

研究種目：特定領域研究

研究期間：2005 ～ 2009

課題番号：17020008

研究課題名（和文） 大量 DNA シーケンシング体制の構築と支援

研究課題名（英文） Large-scale DNA sequencing and its application

研究代表者

小原 雄治 (KOHARA YUJI)

国立遺伝学研究所・所長

研究者番号：70135292

研究成果の概要（和文）：大量 DNA シーケンシング体制構築のために、シーケンシング全ステップを見直し、コストダウンを図り、年間 600 万リード（サンガー法）を可能にした。この結果、16 の生物種の 92 種の組織・発生段階の cDNA ライブラリーについて 340 万リードの EST 解析を支援、野生由来マウス、ギボシムシ、クマムシ、立襟鞭毛虫、細胞性粘菌など 7 種の全ゲノムショットガンシーケンス 1940 万リードを支援、その他合計 2400 万リードを支援した。ゲノム finishing 用ワークベンチの開発、WGS リードより多型を探索するシステム、ビューアを構築した。次世代型シーケンサー解析システム構築を進め、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を整えて、様々な試行を行った。ショートリードより多型を探索するシステム・ビューア、SOLEXA multiplexing データ解析システムの開発、cDNA 全長シーケンスアセンブルシステムを開発した。

研究成果の概要（英文）：We established a large-scale (Sanger) DNA sequencing system through the review, improvement and cost reduction of the sequencing process step by step, leading to allow us to obtain 6 million (Sanger) reads per year. Based on the requests from the researchers in the grant group, we produced 3.4 million EST reads for 92 cDNA libraries from 16 organisms. We also produced 19.4 million reads for whole genome shotgun sequencing of 7 organisms including wild-derived mouse, a hemichordata, water bears, the choanoflagellate *Monosiga ovata*. Altogether 24 million reads were produced during the project period. To analyze the reads more closely, we developed a work-bench for genome finishing and a viewer for polymorphism identification. For the next generation DNA sequencing system, we set up a computer system with a security system that is essential for personal genomics and performed various DNA sequencing. Furthermore, we developed a viewer system for polymorphism identification and a system for cDNA full sequencing. Solexa multiplexing data analysis was also tested.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2005 年度	257,500,000	0	257,500,000
2006 年度	177,400,000	0	177,400,000
2007 年度	177,500,000	0	177,500,000
2008 年度	187,317,017	0	187,317,017
2009 年度	179,100,000	0	179,100,000
総計	978,817,017	0	978,817,017

研究分野：複合新領域

科研費の分科・細目：ゲノム科学・ゲノム生物学

キーワード：(1) ゲノム (2) cDNA ライブラリー (3) 全ゲノムショットガン  
(4) EST (5) DNA シーケンサー (6) アセンブラ

## 1. 研究開始当初の背景

学術分野におけるゲノム研究をさらに発展させるために、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が求められていた。

## 2. 研究の目的

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した。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。

## 3. 研究の方法

### 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センブリ、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。

## 4. 研究成果

### 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、毎年おおよそ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。基本的には直ちに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
アンドンクラゲ	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		2, 110, 426

	山		
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 (Acytostelium)	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。

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) 次世代型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。

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)。

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。

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な一方、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

iv) cDNA 全長シーケンスアセンブルシステムリソース管理

本支援班で解析した DNA クローンはオリジナルの他にレプリカを 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。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 17 件) すべて査読あり

1. Hashimoto, H., Miyamoto, R., Watanabe, N., Shiba, D., Ozato, K., Inoue, C., Kubo, Y., Koga, A., Jindo, T., Narita, T., Naruse, K., Ohishi, K., Nogata, K., Shin-I, T., Asakawa, S., Shimizu, N., Miyamoto, T., Mochizuki, T., Yokoyama, T., Hori, H., Takeda, H., Kohara, Y., Wakamatsu, Y.; 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. K. Sato, T. Shin-i, M. Seki, K. Shinozaki, H. Yoshida, K. Takeda, Y. Yamazaki, M. Conte, and Y. Kohara: Development of 5006 Full-Length cDNAs in Barley: A Tool for Accessing Cereal Genomics Resources. **DNA RESEARCH** 16: 81 - 89 (2009)

3. N. Kobayashi, M. Watanabe, T. Horiike, Y. Kohara, N. Okada: Extensive analysis of EST sequences reveals that all cichlid species in Lake Victoria share almost identical transcript sets. **Gene** 441, 187-191 (2009)

4. The International Silkworm Genome Consortium: The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. **Insect Biochem. & Mol. Biol.**38, 1036-1045 (2008).

