

標準作業手順書

MLSA 法による *Acinetobacter* 属細菌の同定法

平成 28 年 3 月版 (Ver.0.9)

独立行政法人 製品評価技術基盤機構

バイオテクノロジーセンター

目次

序文	1
第 1 章 適用範囲	2
必要なソフトウェアについて	2
第 2 章 用語及び定義	3
第 3 章 解析手順の概要	4
第 4 章 解析に使用する菌株の選抜	6
第 5 章 指標遺伝子の取得	8
5.1 塩基配列取得法の選択	8
5.2 問い合わせ配列 (クエリーシーケンス) の入手	8
5.3 解析株のゲノム配列データからの指標遺伝子塩基配列の取得	10
5.3.1 NCBI web サイトでの遺伝子配列の取得	10
5.3.2 RAST Server で予測した解析株の全 ORF からの遺伝子配列の取得	15
5.3.3 BioEdit の機能を利用した解析株の遺伝子配列取得	20
5.4 解析株のゲノム配列データの用意	27
第 6 章 指標遺伝子の塩基配列のトリミング処理	31
6.1 <i>Acinetobacter</i> 属細菌基準株の指標遺伝子塩基配列の用意	31
6.2 マルチプルアラインメントの実行	32
6.3 Gblocks Server によるトリミング処理	36
第 7 章 6 つの指標遺伝子の塩基配列の連結	38
7.1 連結用指標遺伝子の塩基配列ファイルの用意	38
7.2 塩基配列ファイルの連結	40
第 8 章 分子系統解析	43
8.1 系統樹の作成	43
8.2 系統樹に基づく種同定	46

序文

Acinetobacter 属細菌は土壌、河川水、室内の湿潤環境やヒトの皮膚等の幅広い環境に生息する好気性のグラム陰性桿菌であり、これまでに約 41 の菌種が知られている（2016 年 3 月現在）。*Acinetobacter* 属細菌には、*Acinetobacter baumannii* や *Acinetobacter nosocomialis*、および *Acinetobacter pittii* のように、細菌感染症への抵抗力が低下した患者に対して日和見感染を起こし、様々な疾患の原因となる菌が存在する^{1,2}。特に、近年では抗生物質が効かない多剤耐性 *Acinetobacter* 属細菌による院内感染が世界的な問題となっている。そのため、日本では広域 β -ラクタム剤、アミノ配糖体、フルオロキノロンの 3 系統の薬剤すべてに耐性を示す *Acinetobacter* 属細菌による感染症（薬剤耐性アシネトバクター感染症）は、感染症の予防及び感染症の患者に対する医療に関する法律（感染症法）に基づく 5 類感染症（全数把握）の対象となっている。一方で、*Acinetobacter* 属細菌の中には 3,4-ジクロロアニリン分解菌³や原油分解菌^{4,5}などの環境汚染物質分解能を持つ菌の報告もあり、これら分解菌の環境浄化（バイオレメディエーション）への利用が期待されている。

経済産業省と環境省が共同で制定した「微生物によるバイオレメディエーション利用指針（以下、バイオレメディエーション指針）」および本指針の解説書によれば、外部で培養した微生物を導入して浄化事業を実施する場合には、利用微生物の同定を行い学名から病原性についての既存の情報を調査して利用微生物の安全性を確認することが求められている。従って、*Acinetobacter* 属細菌を利用する際には、種レベルでの同定を正確に行い病原菌種と非病原菌種を区別することが重要となる。

細菌の「種」の同定は 16S rRNA 遺伝子配列の類似度に基づいて行われることが一般的であるが、*Acinetobacter* 属細菌については各菌種間における 16S rRNA 遺伝子配列の変異が少なく、種レベルの同定が困難になることがある^{6,7}。このように 16S rRNA 遺伝子配列で明確に区別できない分類群の細菌については、複数のハウスキーピング遺伝子を用いた Multilocus sequence analysis (MLSA) 法による菌種の同定が推奨されている⁸。

我々は経済産業省受託事業「土壌汚染対策のための技術開発 VOC の微生物等を利用した環境汚染物質浄化技術開発」（平成 22 年度～平成 26 年度実施）においてバイオレメディエーションへの利用が期待される汚染物質分解能を持つ細菌を同属の病原菌と種レベルで識別するための MLSA 法の開発を実施してきた。本手順書は、本事業の一環で開発した *Acinetobacter* 属細菌の MLSA 法について、微生物を利用した産業の促進を図ることを目的に手順の公表を行うものである。

¹ Howard A et al. 2012, Virulence. 3: 243

² Visca P et al. 2011, IUBMB Life, 63: 1048

³ Hongsawata P et al. 2011, J Hazard Mater. 186: 1300

⁴ Sakai Y et al. 1994, Biosci. Biotechnol. Biochem. 58: 2128

⁵ Sugiura K et al. 1996, Environ. Sci. Technol. 31:45

⁶ Kämpfer P et al. 2011, Microbiology Australia. 32: 66

⁷ Chan JZ et al. 2012, BMC Microbiol. 12: 302

⁸ Stackebrandt E et al. 2002, Int J Syst Evol Microbiol. 52: 1043

第1章 適用範囲

本手順書は、*Acinetobacter* 属細菌を6種類の指標遺伝子を用いた MLSA 法により、現在の分類体系に基づき同定するための手順を示したものである。

MLSA 法に用いる遺伝子の塩基配列については、解析対象とする菌株（以下「解析株」という。）について、全ゲノム解析を実施して取得したゲノム配列、もしくは、国際塩基配列データベース（DDBJ/EMBL-Bank/GenBank）に登録されているゲノム配列から抽出することを前提としている。一般的に、MLSA 法に用いる遺伝子は変異が多いため、菌種を超えて増幅可能なユニバーサルプライマーの設計が難しく、解析に十分な長さの遺伝子領域を PCR 法により増幅することは困難である。そこで、近年のゲノム解析費用の低廉化が進んでいる状況を踏まえ、本手順書では、新たに遺伝子配列を取得する方法として全ゲノム配列から遺伝子の塩基配列を取得するケースを想定した。

必要なソフトウェアについて

解析を行う前には、以下のソフトウェアをあらかじめダウンロードし、インストールしておく必要がある。

- ◆ **MEGA: Molecular Evolutionary Genetics Analysis**、URL : <http://www.megasoftware.net/>
(本手順書ではバージョン 5.2.2 を使用)

また、解析手順①（第3章の解析の全体概要参照、詳細については第5章を参照）においてローカル環境で BLAST 検索を行って遺伝子の塩基配列を取得する場合は以下のソフトが必要である。

- ◆ **BioEdit** (オプション)、URL : <http://www.mbio.ncsu.edu/Bioedit/bioedit.html>
(本手順書ではバージョン 7.2.5 を使用)

第 2 章 用語及び定義

2.1 MLSA (Multilocus sequence analysis)法 本手順書内で MLSA 法とは、複数の遺伝子の塩基配列を連結して作製した塩基配列を用いた分子系統解析による同定法をいう。

2.2 指標遺伝子 本手順書内で指標遺伝子とは、*Acinetobacter* 属細菌を同定するための MLSA 法に用いる 6 遺伝子 (*cpn60*, *gltA*, *gyrB*, *pyrG*, *rpoB*, *rpoD*) をいう。

2.3 基準株 (Type strain) 学名 (種名) が提案された際に、その種の性状を表す基準として指定された菌株のこと。解析株の同定を行う際には、基準株との比較を行う必要がある。

2.4 16S rRNA 遺伝子配列 リボソーム小サブユニットに含まれる 16S rRNA をコードする遺伝子。原核生物の全ての種に存在し、種間で配列の保存性が高いことから、細菌の系統分類に用いられる。

2.5 NCBI (National Center for Biotechnology Information) 米国国立生物工学情報センター。分子生物学やバイオインフォマティクスの研究用データベースの構築および運営などを行う研究組織。

2.6 DDBJ/EMBL-Bank/GenBank 国立遺伝学研究所 (DDBJ)、英国の欧州バイオインフォマティクス研究所 (EBI) および米国の NCBI が連携して構築している国際塩基配列データベース。

2.7 Accession 番号 配列データに対する固有な ID として、国際塩基配列データベースが発行する登録番号であり、この番号を用いて配列データの検索することができる。

2.8 問い合わせ配列 (クエリーシーケンス) 塩基配列データベースから相同性を持つ配列を検索の際に使用するユーザーが手元に持っている塩基配列またはアミノ酸配列のこと。

2.9 FASTA 形式 配列名と配列から構成されるテキストデータ。先頭行 (ヘッダ行) には “>” で始まる配列名を書き、改行してから配列を記述する形式のこと。MEGA などの配列解析ツールでアラインメントを行う際には FASTA 形式のファイルを使用する必要がある。

2.10 Multi-FASTA ファイル 複数の FASTA 形式配列で構成されるテキストデータ。

第3章 解析手順の概要

ここに解析の全体概要を示す。各解析手順の詳細については第5章以降で説明する。

【解析手順】

- ① 解析株全ゲノムの塩基配列データから表1. に示した6遺伝子の塩基配列を取得する

※MLSA 法に適した遺伝子は、1) 全種に共通に存在し、2) ゲノム内に1コピーであり、3) 長さが系統解析に十分な情報量 (>900 塩基長) を持ち、4) 水平伝播が起こりにくい遺伝子であるとされ⁹、一般的にはハウスキーピング遺伝子¹⁰が用いられる。本手順書では、前述の1) ~4) の条件を満たすハウスキーピング遺伝子の中から、NITEが検討を行って *Acinetobacter* 属細菌の同定用指標遺伝子として選んだ6遺伝子を用いた解析手順を説明する。

表1. MLSA に利用する遺伝子

	遺伝子名	遺伝子産物名
1	<i>cpn60</i>	60 kDa chaperonin
2	<i>gltA</i>	Citrate synthase
3	<i>gyrB</i>	DNA gyrase subunit B
4	<i>pyrG</i>	CTP synthase
5	<i>rpoB</i>	DNA-directed RNA polymerase subunit beta
6	<i>rpoD</i>	RNA polymerase sigma factor RpoD

- ② *Acinetobacter* 属基準株の遺伝子の塩基配列(第6章で指定するURLからダウンロード)と解析手順①で取得した遺伝子の塩基配列で、MEGA などの配列解析ツールを用いて遺伝子ごとにアラインメントを作成し、各遺伝子の長さを揃えるために Gblocks server でトリミング処理を行う。

Gblocks server: URL : http://molevol.cmima.csic.es/castresana/Gblocks_server.html

- ③ 解析手順②で長さを揃えた遺伝子の塩基配列を、解析株ごとに以下の遺伝子順に連結して1本の塩基配列にし(以下、連結配列、次ページの図参照)、全解析株の連結配列の Multi-FASTA ファイルを作成する。

⁹ Adékambi T et al. 2011, PLoS One. 6: e14792

¹⁰ 微生物に共通に保存されていると考えられる生育に必須な遺伝子

※遺伝子順は必ずしも以下に示した順である必要はないが、解析手順④以降の解析において異なる株の同一遺伝子の塩基配列を比較するために、全解析株について遺伝子順を揃える必要がある。

遺伝子順：*cpn60* - *gltA* - *gyrB* - *pyrG* - *rpoB* - *rpoD*

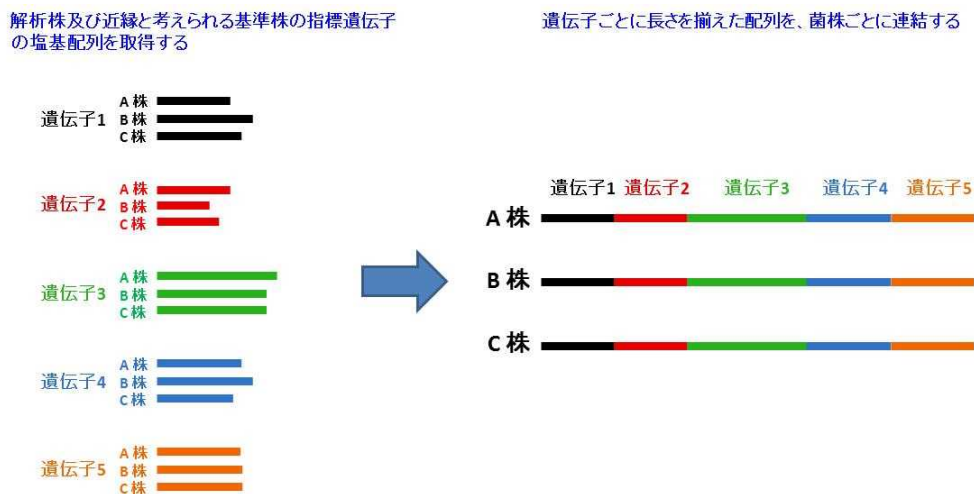


図1. 連結配列の作成

- ④ 解析手順③で作成した連結配列 (Multi-FASTA ファイル) から MEGA (前述) などの配列解析ツールを用いて分子系統樹を描く。
- ⑤ 解析手順④で描いた分子系統樹をもとに、解析株を同定する。

第4章 解析に使用する菌株の選抜

本手順書は、*Acinetobacter* 属細菌について、MLSA 法により種レベルで同定する手順を示したものである。従って、解析株が *Acinetobacter* 属細菌であることをあらかじめ確認しておく必要がある。解析株の同定は、16S rRNA 遺伝子配列を決定し、既知種の基準株 (Type strain) の配列との相同性を比較することで行う。

相同性検索においては、DDBJ/EMBL-Bank/GenBank 等の公共の遺伝子データベースが利用できるが、これらのデータベースには同定が適切でない菌株の配列も含まれるため注意が必要である。そのため、解析株の 16S rRNA 遺伝子配列を用いた同定を行う際には、基準株のみの 16S rRNA 遺伝子配列が登録されたデータベースである EzTaxon¹¹の使用が便利である。

一般的には、既知種の基準株 (Type strain) の 16S rRNA 遺伝子配列との相同性が 98.7% 以上であれば、同種であると考えられるが、解析株の 16S rRNA 遺伝子配列が複数の基準株の配列に高い相同性を示す場合、同定が困難である。このような場合、本手順書に示した MLSA 法を使用することで、解析株の種レベルの同定が可能である。

現在 (2016 年 3 月) までに報告されている *Acinetobacter* 属細菌の学名は表 2 のとおりである。

表 2. *Acinetobacter* 属細菌の学名

学名	基準株ゲノム配列の Accession 番号
<i>Acinetobacter albensis</i>	
<i>Acinetobacter apis</i>	
<i>Acinetobacter baumannii</i>	BBTN00000000
<i>Acinetobacter baylyi</i>	BCMA00000000
<i>Acinetobacter beijerinckii</i>	BBTL00000000
<i>Acinetobacter bereziniae</i>	BBLJ00000000
<i>Acinetobacter bohemicus</i>	
<i>Acinetobacter boissieri</i>	
<i>Acinetobacter bouvetii</i>	BCMB00000000
<i>Acinetobacter brisouii</i>	BBTI00000000
<i>Acinetobacter calcoaceticus</i>	BBTM00000000
<i>Acinetobacter gandensis</i>	
<i>Acinetobacter gernerii</i>	BBLI00000000
<i>Acinetobacter guangdongensis</i>	
<i>Acinetobacter guillouiae</i>	BBRY00000000

¹¹ EzTaxon (URL: <http://www.ezbiocloud.net/eztaxon>)

表 2. (つづき)

<i>Acinetobacter gyllenbergii</i>	BBTK00000000
<i>Acinetobacter haemolyticus</i>	BBSE00000000
<i>Acinetobacter harbinensis</i>	
<i>Acinetobacter indicus</i>	BBSF00000000
<i>Acinetobacter johnsonii</i>	BBTB00000000
<i>Acinetobacter junii</i>	BBSG00000000
<i>Acinetobacter kookii</i>	
<i>Acinetobacter lwoffii</i>	BBSQ00000000
<i>Acinetobacter nectaris</i>	
<i>Acinetobacter nosocomialis</i>	BBSR00000000
“ <i>Acinetobacter oleivorans</i> ” ¹²	BBSS00000000
<i>Acinetobacter pakistanensis</i>	
<i>Acinetobacter parvus</i>	BCME00000000
<i>Acinetobacter pittii</i>	BBST00000000
<i>Acinetobacter puyangensis</i>	
<i>Acinetobacter qingfengensis</i>	
<i>Acinetobacter radioresistens</i>	BAGY00000000
<i>Acinetobacter rudis</i>	BBRX00000000
<i>Acinetobacter schindleri</i>	BCMD00000000
<i>Acinetobacter seifertii</i>	
<i>Acinetobacter soli</i>	BBNM00000000
<i>Acinetobacter tandoii</i>	BBNK00000000
<i>Acinetobacter tjernbergiae</i>	BBND00000000
<i>Acinetobacter townneri</i>	BBNL00000000
<i>Acinetobacter ursingii</i>	BCMC00000000
<i>Acinetobacter variabilis</i>	
<i>Acinetobacter venetianus</i>	BCLZ00000000

