

環境の導入のために必要なポイントをまとめ、失敗しないで導入するための参考としていただきたい。

- GView Server の使い方
- NCBI Blast+の Windows 環境での利用について

### 3. MiSeq を使用する場合のコスト (試算結果例)

MiSeq を使って、ゲノム解析する場合、試薬が高いので、失敗をしないようにすることが重要である。我々は、複数の細菌のドラフト配列を得ているが、最適なクラスターを形成するために、DNA

ライブラリーの定量が非常に重要なステップとなる。複数の株を同時に解析するとどうしても、カバー率のばらつきが出るので、たくさんの株を解析する場合には、同じ DNA ライブラリーを 2 回ランすることとしている。1 回目のランで、クラスター形成のばらつきを判定し、2 回目のランでは、そのばらつきを解消するように DNA ライブラリーのインプット量を調整する。この方法は失敗が少なくなるので、推奨したい。ここではこの方法を用いたときの消耗品、試薬のコスト計算をしたので、紹介する。

表 1. *S. Infantis* 77 株のゲノム配列データのアセンブル後の概要

## Salmonella Infantis 77 株

Miseq Reagent Kit v3 (600 cycle)

		平均値	最大値	最小値
No. of reads		2,047,210	3,835,990	603,348
Total base		490,074,777	825,754,204	158,906,599
Coverage		100	169	33
Contig	Total base	4,891,271	5,158,601	4,588,724
	Count	138	567	64
	Maximum	1,051,615	1,241,747	480,288
	N50	290,813	521,185	159,289

表 2. 比較ゲノム学的手法を用いて、臨床分離 *S. Infantis* 株に特異的に存在することが期待された 56 の候補遺伝子

No.	Seq_ID	Hit Product
1	pj039-01	hypothetical protein
2	pj039-02	hypothetical protein
3	pj039-03	hypothetical protein
4	pj039-04	hypothetical protein
5	pj039-05	hypothetical protein
6	pj039-06	HTH-type transcriptional regulator hdxR
7	pj039-07	SanA protein
8	pj039-08	Cytidine deaminase
9	pj039-09	hypothetical protein
10	pj039-10	hypothetical protein
11	pj039-11	hypothetical protein
12	pj039-12	Ferredoxin
13	pj039-13	Mobile element protein
14	pj039-14	hypothetical protein
15	pj039-15	hypothetical protein
16	pj039-16	hypothetical protein
17	pj039-17	ClpB protein
18	pj039-18	hypothetical protein
19	pj039-19	hypothetical protein
20	pj039-20	Glutamate racemase
21	pj039-21	Glutamate racemase
22	pj039-22	Vitamin B12 receptor BtuB
23	pj039-23	hypothetical protein
24	pj039-24	hypothetical protein
25	pj039-25	5-deoxy-glucuronate isomerase
26	pj039-26	Methylmalonate-semialdehyde dehydrogenase
27	pj039-27	Cellulose synthase catalytic subunit
28	pj039-28	Cellulose synthase, putative
29	pj039-29	Putative cytoplasmic protein
30	pj039-30	hypothetical protein
31	pj039-31	hypothetical protein
32	pj039-32	Preprotein translocase subunit SecE
33	pj039-33	Transcription antitermination protein NusG
34	pj039-34	Cytochrome c heme lyase subunit CcmL / CcmH
35	pj039-35	hypothetical protein
36	pj039-36	hypothetical protein
37	pj039-37	hypothetical protein
38	pj039-38	hypothetical protein
39	pj039-39	Colicin V production protein
40	pj039-40	hypothetical protein
41	pj039-41	hypothetical protein
42	pj039-42	hypothetical protein
43	pj039-43	hypothetical protein
44	pj039-44	hypothetical protein
45	pj039-45	hypothetical protein
46	pj039-46	hypothetical protein
47	pj039-47	Polymyxin resistance protein AmC
48	pj039-48	hypothetical protein
49	pj039-49	hypothetical protein
50	pj039-50	hypothetical protein
51	pj039-51	putative cytoplasmic protein
52	pj039-52	Type III secretion protein SsaB
53	pj039-53	hypothetical protein
54	pj039-54	hypothetical protein
55	pj039-55	hypothetical protein
56	pj039-56	Mobile element protein

表3. 食中毒事例で分離された *C.lari* 株のゲノム配列データのアセンブル後の概要

	カウント数	平均長(bp)	全塩基数(bp)
リード数	2,311,897	277.25	641,005,075
Contigs数	38	42,564	1,617,445 (396 coverages)

表4. 食中毒事例で分離された *C.lari* 株のゲノム配列から推定した主要な病原因子の保有状況

	<i>C.lari</i> Toyama 07	<i>C.lari</i> RM2100	<i>C. jejuni</i> All strains
定着因子	CadF	+	+
	JipA	-	+
	MOMP	+	+
	PEBA	-	+
毒素	CdtABC	+	+
	CiaB	+	+

表5. 腸管出血性大腸菌 O157 堺株のゲノム解析結果の概要

SK01		Summary statistics	Count	average length	Total bases
Contig measurements (including scaffolded regions)		Reads	802,464	296	237,427,329
N75	52,302	Matched	788,226	296	233,224,139
N50	124,190	Not Matched	14,238	295	4,203,190
N25	235,262	Contigs	203	26,429	5,365,229
Minimum	500	Reads in pairs	660,940	487	
Maximum	314,054	broken paired reads	127,286	272	
average	26,430				<b>Coverage=43.5</b>
Count	203				
SK02		Summary statistics	Count	average length	Total bases
Contig measurements (including scaffolded regions)		Reads	941,258	296	278,710,760
N75	84,323	Matched	927,039	296	274,484,347
N50	146,270	Not Matched	14,219	297	4,226,413
N25	224,569	Contigs	183	29,450	5,389,464
Minimum	512	Reads in pairs	786,502	490	
Maximum	374,972	broken paired reads	140,537	271	
average	29,451				<b>Coverage=50.9</b>
Count	183				
SK12 (=SK01+SK02)		Summary statistics	Count	average length	Total bases
Contig measurements (including scaffolded regions)		Reads	1,743,722	296	516,138,089
N75	73,375	Matched	1,716,252	296	507,995,197
N50	148,380	Not Matched	27,470	296	8,142,892
N25	327,284	Contigs	182	29,670	5,399,993
Minimum	537	Reads in pairs	1,442,975	488	
Maximum	374,974	broken paired reads	273,278	272	
average	29,670				<b>Coverage=94.1</b>
Count	182				

