

NGSの新たな利用法：16S rRNAメタゲノム解析の ポイント プロトコールのご紹介

May 9, 2014



小林 孝史
イリミナ株式会社
テクニカルアプリケーションサイエンティスト

© 2012 Illumina, Inc. All rights reserved.

Illumina, illuminaDx, BaseSpace, BeadArray, BeadXpress, cBot, CSPPro, DASL, DesignStudio, Eco, GAllx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, Sentrix, SeqMonitor, Solexa, TruSeq, VeraCode, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks or registered trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners.


illumina®



ILLUMINA'S 16S rRNAメタゲノム解析のポイント プロトコールのご紹介

本資料はwebinar後、下記リンク上にアップロードいたします。

http://www.illumina.co.jp/events/webinar_japan.ilmn?ws=ss



イルミナ株式会社
テクニカルアプリケーションサイエンティスト
小林孝史

微生物を集団で解析する＝メタゲノム解析

さまざまなアプリケーションに応用！

土壌・水質サンプル中の微生物解析

発酵食品中に含まれる微生物ゲノム解析

食品中に含まれる混入微生物のゲノム解析

下水道、工場敷地内の土壌などからの微生物解析

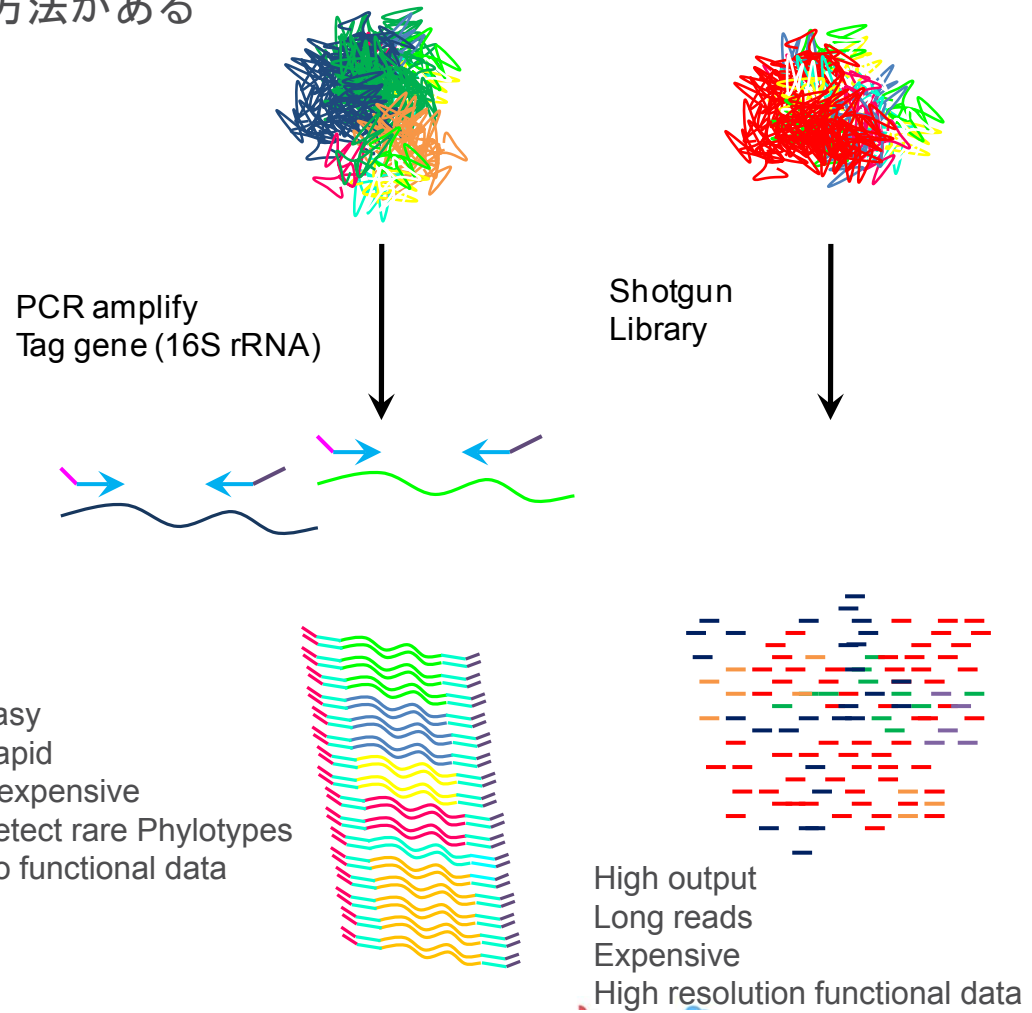
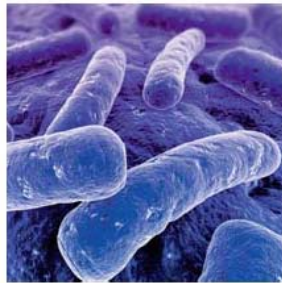
化石サンプルなどから古（微）生物解析

体内臓器・糞便に含まれる微生物解析



メタゲノムアプリケーション

- ▶ 16Sを解析する方法（16Sシーケンス解析）と、
- ▶ ショットガンで全ゲノム解析をする方法がある



なぜ16S rRNAシーケンスを行うのか

- ▶ 16S rRNAはすべてのバクテリアで持つ配列である
- ▶ 16S rRNAは多様性を持つ配列である。
- ▶ 従って、16S rRNA配列を調べることにより、その情報から含まれる微生物を検出・同定することができる（菌叢解析）

Proc. Natl. Acad. Sci. USA
Vol. 74, No. 10, pp. 4537-4541, October 1977
Evolution

Fox GE et al. (1977) PNAS

Classification of methanogenic bacteria by 16S ribosomal RNA characterization
(comparative oligonucleotide cataloging/phylogeny/molecular evolution)

GEORGE E. FOX*[†], LINDA J. MAGRUM*, WILLIAM E. BALCH[‡], RALPH S. WOLFE[‡], AND CARL R. WOESE*[‡]

Departments of *Genetics and Development and [†]Microbiology, University of Illinois, Urbana, Illinois 61801

Communicated by H. A. Barker, August 10, 1977

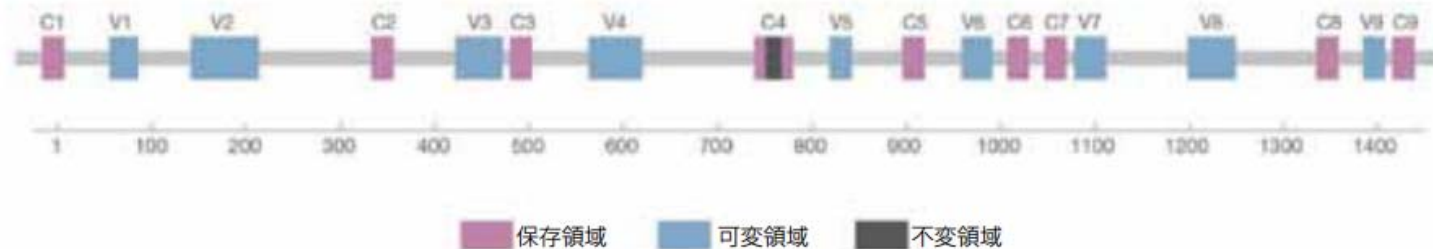
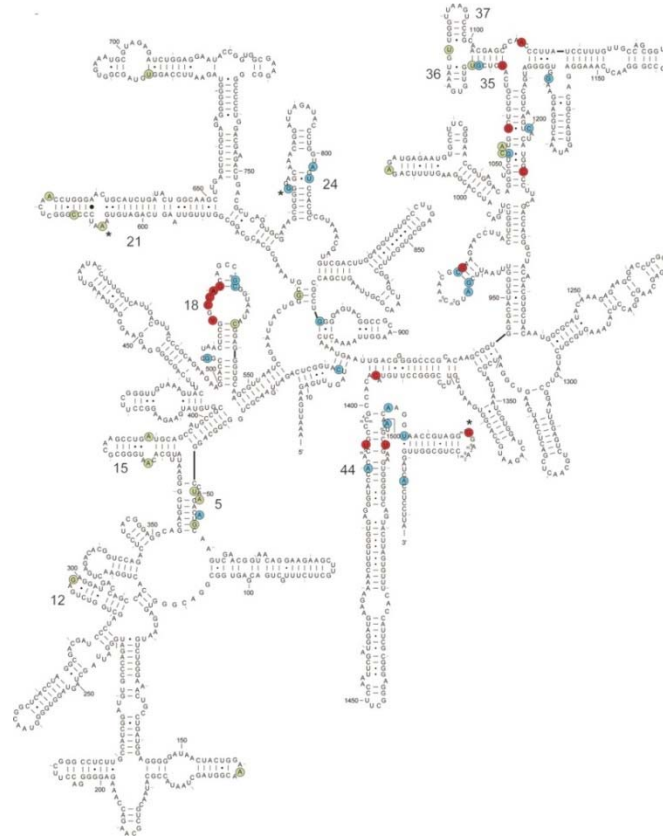
ABSTRACT The 16S ribosomal RNAs from 10 species of methanogenic bacteria have been characterized in terms of the oligonucleotides produced by T₁ RNase digestion. Comparative analysis of these data reveals the methanogens to constitute a distinct phylogenetic group containing two major divisions. These organisms appear to be only distantly related to typical bacteria.

residues represented by hexamers and larger in catalog A and in catalog B and their overlap of common sequences, respectively. The association coefficient, S_{AB} , so defined provides what is generally an underestimate of the true degree of homology between two catalogs because related but nonidentical oligomers are not considered. The matrix of S_{AB} values for each binary comparison among the members of a given set of



16S rRNAシーケンス

- ▶ 16S rRNAは1,542塩基長の原核生物のリボソームRNAの一部である。
(真核生物の場合は18Sとなる)
- ▶ 16S rRNAは種を超えていくつかの保存性の高い領域を持つ。その中で系統樹的に変異が遺伝されている。
 - 従って、この領域を解析することにより含まれる微生物の種類を遺伝学的に同定できる。
- ▶ 微生物の同定やゲノム解析に対して非常に有用なツールとして有用である。



16S rRNA遺伝子の模式図。基準細菌16S rRNAの可変領域 (青色) および保存領域 (紫色) の位置。
灰色部分は、すべての細菌において不変な領域。



Illumina公式 16S rRNA シーケンスプロトコール



ワークフロー

16S rRNA について最適化されたアプリケーション



ターゲット領域とプライマーをどこに設計するか

1)

Open

The ISME Journal (2012), 1–5
© 2012 International Society for Microbial Ecology. All rights reserved. 1712-0202/12
www.ismej.com/content/12

