1. はじめに

ハッシュベースの allele (対立遺伝子) コーリングは、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。したがって、2024年12月31日に AppliedMaths 社が提供するクラウド計算エンジンが BioNumerics のサポート終了後も引き続き実施することが出来ま す。BioNumerics サポート終了に関しては以下のサイトをご覧ください。

(https://www.bionumerics.com/news/bionumerics-phasing-out)

WGS ツールのローカルプラグインに実装されたハッシュベースの allele コーリングアルゴリズムでは、名称受容基準を満たす allele シーケンスがハッシュ関数によって 17 桁の整数に変換されます。ハッシュ関数は決定論的なので、特定の allele シーケンスは常に同じハッシュ値になります。しかし、このようなハッシュ値そのものは、ソフトウェア内で直接使用できません。なぜなら、BioNumerics の Character タイプは 17桁の整数を確実に保存及び比較するようには設計されていないからです。また、ハッシュ値は wgMLST ユーザーが馴染みのある単純な清秋の allele ID とは大きく異なります。

これらの問題を回避するために、WGS ツールのローカルプラグインは、ハッシュを単純な整数 ID に変換す るローカル allele 命名規則を設定します。これらの ID は wgMLST_Local 実験タイプに保存され、自 動的に作成されるので、wgMLST 実験タイプと同期されます。

wgMLST アセンブリベースの allele コーリングジョブの結果が取得されるたびに、wgMLST_Local 実験はローカル命名規則を使ってハッシュベースの allele コールで自動的に更新されます。

異なるラボやデータベース間で wgMLST データを交換するために、ローカル wgMLST プロファイルはハッ シュ値としてエクスポートでき、ハッシュ値からインポートすることも可能です。 (詳細は 6 を参照してください)

ハッシュベースの wgMLST allele コーリングを設定し実行するプロセスについては、Listeria monocytogenes のデモデータベースを使って、このチュートリアルで説明します。

2. データベースの準備

2.1 オプション1:セットアップスクリーンからデモデータベースをダウンロードする

1. BioNumerics スタート画面において、
(Download tutorial databases) ボタンをクリ ックすると、Tutorial database ウィンドウが開きます。 (図 1)

🖆 Tutorial databases					- 0	×
File Database Tutorial Window						
, Tutorial databases						
Ð						
Name	Size (MB)	Min	Max	•		
BandScoring demo database	15	7	8.1	~	Demonstration database containing data for	\sim
DemoBase Connected	4	7.5	8.1		a set of 51 Listeria monocytogenes isolates.	
Neisseria MLST demo database	1	7.5	8.1		Inis database uses publicly available next-	
SNP demonstration database	0	7.5	8.1		Sequence Read Archive (SRA) For each	
WGS_demo_database_for_Brucella_spp	214	7.6	8.1		isolate, NGS reads were de novo assembled	
WGS_demo_database_for_Burkholderia_cepacia	142	8	8.1		into genome sequences. wgMLST alleles	
WGS_demo_database_for_Escherichia_coli	601	7.5	8.1		were called using the assembly-based and	
WGS_demo_database_for_Listeria_monocytogenes	288	7.5	8.1		assembly-free method.	
WGS_demo_database_for_MTBC	279	8	8.1			
WGS_demo_database_for_Salmonella_enterica	618	8	8.1			
WGS_demo_database_for_Staphylococcus_aureus	624	7.6	8.1			
	pases					
	٤		2		I	
Tutorials						
Name				•		
Whole genome MLST typing in the Listeria monocytogenes d	emonstration (data		^	A tutorial for working with the whole genome	
MLST analysis starting from whole genome sequences					Multi Locus Sequence Typing (wgMLST)	
Listeria functional genotyping: predicting phenotypic traits fr	om whole gen	ome			monocytogenes, which allows BioNumerics	
Whole genome MLST typing in the Listeria monocytogenes d	emonstration (data			users to explore the woMLST functionality	
Fast character-based identification					without the need for a wgMLST project or	
Minhashing based cluster analysis of sequences					Calculation Engine credits.	\sim
Hash-based whole genome MLST				\mathbf{v}		