表6 O157 堺株の配列の評価

	EHEC O157 Sakai from Public site				EHEC O157 Sakai by MiSeq		
	Chromosome	pO157	pOSAK1	Total	SK01	SK02	SK12
Length of sequence (bp)	5,498,450	92,721	3,306	<b>5,594,477</b>	<b>5,365,229</b>	<b>5,389,464</b>	<b>5,399,993</b>
Contigs	1	1	1	<b>3</b>	<b>203</b>	<b>183</b>	<b>182</b>
ORF (=CDS, >150)	5,361	83	3	<b>5,447</b>			
RNA	110	0	0	<b>110</b>			
<small>Hayashi, T., et al. <i>DNA Res.</i>, 8:11(2001)</small>							
after Re-annotation by RAST				after annotations by RAST			
ORF (=CDS, >150)	5,208	99	3	<b>5,310</b>	<b>5,091</b>	<b>5,190</b>	<b>5,138</b>
RNA	124	0	0	<b>124</b>	<b>107</b>	<b>137</b>	<b>107</b>



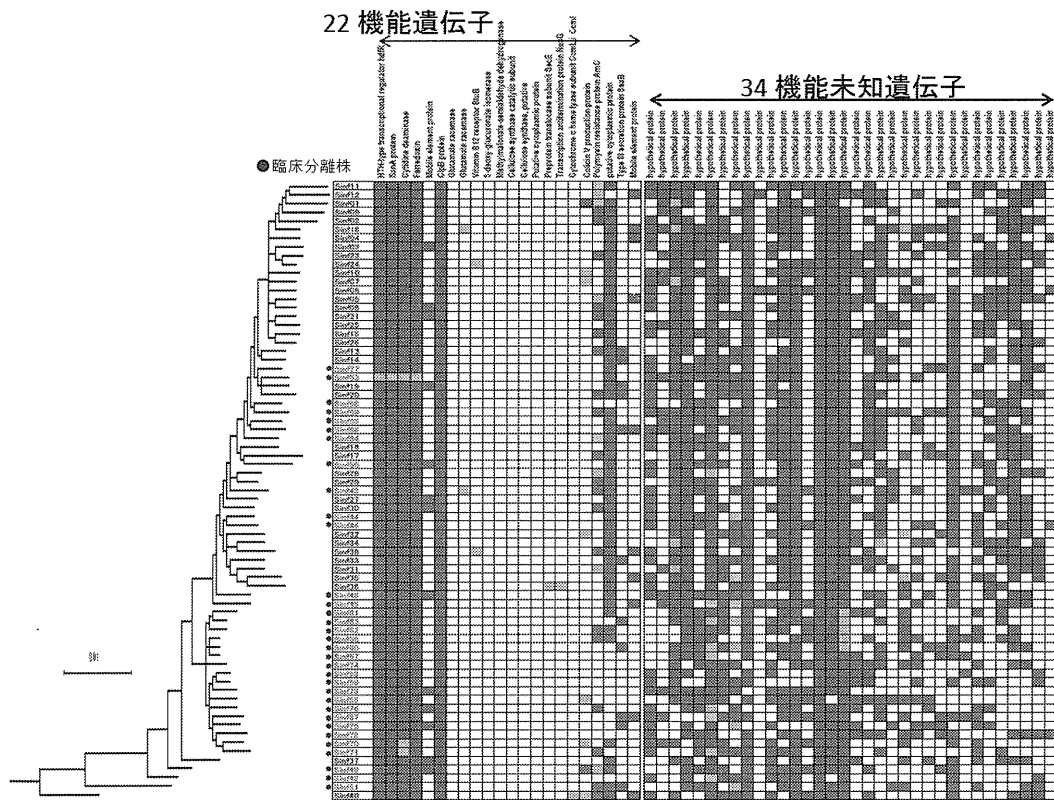


