BTSeq™

An NGS-based Innovative Sequencing Platform with High-fidelity and Low-cost

KEY BENEFITS

- **High fidelity sequencing**
  -High sequencing quality (NGS-based)
  -Specific and clear digitized sequencing result
- **Sequencing with longer reads**
  -Up to 20 kb in a single reaction
  -Short turnaround time (TAT), within a day.
- **No need for sequencing primers**
  -Primer information is not required.
  -Only the template is required for sequencing.
- **Low-cost sequencing**
  -Cost-effective compared with Sanger sequencing

Introduction

BTSeq™ (Barcode-Tagged Sequencing) is a new sequencing platform developed to replace the existing methods, analyzing the DNA sequence rapidly and accurately without the DNA length limitation. It enables to prepare the sequencing-ready library for various length and types of DNA within 2 hours and to determine the DNA sequence using NGS and bioinformatics pipeline within a day. BTSeq™ is compatible with most existing NGS platforms, which is available for various applications and also can be automated easily for the high-throughput.

The Workflow: Fast, Easy, High-fidelity and Low-cost Sequencing

In BTSeq™, DNAs are simultaneously fragmented and tagged with molecular barcodes by a simple enzyme reaction (Figure 1), which significantly reduces the experiment time. No extra equipment is necessary because of DNA fragments are made chemically and its simple workflow also decreases the experimental cost.

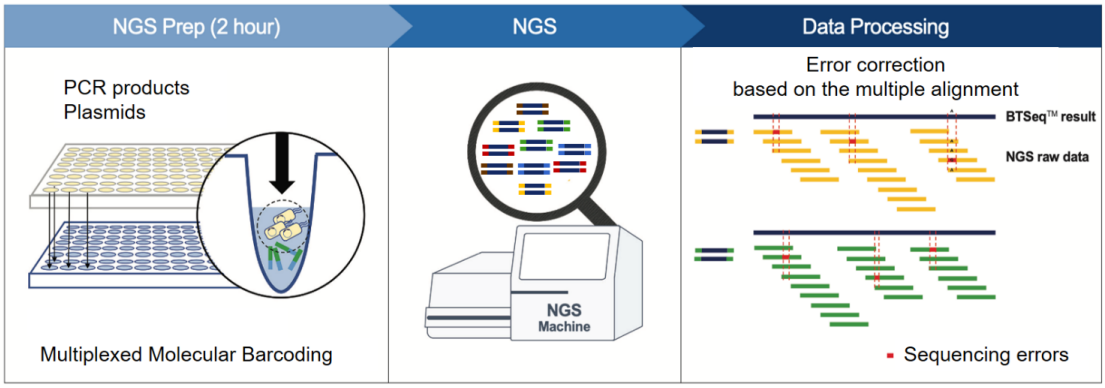
No limitation on DNA Samples

In BTSeq™, the DNA samples to determine their sequence are not limited to their type or length. Both PCR products, plasmids, and even the genomes of bacteria and viruses can be analyzed. Furthermore, sequencing unpurified or small amounts (~0.1 ng) of samples is available with high-quality.
By covering the length from 150 bp up to 20 kb, BTSeq™ can analyze most DNAs that can be sequenced by Sanger method. Moreover, sequencing longer DNA is available in a single reaction. Therefore any consecutive (or repetitive) sequencing such as primer walking is not necessary.
Since BTSeq™ doesn't require sequence-specific primers, it is possible to determine the sequence even if you don't have any available sequencing primers.

Accurate Calls

By using our proprietary bioinformatics pipeline after NGS, the sequencing reads are sorted according to the molecular barcodes, piled up to correct NGS errors, and finally completed the entire sequence of DNA. Therefore, BTSeq™ provides very accurate results with no ambiguity and does not need any reanalysis.

Figure 1. BTSeq™ Workflow



Developed a new enzyme and molecular barcodes, which simplifies sample preparation workflow and reduces turnaround time.

No limitation of samples
- Unpurified sample
- PCR product or plasmid
- Sample amount (>0.1 ng)

Developed a bioinformatics pipeline to reduce NGS errors and to assemble longer consensus sequences up to 20 kbp.

