

# 複雑なサンプル中の 微生物集団を 評価するための ショットガン メタゲノミクスNGS ワークフロー

NextSeq™ 1000およびNextSeq  
2000システムの600サイクルキットが  
提供する種の識別に適した正確さと  
柔軟性

**illumina**®

## 複雑なサンプルのメタゲノム分類

ショットガンメタゲノミクスシーケンスは、16Sや内部転写スペーサー (ITS) リボソームRNA (rRNA) シーケンスなどのアンプリコンシーケンスアプローチに代わる方法であり、複雑なサンプル中の微生物の多様性を評価するために用いられます。アンプリコンシーケンスとは異なり、次世代シーケンサー (NGS) を使用したショットガンメタゲノミクスシーケンスでは、サンプル中に存在するすべての生物の包括的なゲノム情報を取得します。完全なゲノム情報を利用できるということは、ショットガンメタゲノミクスシーケンスがアンプリコンシーケンス<sup>1</sup>では見逃されていた種を同定でき、得られたデータには、アンプリコン法からは得られない機能情報が含まれていることを意味します。<sup>2,3</sup>

本アプリケーションノートは、ハイスループットのショットガンメタゲノム解析に対するNextSeq 1000、NextSeq 2000、およびMiSeq™ システムの性能の類似性を示します。また、実績のあるNextSeq 550システムで生成されたデータを使用して、600サイクルキットがメタゲノム解析アプリケーションで一般的に使用される300サイクルキットよりも優れていることも示します。人工的に合成した集団と実在のサンプルからのデータを示し、600サイクルキットを使用した場合に属と種をより優位に同定できることを実証します。

## 方法

NextSeq 1000/2000 P1 Reagents (600 cycles)キットおよびNextSeq 1000/2000 P2 Reagents (600 cycles)キットは、ショットガンメタゲノミクスシーケンスに最適な仕様を備えたNextSeq 1000およびNextSeq 2000システムの機能とシーケンス出力を拡張します。NextSeq 1000およびNextSeq 2000システムは、装置内にフルイデックスを持たない、ロードしてランするだけの試薬を使用するため、ワークフローのステップ数とサンプル汚染のリスクが軽減されます。ショットガンメタゲノミクスワークフローは、ライブラリー調製、実績のあるイルミナNGS、およびタッチパネル操作の二次データ解析を統合した、マイクロバイオーム探索のための完成されたソリューションです (図1)。

### ライブラリー調製

微生物のゲノムDNAサンプルは2つのソースから入手しました。1つ目のサンプルは、American Type culture collection (ATCC) から市販されている20 Strain Staggered Mix Genomic Material (ATCC、カタログ番号: MSA-1003) です。このATCCサンプルは、グラム染色、GCコンテンツ、および孢子形成などの属性に基づいて選択された細菌株から調製されたゲノムDNAの交互分布 (staggered distribution) で構成される模擬微生物群集です。また、2番目のサンプルセットは、以前に述べた<sup>4</sup>実在する便サンプルを解析用に取得しました。

これらのライブラリーは、Illumina DNA Prep, (M) Tagmentation (24 Samples, IPB) (イルミナ、カタログ番号: 20060060) およびIDT for Illumina DNA/RNA UD Indexes



図1: NextSeq 1000およびNextSeq 2000システムでのショットガンメタゲノミクスNGSワークフロー

Set A, Tagmentation (96 Indexes, 96 Samples) (イルミナ、カタログ番号: 20027213) を用いて調製しました。IDT for Illumina DNA/RNA UD Indexes Set A~Dにより384の16Sライブラリーを生成できます。

### シーケンス

調製したライブラリーをプールし、充填済みのNextSeq 1000/2000 P1 Reagents (600 cycles)キットフローセル、MiSeq Reagent Kit v3 (600-cycle)フローセル (イルミナ、カタログ番号: MS-102-3003)、またはNextSeq 500/550 High Output Kit v2.5 (300 cycles)フローセル (イルミナ、カタログ番号: 20024908) にロードしました。シーケンスは、NextSeq 2000システム、MiSeqシステムまたはNextSeq 550システムでそれぞれ実施しました。すべてのランの代表的なシーケンスデータは、BaseSpace™ Sequence Hubデモデータウェブページでご覧いただけます。

