

# JAMSTEC & in silico biology's



Version 3.4/2.4

## Meta GenomeGAMBLER For Sequencing and Annotation

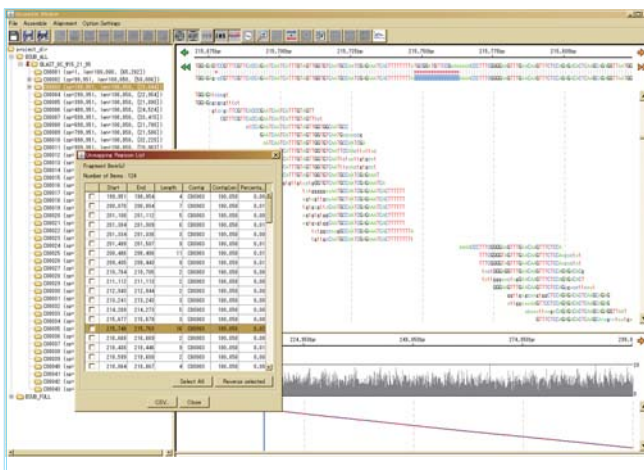
Version 3.4/2.4 for Windows 7 & Mac OS.X.6

### 従来型・新型シーケンサ配列をメタゲノム解析に

MetaGenomeGAMBLER (MGG) は、ゲノムシーケンシング・プロジェクトを支援するためのソフトウェアシステムです。そのうえ、ノート PC の環境でも充分動作するコンパクトな設計となっていますが、このようなメタゲノムプロジェクトを支援するために最小限必要な機能の多くを含んでいます。また、様々な種類の塩基配列データを同時に混在する今日、新型ギガシーケンサーからのデータや、既知のデータベース登録済み塩基配列などをも一緒に解析することができます。既知のゲノム塩基配列へのマッピング機能もプロフェッショナル版に新規搭載されました (マッピングを実行するにはインシリコアSEMBLERが必要です)。

#### MGG からできること

- キャピラリーシーケンサ配列の自動アセンブル機能
- 新型ギガシーケンサーからの塩基配列をマッピング
- 異なる形式の配列データを混在
- 塩基やフラグメントの品質を管理
- 塩基配列のマルチプルアラインメントを表示
- 変異箇所をリスト表示し、アラインメント表示可能
- トレース波形のマルチプルアラインメントが可能
- IMC を起動して、アノテーションを実行
- メタゲノムコンティグの自動分類
- 系統樹から、由来するコンティグを逆引き



