



# 次世代シーケンサーへようこそ

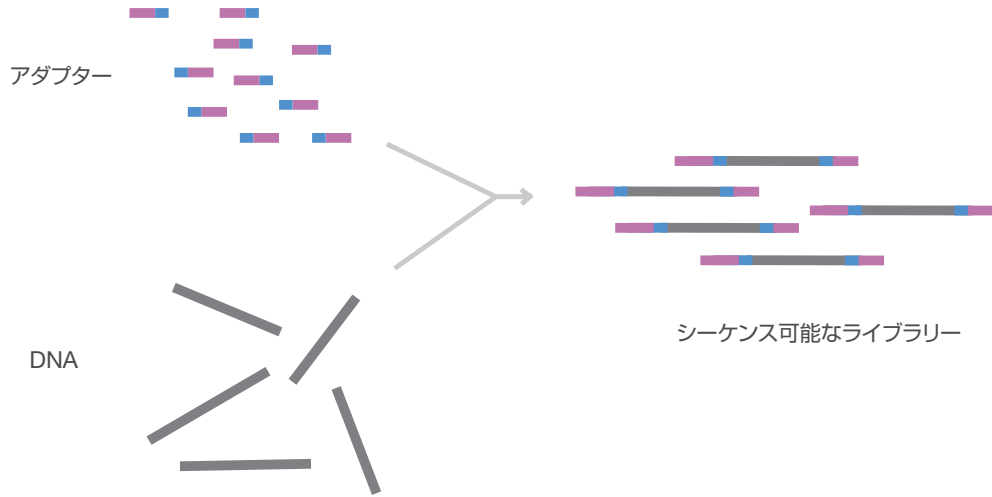
次世代シーケンサーの進歩は、ヒトの疾患研究から環境や進化の研究にまで、研究分野を大きく広げるのに役立っています。ゲノムサイズが小さく、比較的シンプルなデータ解析を行う微生物研究において、次世代シーケンサーは特に有用です。他のゲノム解析法とは異なる次世代シーケンサーを利用する最大の利点は、培養できない生物など予備知識が得られない場合でも、ゲノムのすべての位置において変化を測定できることです。1塩基レベルの分解能は、研究室で保存している株や環境中のサンプルなど、いずれにおいても短期間の微生物適応性の追跡を可能にします。

次世代シーケンサーの出力データ量は、このテクノロジーが誕生してから毎年倍以上という驚異的なペースで増加しており、次世代シーケンサーの急速な普及を証明しています。2007年は1回のランで10億塩基(1Gb)のデータを得ることができました。それが、2012年には1回のランで1000倍の1兆塩基(1Tb)のデータにまで飛躍しました。出力データ量が飛躍的に増加したことによって、微生物のゲノム解析にかかるコストは10万分の1に低下しました。1995年にキャピラリー電気泳動法を用いて180万塩基(1.8Mb)のインフルエンザ菌ゲノムを決定した際には約100万ドル近くのコストがかかり、時間は1年以上かかりました。現在、イルミナの次世代シーケンサーを使えば500万塩基(5Mb)の大腸菌ゲノムを1日で解析することができ、わずかなコストしかかかりません。

AGAAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATT  
TCAACGTACCGTAACGAACGTATCAATTGAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTGC AACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAA  
CCACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAG  
TAACGTACCATTAAAGAGCTACCGTGC AACGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTGC AACGACGAAAAGAAT  
AGAAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATT  
GATTACTTGATCCACTGATTCAACGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAAC  
CGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAAGAGCTACCGTGC AACGAAAAGAATGATA