### データ解析

プールしたライブラリーは、ゲノミクスクラウドコンピューティングプラットフォームのBaseSpace Sequence Hubでデマルチプレックスしました。DRAGEN™ Metagenomicsパイプラインを使用して、NextSeq 2000、MiSeq、およびNextSeq 550システムで生成されたデータを処理しました。メタゲノムは、SPAdes Genome Assemblerを使用してアセンブルしました。分類学的手法は、DRAGEN Metagenomicsパイプラインを通じて実施しました。

600サイクルキットを使用して生成したNextSeq 2000データと300サイクルキットを使用して生成したNextSeq 500データを比較するために、BaseSpace Sequence Hubで利用可能なDRAGEN FASTQ Toolkitを使用してNextSeq 2000リードをトリミングしま

した。サンプル間を比較できるように、DRAGEN FASTQ Toolkitを使用して各サンプルを同じリード数 (3,000万、1,000万、100万) にダウンサンプリングしました。ダウンサンプリングは、アプリケーションからサンプルのサブセットのみを処理できる場合 (例: メモリ制約のある*de novo*アセンブル)、またはサンプル処理に完全なデータセットが不要の場合 (例: さまざまなレベルのゲノムカバレッジで特定の方法を検証する場合) に必要です。

## 結果

### シーケンスの主要メトリクスの改善

NextSeq 2000システムに使用したNextSeq 1000/2000 P1 Reagents (600 cycles)は、MiSeqシステムでランしたMiSeq Reagent Kit v3 (600-cycle)と比較した場合、Q30以上のクオリティスコアをより高い割合で示しました。また、NextSeq 1000/2000 P1 Reagents (600 cycles)は最大1億のシングルエンドパスフィルターリード、または2億のペアエンドパスフィルターリードを出力しました。さらに、NextSeq 1000/2000 P1 Reagents (600 cycles)は、MiSeq Reagent Kit v3 (600-cycle)の約15 Gbと比べて4倍以上となる約60 Gbのデータを生成しました。これに加えて、NextSeq 1000/2000 P1 Reagents (600 Cycles)キットを用いた場合、約34時間でシーケンスランが完了し、MiSeq Reagent Kit v3 (600-cycle)を用いた場合のシーケンスランよりも約20時間短いランタイムとなりました (図2)。

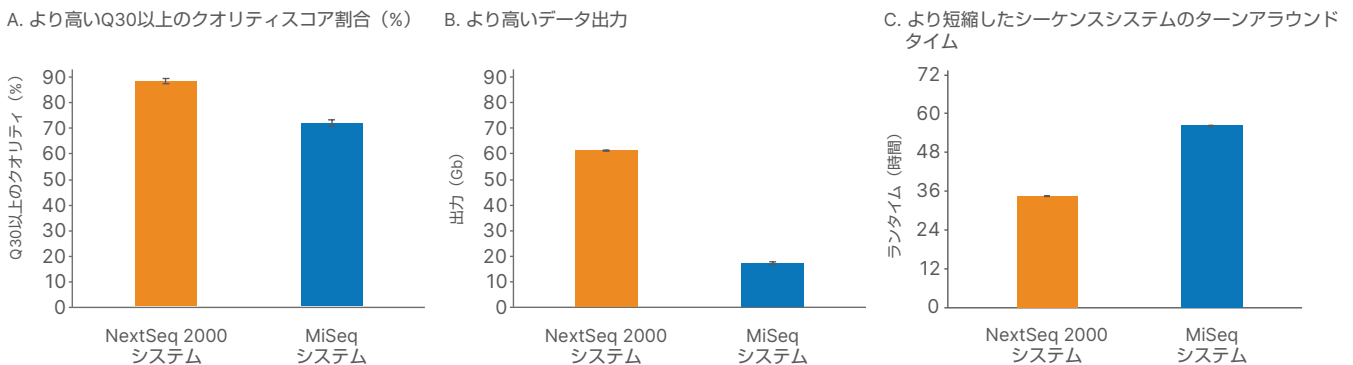


図2: NextSeq 2000およびMiSeqシステムの主要な性能メトリクスの比較: MiSeqシステムでランしたMiSeq Reagent Kit v3 (600-cycle)と比較して、NextSeq 2000システムのNextSeq 1000/2000 P1 Reagents (600 cycles)は、NextSeq P1フローセルを使用した場合、(A) Q30以上のクオリティスコア率が高く、(B) データ出力が4倍高く、(C) 装置ランタイムが約20時間短縮されます。