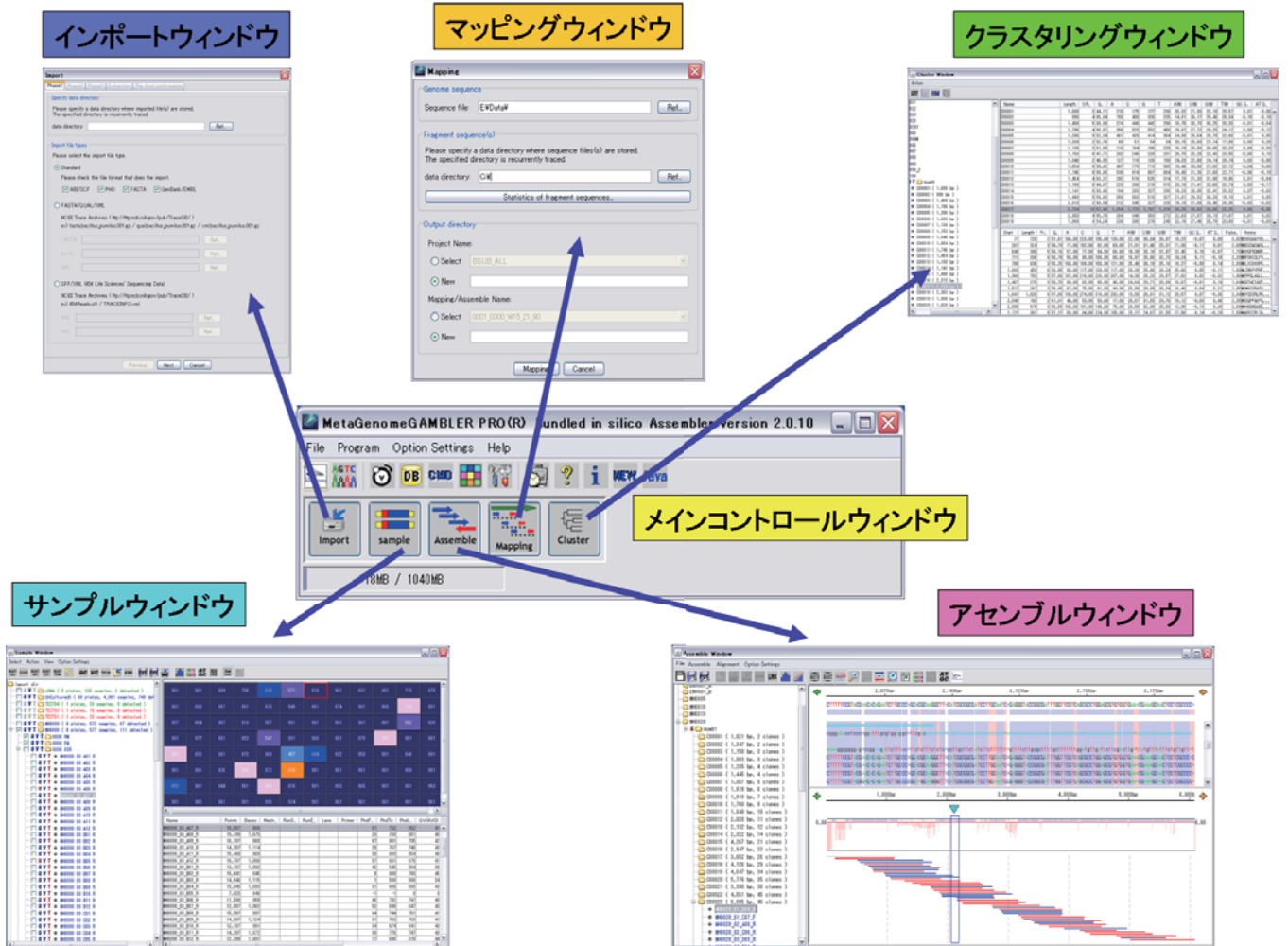
浮動ライセンスが追加されました



このソフトウェアの著作権は独立行政法人  
海洋研究開発機構が保有しています

独立行政法人  
海洋研究開発機構

# MGG の画面構成



## メインコントロールウィンドウから起動できる5つの機能

MGGのメインコントロールウィンドウから5つの機能ウィンドウを開き、それぞれの機能を実行することができます。

### インポートウィンドウ

インポートウィンドウは、アセンブル対象の塩基配列や品質データおよび実験情報などを取り込むための操作ウィンドウです。インポートできる塩基配列ファイルには、ABI/SCF形式ファイル、FastA形式ファイル、GenBank形式ファイルなどがあります。

### サンプルウィンドウ

サンプルウィンドウは、インポートされた塩基配列を対応する品質データや実験情報とリンクさせ、アセンブルに参加させるかどうかを判定するためにあります。

### アセンブルウィンドウ

上図は、メインコントロールウィンドウと5つの機能ウィンドウ。それぞれの機能ウィンドウは、メインコントロールウィンドウの大きなボタンをクリックすることにより、開くことができます。

アセンブルウィンドウの主な役割は、アセンブル結果やマッピング結果の閲覧です。アセンブルウィンドウは、大きく2つの領域に分けられます。左側のペインには、結果の構造が表示されます。プロジェクト、アセンブル、コンティグ、フラグメントの順にツリー構造で結果が表示されます。それぞれのノードをクリックすることにより、プロジェクトの情報（登録フラグメント数など）、アセンブルの情報（コンティグ数、コンティグの総塩基数、シングルトン総数など）の情報が右側のペインに表示されます。

プロジェクトノードをクリックすると、それに属するアセンブル実行結果のリストが表示されます。

アセンブル結果の1つをクリックすると、アセンブルによって生成されたコンティグのリストが表示されます。

コンティグの1つをクリックすると、右側のペインには、上段に塩基配列のアラインメントビューアが表示され、中段にはフラグメント

# MGG 機能の概要

。。前ページからつづく

ビューアが表示されます。

## マッピングウィンドウ

マッピングウィンドウは、次世代シーケンサからの大量塩基配列出力ファイルを指定し、実行パラメータを設定するためのウィンドウです。Solexa と 454 のデータはこのウィンドウを使用してマッピングします。実行前に、入力配列の統計データを表示し、事前にデータをチェックすることが可能です。実行前に、実行時のパラメータ変更用のダイアログが表示されます。