図 1: Tutorial database ウィンドウ

- WGS_demo_database_for_Lysteria_monocytogenes をリストから選択し、
 Database → Download を実行します。
- 3. ダウンロードが完了するとその旨の画面が表示されるので OK ボタンをクリックします。
- 4. File → Exit を選択し、Tutorial database ウィンドウを閉じます。
 WGS_demo_database_for_Listeria_monocytogenes が BioNumerics スター
 ト画面に表示されているのが分かります。
- 5. WGS_demo_database_for_Listeria_monocytogenes をダブルクリックし、データベースを開いてください。

2.2 オプション2:バックアップファイルからデモデータベースをリストアする

Listeria monocytogenes の WGS demo データベースには、BioNumerics バックアップファイルもウ ェブサイトにご用意があります。このバックアップファイルは BioNumerics にリストアすることが出来ます。

 <u>https://www.bionumerics.com/download/sample-data</u> サイトの Demonstration databases リストから "WGS_demo_database_for_Listeria_monocytogenes" を選 択すると、WGS_LMO1.bnbk ファイルがダウンロードされます。

Demonstration databases

- BandScoring demo database (15 MB)
- DemoBase Connected (4.3 MB)
- Neisseria MLST demo database (1.6 MB)
- SNP demonstration database (367 KB)
- WGS_demo_database_for_Brucella_spp (210 MB)
- WGS_demo_database_for_Burkholderia_cepacia (136 MB)
- WGS_demo_database_for_Escherichia_coli (587 MB)
- WGS_demo_database_for_Listeria_monocytogenes (281 MB);
- wGS_demo_database_lot_MTBC (266 MB)
- WGS_demo_database_for_Salmonella_enterica (590 MB)
- WGS_demo_database_for_Staphylococcus_aureus (595 MB)

 一部のブラウザでは、WGS_LMO1.bnbk データベースバックアップファイルの名前を WGS_LMO1.zip に変換することがあります。この場合、手動で.zip 拡張子を .bnbk に置き換える 必要があります。「If you change a file name extension, the file might become unusable」 という警告メッセージが表示されますが、この操作を実施しても大丈夫です。ただし、Windows のフォルダ オプションにおいて「既知のファイルタイプの拡張子を隠す」オプションにチェックをしている場合、.zip という 拡張子が表示されないので、ご注意ください。

- 7. BioNumerics のスタート画面において、 記 (Database management) ボタンをクリックし、 **Restore database** メニューをクリックします。
- 8. Browse ボタンをクリックし、ステップ 6 で保存した bnbk ファイルを選択し、 **Create copy** にチェ ックを入れます。 **Overwrite** にチェックを入れると、既に存在しているデータベースが上書きされま す。
- 9. データベース名を記入します。(例:WGS Listeria demobase)
- 10. OK をクリックするとバックアップファイルからのデータベースリストアが始まります。(図 2)

0.00 0.0070000		-
C:\Users\xyx00700\Dd	whibads/wGS_LMO1.bhbk	Browse
Destination		
Overwrite		
Overwrite the cur	rent data in database	
WGS_demo_datat	ase_for_Listeria_monocytogenes with a prev	ious version of
the data as stored	in a backup. Any new data since the backup v was erased will be recovered	will disappear,
any old data that y		
Create copy		
Create a new data	base from the data stored in a backup. Use th	is option to
have access to be	oth old and current versions of an existing data	base or to
copy a database t	rom anotner computer.	
	WGS Listeria demobase	
New database name:		

図2: バックアップファイル(WGS_LM01.bnbk)からWGSデモデータベースをリストアする

11.リストア終了後、Yes ボタンをクリックするとデータベースが開きます。(図3)