NextSeq 500/550 High Output Kit v2.5 (300 cycles) (イルミナ、カタログ番号: 20024908) は、150 bp × 2のリード長で最大120 Gbの高品質シーケンスデータを約29時間で生成できました。このキットの品質仕様は、Q30以上の塩基が75%以上です(データは示されません)。

### メタゲノム解析の比較

システム間の性能を比較するために、20 Strain Staggered Mix Genomic MaterialをNextSeq 2000、MiSeqおよびNextSeq 550システムでシーケンスしました。分類を明らかにするための下流解析にはBaseSpace Sequence HubのDRAGEN Metagenomicsアプリを使用しました。メタゲノムNGS解析により、模擬細菌群集の予想されるすべての細菌が同定され、3種類すべてのシーケンスシステム間で同等の結果が示されました(図3)。



図3: 微生物組成の比較解析で示された、NextSeq 2000、MiSeq、NextSeq 550システムで解析したATCC 20 Strain Staggered Mixの組成: DRAGEN Metagenomicsアプリ解析による微生物組成より優れた、再現可能な属の同一性と分布が示されています。

ATCCサンプルでは再現性の高い性能を示しましたが、これは模擬サンプルであるため現実世界のサンプルを用いた場合の性能を表しているとは限りません。そこで、実在する便サンプルを使用して微生物検出性能を検証しました。便サンプルのうちの最も代表的な20の属をNextSeq 2000、NextSeq 550およびMiSeqシステムで比較しました(図4)。このより複雑なサンプルタイプでは結果は均一ではありませんが、すべての実在便サンプルのメタゲノム解析のコミュ

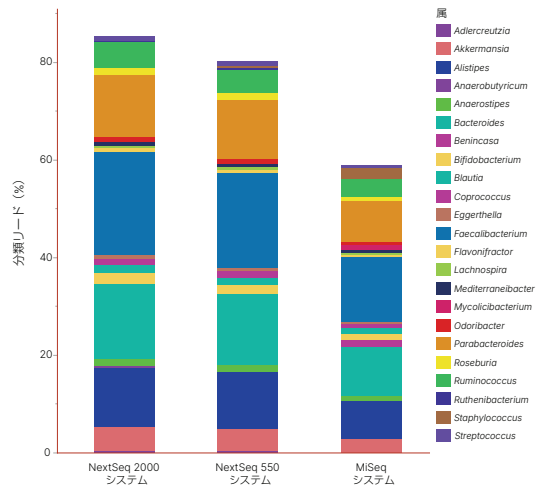


図4: メタゲノム解析で示された、複数のシステムで解析した代表的な実在便サンプルの同様の群集構成: DRAGEN Metagenomicsアプリは、NextSeq 2000、NextSeq 550およびMiSeqシステムでシーケンスした、サンプル内で最も代表的な20属に限定した実在便サンプルの微生物組成を解析します。データは、複雑なサンプルであっても、プラットフォーム間で同様の属をカバーしていることを示しています。

ニティブロファイルでは、NextSeq 2000、MiSeqおよびNextSeq 550システム間で高い一致を認めました。これは、3種類のシステムすべてがさまざまな種類のサンプルに信頼性の高いメタゲノミクスシーケンスを実行できることを示しています。

### リード長によるサンプル特性評価の向上

300サイクルのフローセルも有意義なメタゲノミクスシーケンスデータを提供できますが、600サイクルキットは複雑なサンプル内の個々の種を同定する場合に優れています。メタゲノムアセンブリに対するリード長の効果を実証するために、DRAGEN FASTQ Toolkitアプリを使用してNextSeq 2000システムからのリードを600サイクルから300サイクルにトリミングしました。k-merベースの分類学的分類プログラムであるKraken2\*を使用して、600サイクルまたは300サイクルで各サンプルの分類されたリードの割合を決定しました(図5)。データから、より長いリードの場合、多様な環境サンプルを用いたk-merベースの分類学的手法には、ある程度の改善がもたらされることが示唆されました。