SHORT COMMUNICATION

Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences

David AW Soergel¹, Neelendu Dey^{2,3,4}, Rob Knight³ and Steven E Brenner¹

¹Department of Plant and Microbial Biology, University of California, Berkeley, CA, USA; ²Division of Gastroenterology, Department of Medicine, University of California, San Francisco, CA, USA and ³Howard Hughes Medical Institute and Department of Chemistry and ⁴Biochemistry, University of Colorado, Boulder, CO, USA

Microbial community profiling using 16S rRNA gene sequences requires accurate taxonomy as algorithms. Universally primers target conserved sequences and amplify sequences from many taxa, but they provide variable coverage of different environments, and regions of the rRNA gene differ in taxonomic informativeness—especially when high-throughput short-read sequencing technologies (for example, 454 and Illumina) are used. We introduce a new evaluation procedure that provides an improved measure of expected taxonomic precision when classifying environmental sequence reads from a given primer. Applying this measure to thousands of combinations of primers and read lengths, simulating single-ended and paired-ended sequencing, reveals that these choices greatly affect taxonomic informativeness. The most informative sequence region may differ by environment, partly due to variable coverage of different environments in reference databases. Using our Rax method of classifying paired-end reads, we found that paired-end sequencing provides substantial benefit in some environments including human gut, but not in others. Optimal primer choices for short reads totaling 96 nt provides 82–100% of the confident genus classifications available from longer reads. The ISME Journal advance online publication: 12 January 2012; doi:10.1089/ismej.2011.208

Subject Category: microbial population and community ecology

Keywords: 16S ribosomal RNA; taxonomy; phylogeny; classification; bacteria; sequencing

Variation in 16S ribosomal gene sequences has been used since the mid-1980s to characterize microbial diversity (Stahl *et al.*, 1984). Interest in sequence-based surveys of environmental microbes has exploded in recent years with the availability of sequencing technologies that produce ever-larger data sets at ever-decreasing cost; in particular, the Illumina platform is attractive because of throughput, despite its short reads (Sogin *et al.*, 2006; Lazarovic *et al.*, 2009; Claesson *et al.*, 2012; Caporaso *et al.*, 2011; Dugan and Ochman, 2012). Here, we examine the reliability of assignment of novel sequences to known taxa under thousands of simulated conditions, varying primer choice, read length and environment.

Previous work on taxonomic classification of environmental 16S rRNA gene sequences has focused on whether reference sequences matching a given query share taxonomic annotations (Jonsson *et al.*, 2002; Desnais *et al.*, 2006; Sogin *et al.*, 2006; Wang *et al.*, 2007). Validations of taxonomic classifiers have typically compared a limited range of primers, read lengths and environments (Stundquist *et al.*, 2007; Huse *et al.*, 2008; Liu *et al.*, 2008; Wu *et al.*, 2008; Hays *et al.*, 2009). Reference databases contain many sequences derived from some environments and few associated with others, however (Supplementary Figure S1), leading to substantial variation in classification quality. In addition, the use of ‘best-ones-out’ cross-validation at the sequence level (Stundquist *et al.*, 2007; Wang *et al.*, 2007; Liu *et al.*, 2008; Wu *et al.*, 2008)—where a single sequence with a known annotation is held out from a reference database and classified using the remainder—is problematic: reference sequences matching best-of-sets sequences are likely to originate from the same sample, because natural environments contain ‘microdiverse’ clusters of closely related strains (Achilles *et al.*, 2004). We addressed these issues by simulating taxonomic reads from slight large environmental data

Correspondence: DAW Soergel, Current address: Department of Computer Science, University of Massachusetts, 140 Governors Drive, Amherst, MA 01003-0004, USA. E-mail: soergel@cs.umass.edu

Current address: Center for Genome Sciences and Systems Biology, Washington University School of Medicine, Saint Louis, MO 63108, USA.

Current address: Division of Gastroenterology, Department of Internal Medicine, Washington University School of Medicine, Saint Louis, MO 63108, USA.

Received 20 March 2011; revised 21 November 2011; accepted 3 December 2011

2)

Published online 28 August 2012

Nucleic Acids Research, 2013, Vol. 41, No. 1, e1
doi:10.1093/nar/gks808

Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies

Anna Klindworth^{1,2}, Elmar Pruesse^{1,2}, Timmy Schweer¹, Jörg Peplies³, Christian Quast¹, Matthias Horn¹ and Frank Oliver Glöckner^{1,2,4}*

¹Max Planck Institute for Marine Microbiology, Microbial Genomics and Bioinformatics Research Group, Celsiusstr. 1, 28359 Bremen, ²Jacobs University Bremen, School of Engineering and Sciences, Campus 1, 28759 Bremen, ³Ribicon GmbH, D-28359 Bremen, Germany and ⁴Department of Microbial Ecology, University of Vienna, Althanstr. 14, 1090 Vienna, Austria

Received December 12, 2011; Accepted July 31, 2012

ABSTRACT

16S ribosomal RNA gene (rDNA) amplicon analysis remains the standard approach for the cultivation-independent investigation of microbial diversity. The accuracy of these analyses depends strongly on the choice of primers. The overall coverage and phylum spectrum of 175 primers and 512 primer pairs were evaluated *in silico* with respect to the SILVA 16S/18S rDNA non-redundant reference dataset (SSURF 108 NR). Based on this evaluation a selection of ‘best available’ primer pairs for Bacteria and Archaea for three amplicon size classes (100–400, 400–1000, >1000 bp) is provided. The most promising bacterial primer pair (S-D-Bact-0341-b-S-17/S-D-Bact-0785-a-A-21), with an amplicon size of 464 bp, was experimentally evaluated by comparing the taxonomic distribution of the 16S rDNA amplicons with 16S rDNA fragments from directly sequenced metagenomes. The results of this study may be used as a guideline for selecting primer pairs with the best overall coverage and phylum spectrum for specific applications, therefore reducing the bias in PCR-based microbial diversity studies.

INTRODUCTION

Understanding microbial diversity has been the ambition of scientists for decades. Because diversity analysis by cultivation is problematic for a significant fraction of Bacteria and Archaea, culture-independent surveys have been developed. In the past, the most commonly used approach was cloning and sequencing of the 16S ribosomal RNA gene (rDNA) using conserved broad-range

PCR primers (1). With the advent of massive parallel sequencing technologies, direct sequencing of PCR amplicons became feasible (2–4). In 2006, Roche’s 454 GS 20 pyrosequencing (5) became the first high-throughput sequencing technology to be successfully applied for large scale biodiversity analysis and was key to uncovering the ‘rare biosphere’ (6). The continuous development of the technology, offering read lengths of up to 1000 bp nowadays, further improved throughput and resolution of 16S rDNA sequencing (7). Since then, additional high-throughput sequencing technologies have become commercially available. The attractiveness of Illumina (8) lies in the reduced per base costs and comparatively high sequencing depth (9), despite having short read lengths. While the major advantage of Ion Torrent (10) are its relatively low cost and rapid sequencing speed. Furthermore, Pacific Bioscience (PacBio) now employs the ‘single-molecule real-time’ (SMRT) sequencing technology, designed to achieve average read lengths of more than 3000 bp (11). For a detailed review of sequencing technologies please refer to Loman *et al.* (12). There is no doubt that the rapid development of sequencing technologies has opened a new dimension in biodiversity analysis, but the diversity of technologies also adds complexity to the experimental design of a study. The most critical step for accurate rDNA amplicon analysis is still the choice of primers (4,13). Using suboptimal primers, or more precisely, primer pairs, can lead to under-representation (14) or selection against single species or even whole groups (15–17). Using inappropriate primers consequently leads to questionable biological conclusions (17–19).

In this study, 175 broad range 16S rDNA primers and 512 primer pairs were investigated *in silico* with respect to overall coverage and phylum spectrum for Bacteria and Archaea as well as amplicon length. Primer sequences were compared with all 376 437 16S/18S rDNA sequences

*To whom correspondence should be addressed. Tel: +49 421 2028970; Fax: +49 421 2028580; Email: fog@mpi-bremen.de

© The Author(s) 2012. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by-nc/2.0/>), which permits non-commercial reuse, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com.

1) David AW Soergel, et al., Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *The ISME Journal* (2012), 1–5

2) Klindworth A, et al., Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* (2013), 41(1)



16S Metagenomics Sequencing Library Preparation

ILLUMINA公式プロトコール

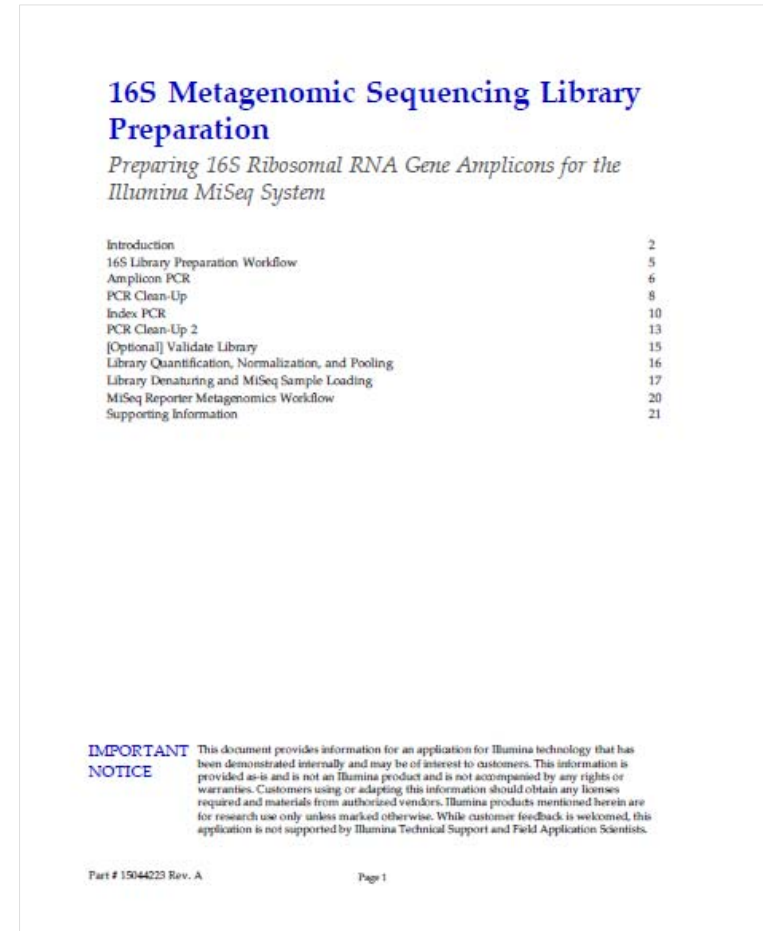
▶ イルミナで公式にリリースした

16S rRNA 菌叢解析プロトコール

- 16S解析用のライブラリー調製プロトコール
- **V3-V4領域を含む460bp**を増幅解析
(以前のプロトコールはV4領域のみ254bp)
- MiSeq V3 kit 2x301 Cycleでのランを想定
(Read長を考慮してV3キットを推奨)
- MiSeq ReporterとBaseSpace (coming soon)
でMetagenomicsワークフローを使用する