図3. 臨床分離株特異的と推定された56遺伝子の確認

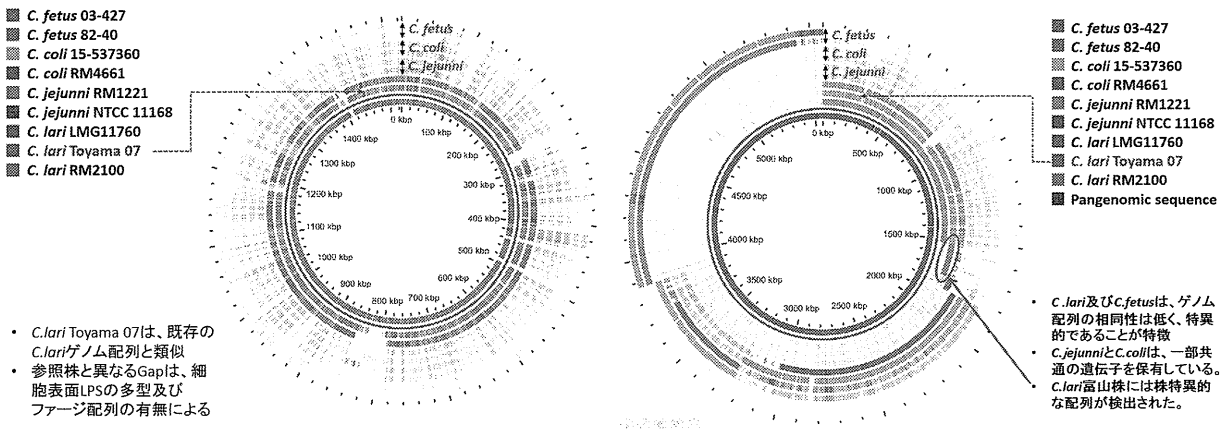


図4. ゲノム解析した *C.lari* 分離株の Blast atlas 結果

図5. ゲノム解析した *C.lari* 分離株の Pangenome blast atlas 結果

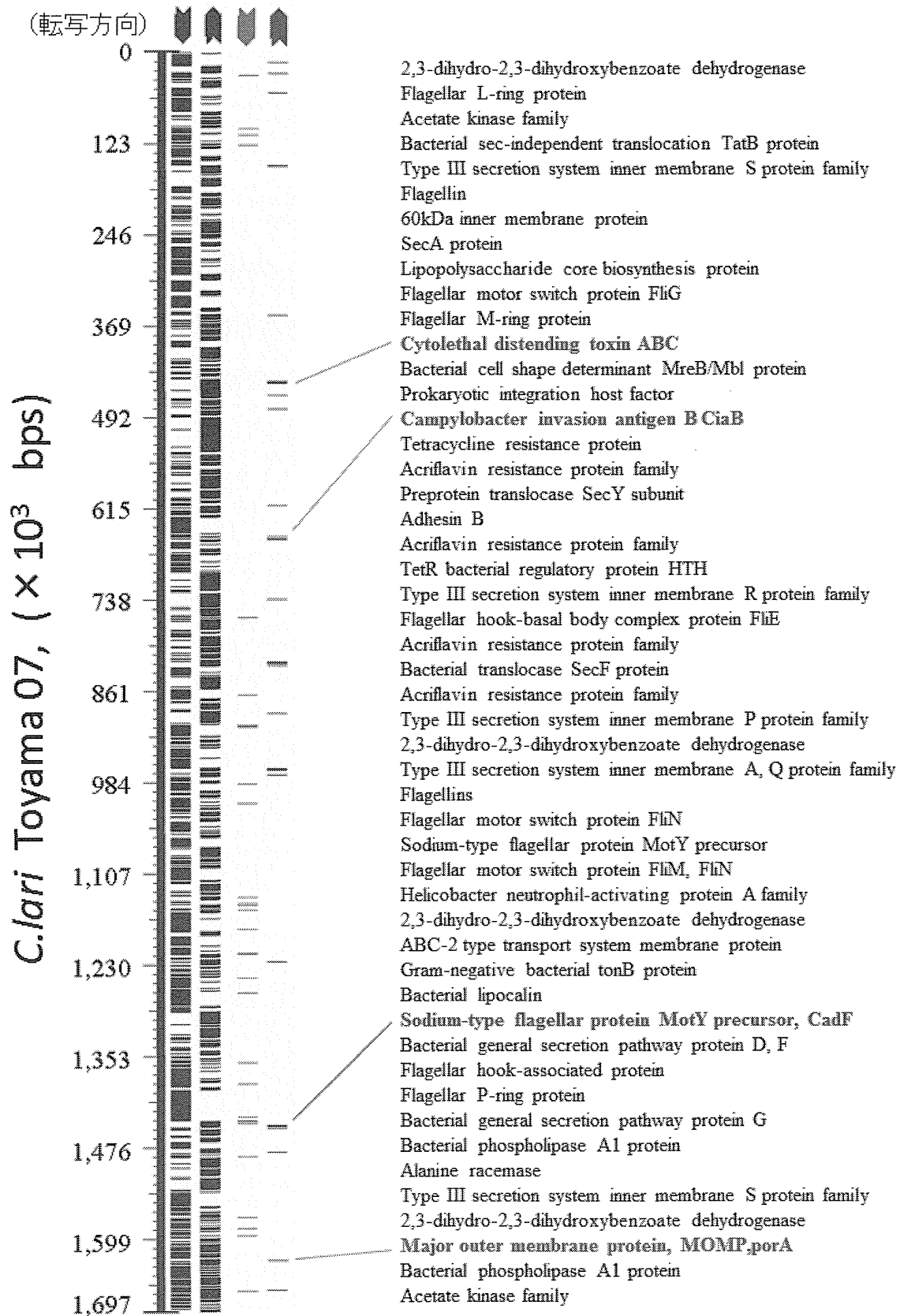
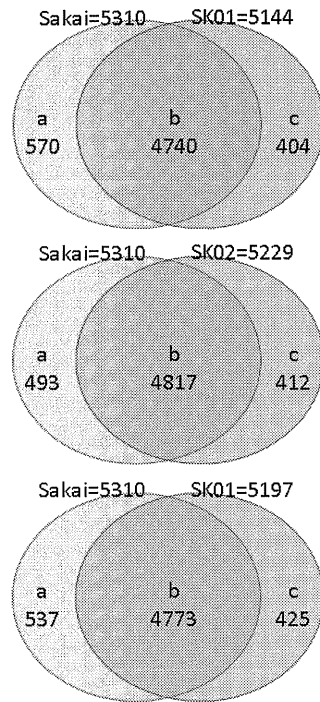


図6. *C.lari* Toyama07 の病原因子の検索結果

左図は、*C.lari* Toyama07 のゲノム上の遺伝子のマッピングしたもので、Virulence Searcher (Web サービス)を利用して検出された53遺伝子のゲノム上の位置

CDS, >150bp



a, Sakaiのみ検出  
b, SakaiとSKに共通  
c, SKのみ検出

Comparison of RAST Annotation Data between Public Sequence and MiSeq Draft Data

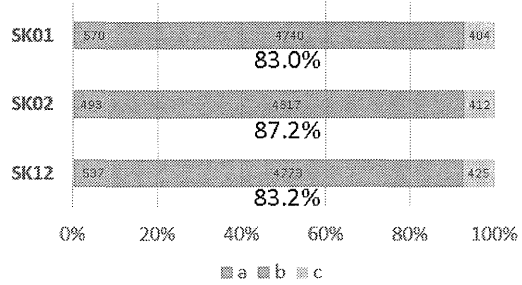


図7. 腸管出血性大腸菌 O157 堺株を用いた MiSeq データのドラフト配列の定量的な理解  
ゲノム配列既知の O157 堺株を MiSeq でゲノム配列を得て、報告されている配列と比較して、  
CDS の一致度を評価した。