図 2：次世代シーケンサー用ライブラリー調製の概要



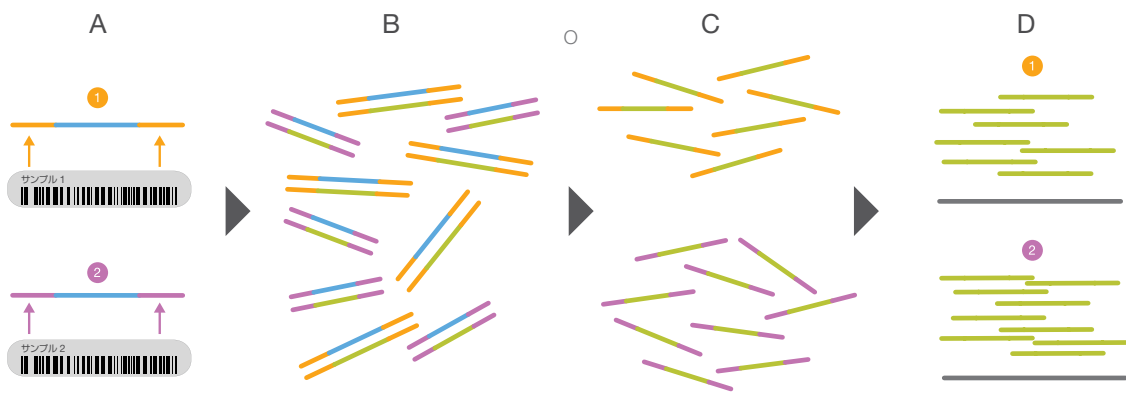
次世代シーケンサー用のライブラリー調製は、断片化したゲノムDNA から直接はじめます。シーケンス時のケミストリー、およびサンプルのインデックスに必要なイルミナシステム独自の特異的オリゴヌクレオチドアダプターを各DNA 断片の末端にライゲーションすることで、シーケンス可能なライブラリーが得られます。

## マルチプレックス法により可能となった拡張性のある研究

微生物やウイルスなどゲノムサイズが小さい生物種の解析を行う研究者は、デスクトップ型の次世代シーケンサーを用いて 1 ランあたりの解析サンプル数を少なくすることもできますし、高いスループットのシステムを活用して大量のサンプルを処理できます。マルチプレックス法は、1 回の実験で多数のサンプルを同時にシーケンスすることを可能にします（図 3）。このためには、各サンプルに固有の「バーコード」配列を付加することで、データ解析時にそれらを識別できるようにします。

マルチプレックス法により、次世代シーケンサーは多数のサンプルの解析においても、データ取得までの時間を劇的に短縮することができます。キャピラリー電気泳動を用いる場合、数百のアンプリコンのシーケンスには数週から数ヶ月かかりましたが、次世代シーケンサーでは同じ数のサンプルをわずか数時間でシーケンスでき、2 日以内に完全な解析が完了します。高度に自動化された使いやすいプロトコルを組み合わせると、実験から成果の発表までをこれまでよりも迅速かつ簡単に行えます。

図 3：マルチプレックス法の概要

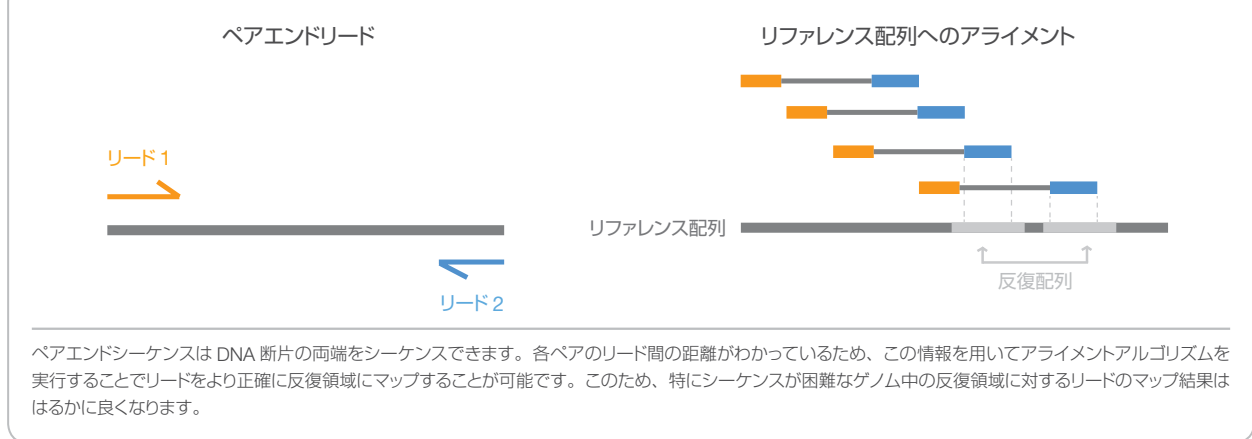


- A. 2 種類の異なるサンプルから得られる DNA 断片のそれぞれに、サンプルの由来を明らかにする固有のバーコードを付加します。
- B. 各サンプル由来のライブラリーを混合し、同時にシーケンスを行います。得られるそれぞれのリードには、DNA 断片の配列とサンプルの由来を明らかにするバーコード配列が含まれます。
- C. バーコード配列情報をもとに、マルチプレックスを分離またはシーケンス情報を識別します。
- D. 各リードセットをリファレンス配列にマップします。