▶ ほかのターゲット領域にも使用できる プロトコール

- **領域特異的なプライマーを用いたTailed-PCR
法**
- 公式のユーザーガイドにはプライマーのTm
値も記載されているので他の領域の増幅にも
適応可能です



プロトコール(pdf): <http://res.illumina.com/documents/products/appnotes/16s-metagenomic-library-prep-guide.pdf>



プロトコール概要:ユーザーガイドを用意しています

プライマーを
注文する

V3とV4領域を含むプライマーにより
25 cycle PCRで460 bpの断片を増幅する。

ライブラリー
調製

Nextera XT index kitを使用して
8 cycle PCRでアダプターを付加する

MiSeq で
シーケンス

384サンプルの同時解析 (MiSeq 1ラン) で、高いクオリティの
全長V3, V4領域のオーバーラップリードがサンプルごと35,000-
60,000リード得られる。(MiSeq V3 kit使用)

MSR あるいは
BaseSpaceで
2次解析

MiSeq Reporterのメタゲノムワークフローを使用すれば
Greengenesデータベースを参照にした分類学的同定が可能。
属・種のレベルまでグラフを含むレポートとして得られる。



16S プロトコール用プライマー配列

V3-V4領域を解析するためのプライマー情報

Forward Primer:

5' **TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG****CCTACGGGNGGCWGCAG**

オーバーハング配列

16S V3-V4領域に特異的な配列

Reverse Primer:

5' **GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG****GACTACHVGGGTATCTAATCC**

緑の配列はTailed-PCRに必ず必要な配列

赤の配列は領域特異的な配列を示す
(16S V3-V4領域内や、対象のアンプリコン領域)

プライマー精製レベルはカートリッジ精製程度で構わない



任意のターゲット領域を増幅させるためのプライマー

16Sの他の領域を増幅させる

Forward Primer:



Reverse Primer:

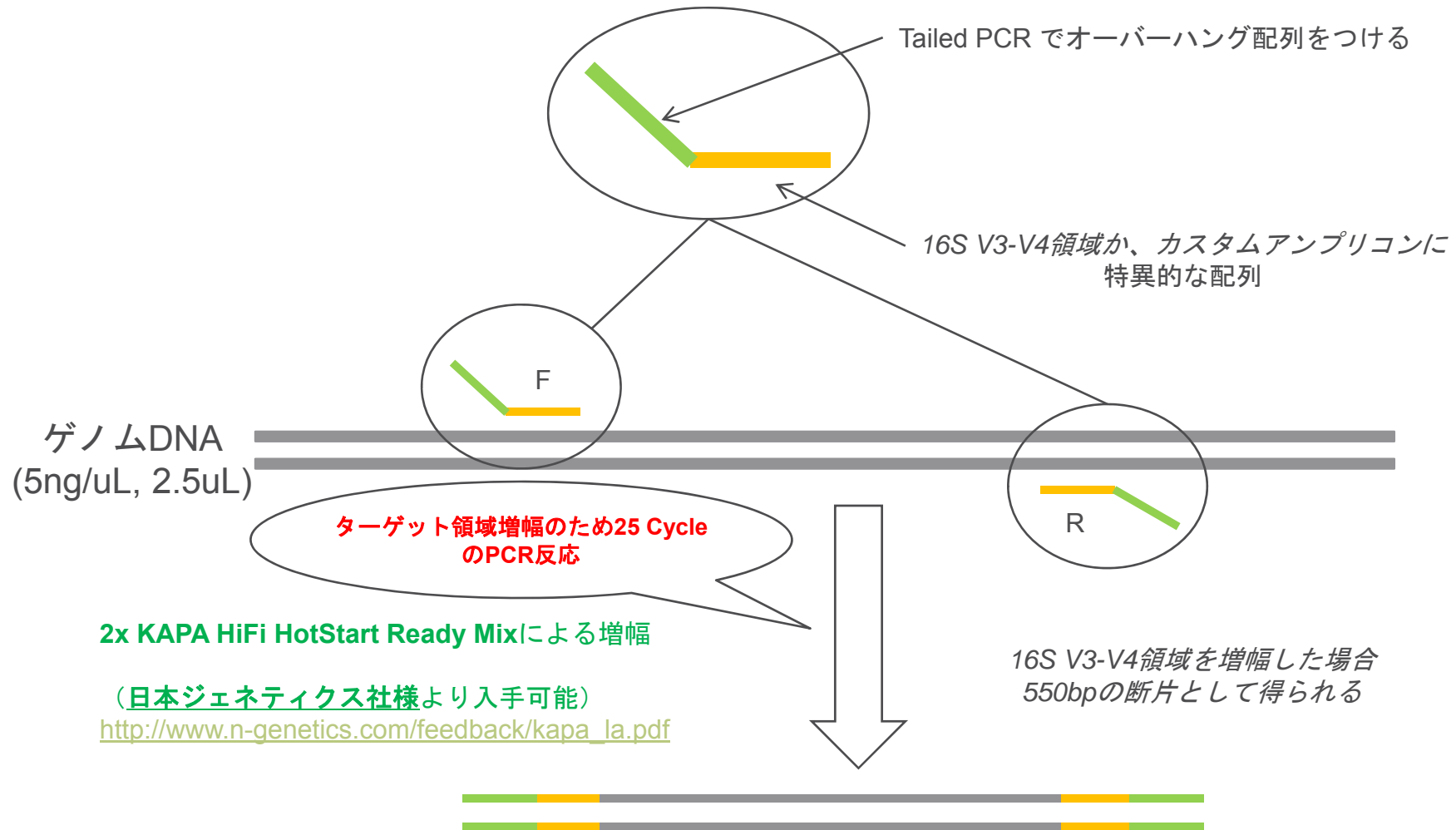


プライマーの5'末端に緑の配列を付加する。ターゲット領域は特定の領域を示す。
ターゲット領域<~550bp PCR断片



Step 1: 目的の領域を増幅する 1 番目のPCR

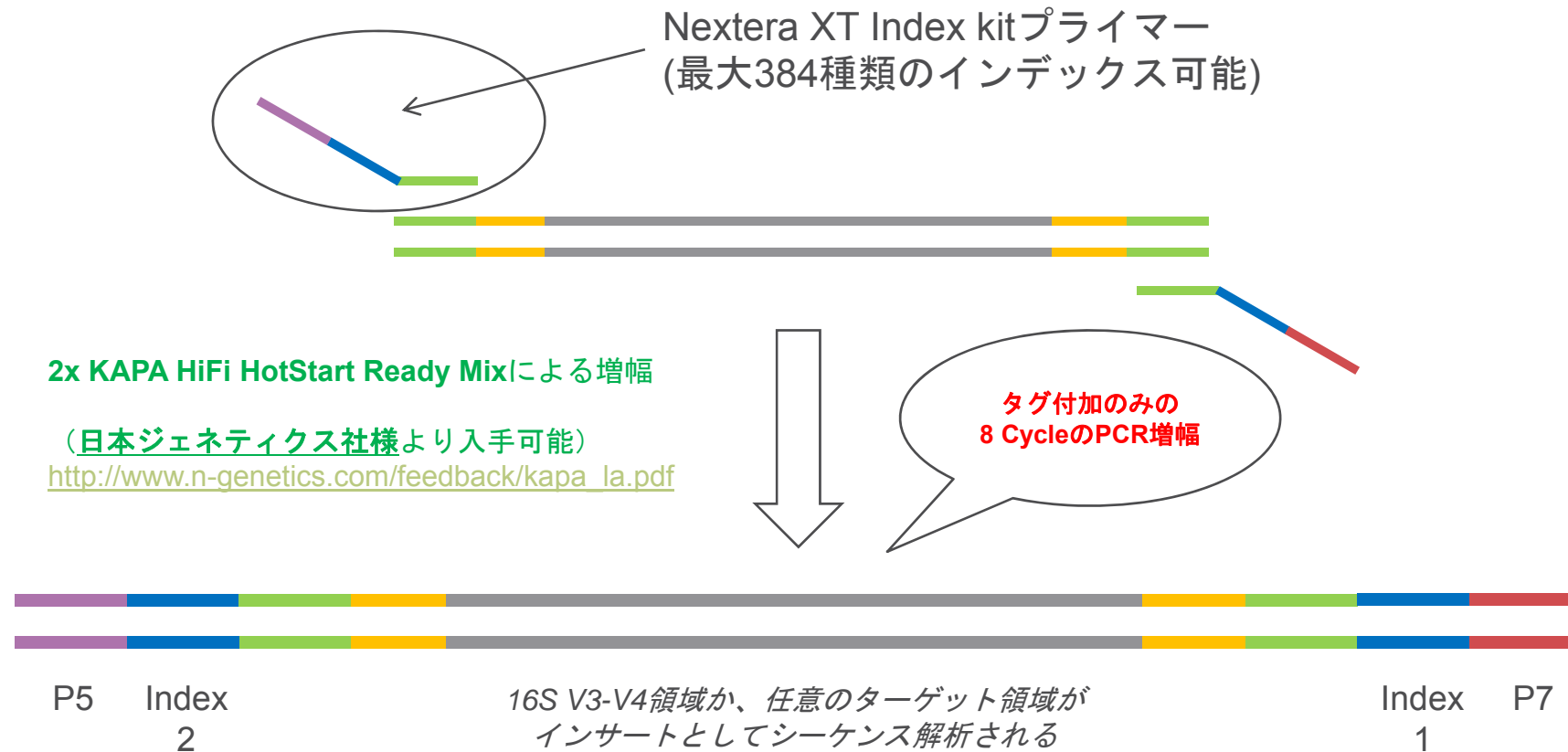
5'末端にオーバーハング配列を含むプライマーを使用



50uLのPCRプロダクト(余剰プライマーを除く AMPure Beads精製後)のうち、5uLを次のPCR反応に用いる



Step 2: イルミナNextera XT Index Kit を使用して、シーケンスに必要なアダプター配列を付加する



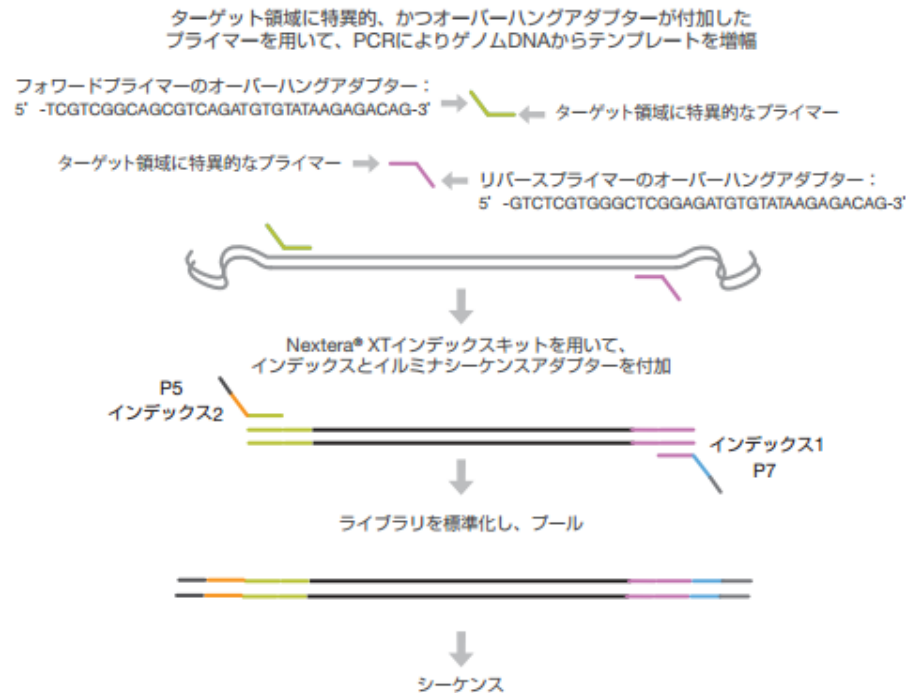
16S V3-V4領域を増幅した場合630bpのライブラリーが調製される

余剰プライマーを除くAMPure Beads精製後25uLの溶液として産物が得られるため、一部を変性してプールする。



全体のワークフローまとめ

16S rRNA シーケンス



(参照) アプリケーションノート

http://www.illumina.co.jp/documents/pdf/appnote_miseq_16s-j.pdf

MiSeq®システムで高速にマルチプレックス16Sアンプリコンシーケンス

わずか2日間のシンプルなワークフローで複雑な微生物群集を属レベルまで分類

はじめに

微生物群集は、深層土壌から海底の熱水噴出孔といったさまざまな環境下で生存しています。またヒトや動物には、体内、体外を問わず多様な種類の微生物が生息しています。メタゲノミクスとは、このような微生物群集から分離した遺伝物質、すなわちメタゲノムに関する研究です。メタゲノミクス調査は通常、原核生物の16SリボソームRNA (rRNA) 遺伝子を対象に行われます。この遺伝子は全長が約1,500塩基で、9つの散在した可変領域と保存領域から構成されています。この領域をシーケンスすることで、微生物を分類することができます。次世代シーケンサーを使い、16S rRNA遺伝子の特定領域をターゲットとしてシーケンスすることで、多様な微生物群集を分類することが可能になります (図1)。

このApplication Noteでは、デスクトップ型のMiSeqシステムを用いた16Sアンプリコンシーケンスの手法と、そこから得られる結果をご紹介します。研究者は、イルミナで確立されたアンプリコンシーケンス用のプロトコルに基づいて、ご希望の分類レベルに応じた、1つ、もしくは複数のゲノム領域をターゲットとするプライマーを設計します。この手法はウイルス研究や変異の検出、または他の微生物関連の研究といった、どのような標的アンプリコンシーケンスでもお使いいただけます。

手法

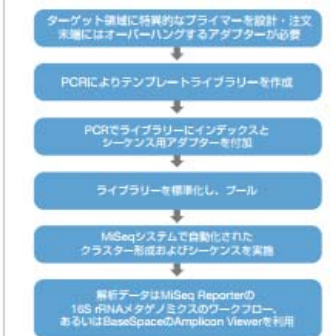
ターゲットの増幅方法

イルミナがご紹介する16S rRNAシーケンスのプロトコルは16S遺伝子のV3およびV4領域をターゲットとし、459塩基対のアンプリコンが得られます (生物により異なる) (図2)。プライマーはターゲット領域周辺の保存性の高い領域をもとに設計されています。MiSeqは250塩基のペアエンド解析ができるので、それぞれのリード末端をオーバーラップさせることで、非常に高品質のV3およびV4領域の全長リードが、たった40時間のランで得られます。そして低サイクルPCRで、イルミナシーケンスアダプターとデュアルインデックスバーコードがアンプリコンに付加されます。このライブラリを定置標準化およびプールしたのち、シーケンスを開始します。この手法は16S遺伝子のさまざまな領域、その他の興味ある遺伝子にもご利用いただけます。MiSeqシステムでは、マルチプレックスで、96サンプルを1ライブラリにまとめ、シーケンスすることが出来ます。

シーケンス

インデックスをつけたアンプリコンをMiSeq試薬カートリッジに投入し、フローセルと共にMiSeqシステムにセットします。その後、自動化されたクラスター形成とデュアルインデックスリードを用いたペアエンドシーケンスを行います。2x250bpを解析するには合計40時間がかかります。

図1: 16Sアンプリコンシーケンスのワークフロー



16Sアンプリコンシーケンスは、MiSeqシステムでお使いいただける確立されたプロトコルです。サンプル調製、シーケンスに加え、装置付属のソフトウェアによる自動データ解析からなっており、分析結果の検討まで行えます。

データ解析

1次解析 (画像解析およびベースコール) はMiSeqシステム上で行います。ランの終了後は、イルミナのフリーのソフトウェア MiSeq Reporter (MSR) を利用して2次解析を実施することができます。MiSeqシステム上またはイルミナの提供するクラウド型ソリューションのBaseSpace、その他ネットワークに接続していないコンピュータ上でもご利用頂けます。このMSRによりメタゲノム群集から属レベルのデータが得られ、それぞれの属が存在する割合まで把握できます。

結果

16S rRNA遺伝子配列と高頻度可変領域は、数多くの生物種で既に決定されています。これらの配列はGreengenes⁴やRibosomal Database Project⁵など複数のデータベースで入手可能です。分類においては、16S rRNA遺伝子の全長の代わりに、高頻度可変領域のみをシーケンスすることで十分な情報が得られます^{6,7}。MSRのメタゲノム16S rRNAワークフローでは、分類のためにGreengenesデータベースを使用しています。