## I. 参考資料 1. WindowsPC環境における有用なNGS解析のためのソフト、アプリケーションのリスト

用途	ソフト名	説明	PC	利用形態	必要条件等	コスト	URL	コメント
NGS用総合解析	CLC genomic workbench	NGSデータを様々な解析が可能。De novo assemble、各種マッピング等可能。また、ゲノム解析にしばしば利用されるLinux系のソフトで利用できるファイル形式を読み込んだり、書き出しなど可能である。このソフトがあれば、ほぼ、なんとかなる。但し、高額である。	Windows	独立		有料	フィルジェン株式会社 <a href="http://www.filgen.jp/">http://www.filgen.jp/</a>	公的データベースや、論文等でも引用されている。
NGS	DDBJ Read Annotation Pipeline	Web環境が必要だが、de novo assemble等が可能。	Windows	ウェブ		無料	<a href="https://p.ddbj.nig.ac.jp/pipeline/Login.do">https://p.ddbj.nig.ac.jp/pipeline/Login.do</a>	登録必要/登録しているが、使用経験なし。
自動アノテーション	RAST	アップロードされたゲノム配列について、遺伝子データベースに対して、blast解析を実施し、蛋白質をコードする領域等を推定し、注釈するサービス。解析結果は、様々な様式で入手可能で、細菌のゲノム解析では有用な解析サービスである。	Windows	ウェブ	firefox	無料	<a href="http://rast.nmpdr.org/">http://rast.nmpdr.org/</a>	登録必要/スタートゴドンの選択が時々、異なっており、注意が必要であるが、様々な様式のファイルがダウンロードできるので、非常に有用なサービス。
自動アノテーション	MiGAP	DDBJ japanが提供している細菌ゲノム配列の注釈サービス。RASTと同じような目的で利用されるWebサービスである。	Windows	ウェブ		無料	<a href="http://www.migap.org/">http://www.migap.org/</a>	登録必要/使用頻度は高くありません。
比較ゲノム解析結果の比較	GviewServer	細菌ゲノムの比較解析を実施してくれるサービス。複数のゲノムを環状、線状形式で、ゲノム上のすべての遺伝子を対象として、視覚化できる。	Windows	ウェブ	firefox	無料	<a href="https://server.gview.ca/">https://server.gview.ca/</a>	登録必要/使用頻度は高い。
ゲノムの可視化	Gview	genebank形式のファイルを見やすい図として表示できる。環状、線状の図として表示し、CDSの色分け等、カスタマイズできる。また、Genbank形式やgffファイルを読み込ませ、より複雑な図として、カスタマイズできる。GVewServerで得られたblast atlasの結果を表示する場合も使われるソフトである。	Windows	独立	Java	無料	<a href="https://www.gview.ca/wiki/GView/Web_Home">https://www.gview.ca/wiki/GView/Web_Home</a>	使用頻度は高い。
手動アノテーション	Artemis	塩基配列の特徴を見やすく表示してくれる。また、遺伝子のアノテーションを行うことができる。EMBL及びGENBANK形式の配列データを読み込むことができる。	Windows	独立	Java	無料	<a href="http://www.sanger.ac.uk/science/tools/artemis">http://www.sanger.ac.uk/science/tools/artemis</a>	使用頻度は高い。
比較ゲノム解析	ACT	Artemisをベースにゲノムワイドな配列を比較し、その結果を表示する。その結果から、大きな領域での逆位、欠失、挿入などを視覚化してくれる。	Windows	独立	Java	無料	<a href="http://www.sanger.ac.uk/science/tools/artemis-comparison-tool-act">http://www.sanger.ac.uk/science/tools/artemis-comparison-tool-act</a>	大きなGAPや逆位等が視覚化できる。
ABIシーケンサーデータのAssemble	Sequencher	ABIシーケンサーで解析した配列の確認、Assembleソフト	Windows	独立		有料	タカラバイオ	MiSeqデータではなく、ABIシーケンサーからのデータの解析のため、使用頻度は高い。



ABIシーケンサーデータのAssemble	GeneStudio	ABIシーケンサーで解析した配列の確認、Assembleソフト。波形を確認して編集可能。系統樹解析も可能。	Windows	独立		無料	<a href="http://genestudio.com/download">http://genestudio.com/download</a>	複数のPCで使用できるので利用価値はあるが、それほど使用頻度は高くない。
相同性解析	NCBI blast+	ペアワイズな塩基配列の局所的なアライメント／相同性検索(ホモロジー検索)を高速に行うプログラム	Windows	独立		無料	<a href="http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&amp;DOC_TYPE=Download">http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&amp;DOC_TYPE=Download</a>	Windows10で使うと非常に使いやすい。使用頻度高い。2つのゲノムを比較するとき、ACTを使用するときに必要なcomparison fileを作成するblasstnを実施できる。 ゲノム解析した配列のアノテーション作業には必須の環境を提供する。
DNA配列のアライメント及び遺伝子の系統解析	MEGA	DNA配列あるいは蛋白のアミノ酸配列の並列化ソフト。配列の編集及び配列の系統分類。	Windows	独立		無料	<a href="http://megasoftware.net/">http://megasoftware.net/</a>	遺伝子の並列化、系統解析等、使用頻度は高い。
細菌ゲノムから病原性解析	The virulence factor database (VFDB)	病原細菌の病原性因子のデータベースが提供されている。	Windows	ウェブ		ウェブ	<a href="http://www.hpa-bioinfotools.org.uk/pise/">http://www.hpa-bioinfotools.org.uk/pise/</a>	各種病原性細菌の病原性遺伝子が確認できる
細菌ゲノムから病原性解析	Virulence Searcher	NGSから得られたde novo assemblyデータから、病原性遺伝子の検索	Windows	ウェブ		ウェブ	<a href="http://www.hpa-bioinfotools.org.uk/pise/virfactfind_small.html">http://www.hpa-bioinfotools.org.uk/pise/virfactfind_small.html</a>	病原性遺伝子の検索

遺伝子解析ユーティリティ	Sequence Assistant	非常に軽い遺伝子解析ソフト。Genbankから、ダウンロードしたファイルから、塩基配列だけ抽出する、相補鎖変換等も可能。	Windows	独立		有料	<a href="http://www2s.biglobe.ne.jp/~haruta/">http://www2s.biglobe.ne.jp/~haruta/</a> <a href="http://www.vector.co.jp/soft/">http://www.vector.co.jp/soft/</a> から入手可能	使用頻度は高い。塩基配列の文字列の抽出、結合、相補鎖変換等が簡単に行える。細菌ゲノムサイズの取扱いも可能。Artemisによる手動アノテーションと一緒に使うと便利。
遺伝子解析ユーティリティ	Format Assistant	遺伝子解析ユーティリティ。他の文字が含まれる文字列から塩基、アミノ酸を表す文字のみを抽出することが出来る。	Windows	独立		無料	<a href="http://www2s.biglobe.ne.jp/~haruta/">http://www2s.biglobe.ne.jp/~haruta/</a> <a href="http://www.vector.co.jp/soft/">http://www.vector.co.jp/soft/</a> から入手可能	塩基配列、アミノ酸配列文字列のの編集に利用。あまり長鎖のDNA配列は利用できないようである(60,00塩基程度が限界?)
遺伝子解析ユーティリティ	秀丸	テキストエディタ	Windows	独立		有料	<a href="http://hide.maruo.co.jp/software/hidemaru.html">http://hide.maruo.co.jp/software/hidemaru.html</a>	ゲノム解析に用いられるファイル形式は、基本的にテキストベースであり、テキストを直接、加工できるテキストエディタは便利である。特に本ソフトは、1億行まで扱えるため、ある程度の巨大ファイルの編集が
遺伝子解析ユーティリティ	7-zip	解凍ソフト	Windows	独立		無料	<a href="https://sevenzip.osdn.jp/">https://sevenzip.osdn.jp/</a> (日本語サイト)	巨大な圧縮ファイルの解凍が出来る。
遺伝子解析ユーティリティ	GENOMICS	ゲノム解析法の紹介	Windows	独立		無料	<a href="http://molbiol-tools.ca/Genomics.htm">http://molbiol-tools.ca/Genomics.htm</a>	ゲノム解析手法のリストが紹介されている。

