

大量 DNA シーケンシング体制の構築と支援

●小原 雄治

情報・システム研究機構 国立遺伝学研究所生物遺伝資源情報総合センター

<研究の目的と進め方>

「基盤ゲノム」は、特定領域ゲノム3領域（「生命情報システム」「比較ゲノム」「応用ゲノム」）と密接な連携のもとに、これまでの特定領域研究等によって整備された施設・設備を有効活用しスケールメリットがありかつゲノム研究に必須のデータ取得を推進することにより、ゲノム研究のさらなる展開をめざすものである。その中において本支援班は平成16年度で終了した特定領域「統合ゲノム」で設置されたDNAシーケンシング施設を活用し、より高効率のシーケンシング体制を構築し、他3領域特に「比較ゲノム」領域との連携のもとにそれぞれの課題におけるゲノム/ESTシーケンシングを集中的におこなうことを目的とする。コストや効率の改良を図り、進化の観点から重要な生物種のゲノムやESTの集中的な解析に応用し、配列・発現比較などから生命システムの解明をめざすものである。

<研究開始時の研究計画>

①大量シーケンシング体制の構築・改良

現有施設を活用して、より高効率のシーケンシング体制を構築する。
・シーケンシングの全ステップを見直し、一層の低コスト化をはかる（これまでの1/2.5をめざす）。

・whole genome shotgun (WGS) 法を最適化するために、偏りの少ないショットガンライブラリー構築、長鎖データ産出の効率的方法を検討する。

・ライブラリー管理からシーケンスアッセムブリ、アノテーション、公開まで一貫した情報処理体制を構築し、処理の高速化と間違いの最小化を図る。この部分では本領域支援班「情報解析と成果公開のための支援活動」との連携をはかる。

②大量シーケンシング支援

・他3領域と連携し、それぞれの課題におけるゲノム/ESTシーケンシングを集中的におこなう。

・限られた能力を最大限有効に利用するために、総括班DNAシーケンシングセンター委員会において支援内容（配列決定の対象、順番など）を決定する。

③リソース管理

・シーケンスをしたクローンは研究者コミュニティで共有する必要がある。これは個々の依頼研究室には大変な手間であるので、集中的にクローン維持配布に応じる体制を構築する。

<研究期間の成果>

1) 従来型シーケンサによる支援

①大量シーケンシング体制の構築・改良

従来型サンガーシーケンサについて、シーケンシングの全ステップを見直し、一層の低コスト化をはかり、2年目までにコストダウンの目標をおおむね達成した。これを用いて次表の通り、毎年おおむね600万リード程度の大量シーケンシング支援を進めた。この結果、メダカドラフトの完成、日本産野生由来マウスなど我が国の特色を生かした研究の支援ができた。基本的には直

にDDBJに登録し、ESTではコムギなど最大のコレクションが得られ、各分野の基盤構築に寄与した。

②大量シーケンシング支援の実績

■EST			
生物種	依頼者	ライブラリー数	EST数
タマーワラビー	情報研・藤山/ 理研・黒木	3	268,835
ポリプテルス	東京慈恵医大・ 岡部	1	45,206
ベニモンアゲハ	東大・藤原	4	204,764
カイコ	東大・嶋田	7	169,145
アンドククラゲ	大阪市大・小柳	1	63,446
ギボシムシ	広島大・田川	15	327,203
ユウレイボヤ	京大・佐藤	5	141,694
近縁線虫 (Diploscapter)	遺伝研・小原	3	193,080
線虫	遺伝研・小原	2	792,454
コムギ	横浜市大・荻原	25	440,798
シャジクモ	金沢大・西山	5	177,968
ヒメツリガネゴケ	基生研・長谷部	4	77,985
ゼニゴケ	京大・福澤	4	255,032
クラミドモナス	京大・福澤	5	107,970
細胞性粘菌 (Acytostelium)	筑波大・漆原/ 桑山	5	82,502
アピコンプレクサ原虫	東大・渡辺	3	63,559
16種	小計	92	3,411,641
■WGS(全ゲノムショットガン)			リード数
野生由来マウスMSM系統	遺伝研・城石		10,249,629
ギボシムシ	京大・佐藤/広 島大・田川		4,340,832
近縁線虫 (Diploscapter)	遺伝研・小原		1,357,824
ヨコヅナクマムシ	東大・國枝		847,872
立襟鞭毛虫	情報研・藤山		2,110,426
細胞性粘菌 (Acytostelium)	筑波大・漆原		482,780
ヌクレオモルフ	筑波大・石田		18,432
7種	小計		19,407,795
			本数
■BAC, fosmid			
	小計	32	146,688
■SAGE tag			
ヒメツリガネゴケ	基生研・長谷部		160,536
ショウジョウバエ	都立大・相垣		32,256
メダカ	東大・武田		120,192
3種	小計		312,984
■fosmid end			
カイコ	東大・嶋田		253,440
クワコ	東大・嶋田		153,216