このうち、基準株ゲノム配列の Accession 番号が記入してある種については、現時点で基準株のゲノム解析が終了している。一方、これら以外の種の基準株についても、今後のゲノム解析の実施状況により、DDBJ 等の国際塩基配列データベースに登録されることから、解析前には DDBJ 等を適宜確認し、新たに基準株のゲノム配列データが登録されている場合は、第 5 章からの手順にしたがって、遺伝子の塩基配列を取得する。

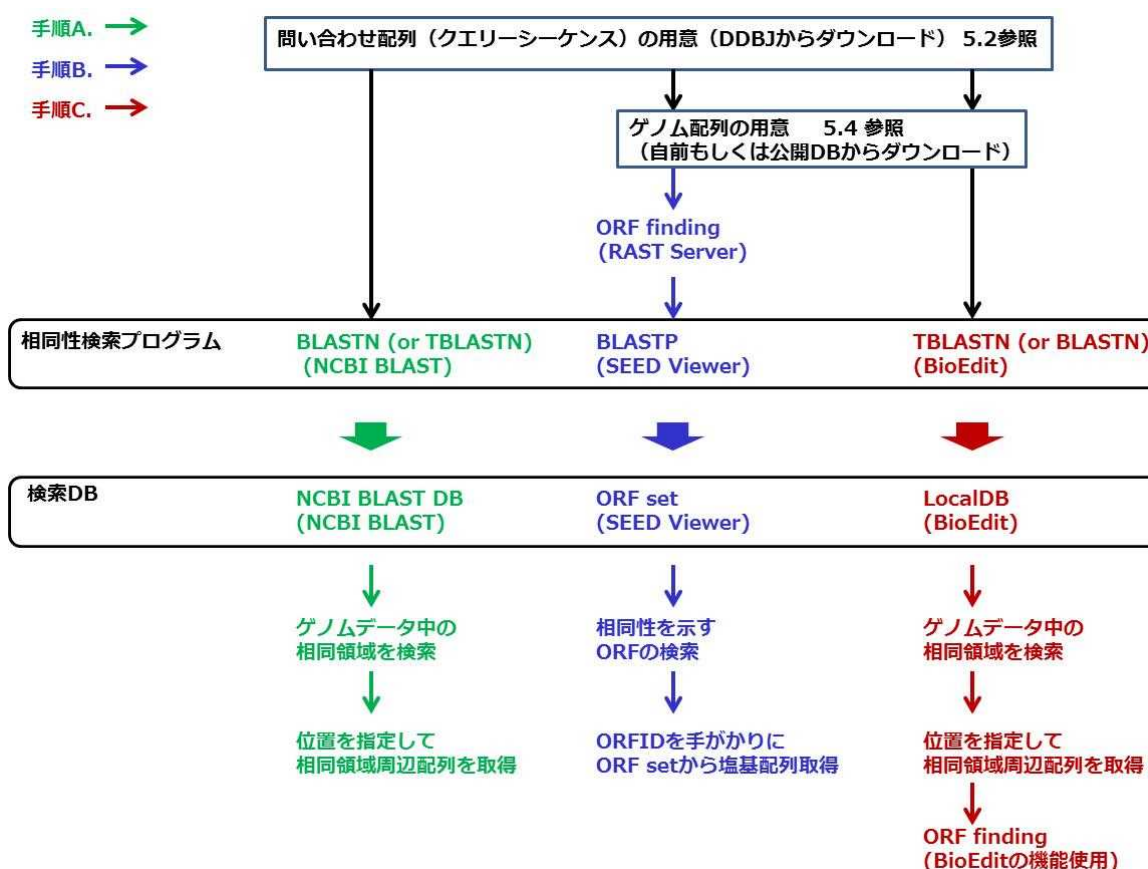
¹² Invalid name : 正式には承認されていない学名

第5章 指標遺伝子の取得

5.1 塩基配列取得法の選択

指標遺伝子の取得方法について、例として3種類の手順を下図に示した。実際の作業においては、いずれかの手順を選択し、指標遺伝子を取得する。

公開されているゲノム配列からの遺伝子塩基配列の取得はNCBI web サイトで行う(手順A、詳細は同章の5.3.1で記述)。手持ちのゲノム配列データから遺伝子塩基配列を取得する場合は、RAST server (手順B、詳細は同章5.3.2で記述)かBioEdit (手順C、詳細は同章5.3.3で記述)のどちらかを使う。外部ネットワークを利用せずに配列を取得したい場合は、BioEditを使うとよい。



5.2 問い合わせ配列 (クエリーシーケンス) の入手

同源性検索に使用する問い合わせ配列 (クエリーシーケンス) を表3.の各 Accession 番号を用いてDDBJの配列検索ページ「getentry」からダウンロードする。上図を参照し、「同源性検索プログラム」がBLASTNである場合は問い合わせ配列に塩基配列を用い、TBLASTNの場合は問い合わせ配列にアミノ酸配列を用いる。(手順A.もしくは手順C.により取得する場合は、同源性検索プログラムとしてBLASTNとTBLASTNのどちらを使用してもよい。)

表 3. 相同性検索に用いる問い合わせ配列の Accession 番号

	遺伝子名	Accession 番号
1	<i>cpn60</i>	LC102559
2	<i>gltA</i>	LC102587
3	<i>gyrB</i>	LC102615
4	<i>pyrG</i>	LC102643
5	<i>rpoB</i>	LC102671
6	<i>rpoD</i>	LC102699

(表 3.の問い合わせ配列は、全て *Acinetobacter baumannii* NBRC 109757^T の配列である。)

問い合わせ配列のダウンロードは、「getentry」(URL:<http://getentry.ddbj.nig.ac.jp/top-j.html>) にアクセスして ID 欄に表 3.の Accession 番号を入力し、出力形式欄は、塩基配列を取得する場合は【CDS 塩基配列 FASTA】を、アミノ酸配列の場合は【CDS アミノ酸配列 FASTA】を選択する。

5.3 解析株のゲノム配列データからの指標遺伝子塩基配列の取得

5.1 にて選択した手順にて塩基配列を取得する。手順 A.の場合は 5.3.1 に、手順 B.の場合は、5.3.2 に、手順 C.の場合は 5.3.3 に進む。

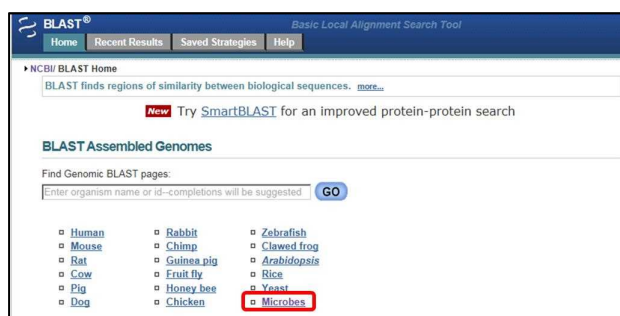
5.3.1 NCBI web サイトでの遺伝子配列の取得

解析株のゲノム配列データに対して、NCBI web サイトで BLASTN もしくは TBLASTN による遺伝子の同源性検索を行い、目的の遺伝子が含まれる遺伝子領域 (ORF) の配列を取得する。BLASTN を使う場合は塩基配列を問い合わせ配列として、TBLASTN を使う場合はアミノ酸配列を問い合わせ配列として用いる。

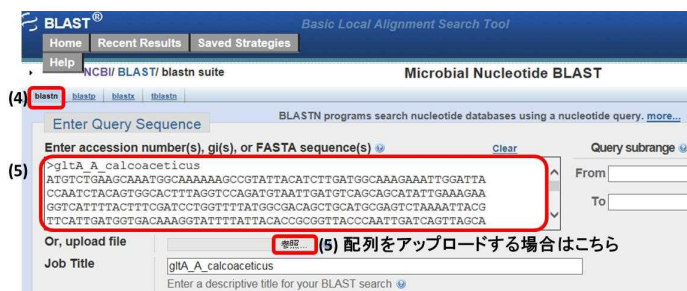
以下に、*Acinetobacter* sp. NBRC 110496, taxonomy ID=1550715 のゲノム配列データ (Accession 番号: BBTF01000001-BBTF01000165) に対して BLASTN で相同領域を探す場合を例に挙げて、手順を示す。

【手順】

- (1) NCBI トップページ (URL : <http://www.ncbi.nlm.nih.gov/>) にアクセスする。
- (2) 画面右側の【Popular Resources】の一覧から【BLAST】をクリックする。
- (3) NCBI BLAST のページでデータベースを選択する (【Microbes】をクリックする)。



- (4) 検索方法が blastn になっていることを確認する (なっていない場合は選択タブの【blastn】をクリックする)。
- (5) 問い合わせ配列 (クエリーシーケンス : Query Sequence) をコピーし、配列入力用のテキストボックスにペーストする。もしくは、配列ファイルをアップロードする (【upload file】の【参照】ボタンをクリックして、問い合わせ配列の保存先を指定する)。



(6) データベースを選択する。【Database】で【All genomes】のオプションボタンをクリックし、プルダウンメニューで【Draft genomes】を選択する（解析株のゲノム配列データが完全長ゲノムである場合は、【Complete genomes】を選択する）。

(7) ゲノム配列データの由来生物名を指定する。【Organism】のテキストボックスにゲノム配列データの由来生物名の一部（例：Acinetobacter sp. NBRC 110496）を入力するとプルダウンメニューが出てくるので、その中からゲノムの由来生物名を選択してクリックする。

※プルダウンメニューが出てこない場合は、テキストボックスに taxonomy ID（例：1550715）の数字だけを入力してもよい。

(8) 検索エンジンの選択をする。ここでは megablast もしくは、discontiguous megablast を選択する。

※megablast は検索時間が速いが、塩基配列の一致率が高く（約 95%以上）ないと検出されない。megablast で検索を実行し、結果が返ってこない場合は discontiguous megablast を使う。

(9) 【BLAST】ボタンをクリックして検索を実行する。

Choose Search Set

Database: Representative genomes only All genomes (6) Genomes: 1

Complete genomes
Draft genomes
Complete bacteriophages

Complete genomesか
Draft genomesを選択する

Organism: Acinetobacter sp. NBRC [x] Exclude [v]

Acinetobacter sp. NBRC 100985 (taxid:...) Only 20 top taxa will be shown. [v]

Acinetobacter sp. NBRC 110496 (taxid:...) (7) ゲノムの由来生物名を選択する

Limit to: Optional

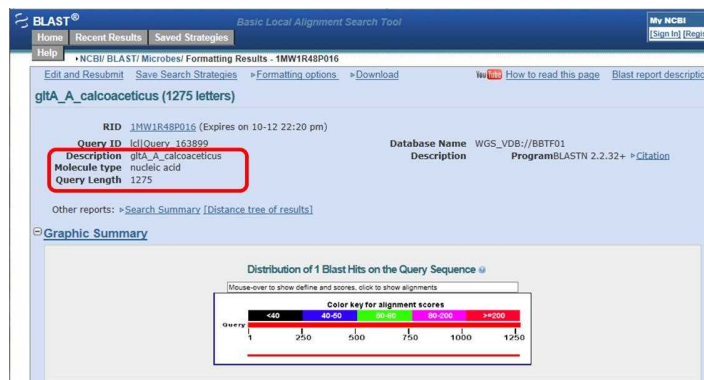
Program Selection

Optimize for (8): Highly similar sequences (megablast)
 More dissimilar sequences (discontiguous megablast)
 Somewhat similar sequences (blastn)
Choose a BLAST algorithm [v]

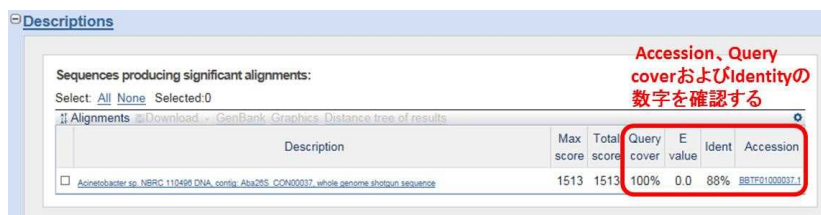
(9) BLAST Search database Draft genomes using Megablast (Optimize for highly similar sequences)
 Show results in a new window

Algorithm parameters

(10) BLAST の結果ページの問い合わせ配列の情報欄で、問い合わせ配列の長さを確認する。



(11) 【Description】欄には、ヒットしたゲノム領域の概要がリストになって示されるので、Accession 番号（下図では BBTF01000037.1）、Query cover（問い合わせ配列の長さに対するヒットしたゲノム領域の長さの割合、下図では 100%）、Identity（ヒットした領域での問い合わせ配列とヒットしたゲノム配列の塩基の一致率、下図では 88%）の数字を確認する。



Description	Max score	Total score	Query cover	E value	Ident	Accession
Acinetobacter sp. NRRC 110498 DNA contig_Aba26S_CCH00037_whole genome shotgun sequence	1513	1513	100%	0.0	88%	BBTF01000037.1

➤ BLAST の結果の見方

BLASTN の結果得られたアラインメントが有意であるかどうかを判断するためのおおよその目安を以下に示す。ただし、問い合わせ配列が表 3. (9 ページ) に挙げた遺伝子の塩基配列であり、*Acinetobacter* 属細菌の全ゲノム配列をデータベースにした場合に限る。

- Query cover（問い合わせ配列に対するアラインメントの長さの割合）：70%以上
- Identities：70%以上

(12) 次に、【Alignments】欄で1対1のアラインメント結果を確認する。ヒットしたコンティグの Accession 番号（次ページの図では BBTF01000037.1）を【Sequence ID】で確認し、アラインメントを見て、問い合わせ配列（Query）の何番目から何番目までの塩基配列が、ヒットしたコンティグの正鎖（Plus）または相補鎖（Minus）の何番目から何番目までの塩基配列にヒットしているかを確認する。

※次ページの図では問い合わせ配列の1番目から1275番目までの塩基配列が、BBTF01000037.1の正鎖の23683番目から24957番目までの塩基配列にヒットしている。

(13) 【Sequence ID】の一行下にある【GenBank】をクリックしてヒットしたゲノム領域の GenBank 形式データを表示させる。

Alignments

Download ▾ GenBank Graphics

Acinetobacter sp. NBRC 110496 DNA, contig: Aba26S_CON00037, whole genome shotgun sequence
Sequence ID: [dbj|BBTF01000037.1](#) Length: 38611 Number of Matches: 1

Range 1: 23683 to 24957 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
1513 bits(819)	0.0	1126/1278(88%)	6/1278(0%)	Plus/Plus

Query 1 ATGCTCTGAAGCAAATGGCAAAGCCGATATTACATCTTGATGGCAAAGAAATGGATTA 60
Sbjct 23683 ATGCTCTGACGCACTGGCAAAGCCGATATTACAGCTTGATGGCAAAGAAATGAATTA 23742

Query 61 CCAATCTACAGTGGCACTTTAGTCCAGATGAATTGATGTCAGCAGCA-TAATGAAAGA 119
Sbjct 23743 CCAATTTACAGCGCACATTGGGCCAGATGAATTGACGTTA-ARGATGTAITGGCAIC 23801

Query 120 AGGTCATTTACTTTCGATCCCTGGTTTTATGGCGACAGCTGCATGGGAGCTTAAATAC 179
Sbjct 23802 TGGTCACCTTTACTTTCGATCCCTGGTTTTATGGCGACAGCGGCTTGGAAITCAAGATCAC 23861

Query 180 GTTCATTGATGGTGACAAAGGTATTTTATTACACCGGGTTACCCAATTGATCAGTTAGC 239
Sbjct 23862 ATTCAITGATGGTGACAAAGGTGTGTTATTACACCGGGTTATCCAATTGATCAATTAGC 23921

Query 240 AACTCAAGCAGACTACCTTGAJAAGTCTGCTACTGCTATTAAATGGCGAGCTTCCAACCGC 299
Sbjct 23922 GACTCAAGCAGACTACTTAGAAACTGCTATTACTTTTAAATGGCGAGCTTCCAATGTC 23981

中略

Query 1078 TACCCTAACGTAGATTTCTATTACAGGATCATCTTAAAGCGAATGGTATCCCTACAGAA 1137
Sbjct 24760 TACCGAAGCTAGACTTCTACTCAGGTATCACTTAAAGCGAATGGTATCCCAACGGGT 24819

Query 1138 ATGTTTACGGTAATCTTTCGCACTTGCACGTACAGTAGGTTGGATCAGCCACTGGTTAGAA 1197
Sbjct 24820 ATGTTTACGGTATCTTTCGCACTTGCACGTACAGTAGGTTGGATCAGCCACTGGTTAGAA 24879

Query 1198 ATGCACAGCGCACCTTACAAAATCGGTCTGCTCTGCGCAGCTTTATCTGGCCCGACTCAG 1257
Sbjct 24880 ATGCACAGCTGCTCTTACAAAGATCGGTCTGCGCAGCTTTACACAGGCTCGACCCAG 24939

Query 1258 CGCGAATCAAAAGCTTAA 1275
Sbjct 24940 CGCGCATCAAAAGCTTAA 24957

(14) (12)でアラインメントが問い合わせ配列の途中から始まっていたり、途中で終わっていたりする場合は、GenBank形式ファイルの右上の【Change region shown】欄の数字を適宜変更して、表示されるゲノム領域が問い合わせ配列と同程度の長さになるように調節する。

Display Settings: ▾ GenBank

Send to: ▾

Acinetobacter sp. NBRC 110496 DNA, contig: Aba26S_CON00037, whole genome shotgun sequence

GenBank: [BBTF01000037.1](#)

FASTA Graphics

Go to: ☺

LOCUS BBTF01000037 1275 bp DNA linear BCT 10-FEB-2015

DEFINITION Acinetobacter sp. NBRC 110496 DNA, contig: Aba26S_CON00037, whole genome shotgun sequence.

