerics ソフトウェア最終バージョンとし、2024 年 12 月 31 日をもって、カスタマーサポートを終了することを意味しています。同時に、Applied Maths Cloud Calculation Engine もオフラインとなり利用不可となります。

Assembly-based wgMLST 解析をローカルで実行する機能は既に導入されていますが、WGS tools プラグインは、特に新しい allele の allele-ID を取得するために、依然として計算エンジンの命名サービスに依存しています。WGS tools local プラグインの目的は、BioNumerics v8.1.1 のユーザーが、Applied Maths Cloud Calculation Engine がストップした後も、ローカル環境で assembly-based wgMLST 解析を実行し続けることが出来るようにすることです。

ハッシュベースの Allele calling は、一元的に管理された名称体系を必要とせずに wgMLST を実行するための有効なオプションです。コンピュータサイエンスでは、ハッシュ関数は任意のサイズのデータを固定サイズの値にマップする数学関数です。ハッシュ関数によって返される値はハッシュ値または播種と呼ばれます。wgMLST の allele calling プロセスに特化すると、命名法の受入基準を満たす allele 配列は、ハッシュ関数によって 17 桁の整数値に変換されます。ハッシュ関数は決定論的であり、明らかに WGS tools ローカルプラグインは常に同じハッシュ関数実装を使用するため、与えられた allele 配列は常に同じハッシュ値になります。しかし、BioNumerics の Character 実験タイプは、17 桁の整数を確実に保存し比較するには設計されていないため、そのようなハッシュ値を直接使うことはできません、加えて、ハッシュ値は wgMLST ユーザーが慣れ親しんでいる単純な整数の allele-ID とは非常に異なって見えます。

 ハッシュ関数で起こりうる懸念は、異なる入力と同じハッシュ値を生成する、いわゆるコリジョンす。WGS ツールのローカルプラグインは、計算エンジンの命名データベースに送信され、格納されるハッシュ値と

BioNumerics v8.1 manual
wgMLST tools local plugin

同じ方法でハッシュ値を計算します。したがって、サルモネラや大腸菌のような大規模な allele データベースでされ、後者のデータベースが使用されている 9 年間、コリジョンが観察されなかったことは心強いところです。

WGS tools ローカルプラグインは、ハッシュが単純な整数 ID に変換される local allele 命名法を設定します。この local allele 命名法は、同じデータベースの全てのユーザーが同じ命名法にアクセスできるように、ソースファイルディレクトリにファイルとして保存されます (see the Reference manual, Chapter The BIONUMERICs relational database)。この命名法では、遺伝子座毎に一意的な配列ハッシュ値に増分整数 ID を割り当てています。ローカル命名法からの単純な整数 allele ID は **wgMLST_Local** 実験タイプに保存され、これは WGS tools ローカルプラグインによって自動的に作成され、**wgMLST_Local** 実験タイプと同期されます。wgMLST_Local 実験タイプは assembly-based wgMLST allele calling のみを考慮し、assembly-free allele calling は考慮されません。

異なるラボやデータベース間での wgMLST データの交換を可能にするために wgMLST tools ローカルプラグインは、ハッシュ値としてローカルの wgMLST プロファイルをエクスポートする機能と、ハッシュ値からローカルの wgMLST プロファイルをインポートする機能を提供します。(Chapter 5)

Chapter 2 : BioNumerics の起動と設定

このマニュアルは、BioNumerics の WGS tools ローカルプラグインのマニュアルとして設計されています。WGS tool local plugin は、wgMLST のためのローカル命名法を提供することにより、外部計算エンジンなしで WGS tools プラグインが動作することを可能にする、WGS tools plugin のアドオンと考えることができます。(Chapter 1)

WGS tools local plugin は、BioNumerics-SEQ 及び BioNumerics-SUITE ユーザーのみ利用可能で、WGS tools plugin のインストール上のみインストール可能です。