細胞性粘菌 (Acytostelinm)	筑波大・漆原	13,824
クラミドモナス	京大・福澤	39,168
ヨコヅナクマムシ	東大・國枝	60,672
近縁線虫 (Diploscapter)	遺伝研・小原	807,840
6種	小計	1,328,160
■その他		
メダカPCR product	東大・武田	6,912
メダカPCR product	基生研・成瀬	13,248
ゼブラフィッシュPCR product	遺伝研・川上	192
カイコsmall RNA	東大・嶋田	21,120
マウスメチル化領域	群馬大・畑田	192
	小計	41,664
	総合計	24,648,932

(2009年12月までのまとめ)

上記の支援推進に伴い種々の関連開発も進めた。

i) ゲノムfinishing用ワークベンチの開発

ゲノム配列解析プロジェクトにおいては、配列決定を以って終了することは稀で、配列はその後遺伝子予測などのアノテーションに供されることになる。したがって、素材としてのゲノム配列自体の完成度の高さ、すなわち、大きな scaffold が得られていること、コンティグ中に誤りが少ないことが要求される。全ゲノムショットガン法ではコンティグ/scaffold 構築の大部分をアセンブラにより自動実行するのが常であるが、最終ステップでは、人手による編集作業、および、個別実験によるギャップ結合が必要である。しかしながら、これらの編集作業を共同で行い、かつ、追加情報を一元管理するワークベンチはこれまで存在しなかった。そこで、「情報解析と成果公開の支援活動」の支援を得て、ゲノム finishing 作業をサポートするプラットフォームを構築した(図1参照)。

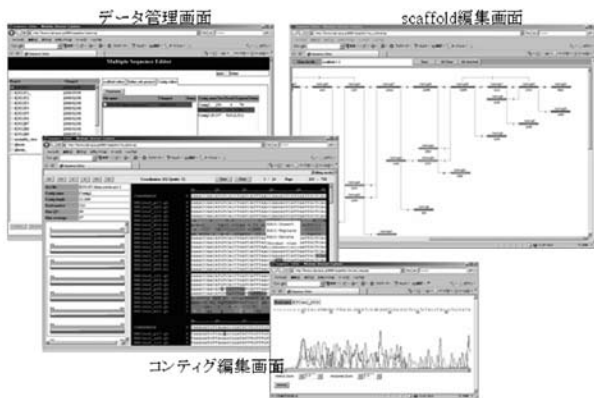


図1. ゲノム finishing ワークベンチ・インターフェース

scaffold 編集機能では、対象となる scaffold の構造をコンティグ間のブリッジ数とともに図示し、それを参照してコンティグ間の不自然なリンクを切断し、また必要な位置にコンティグ間リンクを設定できる。

コンティグ編集機能では、対象となるコンティグのリードアラインメントを表示し、それを参照して各リードの塩基の変更、削除、挿入を行うことができる。また、リードの追加、アセンブリからの削除も可能である。リードアラインメントの表示範囲は、スクロールボタンで変更可能なほか、左側フレームに表示されたコンティグ全体像(リードカバー率が図示されている)より直接選択することも可能である。リードアラインメントの表示では、

低クオリティ領域はバックグラウンドがグレー(クオリティが低いほど黒に近づく)で、コンセンサスと一致しない塩基は赤字で表示され、問題のある領域を視覚的に把握できるよう配慮されている。