ACCESSION [BBTF01000037](#) REGION: 23683..24957

VERSION [BBTF01000037.1](#) GI:756135124

DBLINK BioProject: [FRJDB1754](#)
BioSample: [SAM00020359](#)

KEYWORDS WGS; STANDARD_DRAFT.

SOURCE Acinetobacter sp. NBRC 110496

Change region shown

Whole sequence

Selected region

from: 23683 to: 24957

Update View

Customize view ▾

Analyze this sequence ▾

Related information ▾

Recent activity ▾

(15) 表示したゲノム領域の塩基配列をFASTA形式でダウンロードする。【Send to】をクリックすると開くドロップダウンメニューから出力方式(Choose Destination)として【File】を選択し、ファイル形式(Format)のプルダウンメニューで【FASTA】を選択し、【Create File】ボタンを押す。

Display Settings ▾ GenBank

Acinetobacter sp. NBRC 110496 DNA, contig: Aba26S_CON00037.1 shotgun sequence

GenBank: BBTf01000037.1

[FASTA](#) [Graphics](#)

[Go to](#) 🔍

LOCUS BBTf01000037 1272 bp DNA linear BCT 10-FEB-2015

DEFINITION Acinetobacter sp. NBRC 110496 DNA, contig: Aba26S_CON00037, whole genome shotgun sequence.

ACCESSION BBTf01000037 REGION: 23683..24954

VERSION BBTf01000037.1 GI:756135124

DBLINK BioProject: PRJDB1754

KEYWORDS WGS; STANDARD DRAFT.

SOURCE Acinetobacter sp. NBRC 110496

ORGANISM Acinetobacter sp. NBRC 110496
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Moraxellaceae; Acinetobacter.

REFERENCE 1

AUTHORS Shimodaira,J., Hooyama,A., Noguchi,M., Hirakata,S., Uehara,A., Ohji,S., Ichikawa,N., Kimura,A., Yamazoe,A., Ezaki,T. and Fujita,N.

TITLE Whole genome shotgun sequence of Acinetobacter sp. NBRC 110496

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1272)

AUTHORS Hooyama,A., Kimura,A., Ohji,S., Ichikawa,N., Yamazoe,A. and Fujita,N.

TITLE Direct Submission

JOURNAL Submitted (09-FEB-2015) Contact:Director-General Biological Resource Center National Institute of Technology and Evaluation, NITE Biological Resource Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan URL :http://www.bio.nite.go.jp/

Send to ▾

Choose Destination

File Clipboard

Collections Analysis Tool

Download 1 items.

Format

FASTA [Other]

Analyze this sequence

Related information

Recent activity

(16) 適当な名前をつけてダウンロードした塩基配列を保存して 6 章に進む。

※ファイルの拡張子は「fasta」もしくは「fas」とする。ヘッダ行は”>【遺伝子名】_【解析株名】”とすることを推奨する（例：”>gltA_NBRC_110496”）。

5.3.2 RAST Server で予測した解析株の全 ORF からの遺伝子配列の取得

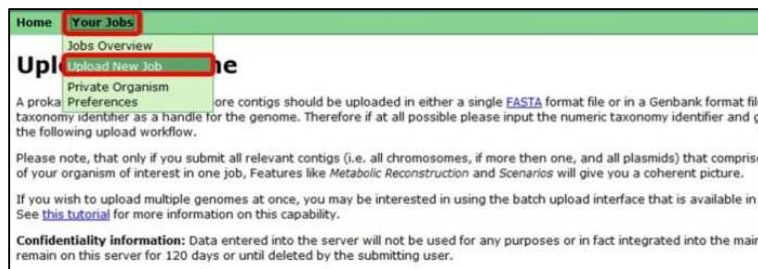
解析株のゲノム配列データ (FASTA 形式) を用意してから (詳細は 5.4 参照)、以下の手順により遺伝子配列を取得する。

外部アノテーションパイプライン RAST Server (Rapid Annotation using Subsystem Technology) を使って、ゲノム配列データから推定した解析株の全 ORF の中から、問い合わせ配列と相同な ORF を探す。検索プログラムは BLASTP を使うため、問い合わせ配列はアミノ酸配列を用いる。

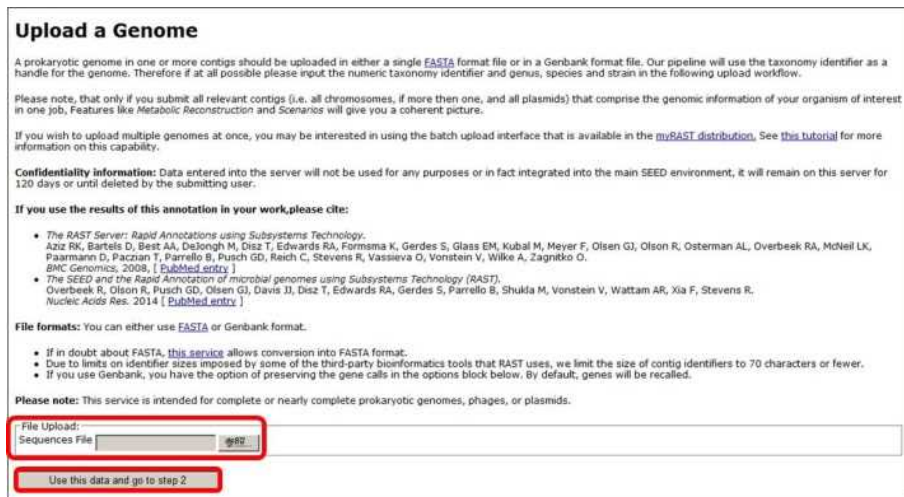
【手順】

5.3.2.1 RAST Server による解析株全 ORF の推定

- (1) RAST Server (URL : <http://rast.nmpdr.org/>) にアクセスし、ログインする。
- (2) メニューバー **【Your Jobs】** から **【Upload New Job】** を選択し、配列アップロードページに移動する。



- (3) **【File Upload: Sequence File】** の **【参照ボタン】** をクリックして、解析したいゲノム配列のファイルを選択し、**【Use this data and go to step2】** のボタンをクリックする。



(4) 【Genome Information】欄で【Taxonomy ID:】のボックスに Taxonomy ID を入力して横の【Look up taxonomy ID at NCBI】のボタンをクリックすると【Taxonomy string】、【Domain】、【Bacteria】、【Genus】、【Species】、【Strain】のテキストボックスに自動的にデータが入力されるので、【Genetic Code:】で 11 番 (Archaea, most Bacteria, most Virii, and some Mitochondria) のオプションボタンをクリックして【Use this data and go to step3】のボタンをクリックする。

Upload a Genome
Review genome data
 We have analyzed your upload and have computed the following information.

Contig statistics

Statistic	As uploaded	After splitting into scaffolds
Sequence size	4648418	4648418
Number of contigs	1	1
GC content (%)	38.3	38.3
Shortest contig size	4648418	4648418
Median sequence size	4648418	4648418
Mean sequence size	4648418.0	4648418.0
Longest contig size	4648418	4648418

Please enter or verify the following information about this organism:

- RAST bases its genome identifiers on NCBI taxonomy IDs.
- If you provide a valid taxonomy-ID, RAST will attempt to fill in the genome metadata for you.
- If you leave the taxonomy-ID field blank, RAST will assign a meaningless taxonomy-ID, and you will need to fill in the below genome metadata manually.
- If you plan on submitting this genome to PATRIC, you will need to provide the most descriptive NCBI taxonomic grouping possible. If you leave the taxonomy-ID field blank, RAST will assign a meaningless taxonomic identifier and the genome will not be suitable for submission to PATRIC. We discuss the motivation and process for submitting your genome to PATRIC in [this document](#).
- You may search for the taxonomy-ID of your organism using the search facilities at the [NCBI Taxonomy browser](#).

Genome information:

Taxonomy ID: [Look up taxonomy ID at NCBI](#)

Taxonomy string:

Domain: Bacteria Archaea Virus

Genus:

Species:

Strain:

Genetic Code: 11 (Archaea, most Bacteria, most Virii, and some Mitochondria)

Code: 4 (Mycoplasmata, Spiroplasmata, Lineoplasmata, and Fungal Mitochondria)

[Use this data and go to step 3](#)

(5) オプション画面でアノテーションの条件設定を行い、【finish the upload】ボタンをクリックしてファイルをアップロードする。

Upload a Genome
Complete Upload
 Please consider the following options for the RAST annotation pipeline:

RAST Annotation Settings:

Choose RAST: Classic RAST GLIMMER-3

Select gene caller: GLIMMER-3

Select F32fam version for this run:

Automatically fix errors? Yes No

Fix frameshifts? Yes No

Build metabolic model? Yes No

Build gaps? Yes No

Turn on debug? Yes No

Set verbose level:

Disable replication Yes No

[Finish the upload](#)

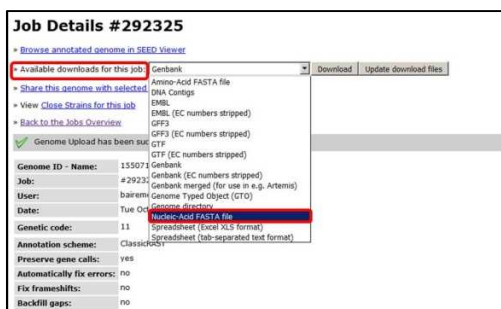
(6) アノテーションが終了するまでは時間がかかる。RAST の Annotation Server からアノテーション終了のメールが届いたら、再度 RAST Server にログインする。

(7) メニューバー【Your Jobs】から【Jobs Overview】を選択し、ジョブ一覧 (Jobs you have access to:) から対象ジョブの【Annotation progress】列の【view details】をクリックする。

Jobs you have access to :

Job	Owner	ID	Name	Num contigs	Size (bp)	Creation Date	Annotation Progress	Status
292325			Acinetobacter sp. NBRC 110496	165	4403056	2015-10-06 22:04:41	<div style="width: 100%; height: 10px; background-color: green;"></div> view details	complete

(8) 【Available downloads for this job:】 のプルダウンメニューから【Nucleic-Acid FASTA file】を選択してダウンロードする。



5.3.2.2 RAST Server で推定した全 ORF への相同性検索

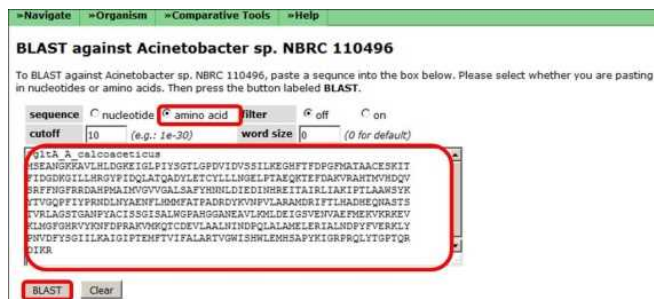
(1) 【Browse annotated genome in SEED Viewer】 をクリックするとアノテーションデータ用のブラウザ【SEED Viewer】が起動する。



(2) 【SEED Viewer】 のメニューバーの【Comparative Tools】 から【BLAST search】 をクリックする。



(3) 配列形式(sequence)のオプションボタン【amino acid】にチェックを入れて、テキストボックスにコピーした問い合わせ配列をペーストし、【BLAST】 ボタンをクリックする。



(4) 検索が終了したら、ヒットした ORF の概要がリストで表示され、その下に問い合わせ配列と各 ORF の 1 対 1 のアラインメント結果が表示されるので、最上位でヒットしている ORF について BLAST の結果を確認する。

➤ BLAST の結果の見方

BLASTP の結果得られたアラインメントが有意であるかどうかを判断するためのおおよその目安を以下に示す。ただし、問い合わせ配列が表 3. (9 ページ) に挙げた遺伝子のアミノ酸配列であり、*Acinetobacter* 属細菌のゲノム配列から推定された全 ORF のアミノ酸配列をデータベースにした場合に限る。