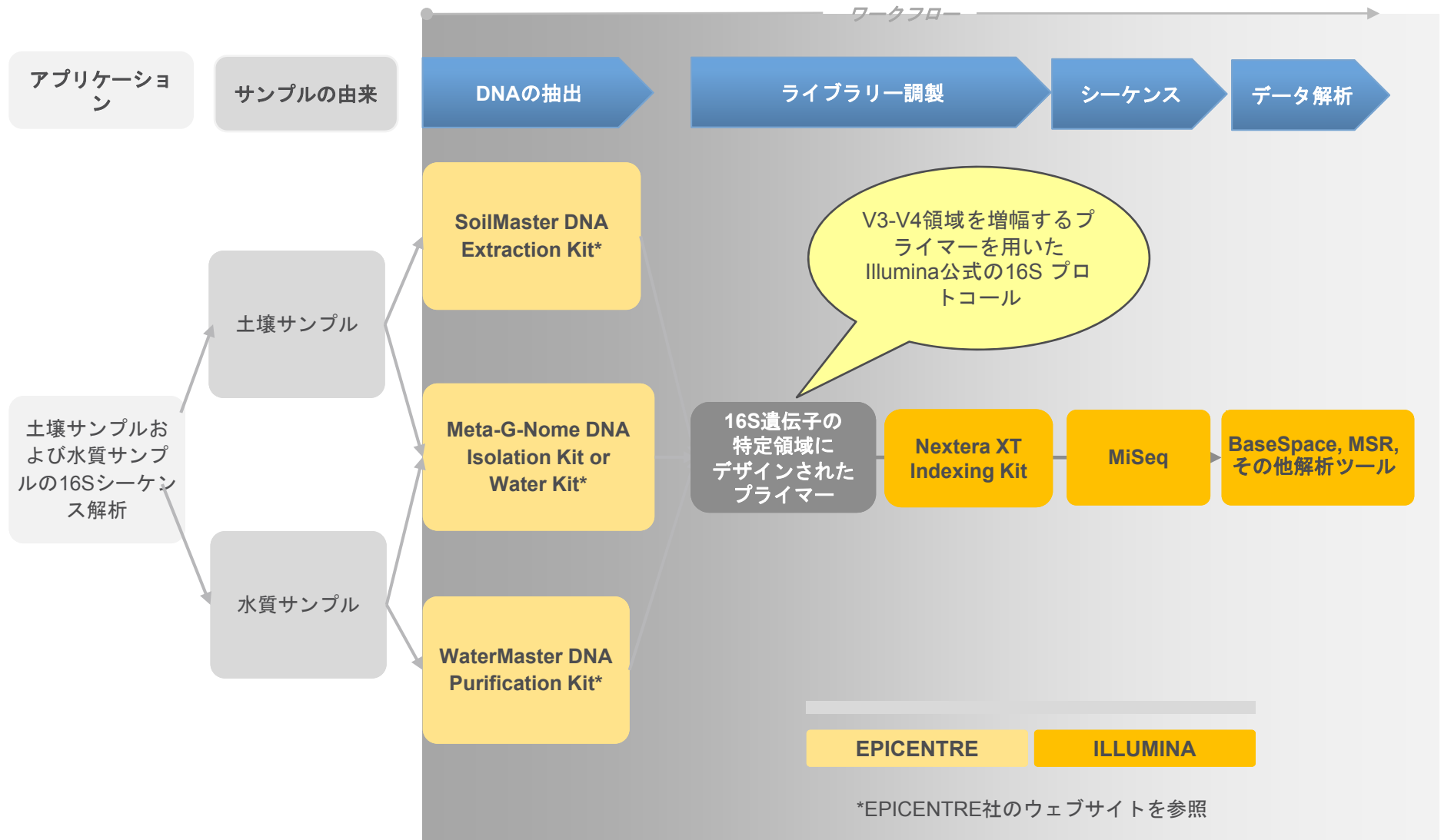


16S rRNAシーケンス、サンプルの抽出方法



どの製品を使用するか？

環境サンプルからDNAを抽出して16Sシーケンスを行う場合

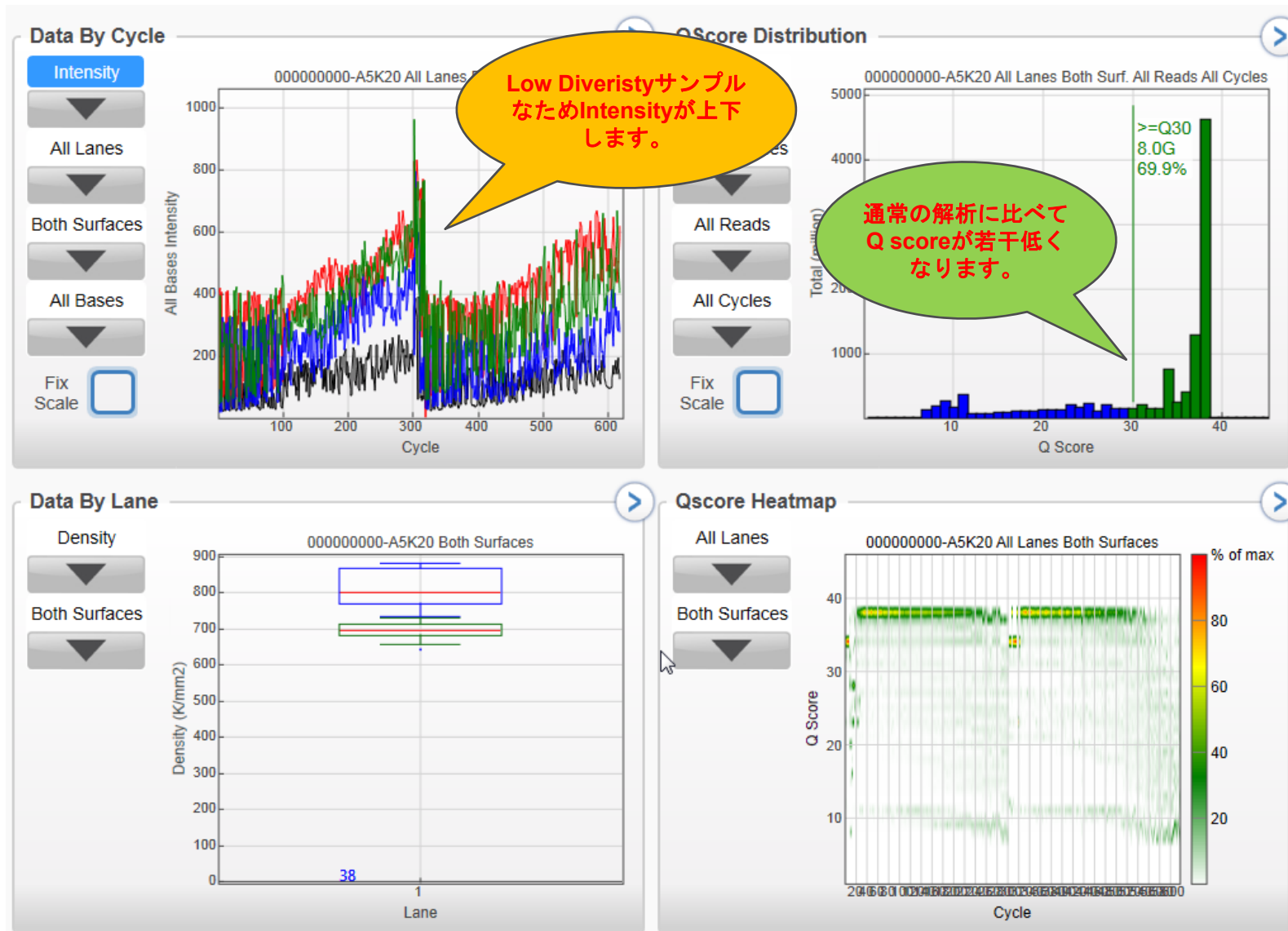


どの製品を使用するか？

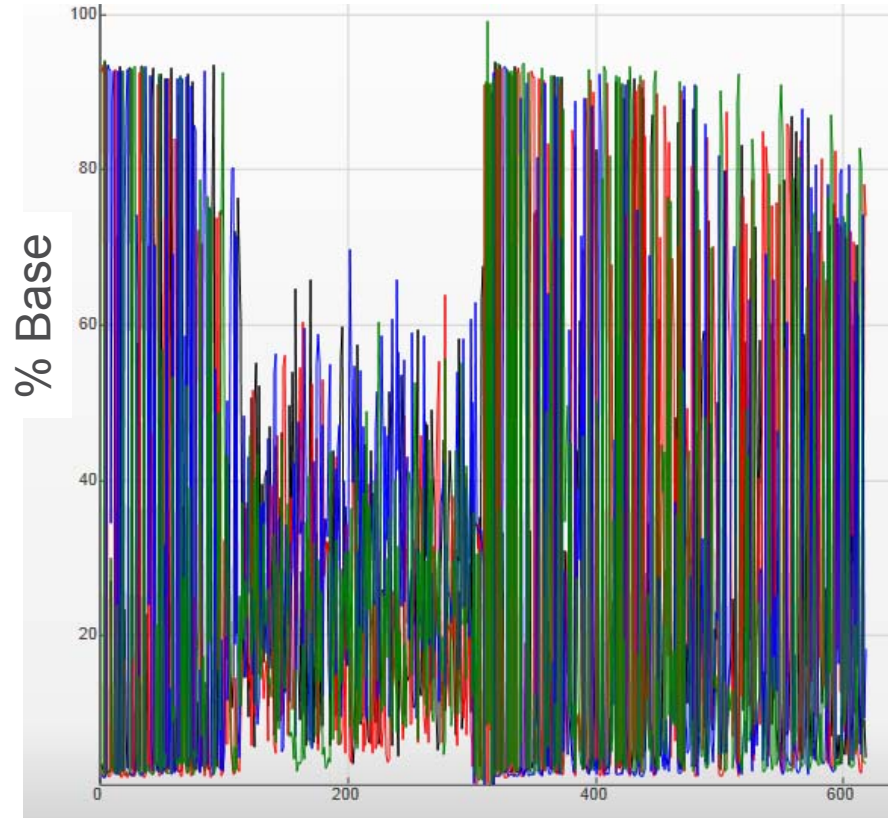
ヒトサンプルからDNAを抽出して16Sシーケンスを行う場合



16S rRNA解析サンプルはLow Diversityサンプルとなります。



Low Diveristyサンプル→PhiX (コントロールDNA) の添加が必須です



MiSeq Control Software v2.3以上を推奨

V3 kit使用の700-800K/mm²にて
PhiX **5%**添加で解析可能

オーバークラスター(1000K/mm²超)で
Qualityを下げやすくなるため
通常はPhiX 20%スタートをお勧めします。
クラスター密度が安定してきたら徐々に
下げていきましょう。

こちらの
サポートウェビナーも
ご参考ください！

2013/12/06 **イリミナサポートウェビナー シリーズ2013**
「Low Diversityサンプルを解析するためのテクニック」
イリミナ株式会社 テクニカル アプリケーション サイエンティスト
小林孝史

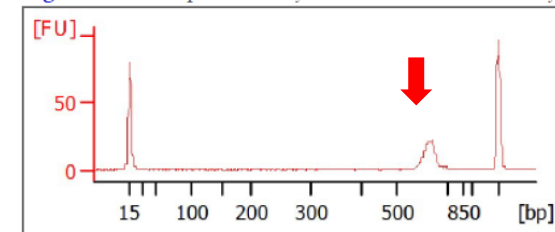
偏った塩基配列を持つサンプルを解析する際のテクニックについてご紹介するセミナーです。



主なトラブルシューットの例

- 収量が小さい→AMPure Beadsのハンドリング
- PCR増幅効率が悪い→Input DNAのQuality/Quantityをチェックする
(公式の配列以外を使用) プライマーの再設計
- クラスター数が得られない (→データ量小さい)
→ng/ul→nMの単位変換を確認。
変性時の0.2N NaOHのpHが13以上であることを確認しましょう。
ライブラリー長のチェックを行いましょう。

Figure 5 Example Bioanalyzer Trace of Final Library



Indexの振り分けがうまくできない→Indexの組み合わせを確認しましょう

Good				Bad			
	Index 1	Index 2		Index 1	Index 2		Index 2
705	GGACTCCT	503	TATCCTCT	705	GGACTCCT	502	CTCTCTAT
706	TAGGCATG	503	TATCCTCT	706	TAGGCATG	502	CTCTCTAT
701	TAAGGCGA	504	AGAGTAGA	701	TAAGGCGA	503	TATCCTCT
702	CGTACTAG	504	AGAGTAGA	702	CGTACTAG	503	TATCCTCT
	√√√√√√√√		√√√√√√√√		√√√√√√√√		√√√√xxxx

√=signal in both color
x=signal missing in one color channel

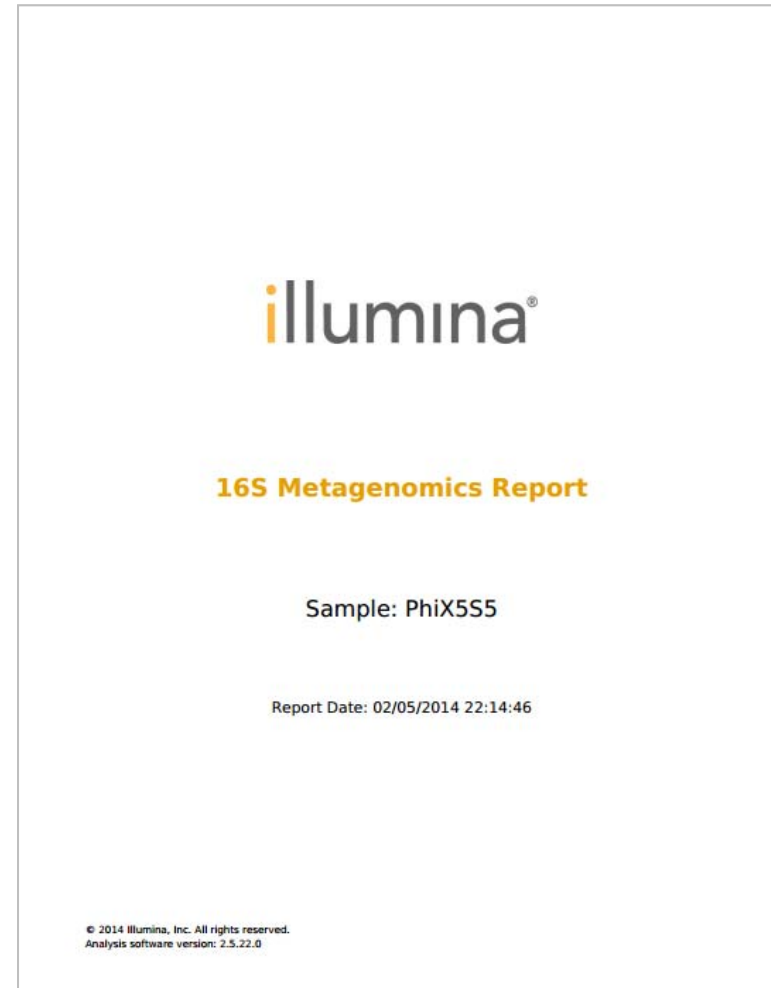


16S rRNA解析: データのアウトプット



MiSeq標準装備のMiSeq Reporterで作成される 16S Metagenomics Report

- ▶ MiSeq Reporterで16S metagenomics workflowで解析された際に出力されるレポート→**html形式**で任意のウェブブラウザで閲覧可能
- ▶ **MiSeq Reporter 2.4からインターフェイスがリニューアル**
- ▶ サンプルごとにRead数と占める割合が表示される
- ▶ 他のサンプルとの集計の比較が可能