## マッピングウィンドウ (SOLiD 用)

SOLiD のデータは特殊な形式であるため、独立したデータインポート・マッピング実行パラメータ設定ウィンドウを使います。入力ファイルとしては、参照ゲノム塩基配列ファイル、および \*.csfasta、\*.qv.qual ファイル、\*.mates ファイルが指定できるほか、GFF ファイルをインポートすることも可能です。フラグメント毎の平均 Quality Value を計算し、一定品質以下のファイルを無視する設定ができます。さらに、大量配列の全件から一部の配列を抽出して、その部分だけをマッピングすることができます。

## クラスタリングウィンドウ

クラスタリングウィンドウは、メタゲノム解析用に用意されています。アセンブルされたメタゲノムのコンティグ毎に、ORF 同定と相同性検索が実行され、トップヒットした参照データベースのエントリに記述された Taxon 情報が転記さえ、各コンティグは Taxonomy Tree 上に展開され、表示されます。

## 配列データインポートと品質管理

ツリー構造の各枝の末端に存在するクローンから 1 つを選択すると、そのクローンのショットガンフラグメントのコンティグのリストが表示され、その中の 1 つをクリックすると、そのコンティグに属するフラグメントがアラインメント表示されます。ここでアラインメントの状況を確認すると同時にコンセンサス配列の閲覧が可能となります。また、フラグメントが ABI や SCF 形式のファイルである場合は、そのトレース波形をマルチプルアラインメント表示可能です。波形は同時にスクロールやズームが可能ですが、特定のフラグメントの波形を独立にシフト・スクロールすることもできます。トレース波形と塩基配列のマルチプルアラインメント表示は同期してスクロールされます。トレース波形は、アラインメントからはずれると、グレー表示され、

主要な波形だけに着目できます。

## 品質チェック機能

MGG に標準で装備されている品質チェック用のソフトウェアの他に、外部の品質チェックソフトウェアを起動し、その結果を自動的に利用することができます。現在使用可能な外部品質管理ソフトウェアは PHRED です。

## アセンブラー起動機能

MGG の本体にはアセンブル機能が搭載されていません。しかし、MGG からは、いくつかの DeNovo アセンブルソフトウェアを起動し、その結果を自動的に取り込むことが可能です。現在 MGG で利用可能な DeNovo アセンブラーは PHRAP およびインシリコアセンブラーがあります。

## マッピング・リシーケンシング機能

MGG では、DeNovo アセンブルの他に、次世代シーケンサから出力される大量の短塩基長フラグメントを参照ゲノム塩基配列上にマッピングし、参照ゲノムとのアラインメントを表示します。近縁種や過去のシーケンシング結果との比較が簡単に実行でき、結果をグラフィカルに表示します。

## インシリコ・アセンブラー (別売)

さまざまな塩基配列ファイルを混在させてアセンブルすることができるソフトウェアです。九州大学との共同研究としてインシリコバイオジー社で開発されました。インシリコアセンブラーは、注釈付き塩基配列、注釈なし塩基配列、波形ファイルなどを一緒にアセンブルできます。

## メタゲノム・クラスタリング

メタゲノム由来の大規模配列であっても、DeNovo アセンブル結果を同様に表示可能です。さらに、コンティグに対して、自動的に注釈付加することにより、Taxonomy Tree 上にコンティグを整列させることが可能です。

## ルート切替機能

ルート切替機能により、ポータブル HDD 上などに保存されたデータを素早く切替ながら、閲覧することができます。ルート以下に保存された解析結果は独立しているため、結果を他の MGG 利用者が閲覧することも簡単に実行できます。