- Identities が 80%以上であること
- 問い合わせ配列のアミノ酸配列長、およびトップヒットの ORF のアミノ酸配列長が、それぞれアラインメント長 (Identities の分母の値) の 70%以上であること。
- 問い合わせ配列の長さに対してトップヒットの ORF のアミノ酸配列長が 80%以上であること (ORF のアミノ酸配列長は“>”に ORF の ID 番号が続く行の 1 行下に表示されている)。

```

BLASTP 2.2.26 [Sep-21-2011]

Reference:
Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
"Gapd BLAST" and PSI-BLAST: a new generation of protein database search
programs", Nucleic Acids Res. 25:3389-3402.

Reference for compositional score matrix adjustment:
Altschul, Stephen F., John C. Nootton, E. Michael Gertz, Richa Agarwala,
Alexandr Morgulis, Alejandro A. Schäffer, and Yi-Xiao Yu (2005) "Protein database
searches using compositionally adjusted substitution matrices", FEBS J. 272:5101-5109.

Query= gltA_A_calcoacetivus 問い合わせ配列のアミノ酸配列長
      (424 letters)

Database:
Acinetobacter sp. NBRC 110496 protein sequences
      4265 sequences: 1,256,475 total letters

Searching.....done

Sequences producing significant alignments:
                                     Score   E
                                     (bits) Value
figI1550715.3.peg.1122  トップヒットのORF          841  0.0
figI1550715.3.peg.2841          174 2e-51
figI1550715.3.peg.97           30  0.39
figI1550715.3.peg.3519         30  0.49
figI1550715.3.peg.2676         27  3.4
figI1550715.3.peg.3388         27  3.9
figI1550715.3.peg.4005         26  4.1

>figI1550715.3.peg.1122  トップヒットのORFのアミノ酸配列長
Length = 424

Score = 841 bits (2122), Expect = 0.0, Method: Compositional matrix adjust.
Identities = 401/424 (94%) Positives = 406/424 (95%)

Query: 1  MSEANGKQVHLHLDGMEIQLPIYSGTIGPDVLDVSSILMKGHFFDPDGFMTAAACESKIT 60
MS A GKKAVL LDGMEI LPIYSGTIGPDVLDV +L GHFFDPPGFMTAAACESKIT
Sbjct: 1  MSAATGKQVQLDQGEIQLPIYSGTIGPDVLDVSKVLSGHHFFDPPGFMTAAACESKIT 60

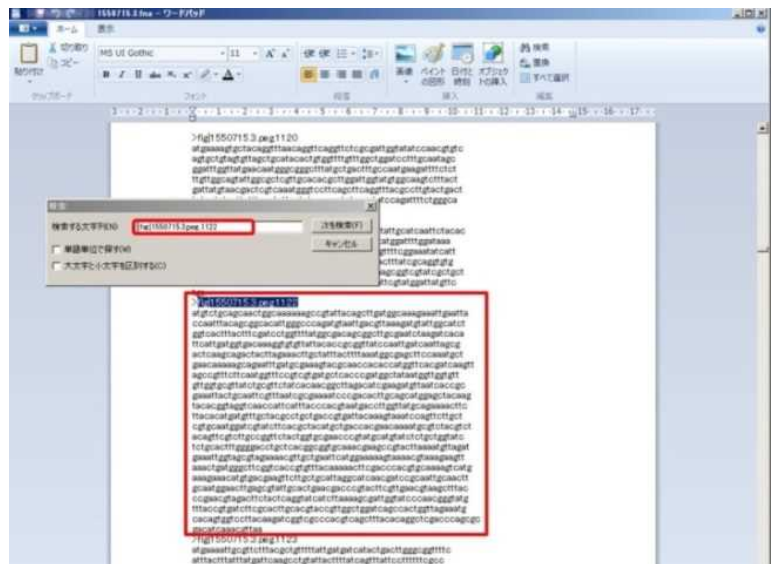
Query: 61  FIDGKGVLLHRGYPIDQLATQADVLEICVLLNGELPFAEQTEFDKVRHMTMHDQV 120
FIDGKGVLLHRGYPIDQLATQADVLEICVLLNGELPFAEQTEFDKVRHMTMHDQV
Sbjct: 61  FIDGKGVLLHRGYPIDQLATQADVLEICVLLNGELPFAEQTEFDKVRHMTMHDQV 120

Query: 121  SREFNFRFRDHPMIMVGVGALSIFYHNNLDIEDINRREITAIRLAKIFILAANSYK 180
SREFNFRFRDHPMIMVGVGALSIFYHNNLDIEDINRREITAIRLAKIFILAANSYK
Sbjct: 121  SREFNFRFRDHPMIMVGVGALSIFYHNNLDIEDINRREITAIRLAKIFILAANSYK 180

```

- (5) 最上位でヒットしている ORF の ID 番号 (図では figI1550715.3.peg.1122) をコピーする。
- (6) 5.3.2.1.(8) で取得した Nucleic-Acid FASTA file から (5) でコピーした ID 番号を持つ塩基配列を取得する。
- (7) 適当な名前をつけて塩基配列を保存して第 6 章に進む。

※ワードパッドなどで開いた全遺伝子配列の中から(5)でコピーした ID 番号を検索し、
 【>で始まる行】から次の【>で始まる行の 1 行上まで】の行を全てコピーして、別ファイルに
 適当な名前をつけて配列を保存する（ヘッダ行は”>【遺伝子名】_【解析株名】”とする
 ことを推奨する。例：”>gltA_NBRC_110496” ファイルの拡張子は、「fasta」もしくは「fas」
 とする。



5.3.3 BioEdit の機能を利用した解析株の遺伝子配列取得

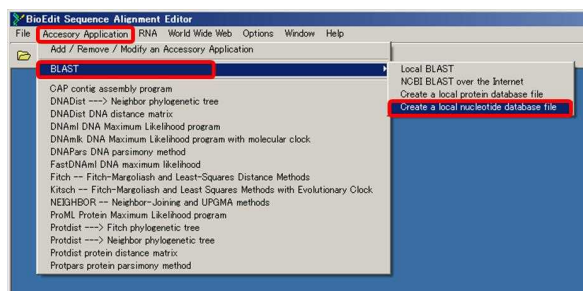
解析株のゲノム配列データ (FASTA 形式) を用意してから (詳細は 5.4 参照)、以下の手順により遺伝子配列を取得する。

解析株のゲノム配列データに対して、無料の配列編集ソフトである BioEdit を用いて Local BLAST (TBLASTN もしくは BLASTN) による相同性検索を行い、目的遺伝子が含まれる遺伝子領域 (ORF) の配列を取得する。TBLASTN を使う場合はアミノ酸配列を問い合わせ配列として、BLASTN を使う場合は塩基配列を問い合わせ配列として用いる。

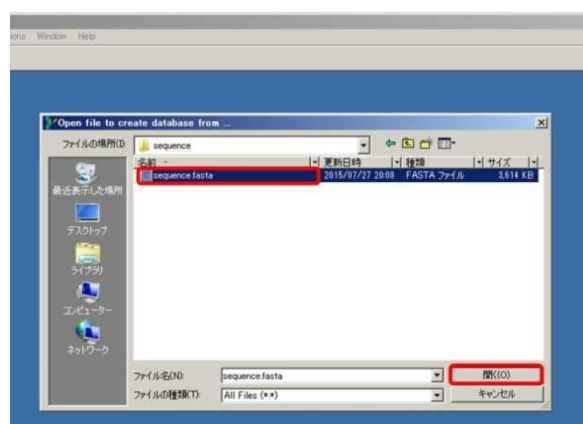
以下にゲノム配列データ (Accession 番号: BBTF01000001-BBTF01000165) に TBLASTN で相同領域を探す場合を例に挙げて、手順を示す。

【手順】

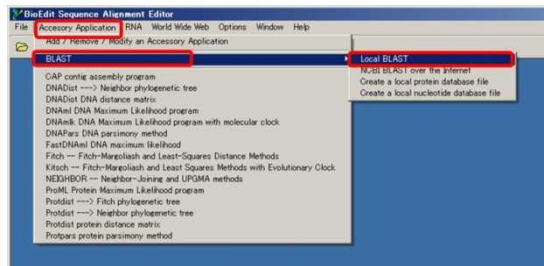
- (1) BioEdit を起動する。以降の作業は、【BioEdit Sequence Alignment Editor】上で行う。
- (2) 解析株のゲノム配列データを用いて Local BLAST 用のデータベースを作成する。メニューバーの【Accessory Application】から【BLAST】 - 【Create a local nucleotide database file】をクリックする。



- (3) 解析株のゲノム配列データのファイル名を指定し、【開く】ボタンをクリックすると Local BLAST 用のデータベースが作成される。



- (4) 次に、再度メニューバーの【Accessory Application】から【BLAST】 - 【Local BLAST】をクリックし、【NCBI Local BLAST】の解析ウィンドウを起動させる。



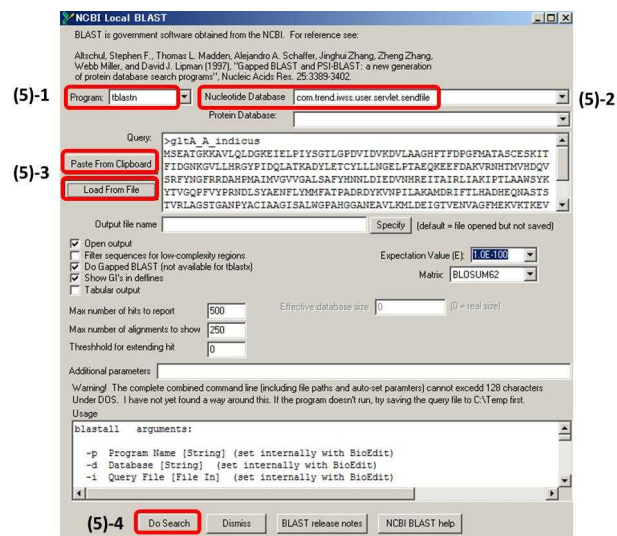
(5) 【NCBI Local BLAST】 解析ウィンドウで以下の設定を行う

(5)-1. Program: tblastn に切り替える。

(5)-2. Nucleotide Database: (2)で作成したデータベースが入っていることを確認する。既に幾つか塩基配列データベースを作成している場合はプルダウンメニューで、解析に用いるデータベースを選択する。

(5)-3. Query: 【Paste From Clipboard】 ボタンをクリックして問い合わせ配列をテキストボックスにペーストするか、【Load From File】 ボタンをクリックして問い合わせ配列をアップロードする。

(5)-4. 【Do Search】 ボタンをクリックして解析を実行する。



(6) BLAST の結果が表示されるので、以下の手順にしたがってゲノム配列中の問い合わせ配列との相同領域を確認する。

(6)-1. 問い合わせ配列のアミノ酸長を確認する（次ページの図では 424aa）。

(6)-2. ”>”の後にあるコンティグの番号を確認する（同図では BBTF01000037.1）。これが、問い合わせ配列に相同領域を持つコンティグの Accession 番号である。

(6)-3. アラインメントの Identities（同図では 96%）とアラインメント長（Identities の分母の値、同図では 424 aa）を確認する。

➤ BLAST の結果の見方

TBLASTN の結果得られたアラインメントが有意であるかどうかを判断するためのおおよその目安を以下に示す。ただし、問い合わせ配列が表 3. (9 ページ) に挙げた遺伝子のアミノ酸配列であり、*Acinetobacter* 属細菌の全ゲノム配列をデータベースにした場合に限る。

- Query cover (問い合わせ配列に対するアラインメントの長さの割合) : 70%以上
- Identities : 80%以上

(6)-4. 次に、アラインメントが問い合わせ配列の何番目のアミノ酸位置から開始し、何番目のアミノ酸位置で終了しているかを、問い合わせ配列 (Query) の開始位置の数字 (下図では 1) と終了位置の数字 (下図では 424) を見て確認する。

(6)-5. 同様にサブジェクト (Subject) 側の配列についても、アラインメントの開始位置 (下図では 23683) と終了位置 (下図では 24954) を確認する。この時にサブジェクト側のアラインメントの開始位置の数字が終了位置の数字より小さい場合は正鎖 (ゲノム配列データと同じ鎖) に、大きい場合は相補鎖 (ゲノム配列データの逆鎖) に相同領域があることを示す。相同領域の存在するストランドの種類で(10)以降の手順が異なる。

```
TBLASTN 2.2.10 [Oct-19-2004]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
"Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs", Nucleic Acids Res. 25:3389-3402.

(6)-1 Query= gitA_haemolyticus
      (424 letters)

Database: D:\Program Files\BioEdit\database\sequence.fasta
      165 sequences; 4,403,056 total letters

Sequences producing significant alignments:

(6)-2 gi|756135124|dbj|BBTF01000037.1| Acinetobacter sp. NBRC 110496 D... 835 0.0
>gi|756135124|dbj|BBTF01000037.1| Acinetobacter sp. NBRC 110496 DNA,
contig: Aba26S_CON00037, whole genome shotgun sequence
Length = 38611

(6)-3 Score = 835 bits (2157), Expect = 0.0
      Identities = 411/424 (96%), Positives = 417/424 (98%)
      Frame = +1

(6)-4 Query: 1 MSAATGKKA...
(6)-5 Sbjct: 23683 MSAATGKKA...

Query: 61 FIDGDKGVLLHRGYPIDQLATQADYLETCYLLNGELPFAEQKAEFDAKVRNHTMVHDQV 120
Sbjct: 23863 FIDGDKGVLLHRGYPIDQLATQADYLETCYLLNGELP AEQKAEFDAKVRNHTMVHDQV 24042

Query: 121 SRFFNGFRDARHPA...
Sbjct: 24043 SRFFNGFRDARHPA...

Query: 181 YTVGQFFIYFRNDL...
Sbjct: 24223 YTVGQFFIYFRNDL...

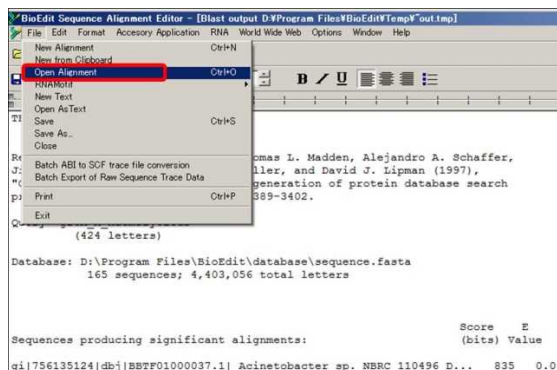
Query: 241 TVRLAGSTGANFYACIS...
Sbjct: 24403 TVRLAGSTGANFYACIS...

Query: 301 KLMFGHVRVYK...
Sbjct: 24583 KLMFGHVRVYK...

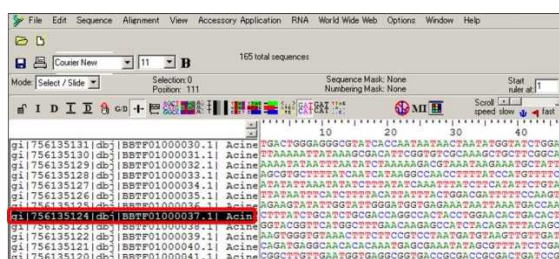
Query: 361 PNVDFYSGIILKAIGI...
Sbjct: 24763 PNVDFYSGIILKAIGI...

Query: 421 DIKR 424 (6)-4
      DIKR
Sbjct: 24943 DIKR 24954 (6)-5
```

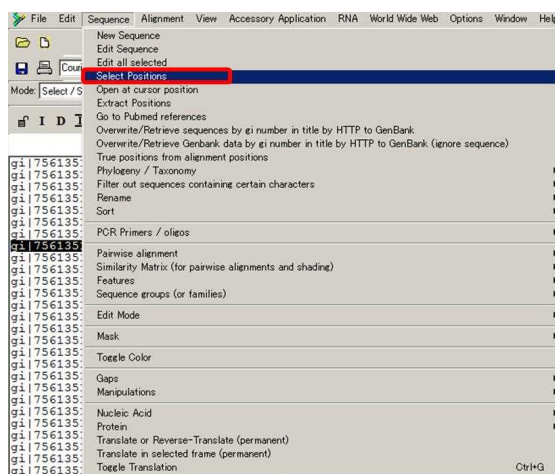

(7) 【BioEdit Sequence Alignment Editor】に戻り、メニューバーの【File】から【Open Alignment】をクリックして、Local BLAST のデータベースに使ったゲノム配列データファイルを選択して開く。



(8) 【BioEdit Sequence Alignment Editor】の画面で、(6)-2.で確認したコンティグ番号をクリックして選択し、そのままメニューバーの【Sequence】－【Select positions】をクリックして配列の領域指定ウィンドウを開く。

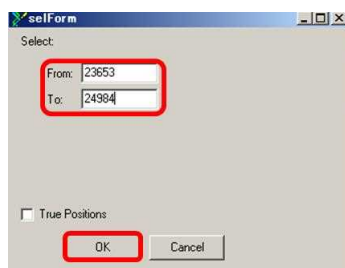


※コンティグの数が多い場合には、メニューバーの【Edit】から【Search】－【Find in Titles】を選択して、出てきた検索ウィンドウでコンティグ番号（例：BBTF01000037）を入力することで探したいコンティグを早く探すことができる。



(9) (6)-5.で確認した数値を参考にして【position data】の【From:】と【To:】に数字を入力して解析株の遺伝子を含んだゲノム領域を指定して【OK】ボタンを押す。

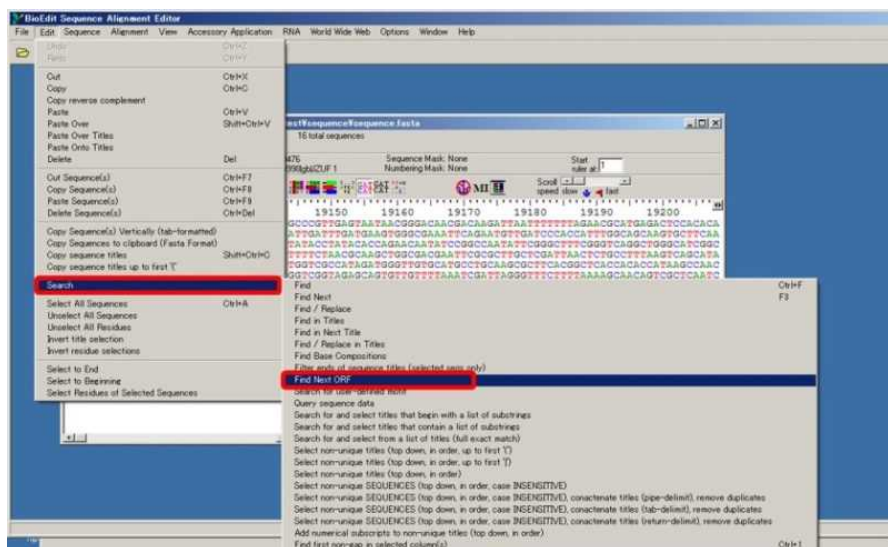
※(10)で開始コドンから終止コドンまでの遺伝子の全長配列を取得するために、相同領域の上流／下流部分について長めに領域を指定しておく（下図では23683～24954までが相同領域である場合に、相同領域の上流／下流部分についてそれぞれ30bp分長めに領域を指定している）。