各リードもしくはリード名を選択することで、その波形データを別ウィンドウで表示することができる。

本システムでは、これら編集作業は、複数のユーザが共同で遂行することを前提にしているため、コンティグ単位、scaffold 単位の排他制御が実現されている。本システムで編集対象コンティグ/scaffold を選択する場合、Explorer 形式の一覧メニューを利用するが、他のユーザによって編集対象であるコンティグ/scaffold では、メニュー中に編集対象である旨が表示される。

本システムは、アセンブリファイルの形式として*.ace フォーマットを採用しており、したがって、phrap, PCAP をはじめとする多数のアセンブラの出力に対応している。

本システムは、複数のメダカ BAC クローンのアセンブリに対して試用し、その有用性が検証された。

ii) WGSリードより多型を探索するシステム、ビューア

キャピラリーシーケンサより得られる研究対象サンプルの WGS (Whole Genome Shotgun) リードを、比較対象となるリファレンスゲノムにマップし、そのアライメント結果からリファレンスとサンプルのゲノム配列間に存在する多型情報 (DNA 塩基配列) を網羅的に算出するシステムを構築した(図2参照)。

本システムは 500bp 程度からなる WGS シーケンス配列、BAC end 配列について、megablast プログラムにより登録されたリファレンスゲノムに対しマップ処理を行う。得られたアライメント候補情報から、リードがユニークにヒットした、または複数個所にヒットしたリードのもっとも高スコアサイトであったサイトについて、bl2seq プログラムにより詳細なアライメント情報を取得する。取得した全リードのアライメント情報をリファレンスゲノム上で整列し、リファレンスゲノムと異なる配列塩基が存在するサイトについて、そのサイトの各リード配列の厚み、リード配列塩基のクオリティを考慮して多型を判定し収集する。またリファレンスゲノムの遺伝子領域情報が存在する場合、収集した多型情報がそれら遺伝子領域に与える影響情報を算出する。

また本多型算出システムにより得られる WGS リードのマップ情報、多型情報を公開するゲノムビューアを WEB システムとして構築した。利用者は WEB ブラウザより、ゲノム上の領域指定、遺伝子名などの文字列検索により、参照したいゲノム領域、多型の影響を受ける遺伝子情報などにたいし、容易にアクセスが可能である。



図2. WGS ゲノムビューア画面フロー

本システムを使用し、日本産野生由来の近交系マウスである MSM 系統についてマップ処理、多型解析を行った。MSM 系統の WGS 配列について、ゲノム配列が決定されている標準系マウス C57BL/6J 系統のゲノム配列に対しマップ処理を行い、これら系統間に存在する多型を算出した。MSM 系統 WGS リード配列、および BAC end 配列の 9,881,203 リードのデータについてマッピング処理を行い、リピート領域を除外したゲノム上領域の 78%、ゲノムカバレッジ 2.45x のマップ結果を得た。これらマップ配列情報から、9,735,746 箇所の多型候補サイトを抽出した。

これら MSM 系統 WGS 配列を C57BL/6J 系統ゲノムにマップした結果は MSM ゲノム DB として現在外部に公開している。

<<http://molossinus.lab.nig.ac.jp/msmdb/index.jsp>>

2) 次世代シーケンサー解析システム構築

「比較ゲノム」支援班と協力して 2008 年度から次世代型シーケンサの試行と解析パイプラインの構築を進めた。また、情報セキュリティ体制の必要な個人ゲノム対応のためのシステムも含めたコンピュータシステム (図 3 参照) を整えた。

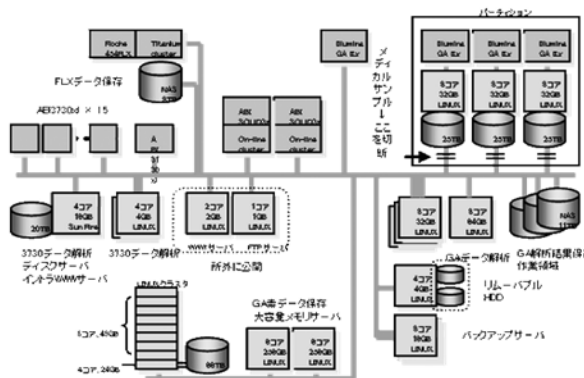


図 3. シーケンサ、解析サーバ/ディスク構成

i) SOLEXA 解析結果管理 WWW サーバ

サーバ上に保管された SOLEXA 解析結果を Web インターフェースから閲覧およびダウンロード可能なシステムを開発した (図 4 参照)。

Web インターフェースのメイン画面は、ラン毎の解析フローをディレクトリツリー形式で表示するリスト画面と、リスト画面で選択されたランの解析結果の表示およびダウンロードが可能なデータ画面の左右 2 画面構成となっている。