## GView Server の使用方法と注意事項

用途：微生物ゲノムを比較解析し、可視化する。環状形式と線状形式、遺伝子の転写の方向、等、カスタマイズ可能である。既に複数の完全ゲノム配列が報告されている細菌種が利用できる状況であり、NGS で得られた細菌の *de novo assemble* した *contigs* を、完全ゲノム配列を参照配列 (Ref. seq) として、遺伝子単位で比較 (*blastn* or *blastp*) する。

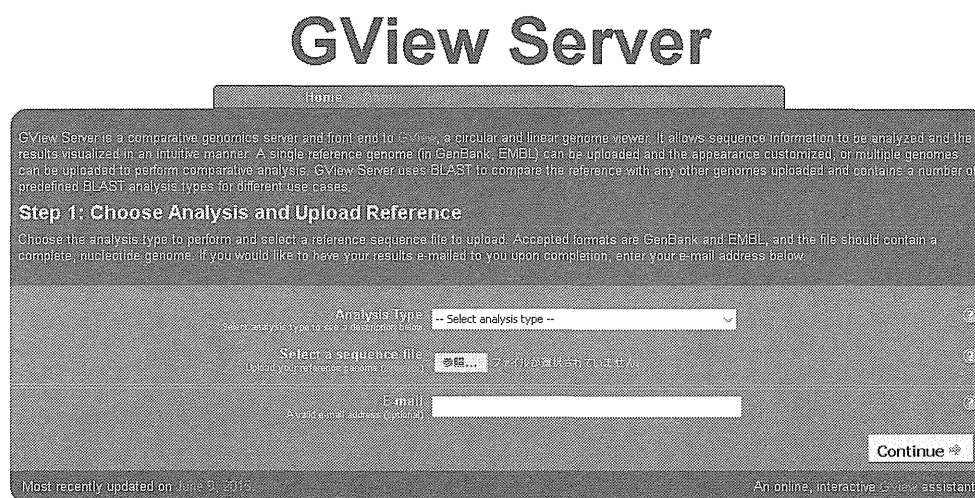
以下の解析が可能である。

1. BLAST Atlas
2. Pangenome Analysis
3. Core Analysis
4. Accessory Analysis
5. Unique Analysis
6. Signature Analysis
7. Reciprocal Analysis
8. Pangenome Phage Analysis

準備：以下の環境を確認

1. OS: Windows7 以上、Windows10 を推奨。インターネット環境必須。
2. ブラウザ：FireFox
3. <https://server.gview.ca/> (以下、トップページ)

**Register** タブから、氏名とメールアドレスを入力し、登録する。



使用方法：

1. Ref seq は、完全配列が報告されている株を用いる。

2. Refseq は、注釈情報が付いた gbk ファイルを用いると後で便利。付いていない場合には、RAST 等を利用して注釈付きの gbk ファイルを作成するとよい。
3. Query は、gbk でも fasta 形式でもよい。
4. gbk file は、RAST で autoannotation を行い、merged タイプを使うと、artemis で読み込んでくれるので、後で遺伝子構成を見たり、遺伝子やアミノ酸配列を抽出するのが楽となる。

GviewServer への file の upload は、以下の図中に示す、4 ステップからなり、条件を入力していく。以下は、Pangenomic Analysis したときの例である。Final Step の入力が終わったら、念のため、Job ID はコピーしておき、log out する。解析が終わると登録したメールアドレスに、連絡がある。そのときの解析の込み具合にもよるが、翌日の連絡が多い。

右図は、記録のため、FileMaker を使い、解析条件等を記録する画面である。Web 上で、ファイルや条件を入力する順番となっている。この例は、Pangenome analysis を選択したときの例示である。その他の解析をする場合には、この図では、解析タイプを選択できるようになっている。

<b>New</b>	Date	2015/11/03
Step 1	Analysis Type	Pangenome Analysis
	Reference genome	Sinf27-merged.gbk
Step 2	Upload Query Files	
	Blast Program & File features	Sinf27 - Sinf_Meat1_18core-pj033.gbk CDS-CDS
Step 3	Blast Parameters	
	Filter Low complexity sequences?	<input type="checkbox"/> on <input checked="" type="checkbox"/> off
	Expect value cutoff	1e-10
	Genetic code	Bacterial and Plant Plastid
	Alignment length cutoff	100
	Percent identity cutoff	80
	Final Step Map Title	SInfantis2015 Pj035
	Description	Meat1 core seeking Pangenome 80%
	Display reference feature	<input checked="" type="checkbox"/> CDS <input type="checkbox"/> -10_signal <input type="checkbox"/> misc_RNA <input type="checkbox"/> protein_bind <input type="checkbox"/> rRNA <input type="checkbox"/> -35_signal <input type="checkbox"/> gene <input type="checkbox"/> repeat_region <input type="checkbox"/> RBS <input type="checkbox"/> tRNA
	Show GC content	<input type="checkbox"/> on <input checked="" type="checkbox"/> off
	Show GC skew	<input type="checkbox"/> on <input checked="" type="checkbox"/> off
	Show COG categories	<input type="checkbox"/> on <input checked="" type="checkbox"/> off
	Show legends	<input type="checkbox"/> on <input checked="" type="checkbox"/> off
	Select a font	Arial
	Options	one strand line
	Backbone	large
	JobID	D96519393986949C7F47FEC56385728

解析が終わったという連絡が来たら、Web からログインして、以下のような画面まで行く。

結果の viewing は、2 通りの方法がある。

WebStart と結果ファイルを download して、中にあるバッチファイルから、表示する方法である。いずれも JAVA が予めインストールされていることが必要。

# GView Server

**GView Results for Job D965193939986949C7F47FEC56385728**

View



Launch GView Webstart  
Circular



Launch GView Webstart  
Linear

The above images are previews of the GView Server analysis. Clicking on the image will display a full size version.

Clicking **Launch GView Webstart** will open the results in GView, an interactive circular and linear genome viewer, where you can pan, zoom, and interact with the results. GView requires Java 1.5 or greater and at least 512MB of memory.

The **Pangenome** analysis type displays a pangenome BLAST Atlas of a set of genomes. The reference genome consists of a pangenome constructed by concatenating all the unique regions among each of the uploaded query genomes. A BLAST atlas is then constructed for each query genome. The features in these slots show regions where there is a BLAST hit between the query genome and the constructed pangenome. An example of how to interpret a Pangenome BLAST Atlas can be found on the examples page at [Pangenome of the E. coli](#).