BioNumerics 最新バージョンがインストールされていることを確認してください。

(<https://www.bionumerics.com/download/software>)

インストールマニュアルは、<https://www.bionumerics.com/download/manuals> からダウンロードできます。

最新版の BioNumerics を起動するとスタート画面が表示されます。(図 2.1)

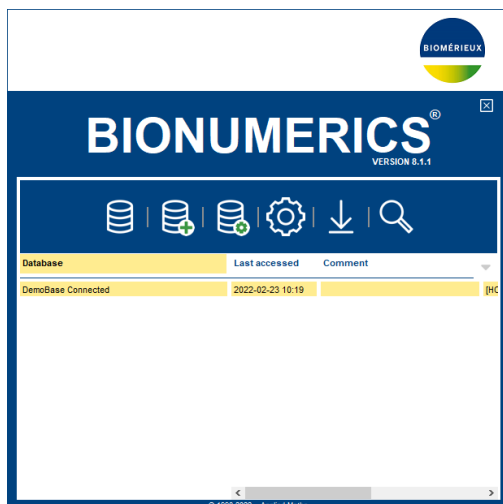




図 2.1 : BioNumerics スタート画面

新たにデータベースを作成する場合は  ボタンをクリックします。

作成済のデータベースを開く場合は、リストから任意のデータベースを指定して  ボタンをクリックするか、ダブルクリックしてください。

Chapter 3 : WGS tools local plugin のインストール

3.1 前提条件

WGS tools local plugin は、BioNumerics 最新版でのみ動作されるよう設計されています。それ以前 WGS tools ローカルプラグインは、最新の BioNumerics で動作するように設計されています。それ以前のバージョンではテストされておらず、互換性の問題がある可能性があります。最新版へのアップデートが利用可能かどうかは、ウェブサイトでご確認下さい。

(<https://www.bionumerics.com/download/software>)

WGS tools ローカルプラグインは、WGS tools プラグインが既にインストールされている BioNumerics データベースにインストールし、CE プロジェクトとパスワードを介して、計算エンジン インスタンスに接続する必要があります。



WGS tools ローカルプラグインのインストールは、Applied Maths Cloud Calculation Engine が稼働している、2024 年 12 月 31 日までに行ってください。

3.2 インストール手順

WGS tools ローカルプラグインは、リレーショナルデータベースにインストールできる、オンラインプラグインとして利用可能です。これには、インストールの際に Windows の管理者権限が必要ないという利点があります。更に、マルチユーザデータベースの設定では、この手順は全てのデータベースユーザが同じプラグインバージョンで動作することを保証します。

以下の手順に従って、WGS tools ローカルプラグインをインストールしてください。

2.1 メインウィンドウにおいて、**File** → **Install/remove plugins** を選択すると、Plugins & Scripts ダイアログボックスが現れます。

2.2 **Manage database plugins** ボタンをクリックすると、Manage database plugins ダイアログボックスが表示されます。

Manage database plugins ダイアログボックスには、リレーショナルデータベースに現在保存されているプラグインが一覧で表示されます。初期状態では空です。

2.3 **Add/Update** ボタンをクリックすると、Add database plugins ダイアログボックスが表示されます。(図 3.1)