5. S. Sasaki, C. C. Mello, A. Shimada, Y. Nakatani, S. Hashimoto, M. Ogawa, K. Matsushima, S.G. Gu, M. Kasahara, B. Ahsan, A. Sasaki, T. Saito, Y. Suzuki, S. Sugano, Y. Kohara, H. Takeda, A. Fire, S. Morishita: Chromatin-associated periodicity in genetic variation downstream of transcriptional start sites. **Science** 323, 401-404 (2009)

6. J-K. Yu, M-C. Wang, T. Shin-I, Y. Kohara, L. Z. Holland, N. Satoh & Y. Satou: A cDNA resource for the cephalochordate amphioxus *Branchiostoma floridae* **Development Genes and Evolution** 218, 723-727 (2008)

7. N. H. Putnam, T. Butts, D. E. K. Ferrier, R. F. Furlong, U. Hellsten, T. Kawashima, M. Robinson-Rechavi, E. Shoguchi, A. Terry, Jr-K. Yu, È. Benito-Gutiérrez, I. Dubchak, J. Garcia-Fernández, I. V. Grigoriev, A. C. Horton, P. J. de Jong, J. Jurka, V. Kapitonov, Y. Kohara, Y. Kuroki, E. Lindquist, S. Lucas, K. Osoegawa, L. A. Pennacchio, A.A. Salamov, Y. Satou, T. Sauka-Spengler, J. Schmutz, T. Shin-I, A. Toyoda, J. J. Gibson-Brown, M. Bronner-Fraser, A. Fujiyama, L. Z. Holland, P. W. H. Holland, N. Satoh, D. S. Rokhsar: The amphioxus genome and the evolution of chordate karyotype, **Nature** 453, 1064-1071 (2008)

8. S. Kawaoka, N. Hayashi, S. Katsuma, H. Kishino, Y. Kohara, K. Mita, and T. Shimada: *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).

9. S. A. Rensing, D. Lang, A. D. Zimmer, A. Terry, A. Salamov, H. Shapiro, T. Nishiyama, P-F. Perroud, E. A. Lindquist, Y. Kamisugi, T. Tanahashi, K. Sakakibara, T. Fujita, K. Oishi, T. Shin-I, Y. Kuroki, A. Toyoda, Y. Suzuki, S. Hashimoto, K. Yamaguchi, S. Sugano, Y. Kohara, A. Fujiyama, A. Anterola, S. Aoki, N. Ashton, W. B. Barbazuk, E. Barker, J. L. Bennetzen, R. Blankenship, S. H. Cho, S. K. Dutcher, M. Estelle, J. A. Fawcett, H. Gundlach, K. Hanada, A. Heyl, K. A. Hicks, J. Hughes, M. Lohr, K. Mayer, A. Melkozernov, T. Murata, D. R. Nelson, B. Pils, M. Prigge, B. Reiss, T. Renner, S. Rombauts, P. J. Rushton, A. Sanderfoot, G. Schween, S-H. Shiu, K. Stueber, F. L. Theodoulou, H. Tu, Y. Van de Peer, P. J. Verrier, E. Waters, A. Wood, L. Yang, D. Cove, A. C. Cuming, M. Hasebe, S. Lucas, B. D. Mishler, R. Reski, I. V. Grigoriev, R. S. Quatrano, J. L. Boore: The Physcomitrella Genome Reveals Evolutionary Insights into the Conquest of Land by Plants, **Science** 319, 64 - 69 (2008)