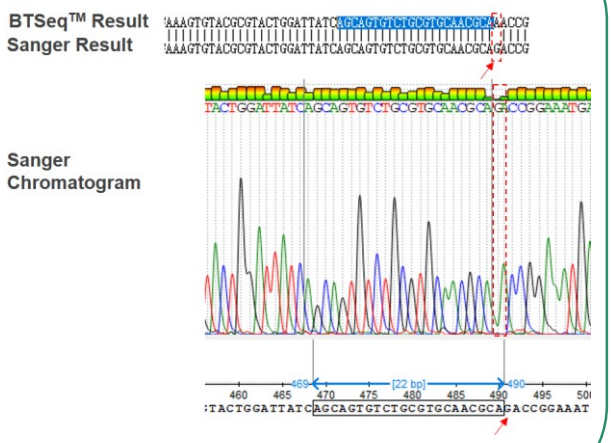
Table 1. DNA sample information in the comparative study

Sample information
▪ DNA type and length PCR product, n=801, 147 bp ~ 2.5 kb Plasmid, n=454, 2.7 kb ~12.9 kb
▪ More than 80% of PCR product samples were not purified.
▪ The range of sample concentration: 0.1ng/ul ~ 200 ng/ul 1 ul of samples (0.1 ng~200 ng) were used for BTSeq™ .

Table 2.
Summary of the comparison study of BTSeq™ and Sanger

# of samples	PCR product	Plasmid	Causes for discrepancy
DNA sample	801	454	
# of match	554	369	
# of discrepancy	247	85	
Causes from BTSeq™	36	0	Difficulty in the exact determination of long repeated sequences
Causes from Sanger	211	85	Poor sequencing results

Figure 3. The discrepancy of a Few Base Pairs due to Poor Quality Base Calls



Comparison of BTSeq™ and Sanger Sequencing

To validate the performance of BTSeq™ as an innovative sequencing platform, a variety of DNAs were sequenced using BTSeq™ and compared with their Sanger sequencing results. A total of 1,285 DNAs were selected, with varying types, lengths, quantities, and purity (Table 1). The whole and a part of respective DNAs were sequenced previously by the Sanger method and we compared both results. Since the starting parts and ending parts of Sanger data tend to be unclear, we only used the middle parts of them excluded from 100 bp region at starting parts and ending parts. After performing BTSeq™ using 1 ul from each DNA, the libraries were pooled and sequenced with paired-end 2×150 reads on the MiSeq system (Illumina). The NGS raw data were analyzed and assembled using bioinformatics pipeline of BTSeq™. Determined sequences were compared with those from Sanger sequencing.

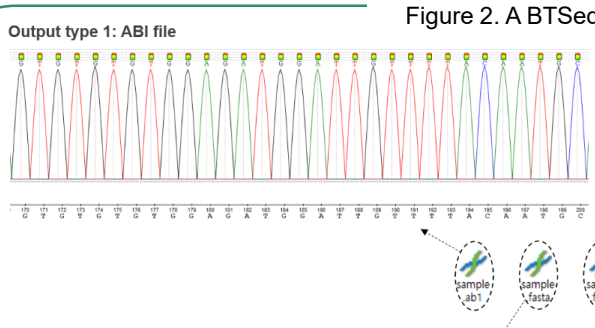
Table 2 shows the result of BTSeq™. All of 1,285 DNA samples were sequenced clearly. Our bioinformatics pipeline produced the high-quality score for every sequence. Only a small amount of DNA (≥0.1 ng) is enough to analyze the various length from 150 bp to 13kb, regardless of its purity. When comparing the determined sequences with Sanger sequencing results (Table 2), 74% of the DNA sequences exactly matched to corresponding Sanger sequencing results. For 26% of DNAs that do not match each other, we analyzed in detail the Sanger chromatogram and the BTSeq™ quality score around the regions of the discrepancy. Poor Sanger sequencing results caused most discrepancies.

The most common pattern of discrepancies was when a few base pairs were mismatched. As shown in Figure 3, there was miscall due to the low-quality result in the Sanger chromatogram (ABI file), while the quality score of BTSeq™ base call for that sequence was high.

User-Friendly Reports

BTSeq™ provides the analysis report composed of a FASTA file and a chromatogram for the DNA sequence (Figure 2). It is similar to that of Sanger sequencing, so very familiar to most customers.