16S rRNA解析:出カデータ

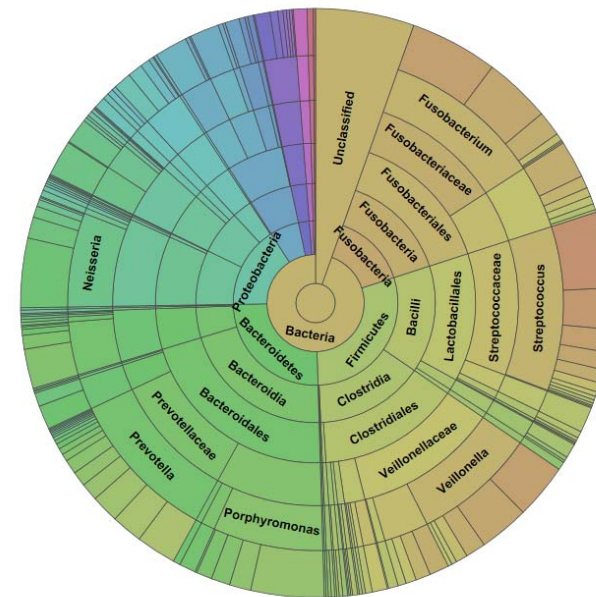
▶ 解析の目的:

- 配列のFASTQファイルの作成
- それぞれのリードに対する分類同定
- サンプルごとに分類結果をグラフ化
- サンプルごとの集計比較をグラフ化

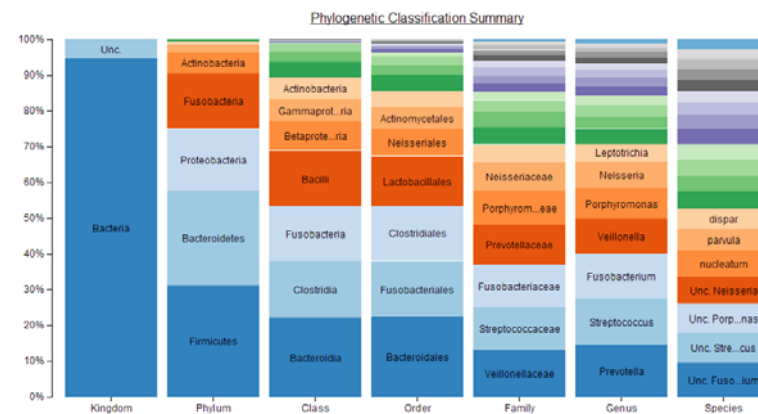
▶ 2つのパイプラインでの解析を推奨:

- クラウドで行うBaseSpace®での解析とレポートの作成 (近日リリース!)
- MiSeq標準装備のソフトウェアMiSeq Reporter

レポート中のSunburst Classification Chart



レポート中の Top 20 Classification Results by Taxonomic Level



MiSeq Reporter (MSR) 上でのデータ閲覧

- ▶ 生データへのアクセスが簡単
- ▶ グラフと実データを並べて閲覧できる

The screenshot displays the MiSeq Reporter 2.4.60 web interface. The browser address bar shows 'localhost:8042/default.htm'. The page title is 'MiSeq Reporter 2.4.60'. The main content area is divided into two sections: a table of samples and a species distribution pie chart.

Samples (63)

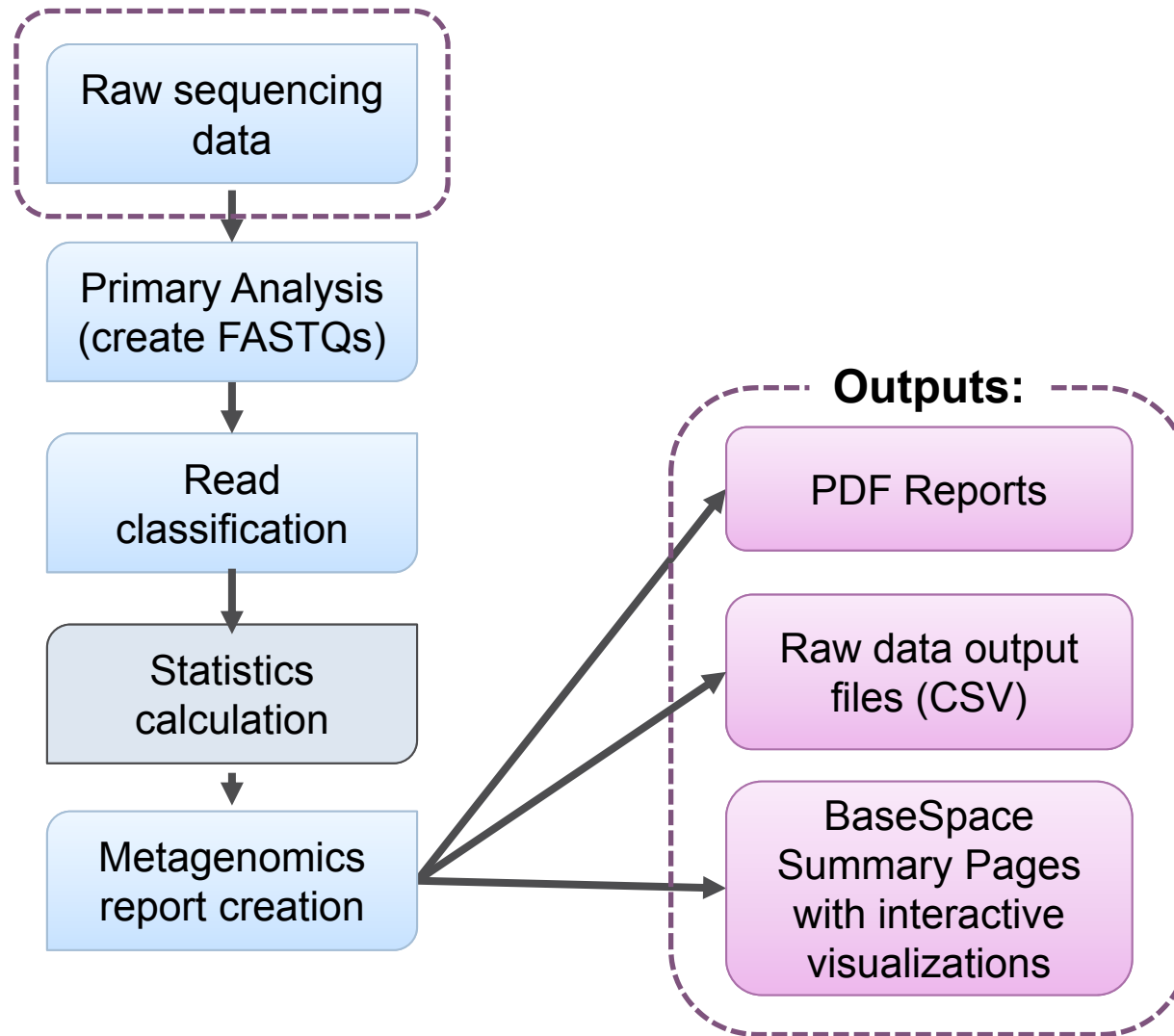
#	Sample ID	Sample Name	Cluster Raw	Cluster PF	Taxomic Level	Clusters Classified
1	Sample1	Sample1	108091	95726	Kingdom	95529
2	Sample1	Sample1	108091	95726	Phylum	95191
3	Sample1	Sample1	108091	95726	Class	95090
4	Sample1	Sample1	108091	95726	Order	94939
5	Sample1	Sample1	108091	95726	Family	92646
6	Sample1	Sample1	108091	95726	Genus	91866
7	Sample1	Sample1	108091	95726	Species	68481
8	Sample2	Sample2	104093	90794	Kingdom	90636
9	Sample2	Sample2	104093	90794	Phylum	90525
10	Sample2	Sample2	104093	90794	Class	90396
11	Sample2	Sample2	104093	90794	Order	90312
12	Sample2	Sample2	104093	90794	Family	89973
13	Sample2	Sample2	104093	90794	Genus	89669
14	Sample2	Sample2	104093	90794	Species	76925
15	Sample3	Sample3	98472	86992	Kingdom	86832
16	Sample3	Sample3	98472	86992	Phylum	86656
17	Sample3	Sample3	98472	86992	Class	86558
18	Sample3	Sample3	98472	86992	Order	86489
19	Sample3	Sample3	98472	86992	Family	83804
20	Sample3	Sample3	98472	86992	Genus	83253
21	Sample3	Sample3	98472	86992	Species	63730

Sample2 - Species

Species	Percentage
Prevotella melaninogenica	23.13%
Haemophilus parainfluenzae	16.22%
Unclassified	15.28%
Rothia mucilaginosa	5.7%
Veillonella dispar	5.57%
Microvirus Enterobacteria phage PhiX174	4.97%
Neisseria cinerea	4.05%
Veillonella atypica	2.48%
Gemella haemolysans	2.46%
Fusobacterium periodonticum	1.97%
Prevotella histicola	1.75%
Neisseria mucosa	1.62%
Granulicatella adiacens	1.25%
Actinobacillus rossii	1.21%
Veillonella parvula	1.2%
Prevotella nanceiensis	1.04%
Prevotella pallens	0.89%
Neisseria flavescens	0.49%
Streptococcus tigurinus	0.48%
Streptococcus pseudopneumoniae	0.47%
Streptococcus vestibularis	0.44%
Gemella cunicula	0.42%
Actinobacillus porcinus	0.4%
Prevotella salivae	0.38%
Streptococcus thermophilus	0.37%
Candidatus Blochmannia rufipes	0.37%
Streptococcus parasanguinis	0.36%
Mannheimia caviae	0.29%
Streptococcus infantis	0.27%
Other Species	4.47%
Total	100%



BaseSpace 16S Analysis Overview (近日リリース!)