10. B. Ahsan, D. Kobayashi, T. Yamada, M. Kasahara, S. Sasaki, T. L. Saito, Y. Nagayasu, K. Doi, Y. Nakatani, W. Qu, T. Jindo, A. Shimada, K. Naruse, A. Toyoda, Y. Kuroki, A. Fujiyama, T. Sasaki, A. Shimizu, S. Asakawa, N. Shimizu, S. Hashimoto, J. Yang, Y. Lee, K. Matsushima, S. Sugano, M. Sakaizumi, T. Narita, K. Ohishi, S. Haga, F. Ohta, H. Nomoto, K. Nogata, T. Morishita, T. Endo, T. Shin-I, H. Takeda, Y. Kohara and S. Morishita: UTGB/medaka: genomic resource database for medaka biology. **Nucleic Acids Research** 36: D747 - D752 (2008).

11. Y. Nakatani, H. Takeda, Y. Kohara, S. Morishita: Reconstruction of the Vertebrate Ancestral Genome Reveals Dynamic Genome Reorganization in Early Vertebrates. **Genome Research** 17(9):1254-65. (2007)

12. F. Gyoja, Y. Satou, T. Shin-i, Y. Kohara, B. J. Swalla, N. Satoh: Analysis of large scale expression sequenced tags (ESTs) from the anural ascidian, *Molgula tectiformis*. **Developmental Biology** 307, 460-482 (2007)

13. M. Kasahara, K. Naruse, S. Sasaki, Y. Nakatani, W. Qu, B. Ahsan, T. Yamada, Y. Nagayasu, K. Doi, Y. Kasai, T. Jindo, D. Kobayashi, A. Shimada, A. Toyoda, Y. Kuroki, A. Fujiyama, T. Sasaki, A. Shimizu, S. Asakawa, N. Shimizu, S. Hashimoto, J. Yang, Y. Lee, K.

Matsushima, S. Sugano, M. Sakaizumi, T. Narita, K. Ohishi, S. Haga, F. Ohta, H. Nomoto, K. Nogata, T. Morishita, T. Endo, T. Shin-I, H. Takeda, S. Morishita, and Y. Kohara: The medaka draft genome and new insights into vertebrate genome evolution. **Nature** 446, 714-719 (2007)

14. K. T. Yamato, K. Ishizaki, M. Fujisawa, S. Okada, S. Nakayama, M. Fujishita, H. Bando, K. Yodoya, K. Hayashi, T. Bando, A. Hasumi, T. Nishio, R. Sakata, M. Yamamoto, A. Yamaki, M. Kajikawa, T. Yamano, T. Nishide, S-H. Choi, Y. Shimizu-Ueda, T. Hanajiri, M. Sakaida, K. Kono, M. Takenaka, S. Yamaoka, C. Kuriyama, Y. Kohzu, H. Nishida, A. Brennicke, T. Shin-i, Y. Kohara, T. Kohchi, H. Fukuzawa, and K. Ohyama: Gene organization of the liverwort *Y* chromosome reveals distinct sex chromosome evolution in a haploid system. **PNAS** 104, 6472-6477 (2007).

15. Jr-K. Yu, Y. Satou, N. D. Holland, T. Shin-I, Y. Kohara, N. Satoh, M. Bronner-Fraser & L. Z. Holland: Axial patterning in cephalochordates and the evolution of the organizer. **Nature** 445, 613-617 (2007)

16. N. Kobayashi, M. Watanabe, T. Kijimoto, K. Fujimura, M. Nakazawa, K. Ikeo, Y. Kohara, T. Gojobori, N. Okada: magp4 gene may contribute to the diversification of cichlid morphs and their speciation. **Gene** 373, 126-133 (2006)

17. E. Shoguchi, T. Kawashima, Y. Satou, M. Hamaguchi, T. Sin-I, Y. Kohara, N. Putnam, D. S. Rokhsar and N. Satoh: Chromosomal mapping of 170 BAC clones in the ascidian *Ciona intestinalis*. **Genome Research** 16, 297-303 (2006)

[学会発表] (計 5 件)

[その他]

ホームページ等

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>>

6. 研究組織

(1) 研究代表者

小原 雄治 (KOHARA YUJI)

国立遺伝学研究所・所長

研究者番号：70135292