

# 全長 cDNA の網羅的シーケンス法の検討

山口 勝司 (基礎生物学研究所 技術課)

YAMAGUCHI Katsushi : Comprehensive sequencing method for full-length cDNA

The next-generation DNA sequencer developed by PacBio utilizes two proprietary technologies: single molecule real-time sequencing (SMRT) and consensus circle sequencing (CCS). These innovations enable both long-read sequencing and high-precision sequencing, which were previously unattainable. This technology has been applied to RNA sequencing, known as ISO-Seq. While cDNA construction from RNA-seq using conventional short-read sequencers requires read assembly, full-length cDNA sequences can be determined at once with high accuracy using this method. However, the average length of cDNA is less than 2 kbase, shorter than that of a genome sequencing library, posing a challenge to cost performance. In this report, a solution to this problem is proposed: a simple ligation of cDNAs via blunt end ligation.

## 1. 目的

Pacbio 社の次世代 DNA シーケンサー Sequel IIe は独自技術である、1 分子リアルタイム sequencing (SMRT) と consensus circle sequencing (CCS) により、従来できなかった長鎖シーケンスと高精度シーケンスを両立させた。この長鎖・高精度シーケンスを RNA シーケンスに適用したのが、ISO-Seq 法として知られている。ISO-Seq 法は生物組織から単離した、全長の mRNA から SMARTer 法により全長の cDNA を作製し、これらをゲノムシーケンスと同様の流れで高精度にシーケンスする方法である。従来の RNA-seq に必要だったシーケンスリードの繋ぎ合わせが不要で、その過程でのミスアセンブルやスプライスバリエーションの不確実性を排除し、全長 cDNA 配列を精度良く決定できる。

一方、cDNA の長さは平均的には 2kb 弱程度で、機器のシーケンス可能な推奨長(15-20kb)と比較して短い。このためスペックに対する無駄があり、シーケンスのコストパフォーマンスが低下するため改善が望まれていた。近年 cDNA 合成の過程で接続突出末端を作製し、それらを連結して複数の mRNA 由来の配列を、1 本の配列として長鎖・高精度シーケンスする方法が報告されている(1)。本報告では cDNA を連結する方法として、もっと簡単にできる平滑末端ライゲーションを用いた方法を検討した結果を報告する。

## 2. 方法

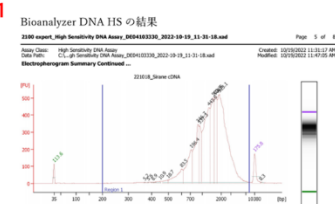
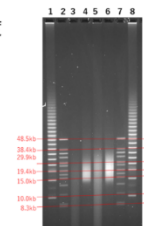
報告された突出末端を作製して連結する方法は、連結による cDNA の向きや順番も制御できる非常に有用な方法であるが、その為に固有配列を使った PCR 増

幅反応が必要など、手間となる過程が多い。そこでもっと簡潔な cDNA 連結法として、平滑末端ライゲーション法を試みた。一般的に広く利用されている SMARTer 法で作製した増幅 cDNA を、そのまま平滑末端化し、5'末端をリン酸化する。これらの反応は NEB 社の Quick Blunting Kit を用いた。次に同様に NEB 社の Blunt/TA Ligase Master Mix により平滑末端ライゲーションを進め、cDNA を連結させた。それらを切断化し、15-20kb 程度に連結された cDNA をサイズセレクト電気泳動で溶出し、常法に従いシーケンスライブラリの作製をおこない、シーケンスした (図 1)。

### 全長cDNAの連結

#### 全長cDNAの連結条件、ライブラリー化条件を検討

1. cDNA作成  
Takara SMARTer kit  
total RNA 1ugから  
逆転写後、PCR 16cycle  
21.7ug cDNA  
平均1,612base  
を確保
2. cDNAの平滑末端連結  
NEB quick blunting kit  
Blunt/TA ligation master mix
3. 連結cDNAの断片化  
megaruptorで20kbase程度  
に断片化
4. 通常のDNAと同様に  
ライブラリー化  
blue pippinで低分子を除去

1 3,5 bule pippin前ライブラリー  
4,6 bule pippin後ライブラリー  
274 ngの完成ライブラリー

シーケンスへ

15-20k DNAを入れたのと同様サイズの連結 cDNAライブラリーができた

図 1. シーケンスリード性能と絞り込み

シーケンス結果を SMARTer 法での PCR 増幅に用いた末端配列を目印に、自作のスクリプトで個別の cDNA に切り分け、cDNA 長や連結 cDNA 数の分布を解析した。また用いた生物種のゲノムに対し、マッピングツールである exonerate を用いてマップさせ、検証した。

### 3. 結果

連結された cDNA 数の分布を図 2 に、個別の cDNA サイズの分布を図 3 に示した。10 程度の cDNA が連結されたシーケンス結果が得られ、cDNA の連結が良好であると確認できた。一方、cDNA サイズの分布は一般的な cDNA サイズの分布に対し、やや短く偏る結果であった。これらの cDNA が全長であるかの評価が必要ではあるが、SMARTer 法での cDNA を用いている部分は変わらないところであるため、平滑末端ライゲーションにより、比較的短い cDNA が優先して連結されてしまうバイアスが考えられた。類似 cDNA を集約し、ゲノムにマップした結果を図 4 に示す。上段に上げる予測遺伝子の大部分を網羅する cDNA 構造が確認でき、発現遺伝子を大方網羅した cDNA が得られた。