# データのインポートと品質管理

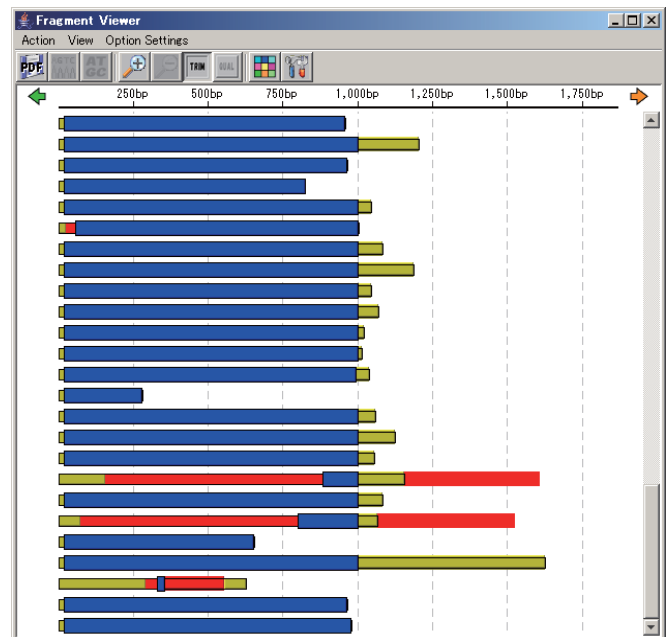
## 様々な形式の塩基配列を混在可能

MGG では様々な形式の塩基配列ファイルを同時に混在させて取り扱うことができます。ABI/SCF 形式はもちろんのこと、PHD 形式ファイルや、FastA、Text、GenBank フォーマットファイルなども同時に処理できます。さらに、圧縮形式のマルチプル FastA 形式ファイルにも対応、このとき Qual ファイルや、実験情報を記載した XML 形式のファイルも同時に取り入れることができます。さらに、最新のギガシーケンサーデータにも対応しつつあります。

シーケンサーデータが格納されているディレクトリを指定すると、その中からシーケンサー関連ファイルを読み込みます。ある親ディレクトリの下に子ディレクトリがあり、その中にそれぞれ別のシーケンシングデータが格納されている場合でも一回の操作ですべてのデータを取り込み可能となります。取り込まれたデータはインポートツリーの下に、Date、Plate、Fragment の順に階層構造が生成されます。これらの情報はファイル名あるいはファイルの内容から自動的に抽出されるため、利用者への負荷はほとんどありません。またこの階層構造は自動的に生成されるため、操作ミスによるデータロスはありません。

DNA シーケンサーから出力される配列データは実験データであるため、様々な誤差を含んでいます。これらの誤差含んだままのデータをそのまま以降の処理で使用すると、解析結果にも誤差を生じる結果となります。このため、最初の段階でこれらの誤差を取り除くことが必要です。Quality Check 機能はこの必要を満たすためにあります。ここでは、1 塩基単位の塩基配列品質を評価することが可能で、全体的あるいは部分的に低品質領域をもつフラグメントデータを取り除くことが可能となります。また、ベクター配列などのコンタミネーションを取り除いたり、Poly-A 配列を除去したりすることも可能です。

フラグメントへの命名規約を登録しておくこと、それに従って情報を抽出し、プレート内のウェルの配置状態で、各フラグメントの品質を閲覧することも可能となります。ここから、Project を登録して、各 Project にデータを割り振ることができます。