BioNumerics v8.1 manual
wgMLST tools local plugin

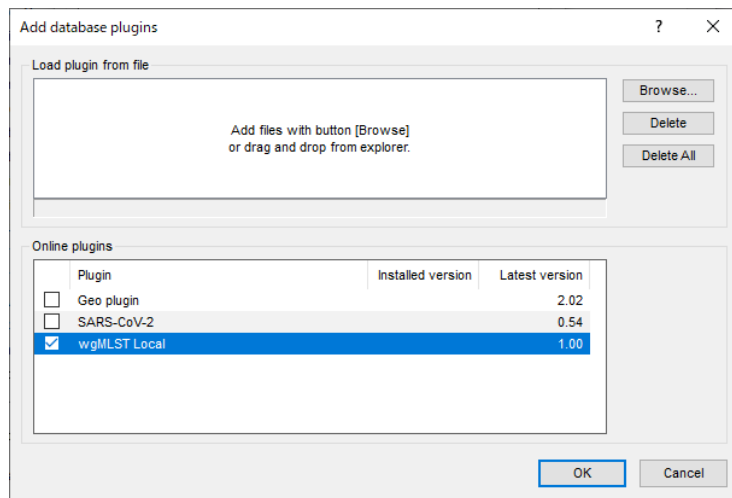


図 3.1 : Add database plugins ダイアログボックス

- 2.4 wgMLST Local プラグインにチェックを入れ、**OK** ボタンをクリックします。データベースの再起動後のプラグインがロードされることを示すメッセージが表示されます。
- 2.5 Add database plugins ダイアログボックスを閉じ、Plugins & Scripts ダイアログボックスも閉じてください。
- 2.6 ウィンドウを閉じ、BioNumerics データベースをリスタートします。
- 2.7 メインウィンドウにおいて、**File** → **Install/remove plugins** を選択すると、Plugins & Scripts ダイアログボックスが再び現れます。

WGS tools ローカルプラグインは、Plugins & Scripts ダイアログボックスの Plugins タブにおいて、リストの一番下に表示され、データベースアイコンが表示されます。(図 3.2)

BioNumerics v8.1 manual

wgMLST tools local plugin

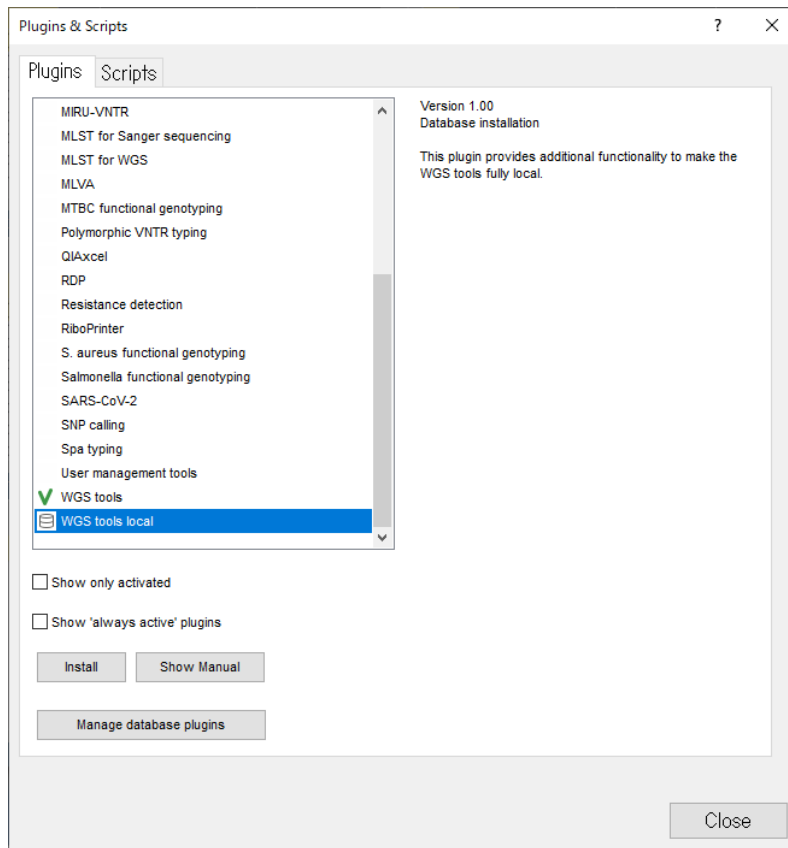



図 3.2 : Plugins &Scripts ダイアログボックスで、WGS tools local プラグインがリストされている

2.8 WGS tools local プラグインをクリックしてハイライトさせ、**Install** ボタンをクリックしてプラグインのインストールを開始します。このプラグインは計算エンジンと同期を行うことにより、自動的に **wgMLST_Local** タイプの実験を作成し、開始します。