perin	ent types		Dat	abase entries	8											9	omparisor	s		
9	+ 13 0 1		. 4	1 + P		P	«All Entries»		10								+ 13	⊗ B.	16	2
-	# Name	Type		Key	Modified date	CollectedBy	Collection	GeoLoc	IsolationSource	Sampleld	serovar	Source	. 1	200	34567	61 T	Name		M	dified dat
5	1 was	Sequence read set types		SRR1112181	2015-12-08 09 40:46	CDC	missing	USA	missing	SAMN02582707	<not prov<="" td=""><td></td><td></td><td></td><td></td><td>× 1</td><td>Isolater</td><td>taly</td><td>201</td><td>15-12-08 1</td></not>					× 1	Isolater	taly	201	15-12-08 1
5	2 WOMLST	Character types		SRR1378346	2015-12-08 09:40:46	missing	2014-01-21	Indonesia	frozen lof raw pe	SAMN02850677	<not prov<="" td=""><td>shrimp</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></not>	shrimp								
5	3 denovo	Sequence types	Ö	SRR1535729	2015-12-08 09:40:46	FDA	2014-06-04	Mexico	avocados	SAMN02934629	<not prov<="" td=""><td>avocado</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></not>	avocado								
	4 quality	Character types		SRR1562163	2015-12-08 09:40:46	CDC	2014-08	USA	Blood	SAMN03013717	1/2a	blood								
-	5 wos TrimmedStat	s Sequence read set types		SRR1566202	2015-12-08 09:40:46	FDA	1993-06	taly	cheese pastry	SAMN02769736	1/2b	cheese pastry								
۲.	6 wgMLST_CalType	s Character types		SRR1566203	2015-12-08 09:40:46	FDA	1993-06	Raly	fruit cake	SAMN02769737	1/26	fruit cake								
8	7 wgsLong	Sequence read set types		SRR1610008	2015-12-08 09:40:46	FDA	1993-06	taly	vol au vent shrimp	SAMN02769735	1/2b	shrimp					100			
		×		SRR1623013	2015-12-08 09:40:46	FDA	09-Sep-2011	USA:CO	blood	SAMN02769790	1/2a	blood					<			
	<	,		SRR1623016	2015-12-08 09:40:46	FDA.	8/5/2011	USA:CO	blood	SAMN02769789	1/2b	blood				10	entitication (roects be	ision net	20005
77	Catilitassa (Salitas)			SRR1623030	2015-12-08 09:40:46	FDA	22-Sep-2011	USA:CO	blood	SAMN02769793	1/28	blood					C	Manager Street	-	Abataadii
10.272	all internet added and			SRR1654984	2015-12-08 09:40:46	Austrian A	2014	Germany	human	SAMN03168687	4b	human					2日 +	D 0	昆,	白
•			10	SRR1656973	2015-12-08 09:40:46	Austrian A	2014	Austria	human liquor	SAMN03168680	4b	human					Name		Mo	dified dat
	Name	Field type		SRR1657003	2015-12-08 09:40:48	Austrian A_	2014	Austria	blood	SAMN03168679	4b	blood					-			
	CollectedDy	Event		SRR1695809	2015-12-08 09:40:46	CDC	Oct-2014	USA	Blood	SAMN03253354	4b	blood								
	CollectionDate	Fired		SRR1695811	2015-12-08 09:40:46	CDC	Oct-2014	USA	Blood	SAMN03253356	4b	blood								
	Geni nchiame	Fired		SRR1695816	2015-12-08 09:40:46	CDC	missing	USA	missing	SAMN03253362	<not prov<="" td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></not>									
	isolation Source	Fixed		SRR1695823	2015-12-08 09:40:46	CDC	Nov-2014	USA	placenta	SAMN03253369	4b	placenta				- H. H.				
к	Samilekt	Fixed		SRR1695834	2015-12-08 09:40:46	CDC	Unknown	USA	CSF	SAMN03253380	4b	CSF								
	AATTIVAT	Fixed		SRR1695836	2015-12-08 09:40:46	CDC	Nov-2014	USA	Blood	SAMN03253382	4b	blood								
ĸ	Source	Figed		SRR1709560	2015-12-08 09:40:46	FDA	1993-06	taly	mover surface	SAMN02769739	1/28	mixer surface					.5			
	wold.ST loci ST	Fiexble		SRR1709629	2015-12-08 09:40:46	FDA	1993-06	Italy	freezer surface	SAMN02769738	1/2b	freezer surface			• •	197	Constants (WARDING.	TO (TT	and prove the
к	MLST ST	Fiexible		SRR1709681	2015-12-08 09:40:46	FDA	1993-06	Italy	blood	SAMN02769734	1/2b	blood					A ADDRESS OF	And the second second		And Descent and
	MI ST Publik ST ST	Fariha		SRR1745438	2015-12-08 09:40:46	CDC	Nov-2014	USA	Blood	SAMN03275454	4b	blood					+ 0	⊗ €.	睑	1 2
				SRR1745445	2015-12-08 09:40:46	CDC	Nov-2014	USA	Sputum	SAMN03275461	4b	sputum					Name		Mo	dified da
	t files Power assemble	Anotations		SRR1745458	2015-12-08 09:40:48	CDC	Nov-2014	USA	Blood	SAMN03275472	1/2b	blood					and and a state of the state of		0000	
n I	I DA O B	0 0		SRR1745464	2015-12-08 09:40:46	CDC	Unknown	USA	CSF	SAMN03275478	4b	CSF								
7	+ [] @ 6	E C C CAll Fingers		SRR1745479	2015-12-08 09:40:46	CDC	01-Dec-2014	USA	Blood	SAMN03275491	1/2b	blood								
1		experiment type 🛛 Link 🔫		SRR1745488	2015-12-08 09:40:46	CDC	01-Dec-2014	USA	Blood	SAMN03275500	4b	blood								
		~		SRR1746767	2015-12-08 09:40:46	CDC	Nov-2014	USA	Hip fluid	SAMN03198339	4b	hip fluid								
				SRR1763833	2015-12-08 09:40:46	FDA	2015-01-05	USA:MO	granny smith apples	SAMN03278638	<not prov<="" td=""><td>apples</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></not>	apples								
		~		SRR1763858	2015-12-08 09:40:46	FDA	2015-01-08	USA:MO	granny smith apples	SAMN03278634	<not prov<="" td=""><td>apples</td><td></td><td></td><td></td><td>~</td><td></td><td></td><td></td><td></td></not>	apples				~				