上図は、ベクターカットの結果をグラフィカル表示したものの。赤で表示されている部分がベクター汚染された領域を示す。

下図は、フラグメントファイル名称からの情報自動抽出機能。ストランドや、ウェル ID、プレート IDなどを抽出することができます。

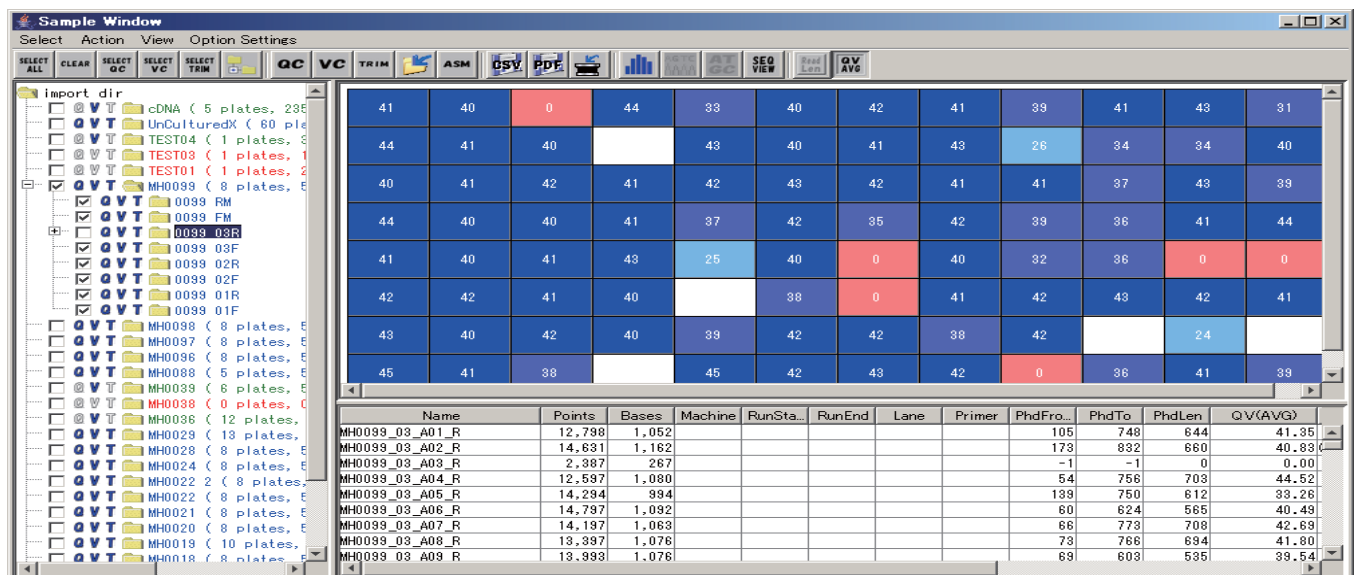
Please drag the mouse and select the corresponding area.

Plate: MH0094\_01\_A01\_F

Strand: MH0094\_01\_A01\_F

Well: MH0094\_01\_A01\_F

下図は、サンプルウィンドウ。左側にはサンプルがディレクトリツリー構造で表示される。中央部に上部には、96 穴プレート配置でのサンプルの品質をカラー分類したものを。



Name	Points	Bases	Machine	RunSta...	RunEnd	Lane	Primer	PhdFro...	PhdTo	PhdLen	QV(AVG)
MH0093_03_A01_R	12.798	1,052						105	748	644	41.35
MH0093_03_A02_R	14.631	1,162					173	832	660	40.83	
MH0093_03_A03_R	2.387	267					-1	-1	0	0.00	
MH0093_03_A04_R	12.597	1,080					54	756	703	44.52	
MH0093_03_A05_R	14.294	994					139	750	612	33.26	
MH0093_03_A06_R	14.797	1,092					60	624	585	40.49	
MH0093_03_A07_R	14.197	1,063					66	773	709	42.69	
MH0093_03_A08_R	13.397	1,078					73	766	694	41.80	
MH0093_03_A09_R	13.993	1,076					69	603	535	33.54	

# アセンブル・マッピングの便利な機能

## アセンブル機能

### [フラグメント選択機能](#)

MGG にインポートされたフラグメントは、さまざまな組み合わせで混合して、DeNovo アセンブルを実行することができます。このため、配列を吟味する必要がありますが、配列品質や長さ、ファイル名などで自由にフラグメントを選択し、アセンブルプロジェクトに送ることができます。

### [アセンブル品質不良領域の表示と除去](#)

アセンブル結果として生成される各コンティグに属する塩基毎の品質を表示し、基準以下の領域を除去してファイル出力することが可能です。

### [IMC に転送してアノテーション実行](#)

生成されたコンティグ全部の塩基配列のうち、品質がよい領域を選択し、IMC に転送することにより、アノテーション作業を連続して実行することができます。

### [グラフィカル表示を PDF ファイル出力](#)

MGG が描画するほとんどのグラフィカル表示は、そのまま PDF ファイルとして保存することができます。大きなゲノム全体のアラインメントマップは、数千ページにも PDF ファイルとして表現できますが、これを瞬時に生成することができます。

### [Mate Pair 情報を用いてコンティグ間の位置を推定](#)

各フラグメントが Mate Pair になっている場合は、この関係を利用して、コンティグ間のリンケージマップを描画することができます。また、1つのコンティグ内に存在する Mate を利用して、コンティグのアセンブル結果が正しいことを実証することができます。