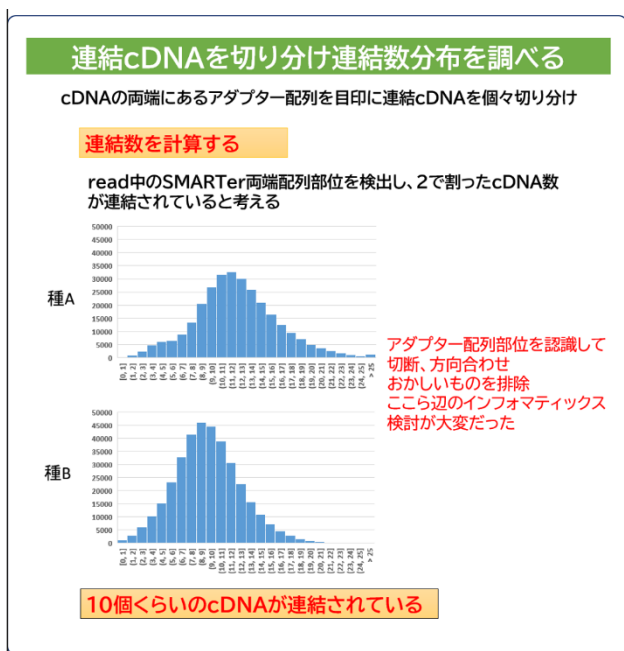


図 2. cDNA 連結数分布

### 4. 考察

本課題は 2021 年に bioRxiv に公開された以下のプレプリント (後に参考文献 1 の正式論文:

<https://www.biorxiv.org/content/10.1101/2021.10.01.462818v1>) を、より簡略化する手法開発を目的に進めた。結果、短い目の cDNA が優先されて連結される傾向はあ

るものの、大部分の cDNA をシーケンス出来ており、また安価・迅速に進められる点でこの手法の優位性は十分あると考えている。

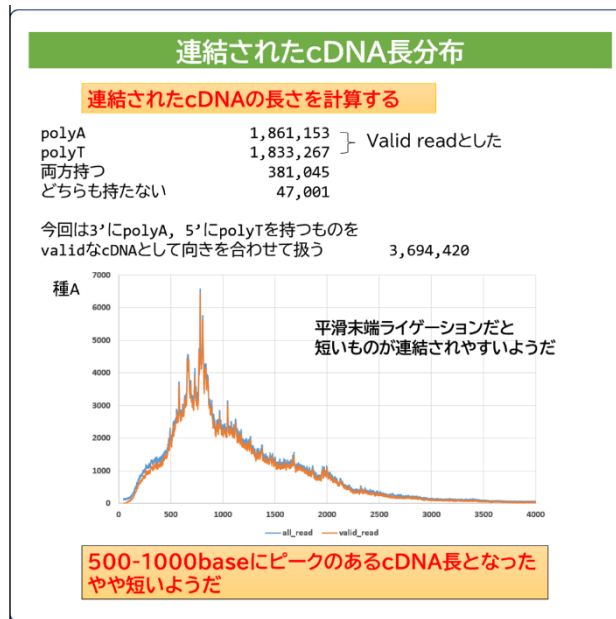


図 3. cDNA サイズ分布

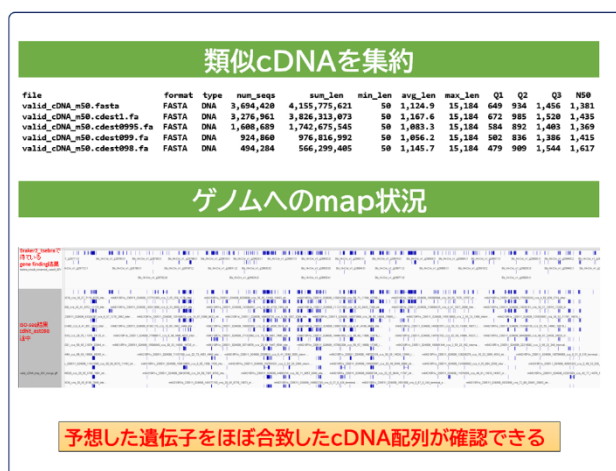


図 4. cDNA の多様性とゲノムへのマップ状況

### 謝辞

本発表は NIBB の統合ゲノミクス共同利用研究の複数課題の実験サンプルを用いておこなっている。共同利用研究の関係諸氏、トランスオミクス解析室の重信秀治教授、技術支援員の秋田朝日氏、松本美和子氏、浅尾久世氏、池田弥華氏に感謝申し上げる。

### 参考文献

1. Al'Khafaji AM et al. High-throughput RNA isoform sequencing using programmed cDNA concatenation. Nat Biotechnol. 2023 Jun 8. doi: 10.1038/s41587-023-01815-7. Epub ahead of print. PMID: 37291427.