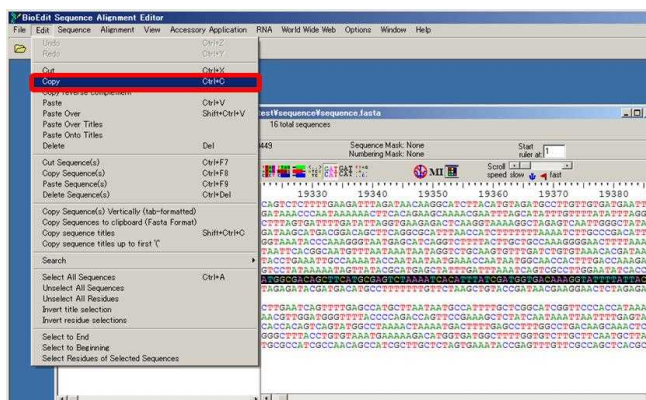
ここから先は、正鎖に相同領域がある場合と相補鎖に相同領域がある場合とで異なる。

【正鎖に相同領域がある場合（例：23683～24954までが相同領域である場合。相補鎖に相同領域がある場合は(13)に進む）】

(10) (9)で指定したゲノム領域を選択したままで、メニューバーの【Edit】から【Search】－【Find Next ORF】をクリックし、選択されたゲノム領域にORFを探す。



(11) ORF を選択したままで、メニューバーの【Edit】から【Copy】をクリックして配列をコピーする。



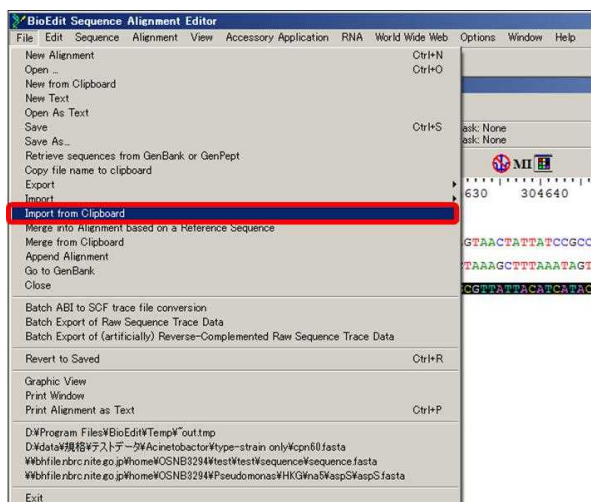
(12) コピーした配列をワードパッドなどにペーストして、ヘッダを付けて保存して5章に進む。

※ファイルの拡張子は「fasta」もしくは「fas」とする。ヘッダ行は>【遺伝子名】_【解析株名】とすることを推奨する。例：”>gltA_NBRC_110496”。

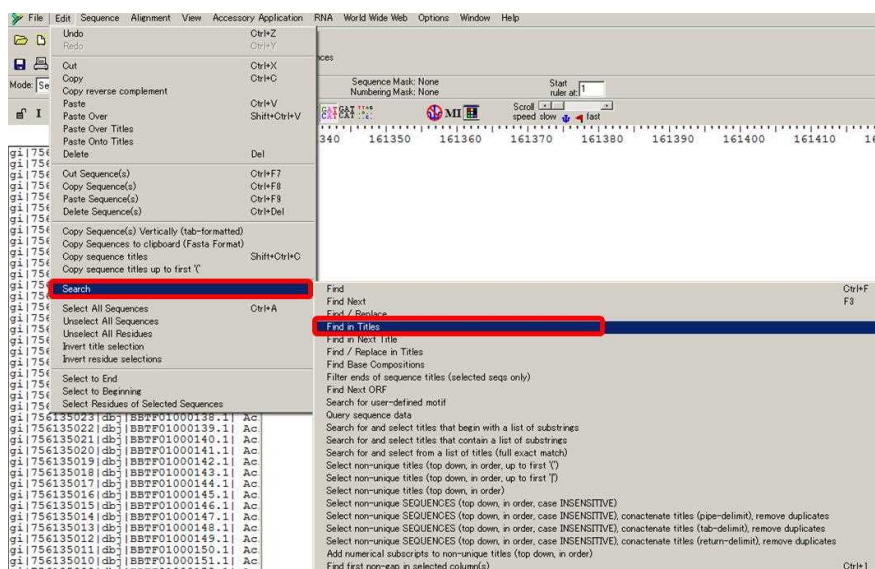
【相補鎖に相同領域がある場合（例：306261～304642 までが相同領域である場合）】

(13) (9)で指定したゲノム領域を選択したままで、メニューバーの【Edit】から【Copy reverse complement】を選択する。

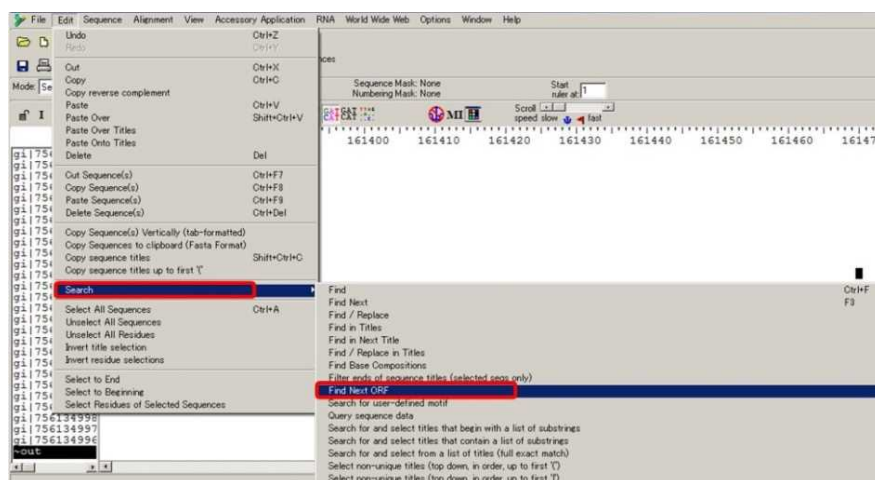
(14) メニューバーの【File】から【Import from Clipboard】をクリックして ORF を含む領域の配列を【BioEdit Sequence Alignment Editor】にインポートする。（～out という配列名でインポートされる。）



(15) メニューバーの【Edit】から【Search】－【Find in Titles】を選択し、出てきた検索ウィンドウに「~out」と入力し(14)でインポートした配列名を検索する。



(16) 配列名「~out」を選択したままで、メニューバーの【Edit】から【Search】－【Find Next ORF】をクリックしてORFを探す。



(16) メニューバーの【Edit】から【Copy】をクリックして配列をコピーする（ORFを選択したままで実施する。(11)参照)。

(17) コピーした配列をワードパッドなどにペーストして適当な名前をつけて保存して6章に進む。

※ファイルの拡張子は「fasta」もしくは「fas」とする。ヘッダ行は>【遺伝子名】_【解析株名】とすることを推奨する。例：>gltA_NBRC_110496”。

5.4 解析株のゲノム配列データの用意

手持ちのゲノム配列データを利用する場合は手順 5.4.1 に進む。公開されているゲノム配列データをダウンロードしてから利用する場合は手順 5.4.2 により塩基配列を取得する。

5.4.1 手持ちのゲノム配列データを利用する場合

配列を FASTA 形式にしておく（“>”の後に配列名を入れ、場合によってはスペースの後に配列に関する解説を記述して改行し、次行以降に塩基配列を記述する。下図参照）。配列は途中で改行してもよい。

```
>AC0601000001 Acinetobacter_baumannii_ATCC_19606
ggatgggtggcgcatgacatcattacaatagatatcaattttgcccaagtgcgtactag
tccaacgattatagtgctcttgtgtaatcaacttattcagtaacaaagcttatggaatc
ccatggccttttttatttcaaaaatttcagaggttgctatggctttaaagtcggaatta
ttaaaagctcagatgttgctcagtggtgcacatttgaactgaaggtggacagccagagt
ttaaattccggggaattgggtataagcccttcaagttgcactagagaaggcaggaacc
aaatcacatccaaaggctatgatgtgatgtaaagatgaaaacgccaagcttaccatg
aattattactggatgcatgtgctgctcacctgattgaagattgaaaaggatagttttt
ctgaggttgtagcggccagccagttgaatcggaaaagccttataccctgagaatgcct
caagcttctcaatcaaggtgacattggatttcaatctggttattcattaaagagcagg
cccagaagattcaggaaagccgacaaggacaaggcttattctgggaaagtcacga
gctctacaaataccaaaaaacgtatgctgcgaaaacgcccagcagaaatcgaacaaatcaa
gttcttaggtggctatctctgatccgccagaatattcttatggcgtgatccattct
ttcggcatttagcactatttcagatcccagcagatagcagggtatccctttatcttt
agatcagcaggcaataaatgtctatgctgagcataatgaattgccagtggtgctcatat
tttaatgactgtatctttacactcagcagatattcttggatgaggcacataagaagc
gagtaaaaaatagaaatctatatgcatttggcaaaaatcaagataaatgatgctca
tactctcactctttgagtaaggccgocggtogaatgctgggcttttttggtaagtg
taattcacctcaaaaggaaagcatgatgaagaaccttactcatgcagaacctatttcat
catgctgtaaaaaagggaactgcaaaagagacttttagataagaacaatgaactattgga
```

※ 配列名には”_”を入れることが可能であるので、スペース入りの配列名を使用したい場合はスペースを”_”で代用すると、“>”の後に続く「最初のスペースまでの文字列」が以降に使用する配列解析ソフト等で配列名として認識される。

（例えば、*Acinetobacter* sp. NBRC 110496 という菌株の名前を配列名にしたい場合、「Acinetobacter_sp._NBRC_110496」とする。）

5.4.2 公開済みのゲノム配列データを利用する場合

国際塩基配列データベース（DDBJ/EMBL-Bank/GenBank）に登録されているゲノム配列データについて、DDBJ もしくは GenBank/NCBI の塩基配列データベースで検索して配列を入手する。以下に、両データベースからのゲノム配列データの入手法を解説する。

5.4.2.1 DDBJ 塩基配列データベースからのゲノム配列データの入手

(1) DDBJ の塩基配列検索ページ「getentry」にアクセスする。

(URL : <http://getentry.ddbj.nig.ac.jp/top-j.html>)

(2) 【出力形式】および【取得方法】のプルダウンメニューで【全塩基配列 FASTA】および【テキスト】をそれぞれ選択し、【ID】のテキストボックスに Accession 番号 (例 CP000521)

を入力後にエンターキーを押して配列をダウンロードする。

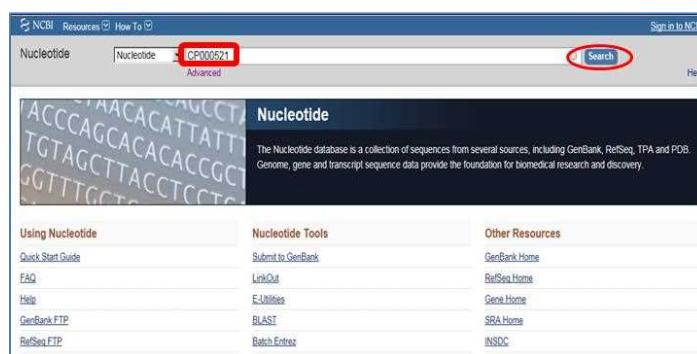
※ドラフト配列の場合は、コンティグ毎に配列が登録されており、複数の配列をダウンロードする必要があるが、例えば各コンティグの Accession が APQI01000001 -APQI01000012 だとすると、APQI01000000 と入力してエンターキーを押せば全コンティグの配列を一括してダウンロードすることができる。ただし、コンティグ数が 10 個を超える場合は、取得するデータ件数の上限値（上限）をコンティグ数に合わせて変更する必要がある。



The screenshot shows the 'getentry' web application interface. At the top, there is a search bar with the text 'アクセッション番号等によるエントリ検索'. Below the search bar, the 'ID:' field contains 'CP000521'. The 'DNA データベース:' section has radio buttons for 'DDBJ / GenBank / EMBL / MGA' and '出力形式:' set to '全塩基配列 FASTA'. The 'Protein データベース:' section has radio buttons for 'UniProt / PDB / DAD / Patent' and '出力形式:' set to 'default'. The '取得方法:' is set to 'テキスト' and the '上限:' is '10' items.

5.4.2.2 NCBI 塩基配列データベース GenBank からのゲノム配列データの入手（完全長配列の場合）ドラフト配列の場合は 5.4.2.3 を参照する。

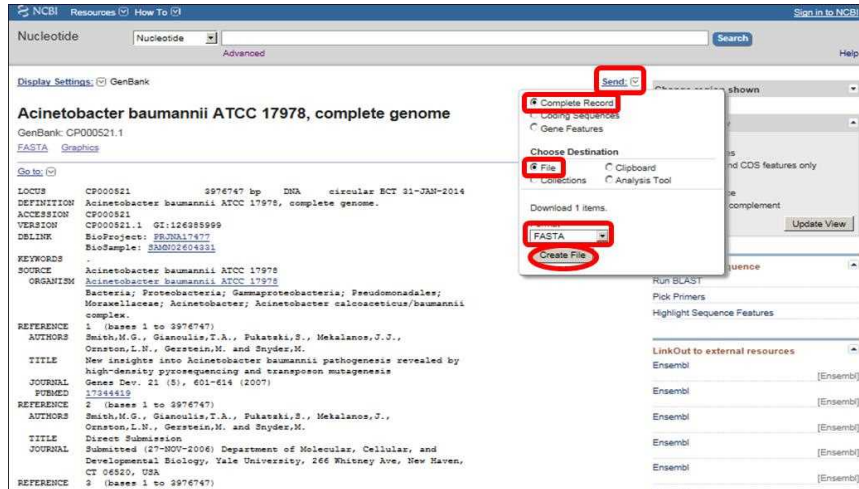
(1) NCBI の Nucleotide database (URL : <http://www.ncbi.nlm.nih.gov/nucleotide/>) にアクセスし、テキストボックスに Accession 番号（例 CP000521）を入力して【Search】ボタンをクリックして検索を実行する。



The screenshot shows the NCBI Nucleotide database search page. The search ID 'CP000521' is entered in the search box, and the 'Search' button is highlighted. The page shows the 'Nucleotide' search results and various navigation links.