リスト画面の初期画面では、サーバ上に保管された SOLEXA 解析結果が、ラン単位で時系列に沿ってリスト表示される。それぞれのランが解析フローのルートディレクトリとなっており、アイコンをクリックすることで、解析フローを順に辿っていくことができる。解析フローは SOLEXA GA Pipeline に類似した構造で表示され、それぞれの解析が 1 つのディレクトリに対応している。具体的には、上位から順に画像解析部の Firecrest 階層、次にベースコール部の Bustard 階層、そしてアライメント部の GERALD 階層の構成で表示される。さらに、これらのディレクトリ構造に加えて、シーケンスを行った生物種や解析に使用したパラメータなどのコメントも付随情報として表示される。これらのコメントに対するキーワード検索機能と、コメントを登録するためのウェブインターフェースも準備した。

データ画面は、リスト画面で選択されたディレクトリ位置が表示されるナビゲーション部、解析結果のサマリーページへのリンクがリスト表示されるサマリー部、マッピング結果や配列データなどのダウンロードファイルがリスト表示されるダウンロード部で構

成される。ダウンロードファイルは通常ファイルと圧縮ファイルに対応しており、各ファイルの横にはファイルサイズが表示される。サマリーページとダウンロードファイルの種類は SOLEXA GA Pipeline のバージョンアップによってしばしば変更されるため、これらのファイルを指定するための設定ファイルを準備することで容易に対応可能とした。

SOLEXA 解析結果の追加または削除は、WWW サーバ上の所定のディレクトリに SOLEXA 解析結果をラン単位で追加または削除することで、自動的にリアルタイムで Web インターフェースに反映される。なお、本システムの運用を開始した 2008 年 6 月から 2009 年 12 月までの間で、68 ランのデータを登録し、総データ容量は 2.7TB となっている。



図 4. SOLEXA 解析結果管理 WWW サーバ

ii) ショートリードより多型を探索するシステム・ビューア

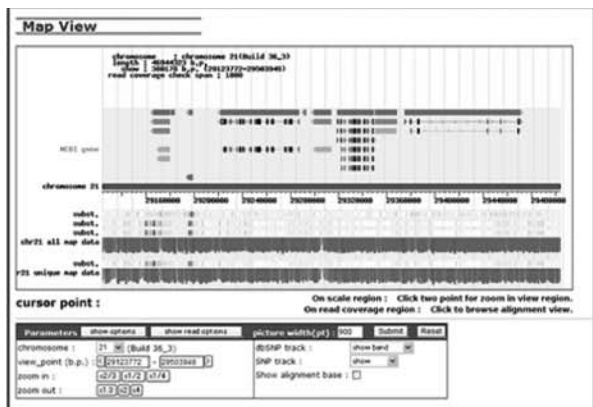
次世代シーケンサは多量の配列解析が可能で一方、これから得られる配列データは、配列長が短く、多量の read 配列件数からなっている。これら多量の short read 配列の re-sequencing による情報の抽出には、現在までのキャピラリーシーケンサより得られる配列に対する解析処理とは異なる、低信頼性だが多量のデータに対応した解析処理が必要となっている。このためこれら short read 配列のリファレンスゲノムへのマップ結果から SNP、Insertion/Deletion などの多型情報を算出することを目的とした、多型情報の算出システムを構築した (図 5 参照)。