## ペアエンドシーケンス

DNA 断片の両端からシーケンスするペアエンド (PE) シーケンスを用いることで、DNA 断片の位置決定が広範囲で可能になります (図 4)。各ペアにおけるリード間の距離がわかっているため、この情報を用いてアライメントアルゴリズムを実行することでリードを正確にアライメントできます。したがって、シーケンスが困難な、あるいは反復配列のあるゲノム領域に対しても優れたアライメントが可能になります。イルミナの次世代シーケンサーは、インサートサイズやリードの長さ (35 ~ 300bp) を柔軟に設定できるため、あらゆるゲノムに対して高解像度の解析が可能です。

図 4: ペアエンドシーケンスおよびアライメント



## イルミナの BaseSpace® クラウドでの解析、保管および共有

シーケンスアプリケーションを考えるうえで、データ解析は重要な要因です。次世代シーケンサーでは、高性能なコンピューター装置、大規模なデータストレージ、高いスキルをもったパイオインフォマティシャンや IT 担当者が必要であり、このことが大きな課題の一つになっています。イルミナでは、この課題を克服するための開発を継続的に行っています。現在のシステムでは、シーケンスランが進行すると同時に、アライメントや変異コールを含む複雑な一次データプロセスがスムーズに実行されています。アプリケーションに応じて、その後のデータ解析のほとんどをシーケンサー内蔵のコンピューターにインストールされている最適化されたソフトウェア、あるいはイルミナ独自のクラウドコンピューティング環境である BaseSpace から直接実行できます。基本的には画面操作によるインフォマティックスソリューションによって解析が簡略化されており、研究者は生物学に集中できます。BaseSpace ユーザーはデータをクラウド上に保管し、解析することで、近隣のあるいは世界中の共同研究者と直ちにデータ共有できます。BaseSpace アプリケーションストアは、多様な商用ソフトウェアツールや学術機関が提供する有名なオープンソースアルゴリズムへのスムーズなアクセスを提供します。これらのツールから得られる生物学的解釈と知見は研究を促進することでしょう。

## End-to-End ソリューション

イルミナの次世代シーケンサーは、アプリケーションに応じた特殊なライブラリー調製、実績があり安定したシーケンス試薬、そして広範囲のデータ解析を提供するシンプルなツールまで、DNA から解析結果の取得までを完全にサポートする唯一のソリューションです (図 5)。

図 5: MiSeq デスクトップ型シーケンサーを用いるイルミナの End-to-End 次世代シーケンサーワークフロー





## 例：de novo シーケンス

ゲノムサイズの小さい生物種のシーケンスでは、ほとんどの生物種でリファレンスゲノムが入手できないことが課題の一つになります。これは全ゲノムシーケンスを *de novo* で行い、リファレンス配列へのアライメントなしでリードをアセンブルしなければならないケースが多々あることを意味しています。ペアエンドリードの活用、そしてリード長が 300bp まで伸びたことにより、反復配列を含む領域のアライメントが向上しました。またコンセンサス配列のギャップをうめることにより *de novo* シーケンスで長いコンティグを生み出すことが可能になりました。その結果、より完全なゲノムのカバレッジが実現されました。キャピラリー電気泳動シーケンス法と比較して、次世代シーケンサーは時間とコストを抑えながら 1 回の実験で多くの株を同時に解析することが可能です (表 1)。

## ターゲットシーケンス

ターゲットシーケンスを行うことで、遺伝子サブセットあるいは特定のゲノム領域のみをシーケンスして、時間、経費、データストレージを最も興味のあるゲノム領域に集中することができます。アンプリコンシーケンスとは、数百塩基の長さには広がるゲノム上の選択した領域をシーケンスする手法です。最新の次世代シーケンサーのアンプリコンライブラリー調製キットは、ゲノム DNA の標的領域を溶液中で迅速に増幅することが可能です。この手法を用いることで、複数のサンプル由来の数千ものアンプリコンを同時に調製し、数時間のうちにインデックス化することができます。1 回のランで多数のアンプリコンとサンプルを処理する能力を備えることで、次世代シーケンサーは 1 回の実験で興味のあるゲノム内の領域をすべて同時に、従来のキャピラリー電気泳動法より少ない時間とコストで解析することが可能です。

表 2. ターゲットアプリケーションに対するイルミナの次世代シーケンサーとキャピラリー電気泳動を使ったサンガーシーケンス法の比較

項目	MiSeq システム	サンガーシーケンス法
プロジェクトのサンプル数	96	96
アンプリコン数	12	12
ターゲットパネルサイズ	~ 5kb	~ 5kb
ライブラリー調製時間	3 時間未満	3 時間未満
シーケンス時間	1 日	6 日
アンプリコンあたりのコスト*	約 150 円	約 500 円
プロジェクトのコスト*	約 25 万円	約 500 万円
アンプリコンあたりのカバレッジ深度	13,000 x 以上	2 x**
装置上でデータ解析は可能か?	できる	できない

