

ハイブリダイゼーションキャプチャ法による SARS-CoV-2 ウイルスゲノムの シーケンス解析

著者

元岡 大祐
大阪大学微生物病研究所
遺伝情報研究センター
ゲノム解析室

石井 善幸、吉崎 史子
アジレント・テクノロジー
診断・ゲノミクス部門

概要

本アプリケーションノートでは、SARS-CoV-2 ウイルスの全ゲノムシーケンスを SureSelect XT Low Input およびカスタムキャプチャパネルを用いて実施した例を紹介しています。Bravo NGS 自動化システムによりサンプル調製を行い Illumina MiSeq でシーケンスを行ったところ、SARS-CoV-2 ウイルスゲノムをカバーするリードが得られアルファ株、デルタ株、オミクロン株ともに検出可能でした。ハイブリダイゼーション法による方法であるため武漢で発見された SARS-Cov-2 レファレンスゲノムから塩基が変異した様々なタイプの変異株にも同じ手法で検出可能となることが期待されます。

方法

キャプチャパネルのデザイン

SARS-CoV-2 ウイルスの全ゲノム配列 (NCBI Accession: NC_045512.2) に対し、120 bp のプローブを 20 bp ずつずらしながらタイリングし、6x 相当のタイリングとしました。アジレント社の SureDesign に配列を Upload する際に、Optimized Performance XT HS/XT HS2/XT LI/QXT boosting (90 min hyb) オプションを選択し GC 含量によるキャプチャバイアスを最適化しました。キャプチャパネルの情報はアジレント社から共有可能です。

cDNA 合成と精製

抽出済み RNA 15 μ L を用いて、Proto Script II First Strand cDNA Synthesis Kit (NEB; part number E6560)、Random Primer (hexadeoxyribo nucleotide mixture; pd (N) 6) (タカラバイオ; part number 3801) によりプロトコルの 2.5 倍量 (Total volume 50 μ L) で 1st strand cDNA 合成を行いました。RNA 抽出は外部施設で行われているサンプルも多いため、RNA 重量ではなく RNA 溶液の容量でスタート量を決定しました。

続けて NEBNext Ultra II NonDirectional RNA Second Strand Synthesis Module (NEB; part number E6111) により 2nd strand cDNA を合成しました。2nd strand cDNA 合成のプロトコルで 1st strand cDNA の持ち込み量はプロトコル上 20 μ L のところ、50 μ L の全量を持ち込み、Nuclease water 量で調整することで 80 μ L の反応液量としました。2nd strand cDNA 合成後の反応液 80 μ L に対して 96 μ L の Agencourt AMPure XP (Beckman Coulter; part number A63880) を用いて精製を行い、8 μ L の Nuclease free water で溶出しました。溶出したサンプルの 1 μ L を用いて Qubit 4 Fluorometer で Qubit dsDNA HS Assay を用いて定量しました。

SureSelect XT Low Input によるライブラリ調製とターゲットエンリッチメント

精製した 2 本鎖 cDNA の全量 (7 μ L) を用い、SureSelect XT Low Input キットおよび SureSelect XT HS and XT Low Input Enzymatic Fragmentation キットを用い、プロトコルの Pre Capture PCR サイクル数を改変しライブラリ調製を行いました。PCR サイクル数は Table 1 に記載されています。

ハイブリダイゼーションの条件は “Utilization of Agilent SureSelect Target Enrichment for Whole Genome Sequencing of Viruses and Bacteria” に従い、アダプタ付きライブラリ 1000 ng をインプットとしました。キャプチャプローブを 1:10 で希釈して使用し、Post Capture PCR は 21 サイクルとしました。SureSelect XT Low Input によるライブラリ調製とターゲットエンリッチメントのステップは、Bravo NGS 自動化システムで行いました。最終ライブラリは Agilent 4200 TapeStation で High Sensitivity D1000 ScreenTape キットを用いてサイズ分布を分析し、Qubit 4 Fluorometer で Qubit dsDNA HS Assay を用いて定量しました。定量結果により等重量ずつ混合し、最終濃度を PCRmax Eco 48 real time PCR system で KAPA Library Quantification kits (Roche, p/n 07960140001) を用いて決定し、MiSeq で 150 bp ペアエンドのシーケンスを行いました。

データ解析

得られた FASTQ ファイルは AGeNT によりアダプタ配列を除去後に BWA-mem を用いてレファレンスゲノム NC_045512.2 にマッピングし、Picard でソート後に Duplicate を判定しインデックスを作成しました。