### [再アセンブルのための部分的フラグメント切り出し](#)

DeNovo アセンブルにおいては、一部の領域のコンセンサスが低く、この領域に属するフラグメントのみを切り出し、少量の配列で再アセンブルすることが有効です。この要求を実現するため、MGG では、任意のアセンブル領域をフラグメント毎切り出し、ファイル保存することが可能です。切り出されたフラグメントを再入力することにより、簡単に部分的再アセンブルを実行できます。

## マッピング機能

### [レファレンスゲノム塩基配列の自動分割によるメモリー節約](#)

ゲノムサイズが大きく、マッピングされる断片配列が大量にある場合は、結果の閲覧に大きなメモリーサイズを必要とする場合があります。これを避けるために、MGG にはレファレンスゲノム塩基配列を指定塩基長・指定オーバーラップ塩基をもつ複数の配列に自動的に分割してから、マッピングを実行することができます。

### [レファレンスゲノムとの不一致箇所を即時同定・表示](#)

レファレンスゲノム塩基配列とフラグメントのコンセンサス配列との相違塩基をすべてリストアップし、指定した位置を直ちに表示することができます。これらの不一致部位は、必要なマッピング冗長度を指定することにより、SNPs 解析用の部位毎のフラグメント塩基組成をリスト表示し、かつ CSV ファイルとして保存することが可能です。

### [マップできなかった領域を再度ローカルマッピング](#)

MGG では第一回目のマッピングでは、通常各フラグメントと参照ゲノム間に比較的高い相同性とオーバーラップ塩基長を要求します。変異がいくつか所に集中しているような領域では、マッピング判定基準を満たすフラグメントが見つからず、結果としてマッピングされない領域が生じます。MGG では、再マッピング機能を使用して、これら領域を少し低い判定基準で埋めることができます。これをローカルマッピングと呼び、結果が良好であれば、メインマップにその結果を挿入することができます。再マッピングの方法には2つあります。1つはウォーキング法で両端の確かなフラグメントからフラグメント同士を結合していくものです。もう1つは、その領域の参照ゲノムを切り取り、参照ゲノムに対して、幾分低い判定基準で再マッピングを実行します。