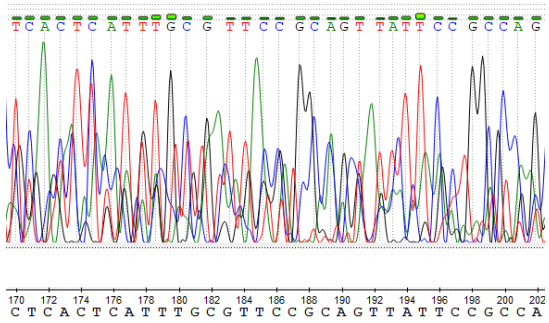
Figure 2. A BTSeq™ Analysis Report



POS	REF	A	C	T	G
170	G	0	0	0	121
171	T	0	0	121	0
172	G	0	0	0	121
173	T	0	0	121	0
174	G	0	0	0	121
175	T	0	0	121	0
176	G	0	0	0	121
177	T	0	0	121	0

The results from each platform often didn't match throughout the whole DNA. Even in this case, the quality of the Sanger result was too low to call the correct bases (Figure 4).

Figure 4. Reaction Failure due to Poor Signals



As a minor error from Sanger sequencing, we observed multiple peaks in certain DNA regions (Figure 5). When homopolymeric sequences exist within the DNA, the results after homopolymers tend to deteriorate in Sanger sequencing (Figure 6). Discrepancies occurred due to this problem. In this case, reanalysis from the opposite direction is required in Sanger to get a clean sequence. On the other hand, BTSeq™ completes the sequencing at once.

Figure 5. Mixed Signals (Multiple Peaks)

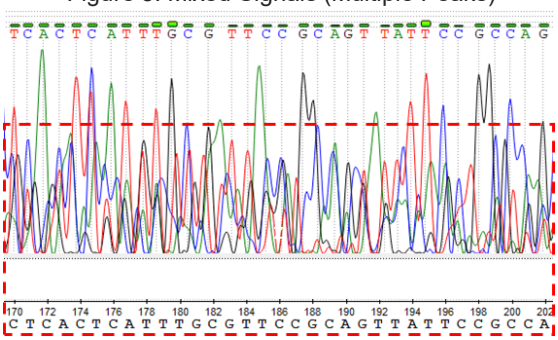
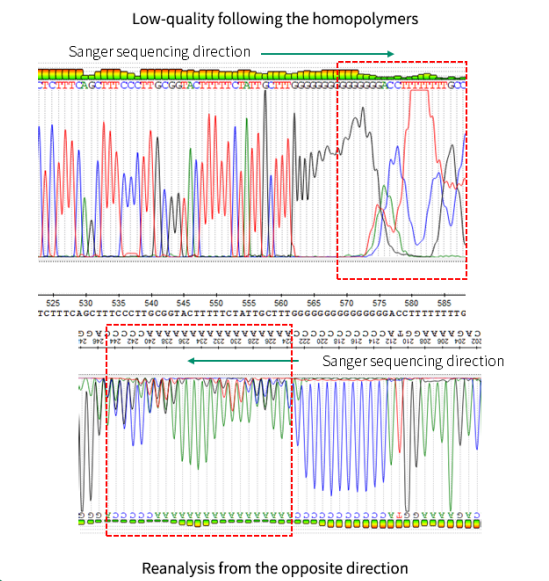


Figure 6. Low-quality Results after Homopolymers



The only case of discrepancy caused by BTSeq™ was when there were long repeated sequences, such as short tandem repeat (STR), within a DNA. These sequences were difficult to determine accurately and should overcome by improvements in our bioinformatics pipelines.

In this study, the performance of BTSeq™ as a powerful sequencing platform was validated. The sequences for all DNAs were determined without any failure or ambiguity, suggesting that BTSeq™ is a sequencing platform that can be used widely as Sanger sequencing. Table 3 shows the key features of Sanger sequencing and BTSeq™. In our comparative study, most discrepancies occurred from the common problems in Sanger sequencing that could be addressed by BTSeq™. When Sanger sequencing was performed again on several DNAs that showed discrepancies, and the results were compared with those of BTSeq™, they perfectly matched (data not shown). Also, the quality of the BTSeq™ result is consistent throughout DNA without deteriorating under specific sequences or regions. Therefore, BTSeq™ is a sequencing platform that can overcome Sanger's limitations.

Table 3. Sanger sequencing vs. BTSeq™

	Sanger sequencing	BTSeq™
Data Type	Analog data	Digitized data
Data Quality	Ambiguous occasionally	Clear
Primer Synthesis	Required	Not required
Template Size	Up to 1 kb	Up to 20 kb
Sample Amount	100 ng	0.1~100 ng

Conclusion

As described so far, BTSeq™ is an NGS-based platform with high-fidelity and low-cost. It overcomes the limitations of existing technology, further can replace Sanger sequencing. As longer DNAs up to 20 kb can be sequenced at once, BTSeq™ can expand its application range, such as genome typing of microbiome and viruses.