結果

2本鎖 cDNA およびライブラリの収量

2本鎖 cDNA の収量は 0.5 ng 以下のサンプルから、200 ng 以上のサンプルまで大きなばらつきがありました。2本鎖 cDNA はすべてライブラリ調製に用い、cDNA 収量にあわせて PCR サイクル数を増やすことにより全サンプルでハイブリダイゼーションに必要な 1000 ng を超える収量のアダプタ付きライブラリが得られました (Table 1)。ハイブリダイゼーション後、全サンプルでシーケンスに十分な収量が得られました。TapeStation によるサイズ確認の一例を Fig.1 に示します。

Table 1. 各サンプルの収量および適用した PCR サイクル数

Sample	Ct 値	cDNA 収量 (ng)	Cycle 数	アダプタ付きライブラリ収量 (ng)	ハイブリダイゼーション後ライブラリ濃度 (ng/μL)
1	27.6	5.306	16	1932	59.3
2	33.3	1.015	18	2028	61.3
3	26.7	0.742	18	1788	50.1
4	30.7	0.497	18	2052	49.1
5	29.6	3.304	16	1968	49.5
6	23.3	140.7	12	1212	61.5
7	22.6	46.69	12	1416	63
8	26.7	241.5	12	1236	20.2
9	22.6	15.47	16	1368	54.4
10	24.4	12.25	16	1488	51.6

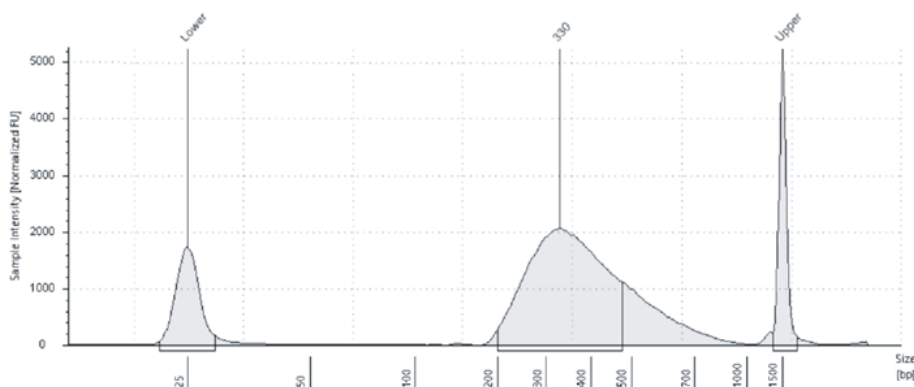


Fig.1 ハイブリダイゼーション後の最終ライブラリを Agilent 4200 TapeStation で High-Sensitivity D1000 キットによる泳動した例

シーケンス結果

インプットとして用いた cDNA の量や、Ct 値から予想されるウイルス量には広い範囲のばらつきがありました。最終的にシーケンスを行い得られたリード数は最大 2.3 倍の範囲となりました (Fig.2)。Ct 値が 33.3 と 10 サンプルの中で最も高く、ウイルス量が最も少ないと予想されるサンプル 2 については Duplicate% が高く (Fig.3) そのため平均カバレッジはほかのサンプルより低くなりましたが、100x 以上カバーされている領域は 99.8% であり変異解析には十分なシーケンス量が得られました。On Target % は全サンプルで 95% を超えており、ヒトゲノム DNA が混入する可能性があるサンプルであってもターゲットエンリッチメントを行うことで効率よくシーケンスを行えることが示されました。SARS-CoV-2 ゲノムの全域にわたってリードが得られ (Fig.4)、デルタ型の特徴的な変異 L452R やアルファ型の特徴的な変異 N501Y も検出が可能でした (Fig.5)。