**Download**

The below table lists all downloadable files containing the resulting data of this analysis. There are two types of data that can be downloaded: the GView zip package, and the BLAST results table. The zip package contains all the necessary information to launch GView offline. The BLAST Results table contains the results of the BLAST analysis that were performed on the uploaded data in a tabular format. To launch GView offline, please execute the appropriate QuickLaunch program, either `QuickLaunch.bat` or `QuickLaunch.sh`.

File	Type	Description
<a href="#">gviewall-D965193939986949C7F47F47FEC56385728.zip</a>	GView Data + Executable	Contains the resulting data plus the GView executable.
<a href="#">gview-D965193939986949C7F47F47FEC56385728.zip</a>	GView Data	Contains only the resulting data from the job, not the gview executable.
<a href="#">blastresults-D965193939986949C7F47F47FEC56385728.csv</a>	BLAST Results Table	A table of the resulting BLAST hits, in csv format. The results may take a while to download for large data sets.
<a href="#">blastresults-D965193939986949C7F47F47FEC56385728.xls</a>	BLAST Results Table	A table of the resulting BLAST hits, in excel format. The results may take a while to download for large data sets.

Logged in as: [M4W4ts](#) Most recently updated on: June 3, 2019. An online, interactive GView assistant.

- ①JAVAstartのスイッチがリンクされている。ここをクリックすると、うまくするとJAVAstartが開始され、結果のファイルが読み込まれて、結果が図で表示される。しかし、PCにより、エラーになる場合が多く、使いにくい。うまく、表示されても、色を変更したりしたい場合には、結果のファイルを一度、PCにdownload (DL)して、Viewing softであるGviewで表示した方が使いやすいので、ほとんどは、②で行っている。
- ②のJobIDをダブルクリックして、DownlordからGview Data+ Excutable typeのfile (ZIP)をPCにDLする。DLしたファイルは圧縮されているので、解凍する。

解凍すると解凍ホルダ内には以下のようなファイル構成となっている。

JAVA上で動く、gviewのソフトが入っているので、インストールする必要はない。勿論、インストールしても利用できる。

名前	更新日時	種類	サイズ
task_1	2015/11/03 18:17	ファイル フォルダ	
gview.jar	2014/01/03 13:09	Executable Jar File	6,746 KB
QuickLaunch.bat	2015/11/03 3:16	Windows バッチ ファ...	1 KB
QuickLaunch.sh	2015/11/03 3:16	SH ファイル	1 KB
results.gff	2015/11/03 0:50	GFF ファイル	1,598 KB
style.gss	2015/11/03 0:50	GSS ファイル	4 KB

Task\_1ホルダの中には、gbkファイルが入っている。

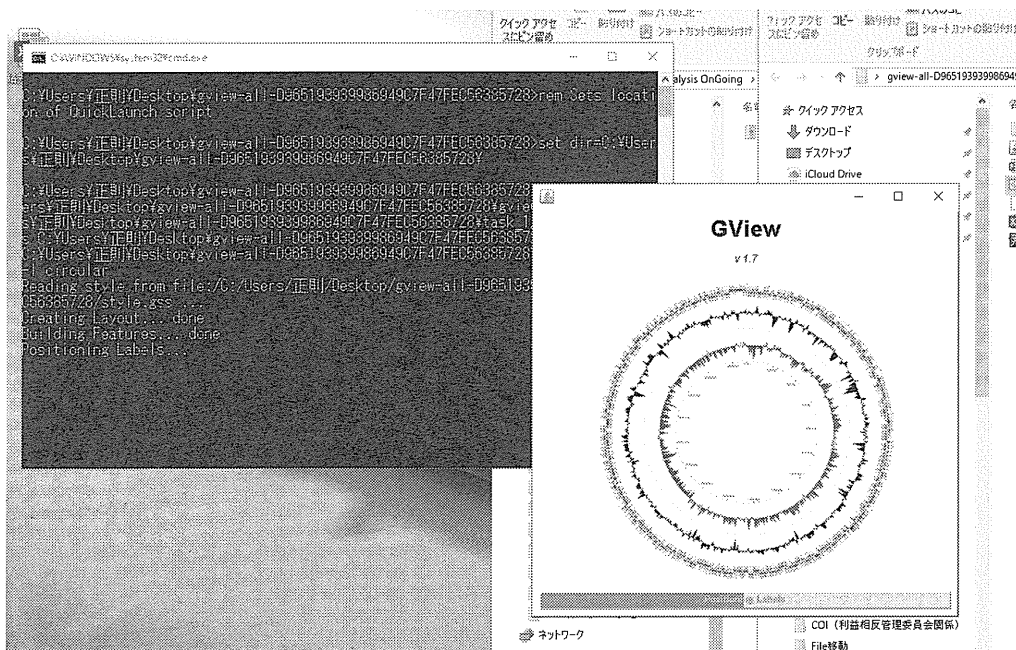
クリックランチャーのバッチファイルが装備されており、まずは、試しにダブルクリックしてみる。運が良ければ、プログラムが稼働し、blast atlasの画面が表示される。

名前	更新日時	種類	サイズ
panggenome.gbk	2015/11/03 0:47	GBK ファイル	8,947 KB

クイックランチャーのバッチファイルをダブルクリックすると、一瞬、何かウインドウが現れるが、直ぐに消える現象が現れる。これが現れるとうまく、表示してくれない。

これは、インストールされている JAVA アプリの状態に依存するものと考えられ、Windows 環境でこれを設定することは、素人には難しい。これは、Linux でもそうであるが、作業ホルダのディレクトリを理解する必要があり、Windows で慣れている人間にはなかなか理解しにくいところ。おそらく、Gview Server から DL したクイックランチャーのバッチファイルには、相対パスではなく、絶対パスで記述されており、これが WindowsPC 上で合致しないとプログラムが動作しないと考えられた。そこで、動作するディレクトリーを探したところ、デスクトップでうまく行くことが分かった。