\* Kraken2 Metagenomics分類学的手法は、イルミナDRAGEN Metagenomics BaseSpaceアプリから利用できます。

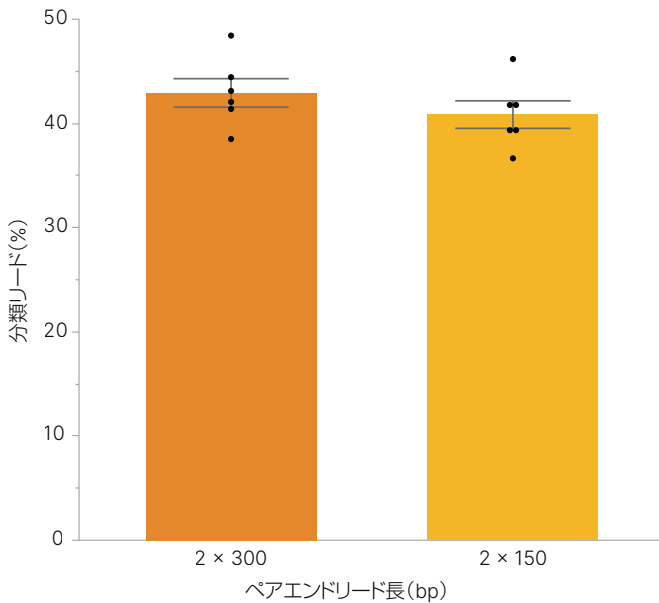


図5：リード長が長いほどNextSeq 2000でシーケンスした実在便サンプルのリード分類が向上します：DRAGEN FASTQ Toolkitアプリを使用して、NextSeq 2000システムからのリードを600サイクル（300 bp × 2）から300サイクル（150 bp × 2）にトリミングしました。次にKraken2を使用して各サンプルの分類されたリードの割合を決定しました。解析の際のリード深度は3,000万リードでした。エラーバーは平均値からの1標準誤差を示します。

同じトリミングされたデータを使用して、サンプルの豊富さ（サンプルに検出される種の数）およびシャノン指数（サンプルに検出される種の相対的存在量を表す指数）に対するリード長の影響も調べました。これらのメトリクスは、リード長が増加すると便サンプルで観察された微生物の多様性が増加する一方、シャノン指数によって定量化される検出種の割合は予想どおり相対的に変化しないことを示しています（図6）。

### シーケンス深度の増加によるサンプル特性評価の向上

次に、実在便サンプルの集団多様性の測定におけるリード深度の重要性を調べました。NextSeq 2000システムからのリード数は、DRAGEN FASTQ Toolkitキットを使用して3,000万、1,000万、100万リードにダウンサンプリングし、豊富さとシャノン指数はDRAGEN Metagenomicsアプリを使用して計算しました。さまざまな種類のメトリクスより、シーケンス深度が増加すると便サンプルの微生物の多様性が増加する一方、シャノン指数によって示された種の割合は相対的に変化しないことが示されました（図7）。

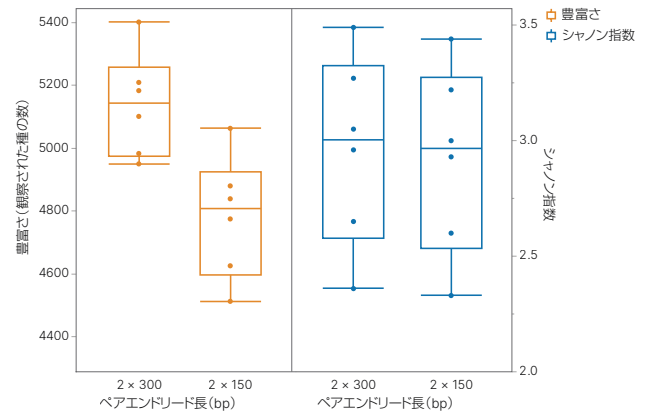


図6：リード長が長いほど便サンプル微生物の豊富さが増加します：DRAGEN FASTQ Toolkitアプリを使用して、NextSeq 2000システムからのリードを600サイクル（300 bp × 2）から300サイクル（150 bp × 2）にトリミングしました。次に、DRAGEN Metagenomicsアプリを使用して、豊富さとシャノン指数を計算し、検出された種の数と比例的な多様性をそれぞれ定量化しました。リードが長くなると豊富さが増加しますが、シャノン指数には相対的な変化は認められません。解析の際のリード深度は3,000万リードでした。