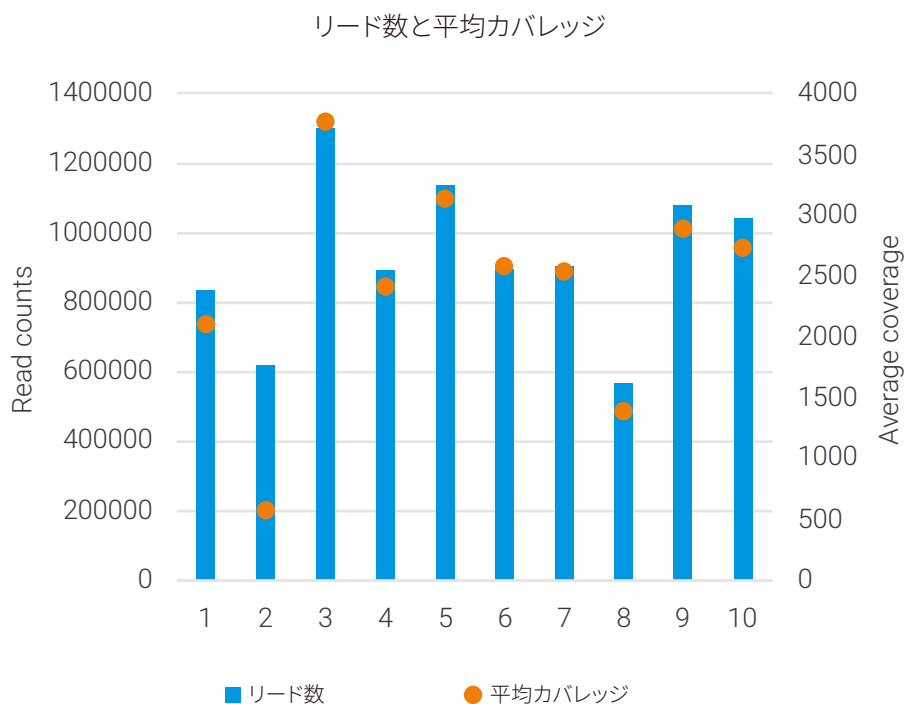


Fig.2 各サンプルのリード数と平均カバレッジ
各サンプルのリード数 (左軸) および平均カバレッジ (右軸)

シーケンスメトリクス

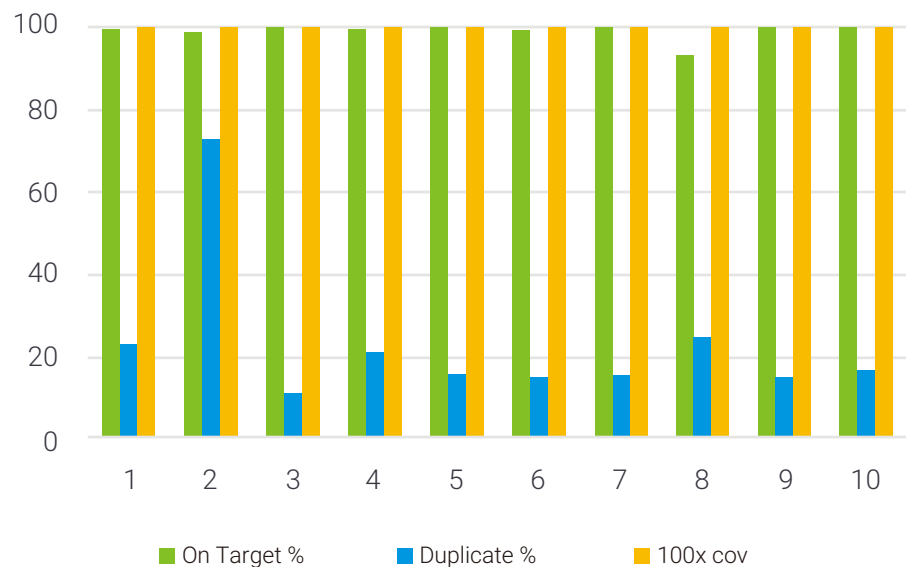


Fig.3 シーケンスメトリクス
SARS-CoV-2 レファレンスゲノムにマップされたリードのパーセンテージ (On Target %)、Duplication のパーセンテージ (Duplicate %)、および 100x 以上の Depth が得られた塩基のパーセンテージ (100x cov) を Picard で算出しました。

ハイスループット化

プレート単位でのサンプル調製を効率化するために、プレキャプチャPCRのサイクル数を18サイクル、ハイブリダイゼーションの持ち込み量は濃度に関わらず12 µLで統一、と2つのパラメータを固定して実施しても安定して同様の結果が得られています。本プロトコルの確立後、6000サンプル以上でライブラリ調製を行い、オミクロン株の同定も可能であることが確認され、また解析不能サンプルは10サンプル未満（いずれもCt値35前後のサンプル）でした。