\* PCR 増幅を除きます。

\*\* 双方向シーケンスを含みます。

## 例：16S メタゲノムシーケンス

微生物の 16S リボソーム RNA (rRNA) 遺伝子の比較は、系統学や分類学の研究で広く用いられる一般的なアンプリコンシーケンス手法です。この手法は、多くの環境下にある微生物の多様性の評価に使用することで、他の手法では研究が困難または不可能なサンプル由来のマイクロバイオームの解析を可能にします。次世代シーケンサーは、数千もの生物種を同時にシーケンスする能力を備えることで特にこのアプリケーションに適しています。次世代シーケンサーは混合したサンプルの場合でも、1 回のランで高いシーケンスカバレッジを得ることができるため、キャピラリー電気泳動シーケンス法を使用した場合には見逃されたり、費用がかかりすぎるような希少な変異株を同定することが可能です (表 2)。表 2 で示された例は、比較を目的としたわずか 96 のサンプルに基づいていることに注意してください。数百あるいは数千種類のゲノムを比較する実際のメタゲノムシーケンス研究にキャピラリーシーケンスを用いるとコストや労力がかかりすぎますが、イルミナの HiSeq システムや NextSeq™ システムのようなハイスループット次世代シーケンサーを用いるとその研究を実現することができます。

## 次のレベルの研究へ

次世代シーケンサーの登場によって、以前は不可能であったレベルでの生物システムを研究できるようになりました。キャピラリー電気泳動を用いたサンガーシーケンス法よりも明らかに有利な点によって、次世代シーケンサーは微生物研究を一変させ、そして新しい探求の道を開くことができるでしょう。www.illumina.co.jp であなたの研究に最適なシーケンスプラットフォームを見つけてください。

## 革新から発表まで

次世代シーケンサーテクノロジーは進化し続けており、研究者は数多くの生物学分野で魅力的な発見を生み出し、すべての研究分野でこれまで不可能であった答えをもたらしています。その結果、多くの雑誌に論文が発表されており、イルミナのシーケンステクノロジーに関する論文は 5,500 報を超えています。微生物学に関連した最近の論文例を以下に掲載します。

### 全ゲノムシーケンス

1. Toprak E, Veres A, Michel JB, Chait R, Hartl DL, et al. (2011) Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat Genet* 44: 101–105.
2. Chua, KYL, Seemann T, Harrison PF, Monagle S, Korman TM, et al. (2011) The dominant Australian community-acquired methicillin-resistant *Staphylococcus aureus* clone ST93-IV [2B] is highly virulent and genetically distinct. *PLoS ONE* 6:
3. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, et al. (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* 30 (5): 434–9.

### de novo シーケンス

4. Chitsaz H, Yee-Greenbaum JL, Tesler G, Lombardo MJ, Dupont CL, et al. (2011) Efficient *de novo* assembly of single-cell bacterial genomes from short-read data sets. *Nat Biotechnol* 29: 915–921.
5. Rodrigue S, Malmstrom R, Berlin A, Birren B, Henn M, et al. (2009) Whole genome amplification and *de novo* assembly of single bacterial cells *PLoS One* (4) 9 e6864.

### メタゲノミクス

6. Caporaso JG, Lauber CL, Walkers, WA, Berg-Lyons D, Lozupone CA et al. (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci USA* 108:4516–22.
7. Mackelprang, R, Waldrop MP, DeAngelis KM, David MM, Chavarria KL, et al. (2011) Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* 480: 368–371.

## イルミナ株式会社

〒108-0014  
東京都港区芝5-36-7 三田ベルジュビル 22階  
Tel (03) 4578-2800 Fax (03) 4578-2810  
www.illumina.co.jp

代理店

本製品の使用目的は研究に限定されます。

© 2014 Illumina, Inc. All rights reserved.

Illumina, IlluminaDx, BaseSpace, BeadArray, BeadXpress, cBot, CSPro, DASL, DesignStudio, Eco, GAlIx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, NextSeq, NuPCR, SeqMonitor, Solexa, TruSeq, TruSight, VeraCode, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks or registered trademarks of Illumina, Inc.

その他の会社名や商品名は、各社の商標または登録商標です。予告なしに仕様を変更する場合があります。

Pub. No. 770-2012-J023 12MAY2014

illumina®