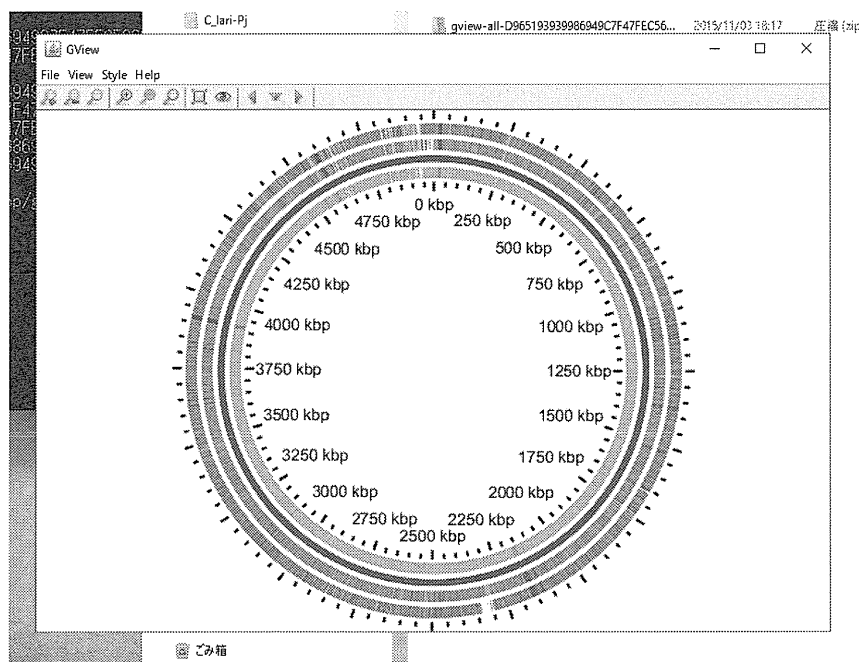
解凍したホルダをデスクトップ画面にドラッグして、クイックランチャーバッチファイルをダブルクリックすると、以下のような画面が現れて、動作が始まる。もし、ここでうまく表示されない場合には、C:ドライブの直下等に解凍したフォルダをコピーして、試してみると開く場合がある。このことは、使用する PC 環境、固有の条件となっている可能性があるため、GViewServer が使用できるディレクトリーを決定しておくことが重要である。



以下のように、うまく表示されるはずですが。

ここまで来れば、後は比較的簡単に、カスタマイズできるようになる。詳細は GView の help を参照のこと。

gbk ファイルを使っているので、カーソルを重ねると遺伝子名が表示される、拡大が出来る、blast 検索のカットオフ値が、ファイルの upload 時に変更できるなど、便利にゲノム全体を比較することが出来る。



(Ver1.1)

## Windows PC を用いたスタンドアロン BLAST のセットアップと Blast 検索

はじめに

NCBI では、BLAST 検索サービスを Web 上のサービスとして提供しているだけでなく、Windows PC を用いて、Web を介さない BLAST 検索ができる環境も提供している。この検索は、DOS コマンドウインドウから、コマンドラインで操作するもので、グラフィカルインターフェイスではない。

これに使用するプログラムはblast+と呼ばれ、NCBI C++ based BLAST programs をWindows PCにインストールして使用する。これまでのWindows バージョンでは、Windows PCのDOSコマンドの入力には、コピー&ペーストは利用できなかったのが、作業領域として、pathを理解する必要があることと正確にコマンドスクリプトをタイプする必要があり、非常に使いにくかったが、最新のバージョンであるWindows10から、コピー&ペーストが出来るようになり、非常に簡単にコマンドが使えるようになった。これまで、blast検索は、blastnやblastpなどを中心として、DDBJやNCBIのWeb service を用いた利用がほとんどであったが、Windows PC単独で、NCBI blast+ packageが使えるようになれば、blast+で提供されているすべてのプログラムが利用できるようになり、その解析の幅が大きく広がることが期待される。また、NGSから得られたリードやContigsを利用した検索に利用できるため、Windows PCを用いたNGSデータの解析の幅が広がることを期待できる。

### 1. ソフトの入手と準備

Blast Packages:

[http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE\\_TYPE=BlastDocs&DOC\\_TYPE=Download](http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&DOC_TYPE=Download)

BLAST+ executables の Installers and source code are available from

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

から、最新版の

ncbi-blast-2.2.30+-win64.exe (64-bit PC 用)をダウンロードする。

(バージョンアップされていれば、そのバージョンとして明記されたファイル名となっています)

今回インストールした PC 環境

Windows 10 Home, Intel Core i3-4160 CPU 3.60 GHz

16.0 GB RAM, 64-bit operating system

インストール

ncbi-blast-2.2.30+-win64.exe をダブルクリックすると自動的にインストールが始まり、終了する。これでWindows 環境下、提供されている blast プログラムがすべて利用できるようになる。

作業フォルダの設定

C:\blast\db を作成する。今後、この db フォルダを使って、blast 検索を実施する。

## 2. プログラム一覧 (一部の説明文は、ウィキペディアから抜粋)

1	blastdbcheck	BLAST データベースの完全性を確認する。
2	blastdbcmd	BLAST データベースから塩基配列やその他の情報を検索する。
3	blastdb_aliastool	
4	blastn	塩基配列 (ヌクレオチドのシーケンス) をクエリシーケンスとして、ユーザが指定した塩基配列データベースから類似するシーケンスを検索する。
5	blastp	アミノ酸配列 (タンパク質のシーケンス) をクエリとして、ユーザが指定したアミノ酸配列データベースから類似するシーケンスを検索する。
6	blastx	塩基配列のクエリシーケンスを、6 フレームで翻訳して、アミノ酸配列データベースから類似するシーケンスを検索する。
7	blast_formatter	
8	convert2blastmask	
9	deltablast	
10	dustmasker	
11	legacy_blast.pl	
12	makeblastdb	blast 検索をするために、FASTA ファイルを BLAST データベース (db) に変換する。
13	makembindex	
14	makeprofiledb	
15	psiblast	プロファイルを利用したホモロジー検索を行う。 <b>blastp</b> よりも遠縁種の検索が得意。アミノ酸配列をクエリとして、ユーザが指定したアミノ酸配列データベースから類似するシーケンスを検索する。最近開発された <b>BLAST</b> プログラムの一つ。このプログラムはタンパク質の遠い系統的関連性を発見するために使われる。まず、クエリのアミノ酸配列と近い関連性をもつタンパク質の一覧を作成する。そしてこの一群のタンパク質を統合して「プロファイル」(平均化したシーケンスのようなもの) を作る。このプロファイルを使ってアミノ酸配列データベースに対する検索を行い、プロファイルのもととなったアミノ酸配列より多い件数のアミノ酸配列を見つける。見つかったアミノ酸配列でさらにプロファイルを作成する。この一連の処理を繰り返し行う。検索を行う際に関連するタンパク質を考慮するため、遠い系統的関連性を発見する場合に、 <b>PSI-BLAST</b> は <b>blastp</b> よりもはるかに正確な処理を行う。
16	rpsblast	blast のアルゴリズムを使ったプロファイル検索プログラム。
17	rpsblastn	
18	segmasker	
19	tblastn	アミノ酸配列をクエリとして、塩基配列データベースを 6 フレームで翻訳しながら、類似するシーケンスを検索する。
20	tblastx	塩基配列のクエリシーケンスを、6 フレームで翻訳して、アミノ酸