### リードの長さによる分類学的識別の向上

多様な環境微生物集団のプロファイリングにおける現在の課題の1つは、多くの希少種や培養不可能な種に対する完全なリファレンスゲノムが不足していることです。一般に、アセンブリの全長が長くなることからわかるように、非常に多様な微生物集団のアセンブルされるコンティグの数は、リード長が長いほど多くなります（図8）。イルミナの301 bp × 2のシーケンスリード長を使用したショットガンメタゲノミクスシーケンスは、通常、環境サンプルからのメタゲノムの*de novo*アセンブルを向上させるため、アセンブルされた各メタゲノムの全体的な完全性に大きく貢献します。

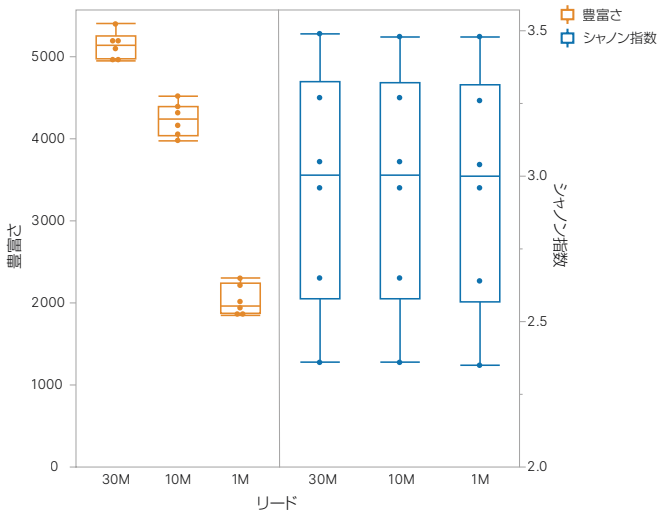


図7：シーケンス深度が増すにつれて便サンプルの微生物の豊富さは増加します：豊富さとシャノン指数は、DRAGEN Metagenomicsアプリを使用して計算し、3,000万、1,000万、100万リードのシーケンスから検出された種の数と観察された比例的多様性を測定しました。予想どおり、ダウンサンプリングされたデータでは豊富さは減少しますが、シャノン指数はほぼ同じままです。

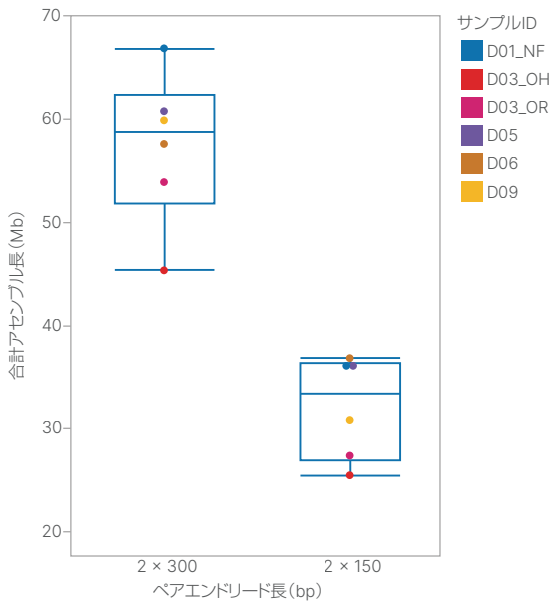


図8：リード長が長いほどNextSeq 2000システムでシーケンスした実在便サンプルの全体的なアセンブル塩基数が増加します：DRAGEN FASTQ Toolkitアプリを使用して、NextSeq 2000システムのメタゲノミクスシーケンスデータを100万リードにトリミングし、300 bp × 2から150 bp × 2のリード長にしました。比較のために、SPAdes Genome Assemblerを使用して、300 bp × 2および150 bp × 2に設定されたシーケンスリードから合計コンティグ長を生成しました。