本多型算出システムは、short read シーケンサより得られる short read 配列について、リファレンスゲノムに対しマッピングを行い、それらアライメント情報から多型情報を算出する。この際、short read においてのベースコール結果の低信頼性 (従来の capillary シーケンスデータと比較してエラー率が高い) を補完するため、read 配列の厚み、および配列の各塩基の quality を考慮して多型の判定を行う。またサンプルとなる short read 配列について、一定の insert 長で作成された paired-end データである場合、これらのマップ位置を利用して、Gap 領域などゲノム上構造の変異を抽出する。また同じリファレンスゲノムにマップされた複数の short read 配列データセット単位で算出された多型情報について、相互の比較により、データセット固有の多型情報の抽出が可能となっている。

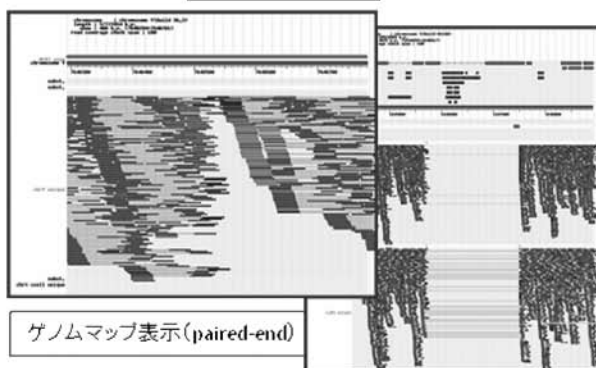
これら short read 配列データセットおよび多型情報のゲノム上機能領域への関与について、直感的かつ容易に情報にアクセスできるシステムとして、WEB 画面より視覚的に確認が行えるゲノムビューアを構築した。本ゲノムビューアにおいては、ゲノム領域の指定、また遺伝子名などの文字列検索により、対象ゲノム領域のマップ情報、塩基配列レベルのアライメント情報の参照が可能である。また paired-end データが登録されている場合、これら end 配列のペアの位置情報を表示し、ゲノム上構造の変異情報が参照可能である。

ゲノムブラウザにおいては、同じリファレンスゲノムへのマッ

プが行われた複数の short read 配列データセットについて、WEB 画面にてそれらのマップ率情報、多型情報の比較参照が容易に行えるよう、複数のデータセットの情報を同時に表示が可能である。



ゲノムマップ表示



ゲノムマップ表示 (paired-end)



アライメント、多型表示



遺伝子検索画面

図5. short readゲノムビューア画面

本システムを使用し、「比較ゲノム」支援班との協同により、ヒト培養細胞 (GM18940) の染色体 21 番の solexa データについて、ゲノム配列へのマップ処理、多型判定処理を行った。染色体 21 番の solexa データ総リード数 209,139,326 read について、NCBI Build37 ゲノムに対しマップ処理を行った結果、カバー領域長 35,084,772bp (GAP を除いたゲノム長の 99.9%)、重複度 93.9x のマップ結果を得た。これらマップ結果について、多型判定を行い、染色体 21 番上に 52,344 サイトの候補サイトを得た。またこれら多型候補の影響を受ける遺伝子 (CDS 領域) について探索し、214 の CDS 領域に多型候補が含まれる、また 3 CDS について exon-intron junction 上の変異、2 CDS について多型候補により STOP 変異が挿入されるという結果を得た。

iii) SOLEXA multiplexing データ解析システムの開発

次世代型シーケンサ SOLEXA では、multiplexing によりひとつのレーンに複数のサンプルをアプライすることが可能である。

具体的には、サンプルフラグメントの一端に既知の 6 ベース (タグ配列) を含むアダプタを結合し、シーケンシングの際にはタグもリードとみなして読み、データ処理ステップでその配列パターンに応じて分類する手法であり、現在のところ、12 種類のタグ配列を使用することが可能である。SOLEXA データの一次解析を実行するシステムである GA Pipeline (Illumina 製) は、本シーケンシング手法に対応していることになっているが、実際には、タグ配列をリードとしてベースコールし、その結果を出力に埋め込んでいるだけで、タグ配列の特性を生かした分類 (2 塩基ミスコールまでは一意に確定できるように設計されている) や、分類可能 / 不可能の判定は行われない。