(2) 右上の【Send:】のボタンをクリックしてドロップダウンメニューを開き、ファイル形式として【Complete Record】を、出力先として【File】を、それぞれのオプションボタンで選択する。

(3) さらに、出力形式 (Format) のプルダウンメニューで【FASTA】を選択し、【Create File】ボタンをクリックして配列をダウンロードする。



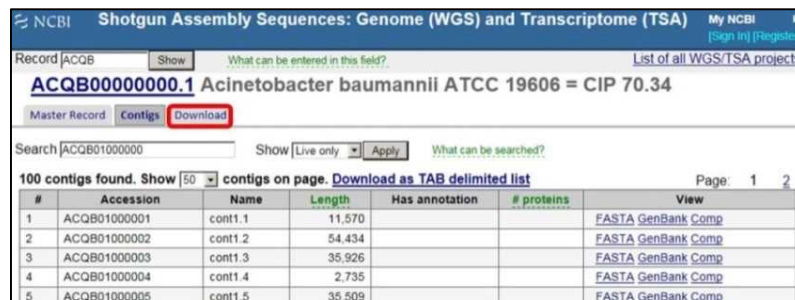
5.4.2.3 NCBI 塩基配列データベース GenBank からのゲノム配列データの入手（ドラフト配列の場合）完全長配列の場合は 5.4.2.2 を参照する。

(1) NCBI の Nucleotide database (URL : <http://www.ncbi.nlm.nih.gov/nucleotide/>) にアクセスし、テキストボックスに Accession 番号（例 ACQB01000000）を入力して【Search】ボタンをクリックして検索を実行する。

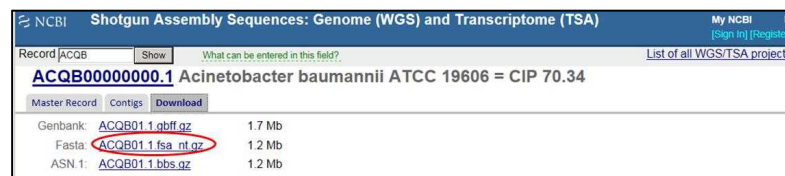
【+Lhaca 等の圧縮・解凍ソフトウェアを使って.gz ファイルの解凍が可能な場合】

(+Lhaca 等の圧縮・解凍ソフトウェアが利用できない場合は手順 5.4.2.3 (5)へ)

(2) 【Download】タブをクリックして表示ページを切り替える。



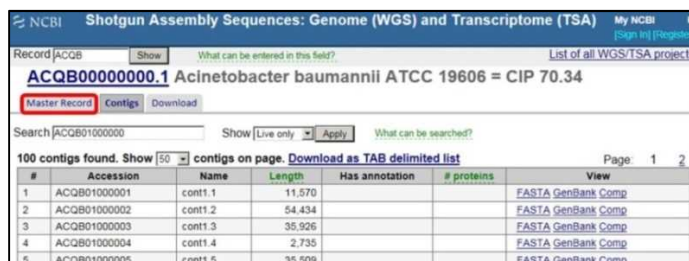
(3) FASTA 形式ファイル（図では、ACQB01.1.fsa_nt.gz）をクリックし、配列をダウンロードする。



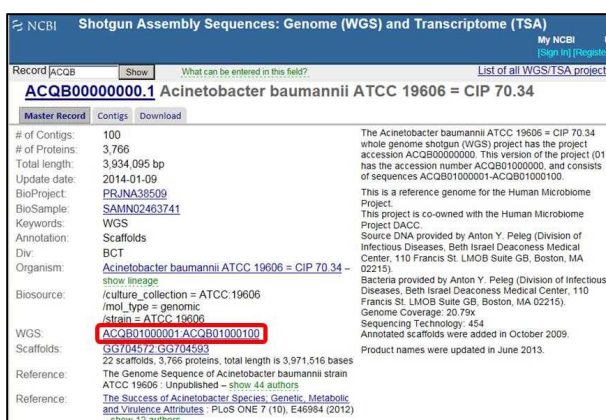
(4) ダウンロードした配列は圧縮・解凍ソフトウェアを用いて解凍する。

【+Lhaca 等の圧縮・解凍ソフトウェアが利用できない場合】

(5) 手順 5.4.2.3 (1)の後に、「Master Record」タブをクリックして表示ページを切り替える。

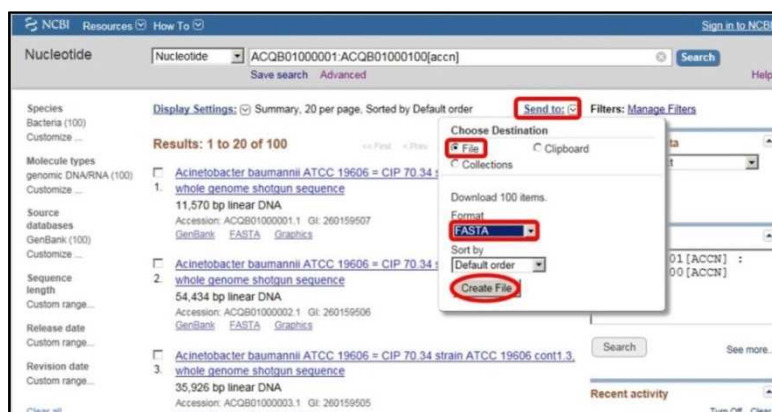


(6) WGS:欄のコンティグ番号 (例 : ACQB01000001:ACQB01000100) をクリックする。



(7) 右上の Send to:のボタンをクリックしてドロップダウンメニューを開き、出力先として【File】をオプションボタンで選択する。

(8) さらに、出力形式 (Format) のプルダウンメニューで【FASTA】を選択し、【Create File】ボタンをクリックして配列をダウンロードする。



※sequence.fasta という名前で作成されるので、必要に応じてファイル名を変更する。

第 6 章 指標遺伝子の塩基配列のトリミング処理

遺伝子ごとに塩基配列のアラインメントを作成し、アラインメント中で挿入や欠損のある場所を削除し、配列の両端の長さを合わせる。配列のトリミング処理は MEGA などの遺伝子塩基配列解析ソフトを用いても可能であるが、ここでは Gblocks Server でトリミング処理を行う場合について紹介する。

6.1 *Acinetobacter* 属細菌基準株の指標遺伝子塩基配列 (Multi-FASTA ファイル) の用意

表 4. に示した Accession 番号を用いて *Acinetobacter* 属細菌基準株の遺伝子の塩基配列をダウンロードして (配列のダウンロードについては 5.2 を参照のこと)、遺伝子ごとに 1 つのファイル (Multi-FASTA ファイル) にまとめ、配列の拡張子を「fasta」もしくは「fas」にして MEGA で読み込めるファイル形式にしておく。

表 4. *Acinetobacter* 属基準株の各指標遺伝子塩基配列の Accession 番号

学名	<i>cpn60</i>	<i>gltA</i>	<i>gyrB</i>	<i>pyrG</i>	<i>rpoB</i>	<i>rpoD</i>
<i>Acinetobacter baumannii</i>	LC102559	LC102587	LC102615	LC102643	LC102671	LC102699
<i>Acinetobacter baylyi</i>	LC102560	LC102588	LC102616	LC102644	LC102672	LC102700
<i>Acinetobacter beijerinckii</i>	LC102561	LC102589	LC102617	LC102645	LC102673	LC102701
<i>Acinetobacter bereziniae</i>	LC102562	LC102590	LC102618	LC102646	LC102674	LC102702
<i>Acinetobacter bouvetii</i>	LC102563	LC102591	LC102619	LC102647	LC102675	LC102703
<i>Acinetobacter brisouii</i>	LC102564	LC102592	LC102620	LC102648	LC102676	LC102704
<i>Acinetobacter calcoaceticus</i>	LC102565	LC102593	LC102621	LC102649	LC102677	LC102705
<i>Acinetobacter gerneri</i>	LC102566	LC102594	LC102622	LC102650	LC102678	LC102706
<i>Acinetobacter guillouiae</i>	LC102567	LC102595	LC102623	LC102651	LC102679	LC102707
<i>Acinetobacter gyllenbergii</i>	LC102568	LC102596	LC102624	LC102652	LC102680	LC102708
<i>Acinetobacter haemolyticus</i>	LC102569	LC102597	LC102625	LC102653	LC102681	LC102709
<i>Acinetobacter indicus</i>	LC102570	LC102598	LC102626	LC102654	LC102682	LC102710
<i>Acinetobacter johnsonii</i>	LC102571	LC102599	LC102627	LC102655	LC102683	LC102711
<i>Acinetobacter junii</i>	LC102572	LC102600	LC102628	LC102656	LC102684	LC102712
<i>Acinetobacter lwoffii</i>	LC102573	LC102601	LC102629	LC102657	LC102685	LC102713
<i>Acinetobacter nosocomialis</i>	LC102574	LC102602	LC102630	LC102658	LC102686	LC102714
“ <i>Acinetobacter oleivorans</i> ”	LC102575	LC102603	LC102631	LC102659	LC102687	LC102715
<i>Acinetobacter parvus</i>	LC102576	LC102604	LC102632	LC102660	LC102688	LC102716
<i>Acinetobacter pittii</i>	LC102577	LC102605	LC102633	LC102661	LC102689	LC102717
<i>Acinetobacter radioresistens</i>	LC102578	LC102606	LC102634	LC102662	LC102690	LC102718
<i>Acinetobacter rudis</i>	LC102579	LC102607	LC102635	LC102663	LC102691	LC102719

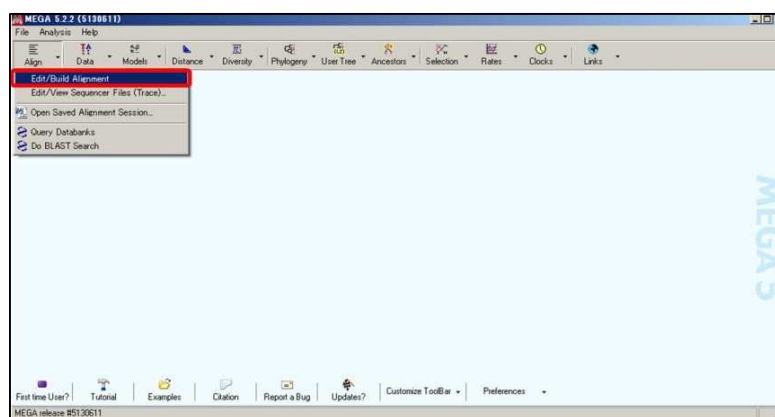
表 4. (つづき)

<i>Acinetobacter schindleri</i>	LC102580	LC102608	LC102636	LC102664	LC102692	LC102720
<i>Acinetobacter soli</i>	LC102581	LC102609	LC102637	LC102665	LC102693	LC102721
<i>Acinetobacter tandoii</i>	LC102582	LC102610	LC102638	LC102666	LC102694	LC102722
<i>Acinetobacter tjernbergiae</i>	LC102583	LC102611	LC102639	LC102667	LC102695	LC102723
<i>Acinetobacter towneri</i>	LC102584	LC102612	LC102640	LC102668	LC102696	LC102724
<i>Acinetobacter ursingii</i>	LC102585	LC102613	LC102641	LC102669	LC102697	LC102725
<i>Acinetobacter venetianus</i>	LC102586	LC102614	LC102642	LC102670	LC102698	LC102726

6.2 マルチプルアラインメントの実行

遺伝子ごとにマルチプルアラインメントを実行する。ここでは **Muscle** を使ってアラインメントを作成する方法を例に挙げて説明する。塩基配列の方がアミノ酸配列より変異が多いため、塩基配列で系統解析を行う方が解像度の高い系統樹が得られる。一方、コドンの位置を考慮したアラインメントを得るためにはアミノ酸配列を用いる方が望ましい。そのため、一度塩基配列をアミノ酸配列に翻訳してアラインメントを作成した後に、得られたアラインメントを塩基配列に戻して系統解析を行う。

- (1) MEGA を起動しメインウィンドウから **【Align】** ボタン— **【Edit/Build Alignment】** をクリックする。

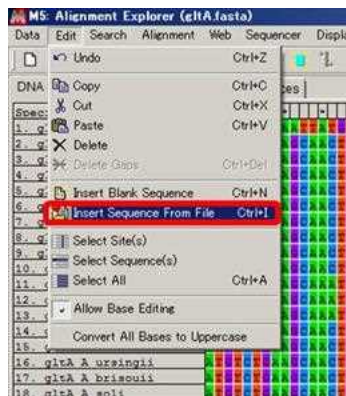


- (2) **【Alignment Editor】** のウィンドウが開くので、**【Retrieve sequences from a file】** を選択して **【OK】** ボタンをクリックする。

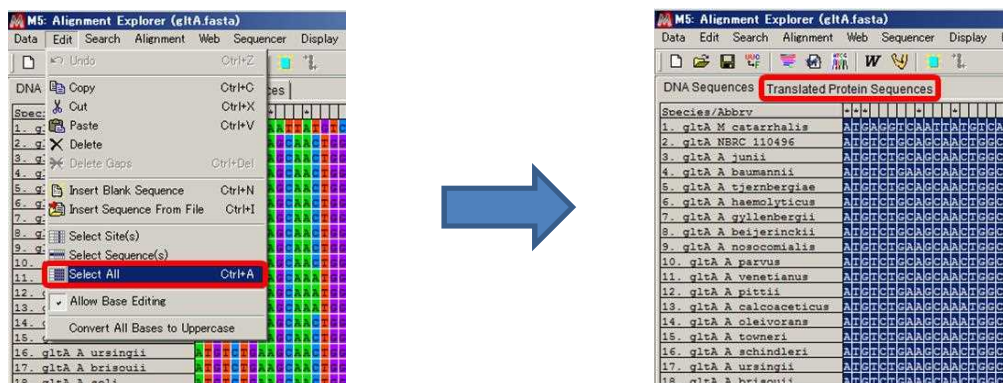


(3) 開くファイル (5.1 でダウンロードしたファイル) を選択すると【Alignment Explorer】のウィンドウが開く。以降の作業は、【Alignment Explorer】上で行う。

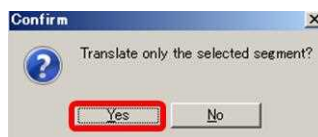
(4) メニューバーの【Edit】から【Insert Sequence From File】をクリックして、第5章 5.3 で取得した遺伝子塩基配列のファイルを選択して開く



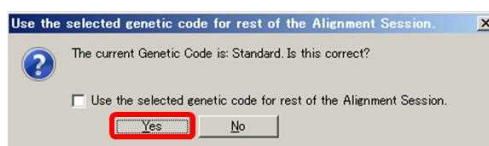
(5) メニューバーの【Edit】から【Select All】をクリックして全配列を選択し、【Translated Protein Sequences】をクリックして配列表示形式をアミノ酸配列に切り替える。



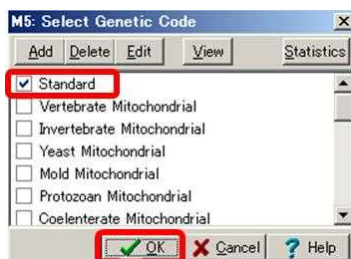
(6) Confirm 画面が立ち上がって、「Translate only the selected segment?」と聞かれるので、【Yes】ボタンをクリックする。



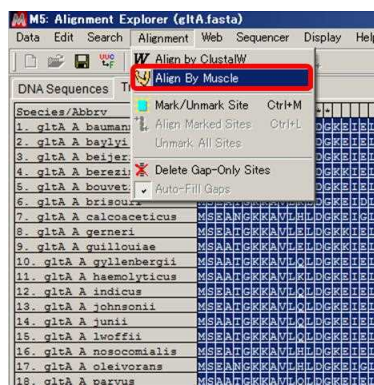
(7) コドン表について聞かれるので、図のように既に【Standard】に設定されている場合は【Yes】ボタンをクリックすると、塩基配列がアミノ酸配列に翻訳される。



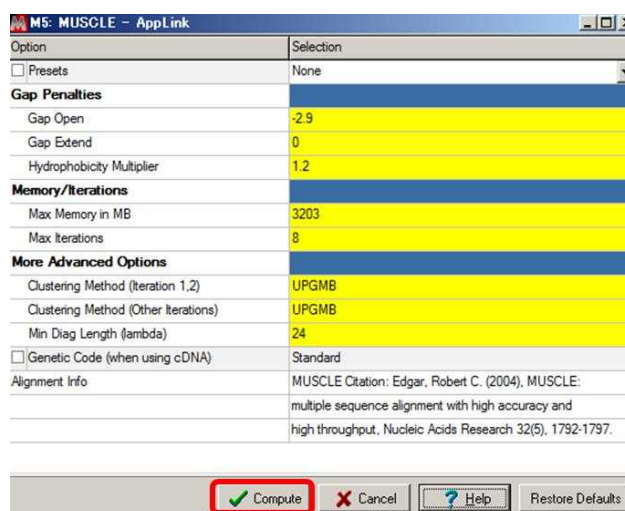
※ここで、コドン表が【Standard】になっていない場合は、【No】 ボタンをクリックして、コドン表選択ウィンドウで【Standard】 を選択して【OK】 ボタンをクリックする。