### [SOLiD の GFF ファイルの利用](#)

MGG では生データだけでなく、SOLiD 付属のコンピューターで解析されたマッピング結果なども直接インポートし、アラインメント表示することが可能です。

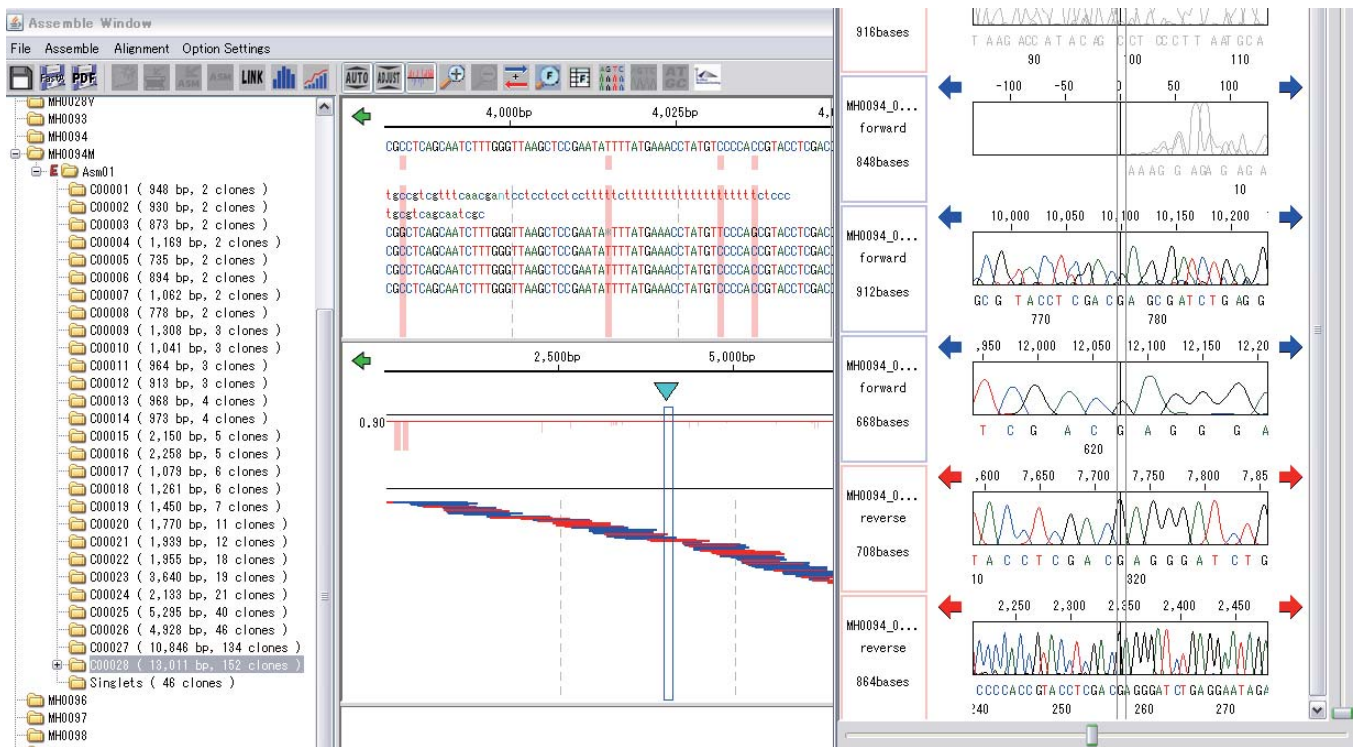
# アセンブル結果の閲覧

## 塩基配列とトレース波形のアラインメント表示

ツリー構造の各枝の末端に存在するクローンから1つを選択すると、そのクローンのショットガンフラグメントのコンティグのリストが表示され、その中の1つをクリックすると、そのコンティグに属するフラグメントがアラインメント表示されます。ここでアラインメントの状況を確認すると同時にコンセンサス配列の閲覧が可能となります。また、フラグメントが ABI や SCF 形式のファイルである場合は、そのトレース波形をマルチアラインメント表示可能です。波形は同時にスクロールやズームが可能です。特定のフラグメントの波形を独立にシフト・スクロールすることもできます。トレース波形と塩

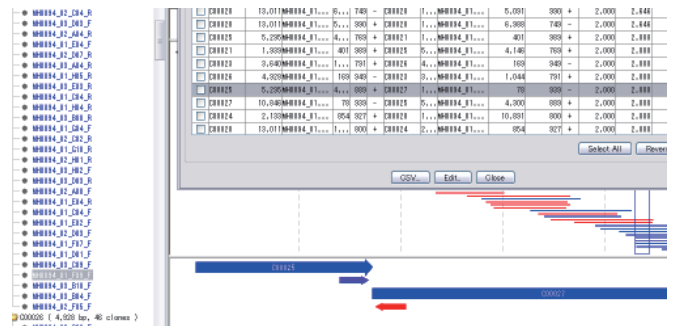
基配列のマルチアラインメント表示は同期してスクロールされます。トレース波形は、アラインメントからはずれず、グレー表示され、主要な波形だけに着目できます。

下図、左側はアセンブル対象のリスト。最上位にはプロジェクト、以下、アセンブル、コンティグ、フラグメントの順にツリー構造で示されます。1つのコンティグを選択した場合は、中央ペイン下にそのコンティグを構成するフラグメントの結合状態がグラフィカルに表示される。さらに、線カーソルを移動すると、上部の塩基配列表示もカーソルの移動につれて、移動します。各フラグメントが ABI フォーマットあるいは SCF フォーマットである場合は、そのトレース波形も同時にアラインメント表示できます。



## コンティグ間の相互位置を表示

フラグメントファイルをインポートする際に、ファイル名や実験情報ファイルから1つのクローンの両端をREADしている関係を記録し、それを利用してコンティグ間の相互位置関係をグラフィカルに表示することが可能です。1つのクローンの両端フラグメントが別コンティグに属するケースをリスト表示し、そのリストをクリックすることにより、直ちに関係するコンティグの相互位置がマップ上に表示されます。相対的な方向が表示できるため、プライマー設計などに便利です。