そこで上記の点を解決し、SOLEXA の標準出力形式である export フォーマットのファイルの内容をタグにより分類し、サブセットを作成できるシステムを構築した。本システムは、「タグ配列分類モジュール」と「出力サブセット構築モジュール」から構成される。

	GA Pipeline		multiplex解析システム	
	リード数	(%)	リード数	(%)
総数	71,302,459		71,302,459	
アサイン可	58,319,928	82	66,917,360	94
アサイン不可	12,982,531	18	4,385,099	6
タグ配列ごとの数				
TGACCA	6,100,371	9	7,150,394	10
GATCAG	4,188,232	6	4,876,986	7
ATCACG	4,489,128	6	5,148,817	7
GCCAAT	4,833,437	7	5,609,370	8
CAGATC	4,244,559	6	4,973,895	7
ACAGTG	4,715,060	7	5,542,358	8
ACTTGA	4,770,459	7	5,523,231	8
GGCTAC	4,462,397	6	4,981,663	7
CGATGT	4,891,219	7	5,618,098	8
CTTGTA	4,812,913	7	5,393,541	8
TAGCTT	4,559,798	6	5,109,495	7
TTAGG	6,252,355	9	6,989,512	10

表2. SOLEXA multiplexing タグ分類能力の比較

タグ配列分類モジュールは、export フォーマット中のタグ配列

を読み込み、参照タグセット（本来読まれるべきタグ配列のセット）のいずれに分類されるか、あるいは分類不可能か判定する。ここで、最大ミスマッチ数、Nの取扱い（全塩基種と一致/不一致）、indelを許容する/しない、を指定できる。分類結果はexportフォーマットの末尾にカラムを追加することで記述される。ここで、読まれたタグ配列、そのクオリティ、分類結果（参照タグ配列、分類不可能な場合はその旨）、ミスマッチ数が分類情報として追加される。

出力サブセット構築モジュールは、上記の分類結果を基に、参照タグ配列ごとのサブセットを作成する。分類不可能であったレコードは、別途ファイルに集約されるので、ひとつのexportファイルから、（使用したタグ数）+ 1種類のファイルが作成されることになる。

12種類のタグを用いた SOLEXA multiplexing リードの4レーン分を分類した事例を（表）に示す。表中の「GA Pipeline」はGA Pipelineによりベースコールされたタグ配列をそのまま用いて分類した結果を、「multiplex 解析システム」は本システムを用いて、最大ミスマッチ2、Nは常にミスマッチ、indelは許容しない条件で分類した結果を示す。前者で82%であった分類率が94%に上昇し、また、どのタグ種類も偏りなくリカバーされたことが分かる。

iv)cDNA全長シーケンスアセンブルシステム

cDNA の全長配列を mix cDNA shutgun 法により決定するプロセスをサポートするシステムを開発した。

本システムでは、cDNA クローンの5' および3' のエンドリードと、mix cDNA shutgun 法により生成されるショットガンリードの混在データを PCAP によりアセンブルし、その結果を元に、5' と3' のエンドリードが整合性を持って同一のコンティグに含まれるクローンを完成クローン候補として抽出する。そのほかの未完成クローンは、5'エンドリードと3'エンドリードが別コンティグに分離する、または複数クローンのエンドリードが混在するコンティグに含まれるなど、アセンブリの状態により分類される。また、完成クローン候補のうち、3' 末端の PolyA 付近のみがカバーされないクローンを対象に、リバースプライマーにより PolyA を超えてシーケンスされたリードをアセンブリに追加する機能も準備した。

本システムにはまた、従来のサンガー法によるショットガンリードのほか、次世代シーケンサ SOLEXA によってシーケンスしたショートリードを使用したアセンブル機能も準備した。この機能では、まずショートリード向けアセンブラ Velvet によりショートリードをアセンブルし、Velvet コンティグを生成する。