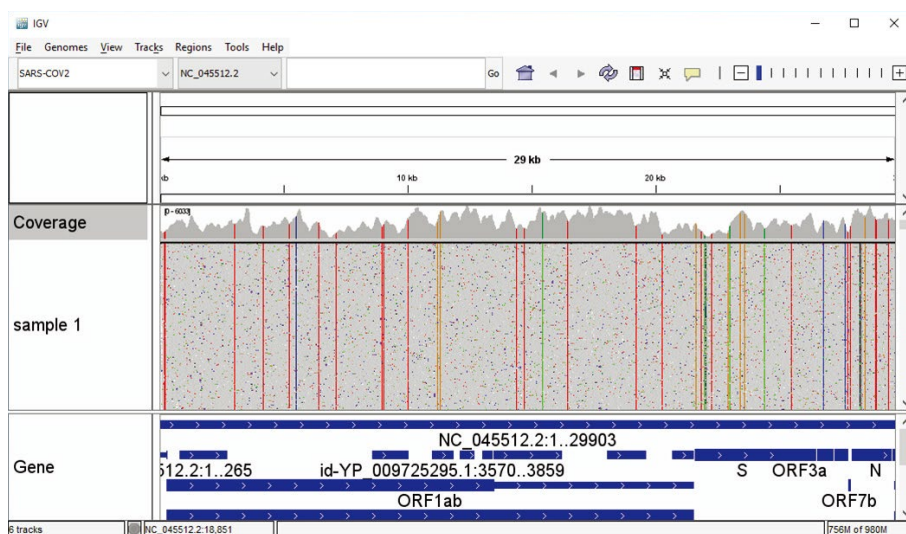


Fig.4 SARS-CoV-2 レファレンスゲノムに対するカバレッジ

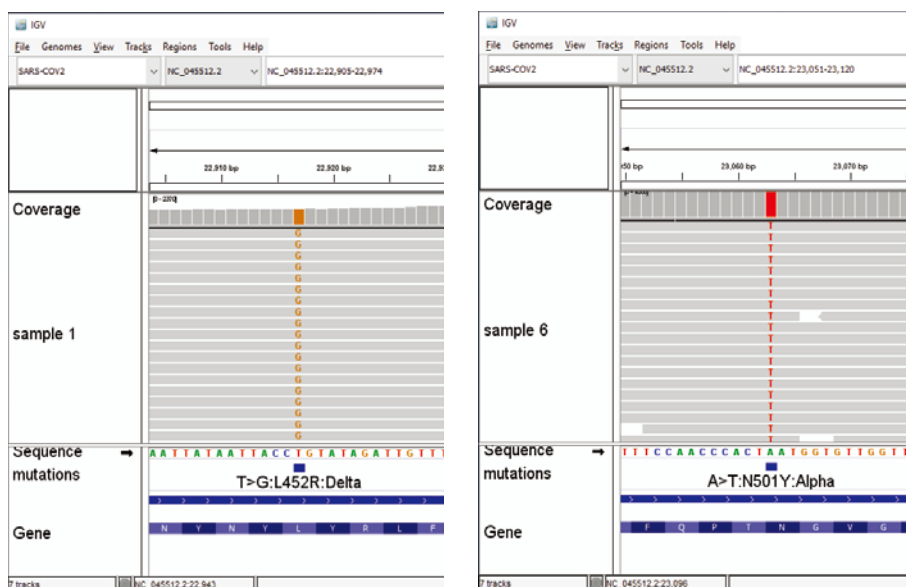


Fig.5 検出された変異の例 (A) デルタ型の特徴的な変異 L452R (B) アルファ型の特徴的な変異 N501Y

結語

本アプリケーションノートで紹介したプロトコルは、広い範囲の Ct 値や二本鎖 cDNA 収量のサンプルに対して安定して使用可能であることが示されました。また、プロトコルの自動化によりハイスループットなウイルスゲノム解析が可能となり、大規模な疫学調査やウイルスの伝播・ゲノム進化などに関する基礎研究に貢献できるものと期待されます。

[お問い合わせ窓口]

アジレント・テクノロジー株式会社

本社 / 〒 192-8510 東京都八王子市高倉町 9-1

●カスタムコンタクトセンター ☎ 0120-477-111

mail : email_japan@agilent.com

※仕様は予告なく変更する場合があります。

※掲載の製品はすべて試験研究用です。

診断目的にご利用いただくことはできません。

<http://www.agilent.com/chem/genomics:jp>

PR7000-8575/G220440

© Agilent Technologies, Inc. 2022

本書の一部または全部を書面による事前の許可なしに複製、
改変、翻訳することは、著作権法で認められている場合を除き、
法律で禁止されています。

Printed in Japan, February 25, 2022

5994-4538JAJP