図 3: Listeria monocytogenes デモデータベースの Main ウィンドウ

3. サンプルデータベースについて

WGS Listeria デモデータベースには、51 の公開されているシーケンシングランに関する NCBI のシーケン スリードアーカイブ(SRA)上のシーケンスリードセットデータへのリンクが含まれています。シーケンスリード セットタイプ実験 wgs には、いくつかの生データ統計と共に、NCBI(SRA)上のシーケンスリードセットへ のリンクが含まれています。

このサンプルセットの完全な wgMLST 分析 (de novo assembly, assembly-based calls, assembly-free calls) は、デフォルト設置と L.monocytogenes wgMLST スキームを使用して、 Applied Maths Calclation engine 上で実施されました。

1. WGS tools \rightarrow Settings を選択します。

計算エンジンプロジェクトは、Listeria monocytogenes アリルデータベースに関連付けられています。 このプロジェクトにはクレジットが割り当てられていないので、外部計算エンジンにジョブを送信することはで きませんが、「Enable running jobs on my own computer」オプションが Calculation engine タブでチェックされているので、皆さんの PC 上でジョブを実行することが可能です。(図 4)

Salculation engine	Organism Experiment types waMLST	
	 Organismi Experiment types wgmcot 	
Remote calculation eng	ine	
Url:	Applied Maths cloud On premises	
	https://wgmlst.applied-maths.com	
Project name:	eval_lmo_demo	
Password:	•••••	
Time out (sec):	120	
Own computer		
Enable running job	os on my own computer	
Interval (min):	1	
interver (min).		

図4: Calculation engine 設定ダイアログボックスの Calculation engine タブ

 wgMLST タブ(図 5)に移動し、Auto submission criteria ボタン(図 6)をクリックします。 デフォルトでは、Use nomenclature acceptance criteria オプションにチェックが入っていて、自 動サブミッション設定は allele データベースのキューレーターによって定義されていることを意味します。

							?	×
Calculation (engine	Organism	Experimer	it types	wgML	ST		
New allele sub	mission							
Lab ID:	eval_In	no_demo						
Submit ne	w alleles a	utomatically						
۵	luto submi	ssion criteria						
Multiple allele c	alls charac	cter values						
Store low	est comm	on allele ID						
⊖ Store as	absent val	ue						
Store as Allele mapping	absent val	ue ts						
Store as Allele mapping wgMLST	absent val experimen _MLST Pu	ue ts						
Allele mapping	absent val experimen _MLST Pu _Core Pas	ue ts bMLST teur						
Allele mapping	absent val experimen MLST Pu Core Pas	ue ts bMLST teur						
Allele mapping	absent val experimen _MLST Pu _Core Pas	ue ts bMLST teur						
Allele mapping	absent val experimen _MLST Pu _Core Pas	ue ts bMLST teur						
Allele mapping	absent val experimen _MLST Pu _Core Pas	ue ts bMLST teur						
Store as Allele mapping wgMLST wgMLST	absent val experimen _MLST Pu _Core Pas	ue ts bMLST teur						