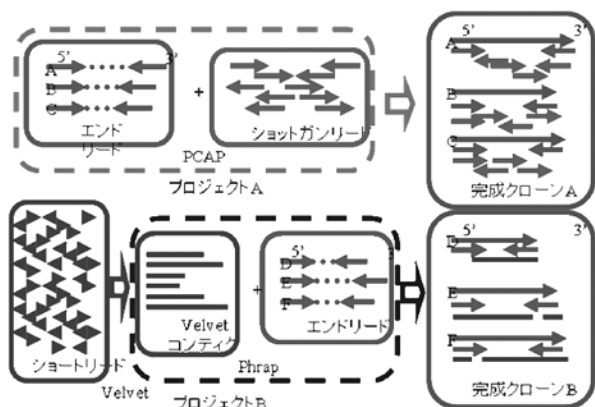


図6. cDNA 全長シーケンスアセンブルシステム

次に、それらの Velvet コンティグとエンドリードを Phrap でアセンブルし、サンガー法と同様に、5' と3' のエンドリードが整合性を持って同一のコンティグを構成するクローンを完成クローン候補として抽出する。さらに、5' と3' のエンドリードがそれぞれ別のコンティグに分離するクローンを完成させるために、コンティグ間のギャップを埋める機能を準備した。この機能では、まずショートリード用マッピングプログラムによりショートリードをコンティグ配列にマップし、マップされたショートリードおよびその Mate リードを抽出する。次に、これらの抽出したショートリードのセットとコンティグを素材にアセンブルすることでコンティグ末端を伸張させる。このマッピングとアセンブルのプロセスを繰り返すことでギャップを埋める。マッピングプログラムには BWA が、アセンブルには Phrap がそれぞれ使用される。

これまでに、本システムを使用してメダカ完全長 cDNA を 13888 クローン完成させた。

③リソース管理

本支援班で解析した DNA クローンはオリジナルの他にレブリカを3セット作成し、支援先の要望に応じて支援班においてクローンの研究コミュニティへの配布を代行してきた。ホヤ及び細胞性粘菌については、期間途中でナショナルバイオリソースプロジェクトに採択されたために全クローンをそちらに移管した。

表3.EST クローン発送実績

		年	2005	2006	2007	2008	2009	累計
ホヤ cDNA								
件数	総数	28	25	12	5	0	70	
	国内	3	3	2	4	0	12	
	国外	25	22	9	1	0	57	
クローン数	総数	115	115	60	34	0	324	
	国内	10	11	8	30	0	59	
	国外	105	104	51	4	0	264	
発送国数		8	8	5	2	0	17	

細胞性粘菌cDNA		年	2005	2006	2007	2008	2009	累計
件数	総数	19	18	10	0	0	47	
	国内	4	2	1	0	0	7	
	国外	15	16	9	0	0	40	
クローン数	総数	140	51	14	0	0	205	
	国内	109	3	1	0	0	113	
	国外	31	48	13	0	0	92	
発送国数		7	6	6	0	0	15	

ナメクジウオ cDNA		年	2005	2006	2007	2008	2009	累計
件数	総数	2	6	4	9	6	27	
	国内	0	2	2	2	2	8	
	国外	2	4	2	7	4	19	
クローン数	総数	2	22	8	81	20	133	
	国内	0	14	6	48	7	75	
	国外	2	8	2	33	11	56	
発送国数		2	4	3	7	5	10	

ギボシムシ cDNA		年	2005	2006	2007	2008	2009	累計
件数	総数					2	2	
	国内					2	2	
	国外					0	0	
クローン数	総数					36	36	
	国内					36	36	
	国外					0	0	
発送国数						1	1	

<国内外での成果の位置づけ>

・米国JGIなど大規模センターを中心に世界的に様々な生物のゲノムシーケンスが進んだなかで、本活動のシェアは大きいとは言えない。

・外国がどちらかといえば有用生物を中心に行っているのに対し、本支援活動では学術研究を中心にして、かつわが国が優位性を持つもの（ホヤ、メダカなど）を重点的に支援したことは特色としてあげられる。これらの分野では海外からの評価は高い。

・また、結果的に国際共同研究になったものの、この支援活動で開始したことが呼び水になったり後押しになった例も、細胞性粘菌やナメクジウオ、ヒメツリガネゴケなど多い。

・いずれの場合も、これまでゲノム科学とは無縁であった基礎生物学のコミュニティに新たなインフラを提供できたことは大きいと考える。

・中型以上のゲノムのWGSでのドラフト作りでは国内では唯一である。

<達成できなかったこと、予想外の困難、その理由>

・ドラフト配列までは進められたが、その先のアノテーションも含めた高精度配列データが進んでいない。このためのワークベンチを構築したが、時間切れになってしまった。特にメダカゲノムは世界中から完成の要望を受けていることから、機会を捉えて、研究コミュニティの協力何とか進めたい。