# マッピング・クラスタリング結果の閲覧

Fragment Item(s)	Start	End	Length	Contig	ContigLen	Percenta
<input type="checkbox"/>	199,951	199,954	4	C00003	100,050	0.00
<input type="checkbox"/>	200,078	200,084	7	C00003	100,050	0.01
<input type="checkbox"/>	201,108	201,112	5	C00003	100,050	0.00
<input type="checkbox"/>	201,304	201,309	6	C00003	100,050	0.01
<input type="checkbox"/>	201,334	201,336	3	C00003	100,050	0.00
<input type="checkbox"/>	201,499	201,507	9	C00003	100,050	0.01
<input type="checkbox"/>	208,486	208,496	11	C00003	100,050	0.01
<input type="checkbox"/>	209,495	209,443	9	C00003	100,050	0.01
<input type="checkbox"/>	210,784	210,785	2	C00003	100,050	0.00
<input type="checkbox"/>	211,112	211,113	2	C00003	100,050	0.00
<input type="checkbox"/>	212,943	212,944	2	C00003	100,050	0.00
<input type="checkbox"/>	213,241	213,243	3	C00003	100,050	0.00
<input type="checkbox"/>	214,269	214,278	5	C00003	100,050	0.00
<input type="checkbox"/>	215,677	215,679	3	C00003	100,050	0.00
<input checked="" type="checkbox"/>	215,748	215,769	16	C00003	100,050	0.02
<input type="checkbox"/>	218,668	218,669	2	C00003	100,050	0.00
<input type="checkbox"/>	218,438	218,446	9	C00003	100,050	0.01
<input type="checkbox"/>	218,599	218,600	2	C00003	100,050	0.00
<input type="checkbox"/>	218,864	218,867	4	C00003	100,050	0.00

## レファレンスゲノムへのマッピング結果閲覧

アセンブル結果の閲覧と同様レファレンスゲノムへのマッピング結果をグラフィカルに表示できます。各断片塩基配列は、ゲノム配列との相同性に従って、由来するゲノム領域に貼り付けられます。

## 次世代シーケンサの大量データをマッピング

454, Solexa などの次世代シーケンサからの膨大な塩基配列を既知のゲノム塩基配列上に、簡単な操作でマッピング可能です。

参照ゲノムとの不一致(変異)部位を直ちにリストアップ、SNPsの解析などに使用します。

参照ゲノム上でマッピングされたフラグメントが存在しなかった領域をリストアップし、ギャップクローニングや新規配列の検出に使用します。

SOLiD GFF ファイルをインポートし、マッピング結果を閲覧できます。

## メタゲノムクラスタリング結果の表示

## 動作環境

### 動作 OS

Mac OS 10.4 以上

Windows xp/Vista

64bit OS にも対応

### 必要メモリー容量 (実装)

MGGL/ MGGLW/MGGPW

512MB 以上 / 512MB 以上 / 2048MB 以上

### ディスク容量 (空き容量)

MGGL/ MGGLW /MGGPW

5GB 以上 / 5GB 以上 / 10GB 以上

### ディスプレイ推奨サイズ

MGGL/MGGLW/ MGGPW

1000W x 800H /1000W x 800H/ 1000W x 800H

### ディスプレイ最小サイズ

320W x 320H

## アセンブラーについて

使用できるアセンブラーは 2 種類あります。インシリコアセンブラーは別売です。ワシントン大学の Phrap を起動することも可能です。Phrap の使用には別途ライセンスの取得が必要です。マッピングにはインシリコアセンブラーが必要です。

## MGG 価格表

### ライト版

浮動 Lic 一般購入価格： 232,000 円

浮動 Lic アカデミック購入価格： 178,500 円

浮動 Lic バージョンアップ料金： 57,750 円

### プロフェッショナル版

浮動 Lic 一般購入価格： 336,000 円

浮動 Lic アカデミック購入価格： 252,000 円

浮動 Lic バージョンアップ料金： 84,000 円

### オプション：インシリコアセンブラー

一般購入価格： 168,000 円

アカデミック購入価格： 126,000 円

バージョンアップ料金： 42,000 円

バージョンアップは、購入後 1 年間は無料です

固定ライセンス版は、特別注文品となります。

## 謝辞

メタゲノムガンブラーは奈良先端大学小笠原教授、九州大学久原教授、海洋研究開発機構高見博士の協力の下にインシリコバイオロジー社で開発され、著作権は海洋研究開発機構が所有するソフトウェアプロダクトです。

## 開発元：インシリコバイオロジー株式会社

取扱代理店

〒 231-0023

横浜市中区山下町 24 番地 8

SOHO STATION 706 号

TEL +81-45-222-0343

FAX +81-45-222-0434

Email: info@insilicobiology.co.jp

URL: http://www.insilicobiology.co.jp/