 WGS tools local プラグインは、計算エンジン設定ダイアログボックスの実験タイプパネルで、wgMLST allele コールを含むよう指定された実験タイプを使い、実験名に **_Local** suffix を追加します。実際の名前が異なっても、実験タイプは簡潔にするために **wgMLST_Local** と呼ばれます。

2.9 同期が完了したら、**OK** ボタンをクリックしメッセージボックスを閉じます。

「Do you want to initialize the local allele nomenclature with the accepted alleles ? 」というメッセージが表示されます。

「No」にすると、各ローカスごとに 1 から始まる allele ID が割り当てられ、ローカル wgMLST 命名規則がゼロから作成されます。

「Yes」にすると、central 命名規則における現在の受け入れられた allele からローカル命名規則が始まり、新しい allele は増加する整数識別子によって追加されます。

BioNumerics v8.1 manual
wgMLST tools local plugin

✎ assembly-based でアクセプトされた allele が次に更新される時、その間にすでに（他の）ローカル allele に既に割り当てられている ID を持つ新しい受入 allele が生じる可能性があります。したがって、初期化以降、allele ID は必然的に分岐し始めます。

Assembly-based wgMLST 検索データが最新ではない場合、検索データの更新が最初に行われます。これには数分かかることがあります。

2.10 **Yes** ボタンをクリックします。このアクションは、ソースファイルディレクトリに wgMLST_Local_Nomenclature.txt ファイルを作成します。処理が完了すると、プラグインがインストールされたことを示す通知が表示され、データベースの再起動を促します。

2.11 Plugins &Scripts ダイアログボックスを閉じ、BioNumerics データベースをリスタートします。

Chapter 4 : ローカル命名法を用いた wgMLST allele calling

4.1 原理

各セッションの開始時に、WGS tools local plugin は、計算エンジンインスタンスへの接続が利用可能かどうかを確認し、それに応じて機能を調整します。

4.2 計算エンジンがシャットダウンする前に (2024 年 12 月 31 日シャットダウン)

計算エンジンインスタンスへの接続が利用可能な場合、WGS tools plugin が提供する全ての機能が引き続き利用可能です。更に assembly-based allele calling 結果が取得されるたびに (それがローカルで実行されるか、計算エンジン上で実行されるかに関わらず)、ローカルの wgMLST 命名法 (Chapter 1) に従って、assembly-based allele コールが、**wgMLST_Local** 実験タイプにインポートされます。

従って、データベース内の全てのエントリに対して **wgMLST_Local** 実験を入力するために、全ての assembly-based の wgMLST コールを再実行する必要があります。ローカル計算エンジンでは、これは CE クレジットを必要としない比較的早い処理です。非常に大きなデータベースについては、最大 1000 エントリのバッチでジョブを提出することをお勧めします。

ローカル命名法ではまだ知られていない allele (ハッシュ値で表される) は、各遺伝子座毎に増分整数として新しい allele ID が割り当てられます。wgMLST_Local 実験タイプでは、Assembly-free allele コールは、たとえそれが存在するか検索されたとしても考慮されません。

計算エンジンへの各リクエストに対して、そのレスポンスはキャッシュされ、リレーショナルデータベースへの DSETTINGS テーブルに保存されます。



計算エンジンのシャットダウン時に予期せぬエラーメッセージが表示されるのを防ぐため、計算エンジンへの接続が可能のうち、WGS tools メニューから各コマンドを少なくとも 1 回実行することをお勧めします。

4.3 計算エンジンシャットダウン後

計算エンジンインスタンスに接続できない場合、Submit jobs ダイアログボックス及び Submit comparison jobs ダイアログボックスでは、外部計算エンジンへのジョブ投入に関する全ての機能がグレースアウトします。(図 4.1)