・期間中に次世代型シーケンサーが実用化され、その能力向上のスピードも目覚ましいものがあった。「比較ゲノム」支援班との協同で対応に努めてきたが、本格的な運用は今後に残された。

<今後の課題、展望>

・今後のシーケンス技術の発展に対応し、ゲノムと表現型の体系的な対応づけが課題である。生物の多様性、発生システムの進化、人多様性の遺伝基盤など、革命的な研究に活用されるであろう。そのための情報解析技術の革新が必須である。

<研究期間の全成果公表リスト>

1) 論文

1. 0912031609

Hisashi Hashimoto et al.; Polycystic Kidney Disease in the Medaka (*Oryzias latipes*) pc Mutant Caused by a Mutation in the Gli-Similar3 (*glis3*) Gene. **PLoS ONE** 4(7): e6299 (2009).

2. 1002062105

Kazuhiro SATO et al.: Development of 5006 Full-Length CDNAs in Barley: A Tool for Accessing Cereal Genomics Resources. **DNA RESEARCH** 16: 81 - 89 (2009)

3.0901131410

The International Silkworm Genome Consortium: The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. **Insect Biochemistry and Molecular Biology** 38, 1036-1045 (2008).

4.0901131406

Shin Sasaki et al.: Chromatin-associated periodicity in genetic variation downstream of transcriptional start sites. **Science** 323, 401-404 (2009)

5. 1002062107

Jr-Kai Yu et al.: A cDNA resource for the cephalochordate

amphioxus *Branchiostoma floridae* **Development Genes and Evolution** 218, 723-727 (2008)

6.0806261537

Nicholas H. Putnam et al.: The amphioxus genome and the evolution of chordate karyotype, **Nature** 453, 1064-1071 (2008)

7.0806262231

Shinpei Kawaoka et al.: *Bombyx* small RNAs: Genomic defense system against transposons in the silkworm, *Bombyx mori*. **Insect Biochemistry and Molecular Biology** (Special Issue on *Bombyx* Genome), 38, 1058-1065 (2008).

8.0801251617

Stefan A. Rensing et al.: The *Physcomitrella* Genome Reveals Evolutionary Insights into the Conquest of Land by Plants, **Science** 319, no. 5859, pp. 64 - 69

9.0801270044

Budrul Ahsan et al.: UTGB/medaka: genomic resource database for medaka biology. **Nucleic Acids Research** 36: D747 - D752 (2008).

10.0801270040

Yoichiro Nakatani et al.: Reconstruction of the Vertebrate Ancestral Genome Reveals Dynamic Genome Reorganization in Early Vertebrates. **Genome Research** 17(9):1254-65. (2007)

11.0707091324

Fuki Gyoja et al.: Analysis of large scale expression sequenced tags (ESTs) from the anural ascidian, *Molgula tectiformis*. **Developmental Biology** 307, 460-482 (2007)

12.0801270031

Masahiro Kasahara et al.: The medaka draft genome and new insights into vertebrate genome evolution. **Nature** 446, 714-719 (2007)

13.0912070936

Katsuyuki T. Yamato et al.: Gene organization of the liverwort Y chromosome reveals distinct sex chromosome evolution in a haploid system. **PNAS** 104, 6472-6477 (2007).

14.0702141053

Jr-Kai Yu et al.: Axial patterning in cephalochordates and the evolution of the organizer. **Nature** 445, 613-617 (2007)

15.0601311708

Eiichi Shoguchi et al.: Chromosomal mapping of 170 BAC clones in the ascidian *Ciona intestinalis*. **Genome Research** 16, 297-303 (2006)

2) データベース/ソフトウェア

NIG DNA Sequencing Center:

<<http://dolphin.lab.nig.ac.jp/>>

マウスMSMゲノムDB

<<http://molossinus.lab.nig.ac.jp/msmdb/index.jsp>>