(8) 全配列を選択したままで、メニューバーの【Alignment】 から【Align by Muscle】 をクリックする。

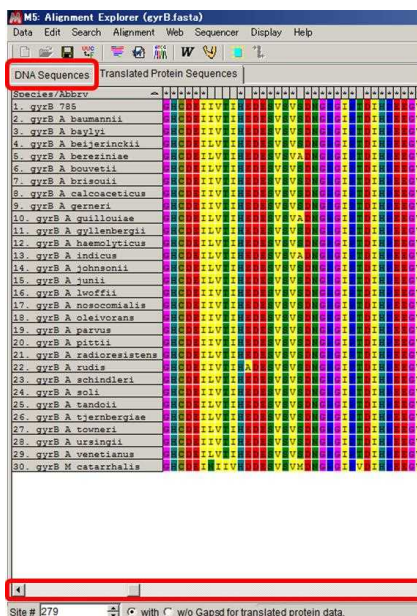


(9) パラメータ設定ウィンドウが立ちあがるので、適宜パラメータを変更し（デフォルトの設定でも構わない）、【OK】 ボタンをクリックすると解析が開始する。

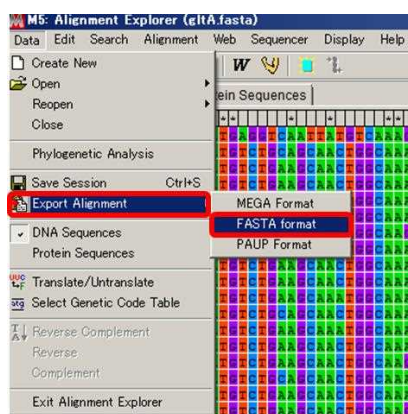


(10) 解析が終了したら、【Alignment Explorer】の画面下部にあるスクロールバーのノブを動かして、アラインメント全体を確認する。解析株の配列だけが他の基準株の配列と大きく異なっているなどの問題がなければ、配列タイプのタブを【DNA Sequences】に切り替える。

※ここで、問題がある場合は第5章5.3で取得した塩基配列が適切でない可能性がある。



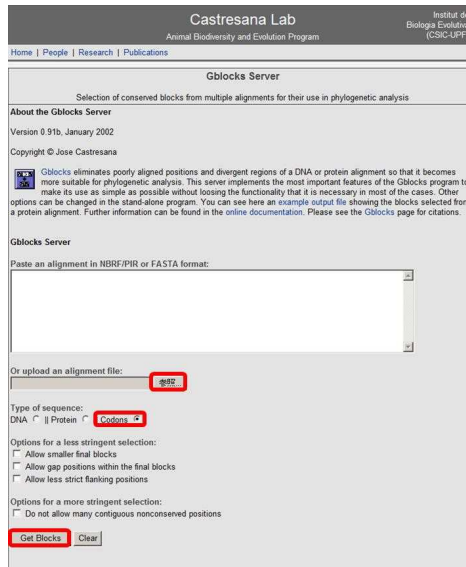
(11) メニューバーの「【Data】から【Export Alignment】－【FASTA format】をクリックして、適当な名前を付けてアラインメントを保存する。（拡張子は「fas」となる）



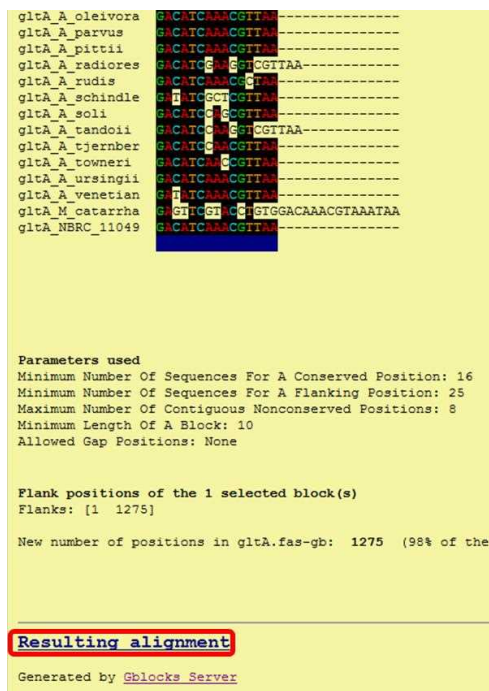
6.3 Gblocks Server によるトリミング処理

Gblocks Server を使ってアラインメント中の挿入や欠損のある場所を削除し、塩基配列の両端の長さを切り揃える

- (1) Gblocks Server (http://molevol.cmima.csic.es/castresana/Gblocks_server.html)にアクセスする。
- (2) 参照ボタンをクリックして 6.2 で作成したアラインメントファイルを選択してアップロードし、配列のタイプとして【Codons】のオプションボタンを選択して、【Get Blocks】のボタンをクリックして解析を開始する。



- (3) 結果が表示されるのを確認し、ページ下の【Resulting alignment】をクリックする。



gltA_oleivora
gltA_parvus
gltA_pittii
gltA_radios
gltA_rudis
gltA_schindle
gltA_soli
gltA_tandoi
gltA_tjernber
gltA_towneri
gltA_ursingii
gltA_venetian
gltA_M_catarrha
gltA_NBRC_11049

Parameters used
Minimum Number Of Sequences For A Conserved Position: 16
Minimum Number Of Sequences For A Flanking Position: 25
Maximum Number Of Contiguous Nonconserved Positions: 8
Minimum Length Of A Block: 10
Allowed Gap Positions: None

Flank positions of the 1 selected block(s)
Flanks: [1 1275]

New number of positions in gltA.fas-gb: 1275 (98% of the ...)

Resulting alignment

Generated by Gblocks Server

(4) トリミング処理後の配列を保存する。(テキスト形式で保存する。)

```
>gltA_baumannii
ATGCTGAAG CAACTGGCAA AAAAGCOGTA TTACATCTTG ATGGCAAAGA AATTGAATTA
CCAATTTACA GTGGCACATT AGGTCOCGAT GTAATCGACG TTAAGATGT ATTGGCCTCA
GGTCACITTA CTTTTGATCC TGGTTTTATG GOGACAGCTT CATGCGAGTC TAAAATCACA
TTTATCGATG GTGACAAAAGG TATTTTATTA CACCCGGGTT ACCCGATTGA CCAAGTTAGCG
ACTCAAGCAG ACTACCTTGA AACCTTGTAT TTATTATTAA ATGGCGAGTT ACCAACTGCT
GAACAAAAGG TTGAGTTGGA TGCGAAAGT CGTGCTCATA CTATGGTTCA TGATCAAGTT
AGCCGTTTCT TCAATGGTIT CCGTCGTGAT GCTCACCCCTA TGCCAATCAT GGTGGTGTG
GTAGGCGCAT TATCTGCTTT CTATCACAA C AACCTTGACA TTGAAGACAT CAACCACOCG
GAAATTA CTG CGATTCGTTT GATTGCTAAA ATTCCAACGC TTGCTGCTTG GAGCTACAAA
TATACTGTAG GTCAGCCATT CATCTATCCA CGTAATGACT TAAATTACGC GAAAACCTC
TTACACATGA TGTITGCAAC TCCTGCAGAC CGTGA CTACA AAGTAAACCC TGTCTTGT
CGTGCAATGG ATCGTATCTT TACGCTTAC GCTGACCACG AACAAAACGC GTCTACTTCT
ACAGTTCGTC TTGCTGGTTC TACTGGTGCG AATCCATATG CGTGTATCTC TGCTGGTATC
TCTGCTCTTT GGGGTCTGCG ACAOGGTGGT GOGAACGAA G CAGTTCTTAA AATGCTTGAT
GAAATCGGTA GCGTTGAAAA TGTGCTGAG TTCATGGAAA AAGTTAAACG CAAAAGAGTT
AAACTTATGG GCTTCGGTCA CCGGTTTAC AAAA ACTTCG ATCCACGCGC TAAAGTGATG
AAGCAAACTT GTGACGAAGT TCTTGAAGCA TTAGGTATCA ATGATCCTCA ATTAGCGCTT
GCTATGGAAC TTGAACGTAT TGCATTGAAC GACCCGTA CT TTGTTGAACG TAAACTTTAC
CCTAACGTAG ACTTCTACTC TGGTATCATC CTTAAAGCGA TTGGTATCCC AACAGAAATG
TTTACCGTTA TCTTGCCTCT TGCACGTACA GTTGGCTGGA TCAGTCACTG GTTAGAAATG
CACAGCGGTC CTTACAAAAT TGGTCGTCTT CGTCAGCTTT AACTGGTGA AGTGCAAAGT
GACATCAAGC GTTAA
```

※ここでは、オンラインツールである Gblocks Server を利用する手順について解説したが、Gblocks program は下記のページよりダウンロードしてローカル環境で使用可能である。

Gblocks program ダウンロードページ : <http://molevol.cmima.csic.es/castresana/Gblocks.html>

第7章 6つの指標遺伝子の塩基配列の連結

同じ菌株の指標遺伝子の塩基配列を *cpn60-gltA-gyrB-pyrG-rpoB-rpoD* の順に連結して1つの配列にして、全解析株の連結配列から構成される Multi-FASTA ファイルを作成する。各配列のヘッダ部分は、">(解析株名)"とする。ただし、連結する遺伝子塩基配列は、全菌株間で遺伝子ごとに長さが揃ったものとし、揃っていない場合は解析手順②（第3章の解析の全体概要参照、詳細は第6章に記述）に戻り必要なトリミング処理を行い配列の長さを揃える。

以下に、配列の連結方法の1つとして、Microsoft Excel の機能を利用した連結方法を紹介する。

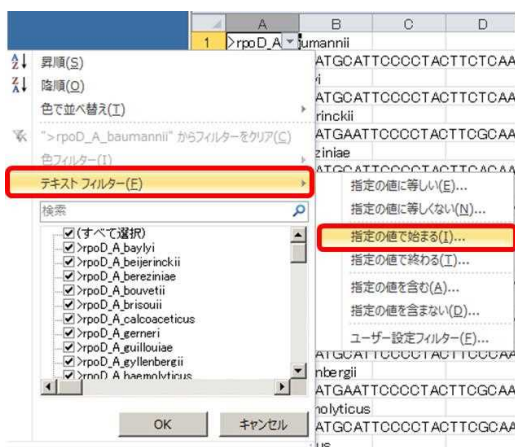
7.1 連結用指標遺伝子の塩基配列ファイル（Microsoft Excel 形式）の用意

A 列にヘッダ（">菌株名"）、B 列に遺伝子塩基配列を貼り付けた Excel 形式のファイルを用意する（下図参照）。以下にその手順を説明する。

	A	B	C	D	E	F	G	H	I																
1	>M_catarrhalis	GC	AAA	AGAT	GT	AAG	TTT	GGCT	CAAT	GC	CC	CG	GAAAA	AAT	GAT	TG	AT	GGT	GTT	AAC	ATTT				
2	>A_venetianus	GCT	AA	GAC	GT	AAAA	TTT	GGT	GAT	TC	AG	TC	GC	TC	AAAA	AAT	GAT	TG	CG	AGG	CGT	AA	AC	GT	AC
3	>A_ursingii	GCT	AA	GAC	GTT	AA	AT	TC	GGT	GAT	TC	AG	CG	CGT	TC	AAAA	AAT	GAT	CG	AGG	GT	CA	AT	GTT	C
4	>A_towneri	GCT	AA	GAC	GT	AAAA	TTT	GGT	GAT	TC	AG	TC	GC	CAC	AAAA	AAT	GAT	CG	AGG	GT	AA	AC	GT	AC	
5	>A_tjernbergiae	GCT	AA	GAC	GT	AAAA	TTT	GGT	GAT	TC	AG	TC	GC	TC	AAAA	AAT	GAT	CG	AGG	CGT	AA	AC	GT	AC	
6	>A_tandoii	GCT	AA	GAT	GT	AAAA	TTT	GGT	GAT	TC	AG	TC	GC	GC	AAAA	AAT	GAT	TG	CG	AGG	CGT	AA	CG	TT	C
7	>A_solii	GCT	AA	GAC	GT	AAAA	TC	GGT	GAT	TC	AG	CG	CGT	TC	AAAA	AAT	GAT	CG	AGG	GT	AA	AC	GT	TC	
8	>A_schindleri	GCT	AA	GAC	GT	AAAA	TTT	GGT	GAT	TC	AG	TC	CGT	GC	GAAAA	AAT	GAT	CG	TGGT	GT	GA	AC	AT	CC	
9	>A_rudis	GCT	AA	GAC	GT	AAAA	TTT	GGT	GAT	TC	GC	AC	GT	TC	AAAA	AAT	GAT	CG	AGG	CGT	AA	AC	GT	AC	
10	>A_radioresistens	GCT	AA	GAC	GT	AAAA	TC	GGT	GAT	TC	AG	CC	CGT	AC	AAAA	AAT	GAT	TGC	TGGT	GT	AA	AC	ACT	C	

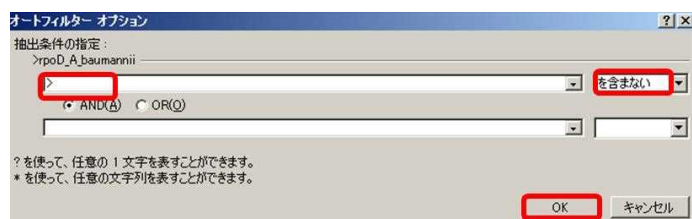
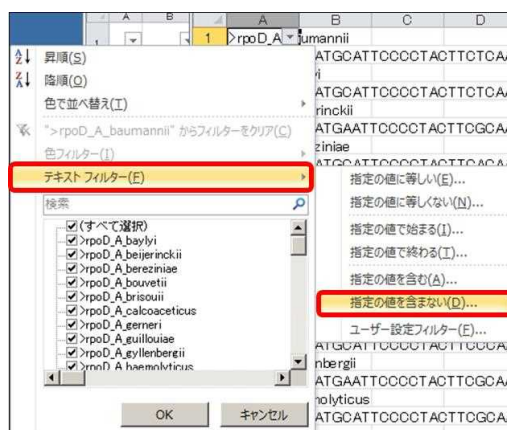
- 解析手順②（全体概要参照）でトリミング処理を行った後の Multi-FASTA ファイル (.txt ファイル) の拡張子を「fasta」もしくは「fas」に変更して MEGA で読み込めるファイル形式にする。
- MEGA を起動し、メインウィンドウから【Align】ボタン－【Edit/Build Alignment】をクリックする。
- 【Alignment Editor】のウィンドウが開くので、【Retrieve sequences from a file】を選択して【OK】ボタンをクリックする。
- 開くファイル（(1)で拡張子を変更したファイル）を選択すると【Alignment Explorer】のウィンドウが開く。
- 【Alignment Editor】メニューバーの【Data】から【Export Alignment】－【FASTA format】をクリックして、再度ファイルを FASTA 形式で保存する。
- エクセルを起動し、(5)の FASTA ファイルを開く（【Ctrl】キー＋【O】キーを押して、開くファイルを選択する）。
- A 列にフィルターを設定する（A 列を選択して【Ctrl】キー＋【Shift】キー＋【L】キーを押す）。
- A 列のオートフィルターメニューから【テキストフィルター】→【指定の値で始まる】

をクリックしてオートフィルターオプションのダイアログを開き、テキストボックスに“>”を入力して【OK】ボタンをクリックして、ヘッダ行だけを表示させる。



(9) A 列の全行を選択してコピーし、別シートの A 列に貼り付ける。

(10) 次に、A 列のオートフィルターメニューから【テキストフィルター】－【指定の値を含まない】をクリックしてオートフィルターオプションのダイアログを開き、テキストボックスに“>”を入力して【OK】ボタンをクリックして、配列行だけを表示させる。



- (11) A 列の 2 行目以降を選択してコピーし、(9)で作成したシートの B 列に貼り付ける。
- (12) (9)で作成したシートで、A 列の遺伝子名と”_”を全て削除する。(A 列を全行選択して【Ctrl】キー+【H】キーを押して【検索と置換】のダイアログボックスを開き、【検索する文字列】に遺伝子名と”_”(例：rpoD_)を入力し、【置換後の文字列】には何も入力せずに【すべて置換】をクリックして一斉に遺伝子名を削除する。)



- (13) (12)の A 列、B 列のデータをコピーして、新しいエクセルファイルに貼り付け、シート名を遺伝子名にしておく
- (14) 他の遺伝子についても(1)~(12)の作業を行い、(13)で作成したものと同一エクセルファイルの別シート (シート名は遺伝子名と同じにする) にできあがったデータを貼り付ける。