BioNumerics v8.1 manual wgMLST tools local plugin

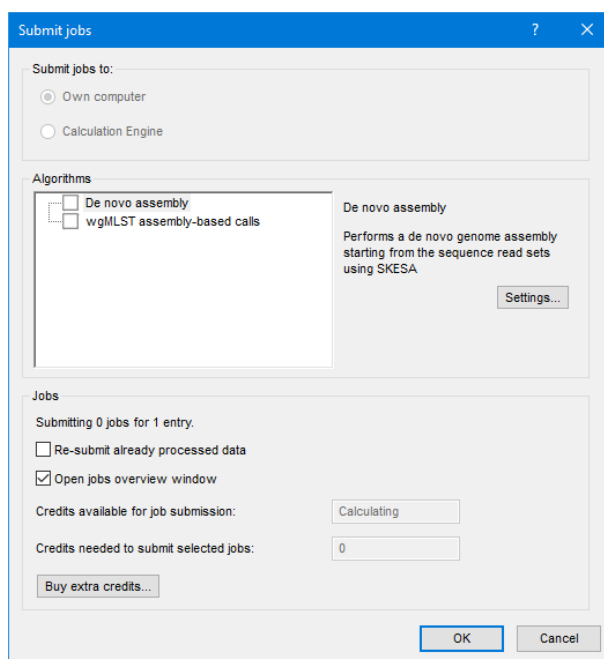


図 4.1 : Submit jobs ダイアログボックス、外部計算エンジンに Job を送信するオプションがグレーアウトされています

結果として、Job は自分のコンピュータ上のローカル計算エンジンのみを送ることが出来ます。

ローカルの wgMLST assembly-based calling ジョブの結果が検索されると、**wgMLST** と **wgMLST_Local** 実験タイプの両方が更新されます。**wgMLST** 実験は assembly-free allele calling ジョブが同じエントリの為に以前に実行された場合のコンセンサスコールを含みます。新しい allele が見つかった場合、これらの新しい allele は計算エンジンの命名データベースには提出できないので、wgMLST に保存された allele プロファイルは不完全となります。**wgMLST_Local** 実験に保存された allele プロファイル (ハッシュキーに allele コーリングによって得られる) は、WGS tools local plugin によって維持されたローカル命名法に基づいて完全なものになります。

計算エンジンへのどのようなリクエストに対しても、利用可能であればキャッシュされた応答が提供されます。そうでない場合は、エラーメッセージが表示されます。いくつかの wgMLST 機能は、特定の命名法サービスに依存しているため、利用できなくなります。

- **WGS tools > Assign wgMLST sequence types**

「The “Assign sequence types” nomenclature service is no longer available」と表示されるようになります。MLST sequence タイプは、MLST for WGS プラグインでアサインすることが出来ます

- **WGS tools > View wgMLSt reports**

キャッシュされたレポートのみを表示する「The “Update reports” nomenclature service is no longer available」と表示されるようになります。同様のレポートは MLST for WGS プ

ログインでも生成できます。

- **WGS tools > Store wgMLST locus sequences**

「The “Get allele sequences” nomenclature service is no longer available」と表示されるようになります。WGS プラグインは、オプションで MLST loci の配列を保存することが出来ます。

- **WGS tools > Get alleles mapping**

「The “Get allele mapping” nomenclature service is no longer available」と表示されるようになります。小規模なスキームでは、MLST for WGS プラグインが代替となるでしょう。

- **Alleles > Open alignment**

WGS tools > wgMLST quality assessment を実行して表示される、wgMLST Quality Assesment ウィンドウ内の機能ですが、シャットダウン後は「The “Get allele sequences” nomenclature service is no longer available」と表示されるようになります。Allele 配列のアラインメントは、assemble-based allele calling 検索データから accepted alleles ファイルから allele 配列をインポートした後に、手動でのみ生成することが出来ます。