図 5 : Calculation engine 設定ダイアログボックスの wgMLST タブ

Auto submission criteria		?	×
Use nomenclature accept	tance c	riteria	
O Set custom submission o	riteria		
Require start/stop codons	(for CE)S loci (only)
Allow internal stop codon	s (for C	DS loci	only)
Minimum homology	85.0	%	
Maximum number of gaps	999		
ок		Can	cel

図 6 : Auto submission criteria ダイアログボックス

3. Cancel を 2 回クリックし、Calculation engine 設定ダイアログボックスを閉じます。

wgMLST 解析にリンクされた実験タイプは、各エントリのデータベースに存在し、Experiment types パネルに表示されています。

Experim	ient ty	pes			
କ୍ଷ	-	+ 🗗 🛛 民	Ê <u>.</u> 1	Ļ	<all experiment="" th="" typ<=""></all>
	#	Name	Туре		-
₹ [1	wgs	Sequence read set types		^
	2	wgMLST	Character types		
🗆 🛟	3	denovo	Sequence types		
	4	quality	Character types		
≣ ≩	5	wgs_TrimmedStats	Sequence read set types		
	6	wgMLST_CallTypes	Character types		
≣ ≣	7	wgsLong	Sequence read set types		
					~

図 7: Main ウィンドウの Experiment types パネル

- Character 実験タイプ wgMLST には、各サンプルで検出された loci の allele コールが含まれ、 assembly-based 及び assembly-free コーリングからのコンセンサスの結果、単一の allele ID が得られました。
- Sequence 実験タイプ denovo には、de novo assembly アルゴリズムの結果、すなわち連結された de novo コンティグ配列が含まれます。
- Character 実験タイプ quality には、生データ、de novo assembly、様々な allele 同定ア ルゴリズムの品質統計が含まれます。
- Sequence read set 実験タイプ wgs_TrimmedStats: トリミング後に保持されたリードに 関する統計データです。
- Character 実験タイプ wgMLST_CallTypes : コールタイプの詳細です。

Enty info fields (CollectionDate、CollectedBy、Serovar など) に追加された追加情報は、 対応する出版物から収集され、デモデータベースに追加されました。

データベースのエントリー横にある緑色の●をクリックすると、対応する結果を別ウィンドウまたは character data タイプの実験カードで見ることが出来ます。 4. 最初のエントリの一番左(1番)の緑色の●をクリックしてください。(今回のデモデータベースの場合、Sequence rad set types : wgs 実験の情報が表示されます)

Data	base entries														
₽ _₽ ₽] + 🖻	⊗ €	₹ I	ß	∽	<all entries<="" th=""><th>•</th><th>1</th><th>ย</th><th></th><th></th><th></th><th></th><th></th><th></th></all>	•	1	ย						
	Key		Collect	tedBy	Collection	Geo	Isolation S	SampleId	serovar		T	1	2		4 5
	SRR1112181		CDC		missing	USA	missing	SAMN025	<not provided=""></not>			•	•	•	• •
	SRR1378346		missing		2014-01-21	Indon	frozen iqf r	SAMN028	<not provided=""></not>	shrimp		٠	•	•	• •

ウインドウには、NCBI (SRA) のシーケンスリードセットへのリンクと、シーケンスリードセットの特徴の要約が表示されます。Read set size、Sequence length statistics、Quality statistics、Base statistics (図 8) 。



図 8: エントリの sequence read set 実験カード

5. Sequence read set 実験ウィンドウを閉じます。

 最初のエントリの 2 番目の緑色の●をクリックしてください。(今回のデモデータベースの場合、 Character types : wgMLST 実験の情報が表示されます)

Character type 実験 **wgMLST** には、各サンプルで検出された loci の allele コールが含まれ、 assembly-based と assembly-free コーリングからのコンセンサス結果、単一の allele ID が得られ ました。 (図 9)

SRR1112181			
Character		Mapping	-
LMO_1	2	<+>	^
LMO_10	1	<+>	
LMO_11	1	<+>	
LMO_13	1	<+>	
LMO_14	1	<+>	
LMO_18	1	<+>	
LMO_19	3	<+>	
LMO_21	1	<+>	
LMO_48	2	<+>	
LMO_73	2	<+>	
LMO_104	12	<+>	
LMO_106	2	<+>	~
Press Insert to add chara	cter		1.