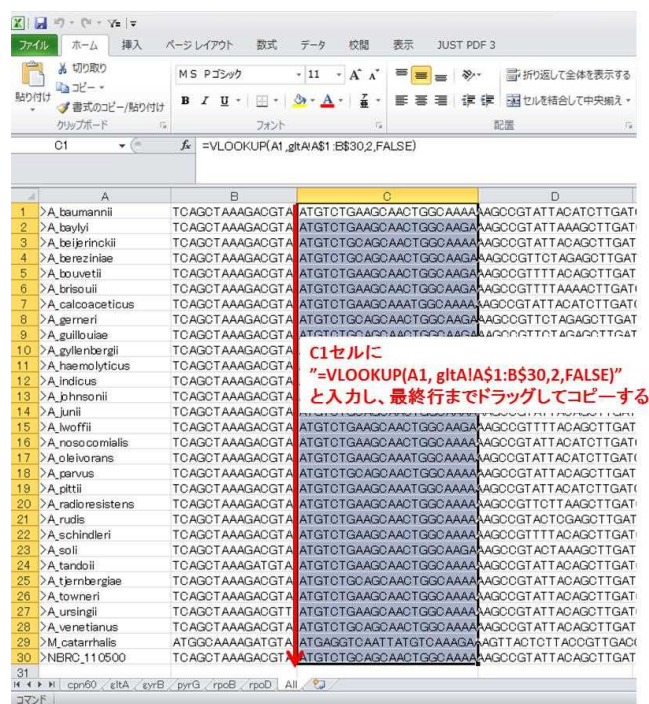
24	>A_tandouii	TCAGCTAAAGATGTAAAATTTGGTGATTCAGCTCGCGCAAAAATGATTGCAGGCGTTA#
25	>A_tjernbergiae	TCAGCTAAAGACGTAAAATTTGGTGATTCAGCTCGCTCAAAAATGATCGCAGGCGTAA#
26	>A_towneri	TCAGCTAAAGACGTAAAATTTGGTGATTCAGCTCGCACAAAATGATCGCAGGCGTAA#
27	>A_ursingii	TCAGCTAAAGACGTTAAAATTCGGTGATTCAGCGCGTTCAAAATGATCGCAGGCGTCA#
28	>A_venetianus	TCAGCTAAAGACGTAAAATTTGGTGATTCAGCTCGCTCAAAAATGATTGCAGGCGTAA#
29	>M_catarrhalis	ATGGCAAAAGATGTAAATTTGGCTCAAATGCCGCGCAAAAATGATTGATGGTGTAA#
30	>NBRC_110500	TCAGCTAAAGACGTAAAATTTGGTGATTCAGCGCGCTCAAAAATGATTGCAGGCGTAA#
31		

コマンドバー: cpn60 / gltA / syrB / pyrG / rpoB / rpoD

7.2 塩基配列ファイルの連結

7.1 で作成した配列ファイルを連結する。

- (1) 7.1 で作成した遺伝子塩基配列ファイルについて、シート「cpn60」で全データを選択して (【Ctrl】キー+【A】キーを押す) コピーし (【Ctrl】キー+【C】キーを押す)、新しい別シートを作って (以下、連結データ用シートという。)、貼り付ける (【Ctrl】キー+【V】キーを押す)。
- (2) 連結データ用シートの C1 セルで”=VLOOKUP(A1,gltA!A\$1:B\$30,2,FALSE)”と入力してエンターキーを押す (菌株数が全部で 30 の場合)。
- (3) 連結データ用シートの C1 セルを選択してそのまま 30 行目 (最終行目) までドラッグしてコピーする。これで C1 セルに *gltA* の塩基配列が入力される。



(4) 連結データ用シートの D1 セルで”=VLOOKUP(A1, gyrB!A\$1:B\$30,2,FALSE)”と入力してエンターキーを押す（菌株数が 30 の場合）。

※VLOOKUP 関数の詳細については他の参考書等を参照のこと。

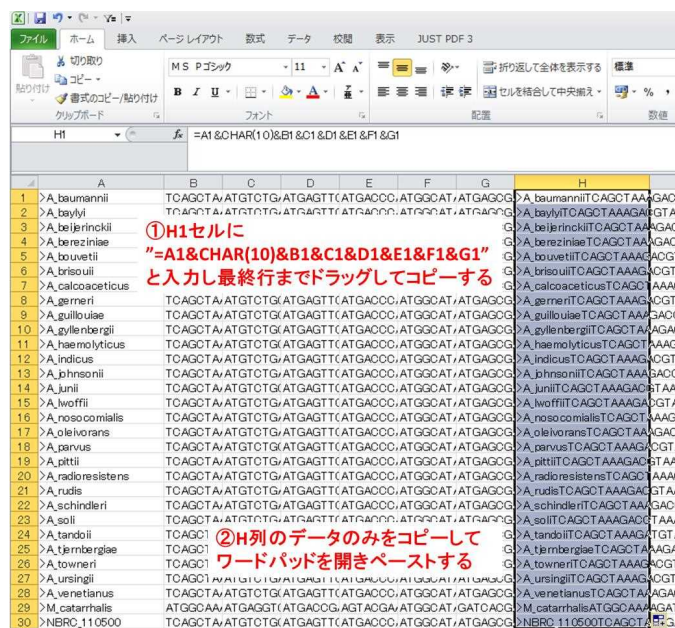
(5) 連結データ用シートの D1 セルを選択してそのまま 30 行目（最終行目）までドラッグしてコピーする。これで D1 セルに *gyrB* の塩基配列が入力される。

(6) 同様に、連結データ用シートの E 列に *pyrG* を、F 列に *rpoB* を、G 列に *rpoD* の塩基配列を入力する。

(7) 連結データ用シートの C 列～G 列に塩基配列データを埋め終わったら、H1 セルに”=A1&CHAR(10)&B1&C1&D1&E1&F1&G1”と入力してエンターキーを押す。

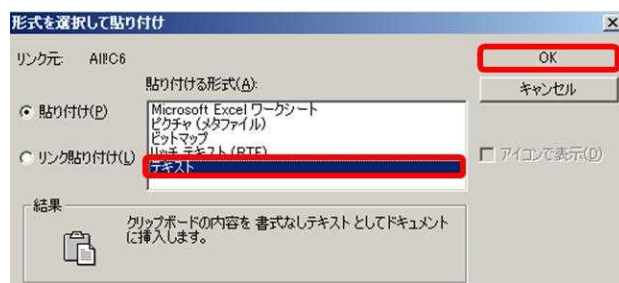
※「CHAR(10)」は改行の意味。菌株名の後に改行が来るようにしている。

(8) 連結データ用シートの H1 セルを選択してそのまま 30 行目（最終行目）までドラッグしてコピーする。これで、H 列に各菌株の連結塩基配列が入力されるので、H 列のデータのみをコピーする。



(9) ワードパッドを開く (【スタート】メニューから【すべてのプログラム】 - 【アクセサリ】 - 【ワードパッド】をクリックする)。

(10) 【Alt】キー+【Ctrl】キー+【V】キーを押して、貼り付けの形式として【テキスト】を選択して【OK】ボタンをクリックし、コピーしたデータを貼り付ける。



(11) ワードパッド上で【Ctrl】キー+【H】キーを押し、【検索と置換】ダイアログで【検索する文字列】に「」を入力、【置換する文字列】には何も入力せず、【すべて置換】をクリックする。ファイルの種類をテキストドキュメント (*.txt) にして、拡張子を「fasta」にした任意の名前でファイルを保存する。

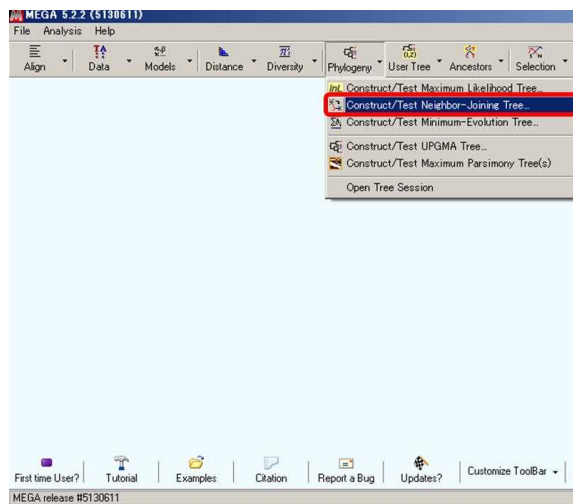


第 8 章 分子系統解析

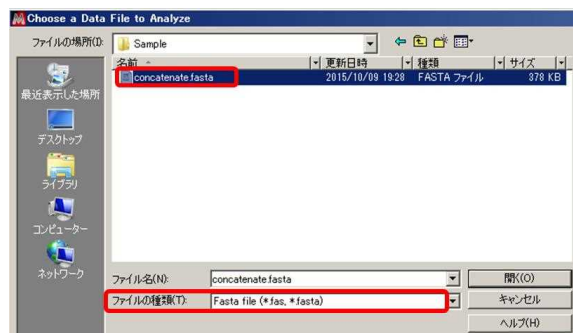
8.1 系統樹の作成

連結塩基配列を用いて近隣結合 (Neighbor-Joining: NJ) 法で分子系統樹を作成する。

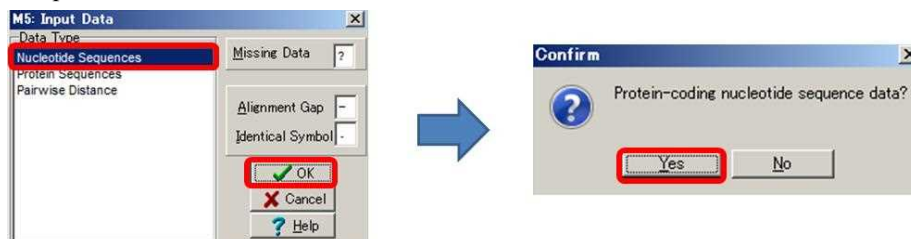
(1) MEGA を起動し、メインウィンドウの【Phylogeny】から【Construct/Test Neighbor-Joining Tree】をクリックする。



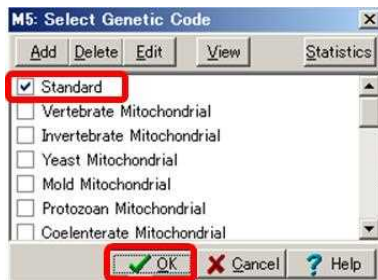
(2) ファイルの種類として【Fasta file (*.fas, *.fasta)】を選択して、開くファイル (第 7 章で作成した連結配列) を選択する。



(3) 選択した配列の種類 (塩基配列/アミノ酸配列) を聞かれるので、【Nucleotide Sequences】を選択して【OK】ボタンをクリックすると、Confirm 画面が立ち上がり、「Protein-coding nucleotide sequence data?」と聞かれるので、【Yes】ボタンをクリックする。



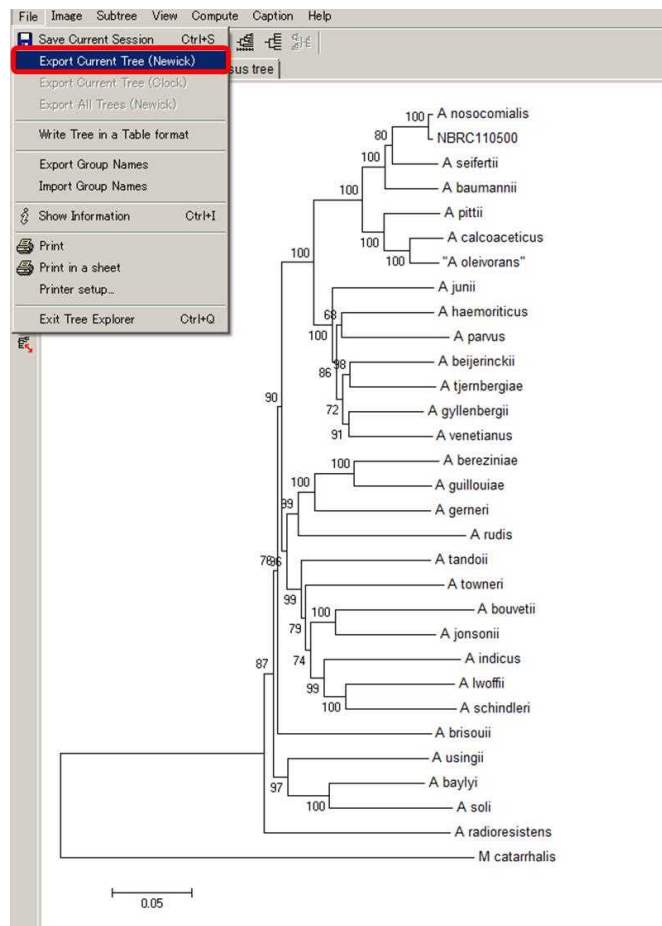
(4) コドン表が出てくるので【Standard】を選択して【OK】ボタンをクリックする。



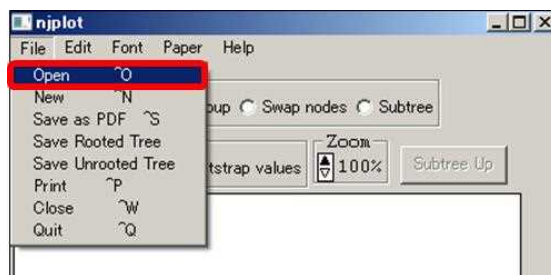
(5) パラメータ選択画面が開くので【Test of Phylogeny】は【Bootstrap method】（通常は1000に設定）を選択し、【Select Codon Positions】の【1s】から【3rd】までチェックを入れて、その他のパラメータも適宜変更して【Compute】ボタンをクリックする。



(6) 【Tree Explorer】が立ち上がって系統樹が表示されるので、【Tree Explorer】のメニューバーの【File】から【Export Current Tree (Newick)】をクリックしてファイル名を付けて系統樹を保存する。(Newick フォーマットで系統樹が保存される。拡張子は「nwk」となる。)



(7) (オプション) NJplot で系統樹を開く場合は、NJplot を立ち上げ、メニューバーの【File】から【Open】をクリック、保存しておいた系統樹ファイル (nwk ファイル) を開く。



(8) Display のメニューの中から【Bootstrap value】のボックスにチェックを入れるとブートストラップ値が表示される。

※ブートストラップ値は MEGA では%で表示されるため、NJplot で系統樹を描画した場合には値が 1 以下で表示される。この場合、各値に 1000 を掛けた値が本来のブートストラップ値となる。

※系統解析の結果の見方については以下に挙げる参考書を参照のこと。

1. 宮田隆 (1998). 分子進化—解析の技法とその応用 共立出版
2. 根井正利、クマー, S (2006). 分子進化と分子系統学 培風館
3. 斎藤成也 (2007). ゲノム進化学入門 CD-ROM 付 共立出版
4. バイオメトリックス 第 12 回 系統樹 東京大学 バイオメトリックス (応用生物学専修) 講義関連資料 URL: http://lecture.ecc.u-tokyo.ac.jp/~aiwata/biomet/biometrics_lec12.pdf
5. 田邊晶史 分子系統学演習 - データセットの作成から仮説検定まで
URL: <http://www.fifthdimension.jp/documents/molphytextbook/molphytextbook.ja.pdf>

8.2 系統樹に基づき種同定

8.1 で描いた系統樹に基づき種同定を行う¹³。解析株が既知種の基準株とクラスターを形成し、かつ、そのクラスターの枝のブートストラップ値が 70%以上である場合はその既知種と近縁関係にあると考えられる。例えば NBRC 110500 株は、16S rRNA 遺伝子塩基配列の比較では、*A. nosocomialis* に 99.83%、*A. pittii* に 99.57%、*A. seifertii* に 99.48%の相同値を示すが、MLSA の結果 (前ページの系統樹) では、ブートストラップ値 100%で *A. nosocomialis* とクラスターを形成するので、*A. nosocomialis* に近縁であると判断できる。

¹³ 種は増えたり、再分類される可能性があるため、試験を行う前に LPSN (List of prokaryotic names with standing in nomenclature, <http://www.bacterio.net/>) や International journal of Systematic and Evolutionary Microbiology 誌上の Validation List で最新の情報を確認する必要がある