Chapter 5 : Hash-based コールのインポート&エクスポート

5.1 原理

ローカルな命名法が使用されているため、異なる BioNumerics データベースの **wgMLST_Local** 対立遺伝子プロファイルを直接比較することはできません。他のデータベースやラボとデータを交換し比較するためには、allele ID を一意に参照できるもの、即ち allele 配列から得られるハッシュ値に変換する必要があります。

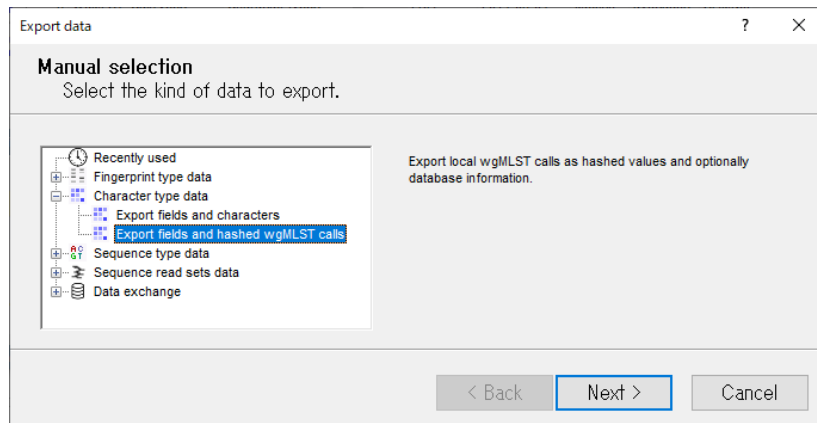
インポート及びエクスポートウィザードには、ハッシュ化された wgMLST コールをインポート及びエクスポートするためのメソッドが Fields and characters に追加されています。見た目はデフォルトのインポートやエクスポートと同じですが、character 実験が **wgMLST_Local** 実験タイプに制限されていることと、全ての (local) allele id が allele ハッシュと変換されることを除きます。

ローカルの命名法にまだ存在しない allele ハッシュをインポートする場合、それは追加され（すなわち、新しい allele id が割り当てられ）、同じ allele がその後のローカルでの検出は、この新しい allele id が割り当てられます。

5.2 プロファイルのエクスポート

一組のエントリの allele ハッシュをエクスポートするには、以下の手順に従ってください。


- 2.1 エクスポートしたい任意のエントリを選択し、**File / Export** を実行します。
- 2.2 Export ダイアログボックスにおいて、**Character type data** を開き、**Export fields and hashed wgMLST calls** を選択して、**Next** ボタンをクリックします。



- 2.3 次に表示される Export ダイアログボックスでは、エクスポートしたい **Fields** アイテムと **wgMLST_Local** 実験のサブセット (**Characters** 欄) を選択します。完了したら **Next** ボタンをクリックします。

2.4 次に表示される Settings ダイアログボックスでは、欠損値をどのように表記するか、及びエクスポートをアクティブな character のみに制限するかどうかを指定し、**Finish** ボタンをクリックします。

この操作により、データベースディレクトリに exoprt.csv ファイルが作成され、あなたの P C のデフォルトエディタで csv ファイルが開きます。

ほとんどの PC でデフォルトの csv エディタである MS Excel は、allele ハッシュを科学表記（例：5.83E+16）で表示します。Categorical データの場合、これは実際には意味を成しません。

5.3 プロファイルのインポート

エクスポートしたテキストファイルから allele hash を含むエントリをインポートするには、以下の手順に従います。

3.1 **File / Import** を選択します。

3.2 先にエクスポートしたテキストファイル (*.csv または *.txt) を参照し、**Import fields and hashed wgMLST calls (text file)** オプションをハイライトし、**Finish** ボタンをクリックします。

3.3 **Next** ボタンをクリックします。

残りの手順は、**wgMLST_local** character 実験タイプだけが、ターゲットとして選択できることを除き、テキストファイルから fields and character データをインポートするのと同じです。