図9: エントリの character 実験カード

- 7. 左上の三角をクリックすると、character experiment カードが閉じます。
- 最初のエントリの 3 番目の緑色の●をクリックしてください。(今回のデモデータベースの場合、 Sequence types : denovo 実験の情報が表示されます)

Sequence editor ウィンドウが開き、de novo assembly アルゴリズムの結果、即ち連結された de novo コンティグ配列が表示されます。(図 10)



図 10 : Sequence editor ウィンドウ

9. Sequence editor ウィンドウを閉じます。

```
10.最初のエントリの 4 番目の緑色の●をクリックしてください。(今回のデモデータベースの場合、
Character types : quality 実験の情報が表示されます)
```

Quality character カードには、生データ、de novo アセンブリ、及び様々な allele 同定アルゴリズム に関する品質統計が含まれています。(図 11)

Character	Value	Mapping	
AvgQuality	36	<+>	^
AvgReadCoverage	73	<+>	
SrsQ30Freq	92	<+>	
SrsQ30Freq_1	96	<+>	
SrsQ30Freq_2	88	<+>	
AvgQualityTrimmed	36	<+>	
AvgReadCoverageT	73	<+>	
SrsQ30Freq_Trimmed	92	<+>	
SrsQ30Freq_1_Tri	96	<+>	
SrsQ30Freq_2_Tri	88	<+>	
N50	1458828	<+>	
NrContigs	13	<+>	~

図 11: エントリの character 実験カード

11. 左上の三角をクリックすると、character experiment カードが閉じます。

4. WGS ツール ローカルプラグインのインストール

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。

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。

- 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. **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。

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 12)

		Add files with button [I	Browse]		Delete			
	or drag and drop from explorer.							
Inline	plugins							
	Plugin		Installed version	Latest version				
	Geo plugin			2.02				
	SARS-CoV-2			0.54				
\checkmark	wgMLST Local			1.00				

図 12 : Add database plugins ダイアログボックス

- 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。
- 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。
- 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。
- 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 13)

Plugins Scripts				
INGUINS SCITPIS MIRU-VNTR MLST for Sanger sequencing MLST for WGS MLVA MTBC functional genotyping Polymorphic VNTR typing QIAxcel RDP Resistance detection RiboPrinter S. aureus functional genotyping SARS-CoV-2 SNP calling Spa typing User management tools ✓ WGS tools Show only activated Show valways active' plugins Install Manage database plugins Manage	Ť	Version 1.00 Database installation This plugin provides additional functionality to WGS tools fully local.	o make the	

図 13: 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

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いても、実験タイプは簡潔にするために wgMLST Local と呼ばれます。

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。

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実行されます。これには数分かかることがあります。

- 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。
- 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。

5. アセンブリベースの wgMLST コールの(再)実行

メインウィンドウの Experiment presence パネルから、現在 wgMLST_Local 用の実験が存在して いないことがわかります。データベース内の全てのエントリに対して wgLST_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。

- 1. Edit / Select all (Ctrl+A) コマンドを使い、Listeria monocytogenes デモデータベース の全エントリを選択します。
- 2. WGS Tools → Submit jobs コマンドを実行します。
- 3. Submit jobs ダイアログボックス(図 14)において、**Own computer** にチェックが入っているこ とを確認し、CE クレジットを必要としないローカル計算エンジンでジョブを実行します。
- 4. Algorithms 設定において、wgMLST assembly-based calls にチェックを入れ、jobs 設 定では Re-submit already processed data にチェックを入れます。

ubmit jobs	? ×
Submit jobs to:	
Own computer	
O Calculation Engine	
Algorithms	
De novo assembly	wgMLST assembly-based calls
	Performs a BLAST-based wgMLST allele calling on the assembled genome sequence
	Settings
Jobs	
Submitting 51 jobs for 51 entries.	
Re-submit already processed data	
Open jobs overview window	
Credits available for job submission:	0
Credits needed to submit selected jobs:	0
Buy extra credits	
	OK Cancel

図 14 : Submit jobs ダイアログボックス、ローカル計算エンジンでの wgMLST assemblybased allele コーリングジョブの再計算設定

5. OK ボタンをクリックし、ジョブを進めます。

Local assembly-based wgMLST 検索データが最新ではない場合、それらが最初に更新されます。 これには数分かかるかもしれません。次に、ジョブがローカル計算エンジンに投入されます。

6. ジョブが始まると自動的に Overview ウィンドウが開きますが、開かない場合は WGS tools / Jobs overview を実行すると、job overview ウィンドウが開きます。