## まとめ

本アプリケーションノートの目的は、NextSeq 2000システムおよびMiSeqシステム上での600サイクルキットの同等の性能を示すことです。NextSeq 550システムも正確なメタゲノム解析が可能です。600サイクルのオプションにはありません。3種類すべてのシステムは、使用した装置にかかわらず高精度な*de novo*アセンブルを達成しました。一致するメタゲノムプロファイル、特に観察された属の比例的多様性は、3種類のシステムすべてで達成されました。

全体として、NextSeq 2000システムは、リード長とリード深度の強みがあるため、多様な、培養なしのサンプルからサンプル豊富さの詳細を解析するという点で、MiSeqおよびNextSeq 550システムよりも優れています。<sup>†</sup> これらの利点は、便や環境サンプルなどの複雑なメタゲノミクスサンプルにとって特に重要です。

まとめると、NextSeq 1000システムおよびNextSeq 2000システム上でのNextSeq 1000/2000 P1 Reagents (600 cycles)およびNextSeq 1000/2000 P2 Reagents (600 cycles)キットは、MiSeqシステム上でのMiSeq Reagent Kit v3 (600-cycle)と比べて、迅速なターンアラウンドタイムで高品質なシーケンスを提供します。NextSeq 1000およびNextSeq 2000システムの600サイクルキットは、このアプリケーションノートで実証されているように、高いQ30スコアと低いエラー率を維持します。さらに、NextSeq 1000/2000 P1 (600 cycles)およびNextSeq 1000/2000 P2 Reagents (600 cycles)は、小規模および中規模のゲノムシーケンス実験に最適な柔軟なデータ出力を提供します。最後に、NextSeq 1000およびNextSeq 2000システム用の600サイクルキットは、実績のあるMiSeqシステムで確立されたデータ品質を維持しつつ、アプリケーションの拡大および操作の簡便性を実現します。

<sup>†</sup> NextSeq 1000システムは機能的にNextSeq 2000システムと類似しており、同じ600サイクルキットを使用した場合NextSeq 2000システムと同様の結果をもたらします。

## 詳細はこちら

[イルミナシーケンスプラットフォーム](#)

[NextSeq 1000およびNextSeq 2000システム](#)

[NextSeq 1000/2000試薬](#)

[Illumina DNA Prep](#)

[Illumina DNA Prep crude lysate protocol for NGS](#)

## 参考文献

1. Peterson D, Bonham KS, Rowland S, Pattanayak CW; RESONANCE Consortium, Klepac-Ceraj V. [Comparative Analysis of 16S rRNA Gene and Metagenome Sequencing in Pediatric Gut Microbiomes](#). *Front Microbiol.* 2021;12:670336. Published 2021 Jul 15. doi:10.3389/fmicb.2021.670336
2. Durazzi F, Sala C, Castellani G, Manfreda G, Remondini D, De Cesare A. [Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota](#). *Sci Rep.* 2021;11(1):3030. Published 2021 Feb 4. doi:10.1038/s41598-021-82726-y
3. Stothart MR, McLoughlin PD, Poissant J. [Shallow shotgun sequencing of the microbiome recapitulates 16S amplicon results and provides functional insights](#). *Mol Ecol Resour.* 2023;23(3):549-564. doi:10.1111/1755-0998.13713
4. Illumina. 16S rRNA sequencing on NextSeq 1000 and NextSeq 2000 Systems. [illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/nextseq-600c-16s-rrna-application-note-m-gl-01146/nextseq-600c-16s-rrna-application-note-m-gl-01146.pdf](https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/nextseq-600c-16s-rrna-application-note-m-gl-01146/nextseq-600c-16s-rrna-application-note-m-gl-01146.pdf). Published 2023. Accessed February 6, 2024.

## 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

[jp.illumina.com](http://jp.illumina.com)

 [www.facebook.com/illumina](https://www.facebook.com/illumina)

販売店

本製品の使用目的は研究に限定されます。診断での使用はできません。 販売条件：[jp.illumina.com/tc](http://jp.illumina.com/tc)

© 2024 Illumina, Inc. All rights reserved.

すべての商標および登録商標は、Illumina, Inc. または各所有者に帰属します。  
商標および登録商標の詳細は [jp.illumina.com/company/legal.html](http://jp.illumina.com/company/legal.html) をご覧ください。  
予告なしに仕様および希望販売価格を変更する場合があります。