		配列に翻訳された塩基配列データベースから類似するシーケンスを検索する。このプログラムは BLAST のプログラムの中で最も処理速度が遅い。このプログラムの使用目的は、塩基配列のシーケンス間の非常に遠い関連性を発見することである。
21	update_blastdb.pl	
22	windowmasker	

空欄は、ほとんど使うことはないが、今後、利用法について研究したい。また NCBI の BLAST データベースは日々、アップデートされているらしい。

#### Databases

BLAST databases are updated daily and may be downloaded via FTP from <ftp://ftp.ncbi.nlm.nih.gov/blast/db/>. Database sets may be retrieved automatically with `update_blastdb.pl`, which is part of the BLAST+ suite. Please refer to the [BLAST database documentation](#) for more details.

### 3. 公的データベースのダウンロード(公的データベースの利用が前提の場合)

1. 次のリンクをクリックして NCBI の ftp サーバーにアクセスする。  
<ftp://ftp.ncbi.nih.gov/blast/db/>
2. ダウンロードしたいデータベースファイル (たとえば 16SMicrobial.tar.gz) をダブルクリックする。
3. ファイルを C:\blast\db に保存する。
4. ダウンロードしたファイルを展開する。 解凍ソフトとして 7-zip が使いやすい (無料)。

解凍したファイルは、

16SMicrobial.nsq  
16SMicrobial.nsi  
16SMicrobial.nsd  
16SMicrobial.nog  
16SMicrobial.nni  
16SMicrobial.nnd  
16SMicrobial.nin

16SMicrobial.nhr の 8 つファイルが存在する。blastn を実施するとき、この 8 つの一つのセットとして、そのまま使用可能である。

### 4. In-house データベースの作成と blast サーチ

#### 準備

1. blast 検索に使用したいデータベースは、FASTA ファイル形式として作成する。このままでは、blast のデータベースとしては使用できないので、データベースとして使用できるように様式を変換する必要がある。
2. コマンドプロンプトを開く。(「管理者用のコマンドプロンプト」も存在するが、これは使用しない)
3. [cd 半角スペース]入力し、C:\blast ウィンドウ内のフォルダ db をコマンドプロンプト・ウィンドウにマウスを使って、ドラッグする。Enter
4. C:\blast\db>  
表示になる。
5. 点滅カーソル状態で、[dir] Enter
6. XXXXX.fas が表示されていることを確認する。
7. この後、query 配列とデータベースは全て、このディレクトリを使用する。また、blast の結果は、全てこのディレクトリに作成される。

## データベースの作成

1. DNA 配列かアミノ酸配列の FASTA ファイルとして作成する。ファイル名は、DNA の場合には、XXXXXnuc.fas、蛋白質の場合には、XXXXXaa.fas などと区別するように命名するとよい。ここで、作成した FASTA ファイル、例えば、XXXXXnuc.fas は、C:\blast\db フォルダにコピーする。

次にこの“XXXXXnuc.fas”という名前の fasta ファイルをデータベースとして利用する様式に変換する。

2. “makeblastdb -in XXXXXnuc.fas -out XXXXXnucDB -dbtype nucl”と入力し、**Enter**を押す。
3. 「XXXXXnucDB」は、データベース名である。「-dbtype nucl」は、入力必須項目である。因みに蛋白質の場合は、「-dbtype prot」と入力する。
4. “db” ディレクトリに以下の3つのファイルが作成されたことを確認する。  
XXXXXnucDB.nhr  
XXXXXnucDB.nin  
XXXXXnucDB.nsq

5. blastn

query に” dna-query.fas” ファイルを使用する場合、

“C:\blast\db>”の後に、以下のように入力する。

```
blastn -query dna-query.fas -db XXXXXDB -out dna-query.out
```

と入力して、**Enter**を押すところであるが、この入力に、コピー&ペーストを利用する。

C:\blast\db>blastn -query まで入力して、db フォルダ（のアクティブウインドウ）から、「dna-query.fas」をドラッグすると、以下ようになる。

```
C:\blast\db>blastn -query C:\blast\db\dna-query.fas
```

、そこで続けて

```
C:\blast\db>blastn -query C:\blast\db\dna-query.fas -db
```

と入力し、db フォルダから、データベースとして、作成した上記3つの「XXXXXDB」のうち、いずれか一つをドラッグする（今回は、「XXXXXnucDB.nhr」を用いる）と、

```
C:\blast\db>blastn -query C:\blast\db\dna-query.fas -db
```

```
C:\blast\db\XXXXXDB.nhr
```

となる。そこで、データベースファイルの3つに共通な部分である

「XXXXXuncDB」だけを残すため、「.nhr」の文字を `BackSpaceKey` で削除する。

そして、その後に結果の出力ファイル名として、「-out dna-query.out」と入力して、`Enter`を押す。

“db”内に、瞬時に”dna-query.out”が作成される。

このコマンドの意味は、クエリー (-query) に「dna-query.fas」という fasta 形式のファイルにある配列 (単独でもマルチでも可) を用いて、データベース (-db) として、[XXXXXDB] (上記の3つのファイルを指定するために、最初の共通の文字列のみを入力する) を使用して、その結果を (-out) [dna-query.out] というファイル名として、書き出す、ことである。

#### Comments & Options

① option switch を指示しなければ、デフォルトでは、500 行と 250 アラインメントの表示となる (従って、長大なファイルとなる)。これを変更する場合には、以下の通り、表示 option スイッチと表示数を加える

```
>blastn -query [dna-query.fas] -db [XXXXXDB] -out [dna-query.out] (-num_descriptions <500> -num_alignments <250>)
```

②表形式結果表示 ⇒ エクセルでの表形式で結果を示すことができるので、利用価値は高い。

```
>blastn -query [dna-query.fas] -db [XXXXXDB] -out [dna-query.out] -outfmt [String]
```

\*String=6: tabular (ACT で利用する Comparison file の作製に必要)

\*String=7: tabular with comment lines

\*String=10: Comma-separated values

その他、多数の formatting option が利用可能である。この説明は、[-help] で確認することができる。

(Version 1.2 2016.02.29)