Assembly-based allele コーリングジョブは比較的早く完了します。私たちのテストセットアップでは、 51 エントリのジョブが約 25 分で終了し、ジョブ結果を取得して BioNumerics データベースに保存する のに、更に 5 分かかりました。

- 7. View → Refresh (F5) を選択し、実行の進捗状況を確認します。
- ジョブが完了したら、job overview ウィンドウでそれらを選択し、Jobs / Get results を実行す ると結果が取得できます。

Experiment presence パネルに、データベース内の各エントリの wgMLST_Local experiment type の緑色の●が表示されます。

9. 任意のエントリにおいて、wgMLST_Local 実験の●と、同じエントリの wgMLST 実験の●をク リックすると、実験カードウィンドウが表示されます。(図 15)

Central 命名規則の受け入れられた allele からローカル命名規則を開始したため(ステップ 4)、両 方の wgMLST プロファイルはほぼ同一に見えます。wgMLST では assemble-free の allele コー ルが考慮されるのに対して、wgMLST_Local では assembly-based の allele のみが使用される ため、(locus LMO49 を例とした図 15 の通り)、わずかな違いが生じることがあります。 時間が進 むにつれて、両方の命名法は乖離し始め、異なる ID が割り当てられる新しい allele のために、両方の 実験間でより多くの差異が観察されるようになります。計算エンジンがオフラインになると、新しい allele は wgMLST ではコールされる、プロファイルは不完全となります。対照的に、wgMLST_Local に格 納された wgMLST プロファイルは、パッシュベースの allele コールが新しい allele をコールし続けるので、 完全なまま保持されます。



図 15: エントリ: SRR1657003の wgMLST (左)と wgMLST_Local (右)の実験カードウ ィンドウ。 Local LMO49 は、両プロファイルで異なるコールとなっていることに注意。

wgMLST_Local 実験は、wgMLST 実験と同じ方法で分析でき、結果は同じように解釈できます。

6. Hash-based コールのインポート&エクスポート

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。

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。

- 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. Export ダイアログボックスにおいて、 Character type data を開き、 Export fields and hased wgMLST calls を選択して、 Next ボタンをクリックします。

Export data	?	×
Manual selection Select the kind of data to export.		
Recently used Fingerprint type data Fingerprint type data Character type data Fingerprint fields and characters Fingerprint fields and hashed wgMLST calls Port Sequence type data Port Sequence read sets data Port Data exchange	Export local wgl/ILST calls as hashed values and optionally database information.	
	< Back Next > Cancel	

- 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 Fileds 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。
- 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。

インポートは、WGS tools プラグインと WGS tools local プラグインの両方がインストールされている、 別の BioNumerics データベースで行う必要があります。 WGS tools プラグインのインストール手順は、 既に「4. WGS ツール ローカルプラグインのインストール」で説明済ですので、ここでは手順を繰り返し説 明はしません。

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。

- 5. File / Import を選択します。
- 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。
 Next ボタンをクリックします。
- Import template はまだ用意されていないので、まずは作成する必要があります。このチュートリアルでは、他のデータベースも SRA accessions をデータベースキーとして使用していると仮定します。
- 8. Import rules ダイアログボックスにおいて「Key」をダブルクリックし、Edit data destination ダイア ログボックス(図 16)で保存先として Key を選択し、OK ボタンをクリックします。



図 16 : Edit data destination ダイアログボックス、Key フィールド選択

- 9. Import rules ダイアログボックスに戻るので、全ての wgMLST loci をハイライトし、**Edit destination** ボタンをクリックします。
- 10. Edit data destination ダイアログボックスにおいて、**Hashed wgMLST call** の下の wgMLST_local を選択します。(図 17)



図 17: Edit data destination ダイアログボックス、wgMLST_Local を選択

- 11. Import rules ダイアログボックスにおいて、Next ボタンをクリックします。
- 12. Import links ダイアログボックスにおいて、**Key** にチェックを入れたまま、**Finish** ボタンをクリックしま す。
- 13. Import template の Name に任意の名前を入力し(例: Hashes)、OK ボタンをクリックします。

Import template が作成されたので、ハッシュ化された wgMLST allele コールのインポートに使うことが出来るようになります。

- 14. Import template ダイアログボックスにおいて **Next** をクリックし、**Hashes** import template を適用します。
- 15. Finish ボタンをクリックすると、 hash-based allele コールがインポートされます。