16S rRNA 菌叢解析实用例:1

▶ Melissa Spears



16S rRNA菌叢解析の実例



ユーザー:**GEORGE WEINSTOCK/ERICA SODERGREN**

The Genome Institute at Washington University, St. Louis, MO

研究対象:

臨床サンプルからの16S rRNA シーケンス解析（特に可変領域の解析 (V1-V3 and V3-V5)）

今回のプロジェクト:

既知の微生物サンプルを用いてMiSeq V3kit (2x301 cycle)とRoche社454とのパフォーマンスに比較を行った。

→今まで使用していた454との時間とコスト面での比較



16S rRNA菌叢解析の実例、結果について



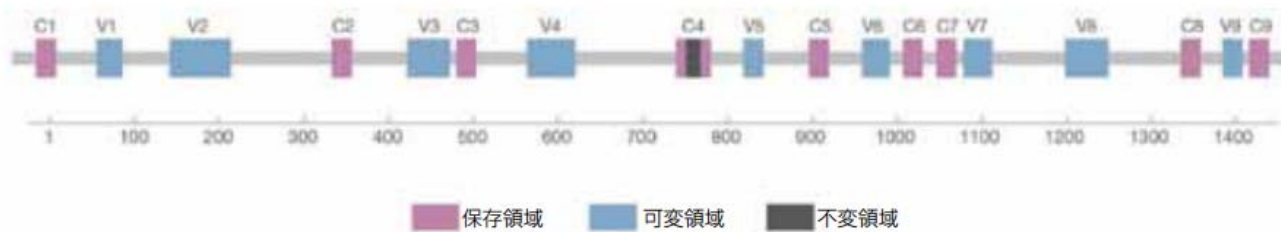
ユーザーからのフィードバック

- 操作性: MiSeqは454に比べて操作性に優れている（簡単！）
~88%のリードでRead1とRead2のアッセンブルが可能
- コスト: MiSeqで最終的にかかったサンプル単位のコストは454
使用時のなんと1/ 12-20で計上された！
- アウトプット量: 454に比べて~20xのカバレッジが得られた



ターゲット領域の検討

- ▶ Human Microbiome ProjectはV3–V5領域を使用している。
 - V1–V3領域はほとんど同一配列であり、現在の解析ではあまり使用されることはない。
- ▶ V3–V5領域は若干長いため、V1–V3領域に比べてMiSeq V3キットの2x300bpでのアセンブリー配列の作成に向かない
 - V3–V5, 569bp, 2x300bpでの解析には若干長い（オーバーラップ領域が短い）
 - V1–V3, 507bp, 長さ的には適している
- ▶ ほかの領域はどうか？:
 - V4領域 (イルミナシーケンサーでよく使用される配列)
 - V1–V2領域 (増幅されにくいサンプルによく使用される短い配列)
 - V6–V9領域 (HMP種などの同定には向かない配列)
- ▶ TGI 推奨: V1–V3領域 (複数のコミュニティとのディスカッションを経る)



16S rRNA遺伝子の模式図。基準細菌16S rRNAの可変領域（青色）および保存領域（紫色）の位置。
灰色部分は、すべての細菌において不変な領域。



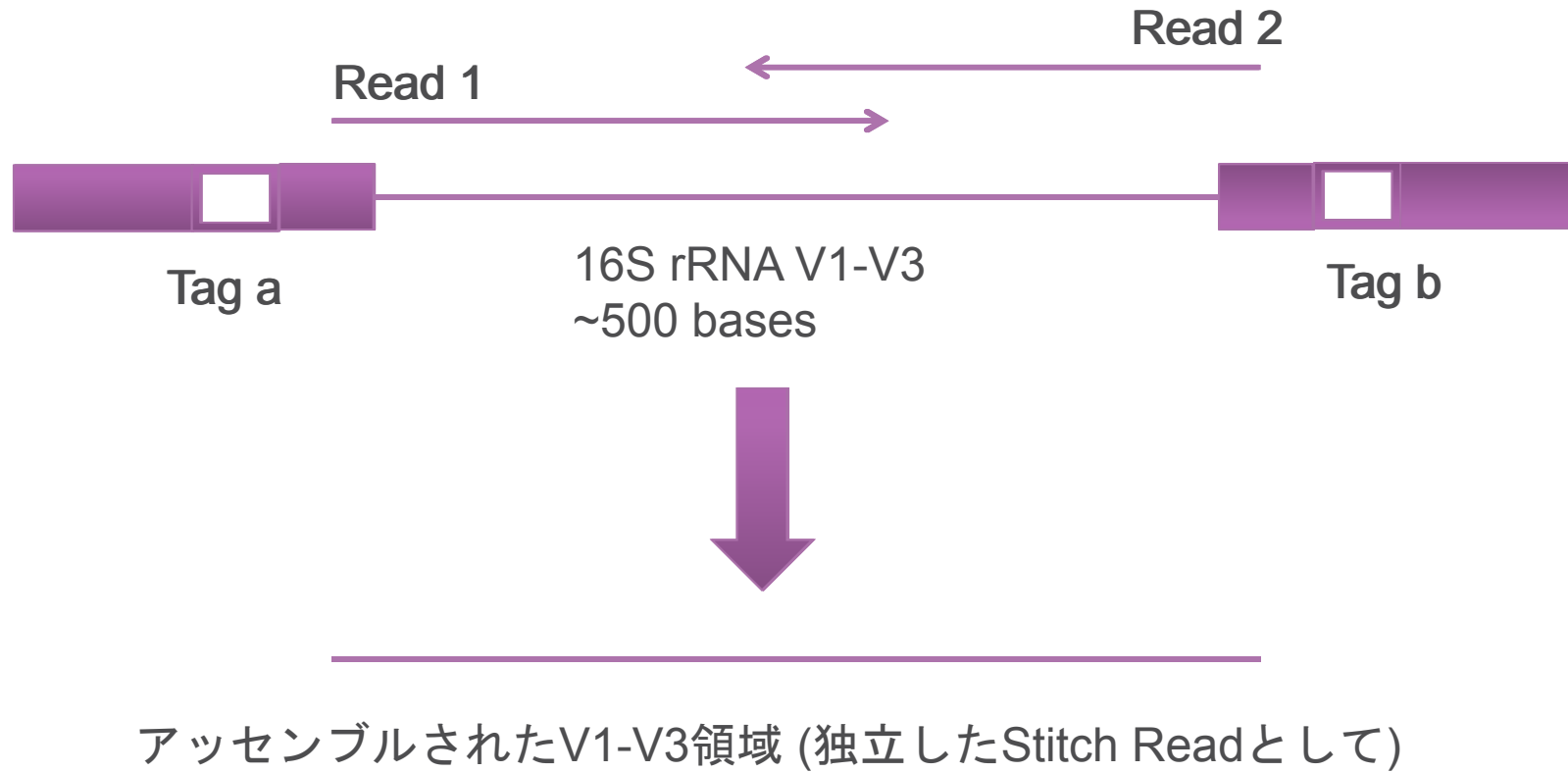
ターゲットの比較

ターゲット	種レベルでの正確性	属レベルでの正確性
全長	92%	94%
V1-V3 (507bp)	91	93
V3-V5 (569bp)	86	92
V6-V9 (524bp)	88	93
V1-V2 (310bp)	90	92
V4 (219bp)	75	83

Thanks to Bo-Young Hong and Patricia Diaz, Univ. Conn. Health Center.
830 sequences from the Human Oral Microbiome Database classified
using mothur/RDP.



MiSeq V3 kit (2x300bp) 16S rRNA解析データ



16S rRNA菌叢解析のワークフロー



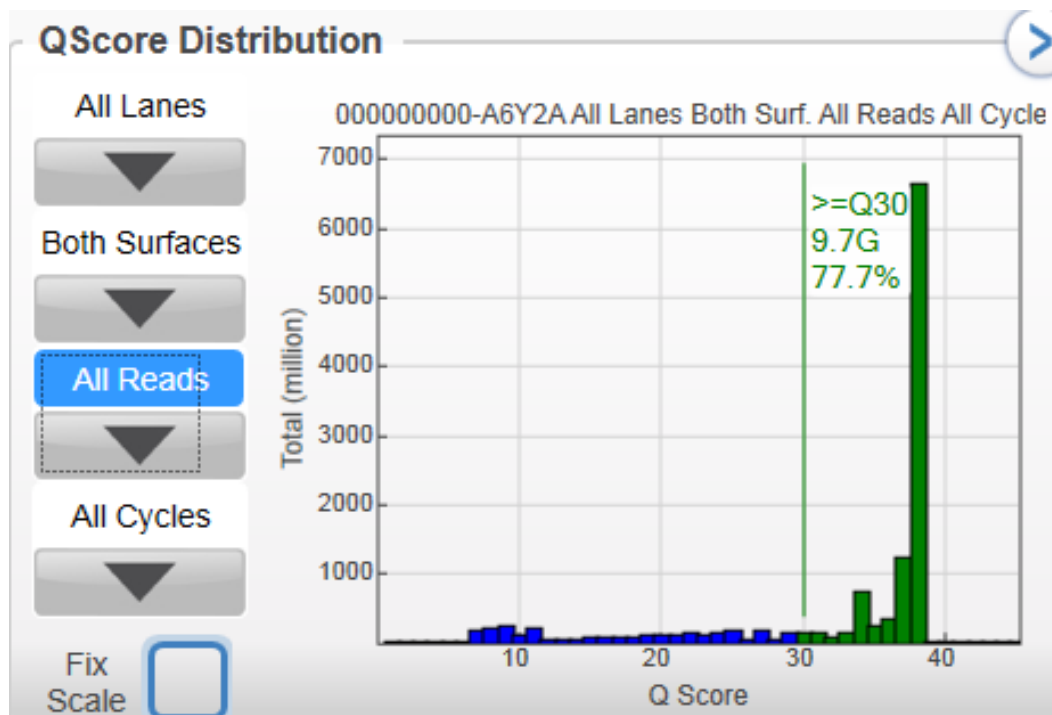
サンプル	属レベルで同定されたRead	%
0.5 CONFIDENCE		
MiSeq Assembled	43,5271	96.1
454 Pool 1	3,589	96.5
454 Pool 2	4,697	97.8
0.6 CONFIDENCE		
MiSeq Assembled	427,930	94.5
454 Pool 1	3,477	93.5
454 Pool 2	4,578	95.3
0.7 CONFIDENCE		
MiSeq Assembled	419,751	92.7
454 Pool 1	3,358	90.3
454 Pool 2	4,400	91.6
0.8 CONFIDENCE		
MiSeq Assembled	409,454	90.4
454 Pool 1	3,200	86
454 Pool 2	4,177	86.9

結論として、454に比べてMiSeq V3 kit使用時の方がより正確に多くの微生物の同定に成功した。



結果:

- ▶ V1-V3 アンプリコン (~500bp)
- ▶ 20 M ペアエンドリード
- (パスフィルター後)
- ▶ 9.7 Gb > Q30
- ▶ 95.85% Clusters PF



Analysis | Imaging | Summary | **Tile Status** | TruSeq Controls | Indexing

Run Summary

Level	Yield Total (G)	Projected Total Yield (G)	Yield Perfect (G)	Yield <=3 errors (G)	Aligned (%)	% Perfect [Num Cycles]	% <=3 errors [Num Cycles]	Error Rate (%)	Intensity Cycle 1	% Intensity Cycle 20	% >= Q30
Read 1	6.2	6.2	0.6	0.8	13.59	67.4 [300]	95.3 [300]	0.92	25	118.1	84.1
Read 2 (I)	0.1	0.1	0.0	0.0	0.00	0.0 [5]	0.0 [5]	0.00	518	0.0	81.6
Read 3	6.2	6.2	0.5	0.8	13.52	58.9 [300]	91.6 [300]	1.06	38	91.2	71.2
Total	12.5	12.5	1.1	1.6	13.55	63.2	93.5	0.99	194	104.6	77.7



結論



- ▶ ワシントン大学ではHuman Microbiome Projectにおいてプライマーを独自に設計し、16Sシーケンス解析のプロトコルを検証を行った。
- ▶ ユーザーはMiSeqを導入することにより、最終的に時間とコストの節約に成功した
- ▶ 以前までの454を用いた方法に比べて
 - 良好なクオリティー
 - 大きなアウトプット量

協力、謝辞:

- ▶ Weinstock Lab
 - George Weinstock
 - Erica Sodergren
 - Brandi Herter
 - Kathie Mihindukulasuriya
- ▶ TGI Illumina Production
 - Matt Cordes



16S rRNA シーケンスの実用例: サバクゴファーガメの鼻腔に含まれる菌叢解析

- ▶ Gary Nunn, Ph.D



サバクゴファーガメの例

- ▶ SanDiego動物園ではモハーヴェサバクゴファーガメの保護活動を行っている。
- ▶ **マイコプラズマ菌の感染**による上気道炎が保護活動および生活地域の移転作業を妨げていた
- ▶ マイコプラズマ菌 (*Mycoplasma*) : グラム陰性菌, 人においてもマイコプラズマ肺炎を引き起こす種類もある。



サバクゴフアーガメの保護に向けたSanDiego動物園との共同研究

WILDLIFE DISEASE LABORATORIES

Bruce Rideout, Director

Josephine Braun, Scientist

- ▶ 目的は保護活動の妨げになる疾患を防ぐこと。
- ▶ 具体的に下記のプロジェクトを行った。
 - 疾患の同定を行った
 - 疾患に対する診断方法を確立した
 - 下記の施設においてアウトブレイクの危険性を伴う病原菌の同定を行った。
 - San Diego Zoo
 - San Diego Zoo Safari Park
 - Field conservation programs



健康なサバクゴファーガメから同定された微生物 (1)

Rank	Species	Total # Sequences in sample	Total % Sequences in Sample	Notes
1	<i>Calothrix parietina</i>	106,004	38.11%	Blue-green algae
2	<i>Rickettsia</i> sp.	15,631	5.62%	Tick, flea, lice borne
3	<i>Acinetobacter</i> sp.	5,764	2.07%	Soil bacteria
4	<i>Symploca atlantica</i>	4,917	1.77%	Algae
5	<i>Proteobacteria</i>	4,109	1.48%	Variable (some pathogen known)
6	<i>Thiomonas thermosulfata</i>	3,847	1.38%	Extremophile
8	<i>Blautia</i> sp.	3,705	1.33%	Gut bacteria
9	<i>Blautia coccooides</i>	3,330	1.20%	Gut Bacteria
10	<i>Pedobacter</i>	2,881	1.04%	Soil Bacteria
Other Species of note	<i>Mycoplasma agassizii</i>	138	0.05%	



疾患を持ったサバクゴフアーガメから同定された微生物(6)

Rank	Species	Total # Sequences in sample	Total % sequences in sample	Notes
1	<i>Mycoplasma agassizii</i>	31,924	18.2%	Etiologic agent*
2	<i>Myroides odoratus</i>	11,654	6.6%	
3	<i>Flavobacteriaceae sp.</i>	11,481	6.5%	
4	<i>Chelonobacter sp.</i>	11,384	6.5%	Associated diseased tortoise **
5	<i>Pedobacter sp.</i>	7,894	4.5%	Soil bacteria
6	<i>Flavobacterium swingsii</i>	5,492	3.1%	
7	<i>Proteus penneri</i>	5,263	3.0%	Invasive pathogen
8	<i>Pseudomonas brenneri</i>	4,538	2.6%	
9	<i>Pseudomonas sp.</i>	4,193	2.4%	Water borne biofilms
10	<i>Pseudomonas marginalis</i>	4,160	2.4%	

* Brown et al., 1994

** Gregersen et al., 2009



疾患を持ったサバクゴファーガメから同定された微生物(8)

Rank	Species	Total # Sequences in Sample	Total % sequences in sample	Notes
1	<i>Chelonobacter sp.</i>	141,734	57.6%	Associated with disease tortoise **
2	<i>Chelonobacter oris</i>	37,883	15.4%	
3	<i>Mycoplasma agassizii</i>	7,639	3.1%	Etiologic agent *
4	<i>Granulicatella adiacens</i>	5,825	2.4%	Normal human mucosal membranes
5	<i>Flavobacteriaceae</i>	5,443	2.2%	Water borne
6	<i>Myroides odoratus</i>	4,855	2.0%	Hm nosocomial infection
7	<i>Pedobacter sp.</i>	4,108	1.7%	Soil bacteria
8	<i>Deinococcus sp.</i>	2,051	0.8%	Soil, water
9	<i>Flavobacterium swingsii</i>	1,800	0.7%	Water borne
10	<i>Proteus penneri</i>	1,625	0.7%	Invasive pathogen

* Brown et al., 1994

** Gregersen et al., 2009



結論

サバクゴフアーガメの鼻腔内メタゲノム解析

- ▶ サバクゴフアーガメのほとんどは*Mycoplasma agassizii*感染が見られた
 - 9匹中8匹で病原菌を検出した。
 - 9匹中7匹の疾患を持つサバクゴフアーガメの鼻腔には病原菌の占める割合がトップ10に含まれていた。
 - Tortoise-16025：一番多く含まれているバクテリア(18.2%)
 - Tortoise-18963：二番目に多く含まれているバクテリア(13.9%)
 - Tortoise-13498：三番目に多く含まれているバクテリア(3.1%)
 - 以前から上気道炎の原因となる病原体であることが知られていた
- ▶ *Chelonobacter* 種も多くのサバクゴフアーガメの鼻腔で検出された。
 - 10匹中8匹のサバクゴフアーガメの鼻腔でこれら病原菌が多い方から10種類に含まれていた
 - *Chelonobacter* についても以前より上気道炎との関連が報告されていた
- ▶ 健康なサバクゴフアーガメは土壌や水中に含まれる多様なバクテリアを有していた
 - 健康なサバクゴフアーガメには0.05%程度しか*Mycoplasma agassizii*が含まれていなかった



その他



- ▶ www.epicentre.com
DNA抽出キットその他



- ▶ http://support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.ilmn
プロトコール、デモデータなど



- ▶ <https://basespace.illumina.com/home/index>

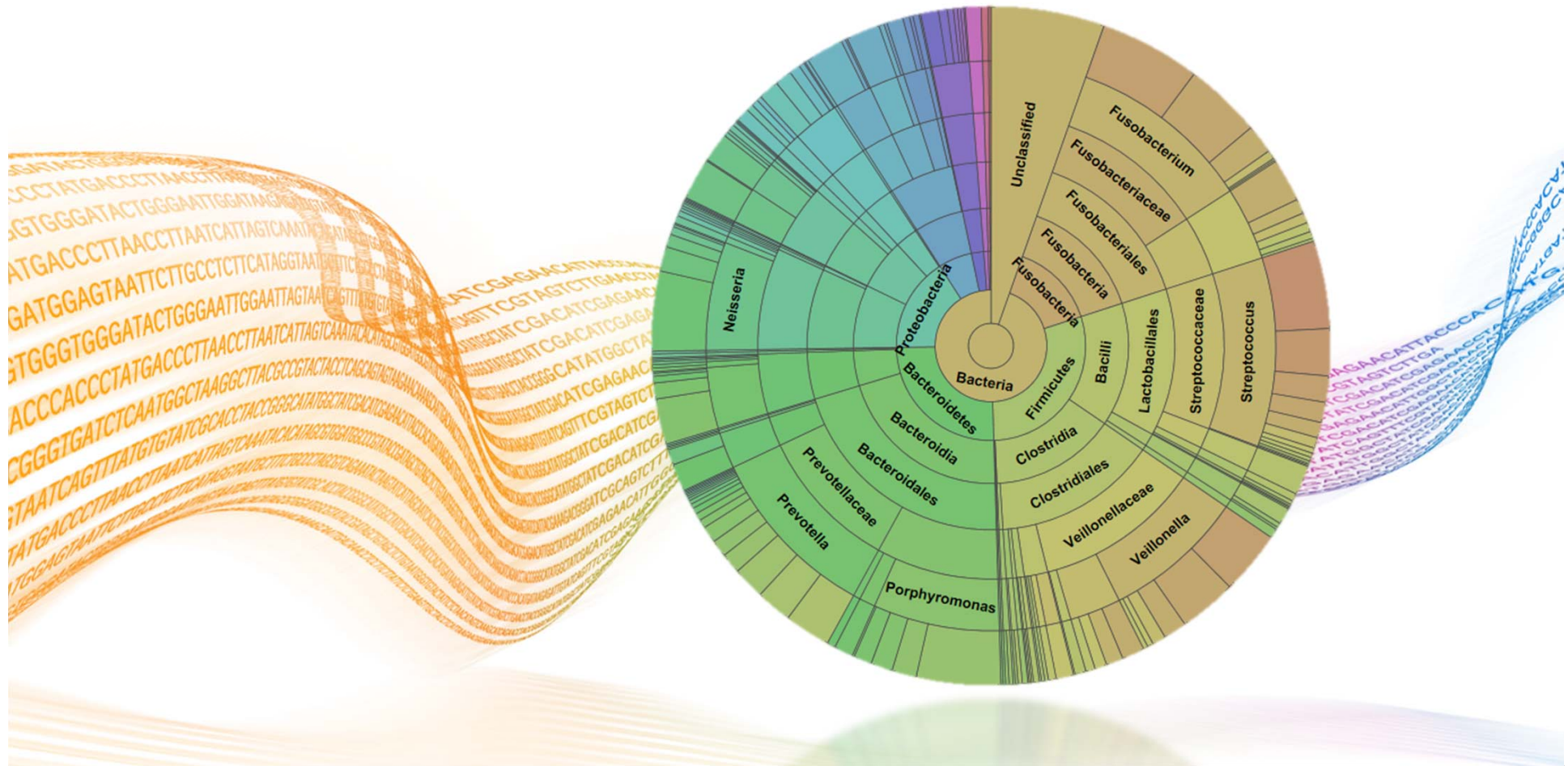
本資料はwebinar後、下記リンク上にアップロードいたします。

http://www.illumina.co.jp/events/webinar_japan.ilmn?ws=ss

プロトコールは、原則として弊社キットを使用した場合に比べてサポートの範囲が限られます。ご利用にあたりましてはご了承の上ご利用ください。



ご清聴頂き誠にありがとうございました。
しばらくご質問をいただく時間を設けます。



サポートウェビナーにご参加いただき
ありがとうございました。

本日のセッション終了後のご質問は、
techsupport@illumina.com
で承ります。

