HITACHI
Inspire the Next

Improvement of reading accuracy of short DNA using the DS3000 Compact CE Sequencer

Overview

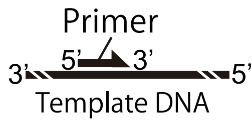
In Sanger sequencing, the quality of the peak shapes may not be good enough to identify the correct base in the beginning of the electropherogram. This is because the sieving matrix for electrophoresis (polymer) filled in the capillary may not be able to detect and separate short DNA properly. This report introduces an example of improving the data quality in the beginning of the electropherogram using a sequencing primer equipped with a tail on the DS3000 Compact CE Sequencer[1]. The tail is a polynucleotide with any sequence and is added to the 5' end of the sequencing primer. The tail lengthens the DNA molecules for sequencing, allowing them to be separated by electrophoresis. Adding a 40-base tail enabled error-free reading from the DNA at the 1st base downstream of the 3' end of the sequencing primer. On the other hand, it was shown that the read length is shortened by the length of the tail. When adding a tail to the sequencing primer, please refer to this report and check the required analysis sections in advance.

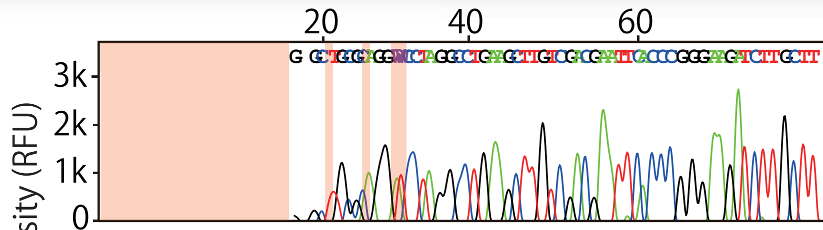
DS3000
Compact CE Sequencer

Results and Discussion

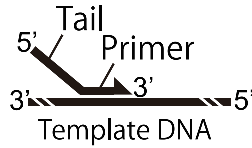
We sequenced plasmid DNA (pTS1, Nippon gene) using a T3 primer. The tail is 40 bases long. First, the effect of the tail on the peak shapes was confirmed. (Please note that the size of the DNA shown in Figure 1 is determined from the 3' end of the primer, with 0 as the starting value.) A primer without a tail does not show peak shapes for 17 bases from the 3' end of the primer. Peak shapes become clear after the 20th base, but there is a definite base call error (Figure 1A, red shaded areas). In contrast, the tailed primer shows the peak of the 1st base from the 3' end of the primer. Furthermore, there is no base call error within the range shown in the figure. We then compared the start positions of the base callings (Figure 2). The DS3000 implements two types of assays for sequencing. Generally, Fast sequencing (Fast-seq) is used when you want to obtain the longest possible read length in a short time, and Standard sequencing (Std-seq) is suitable when you want to obtain a longer read length than Fast-seq. With either Fast-seq or Std-seq, the base callings can be obtained from a region close to the primer by adding the tail. Notably, when using Std-seq, base callings were obtained from the 1st base in all 16 replicates. Finally, we compared the end positions of the sequencing. The end position shown in Figure 3 indicates the location of the last base where the reliability of the identified base remains greater than 99% counting from the 3' end of the sequence primer. In both Fast-seq and Std-seq, the tail shortened the end position (Figure 3). As a characteristic of electrophoresis, the longer the DNA, the lower the separation performance, making it difficult to correctly identify the base. For DNA molecules with an added tail, the base length is increased by the size of the tail, and the separation performance on the lengthened base side decreases. Consequently, the reading end position is shortened.

A. Without tail





B. With tail



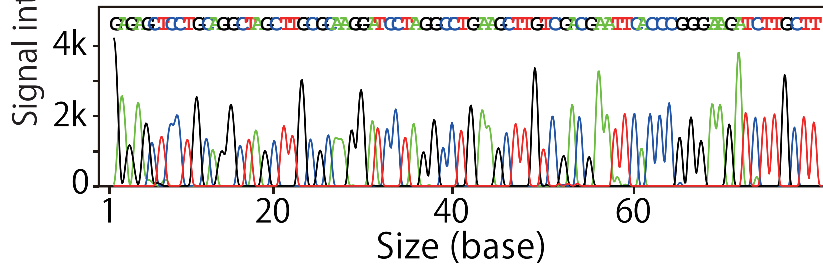


Figure 1: Comparison of peak shapes

The peak shapes obtained using a primer without a tail (A) and a primer with a tail (B) are shown.
The size was counted from the 3' end of the primer starting at 0. Please note that is different from a general sequence viewer.
The red shaded areas represent the part where a result different from the correct sequence was obtained.
The data were obtained with standard sequencing (Std-seq).

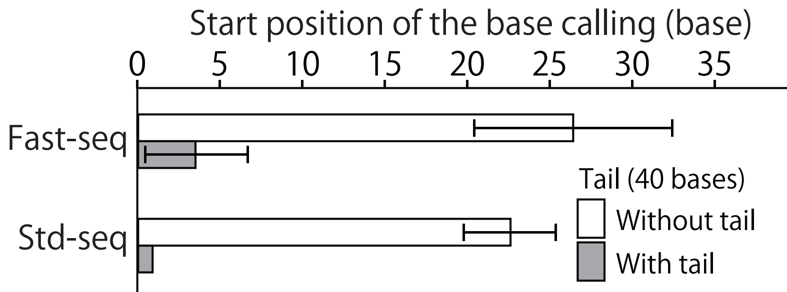


Figure 2: Comparison of the base calling start positions

The 3' end of the sequencing primer is assumed to be 0 to indicate the position from which the DS3000 starts the base calling.
Sixteen trials were run under each condition. The bar graph shows the mean values, and the error bars represent the standard deviations.
In Std-seq under the conditions with a tail, there are no error bars because the base calling started from the 1st base in all trials.

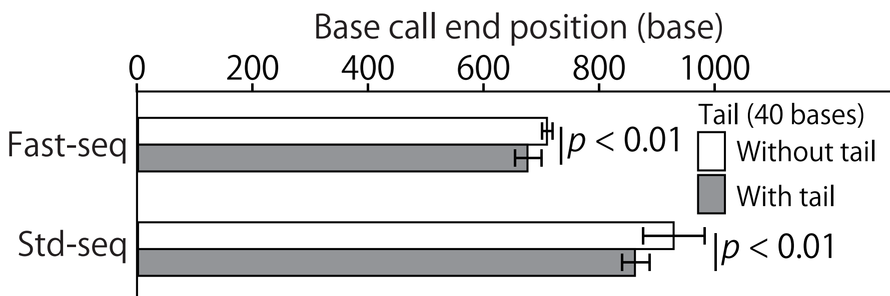


Figure 3: Comparison of reading end positions

The figure shows the end position obtained for each condition. The reading end position indicates the last base position where the reliability of the identified base remains greater than 99%. The 3' end of the sequencing primer was counted as 0.
The bar graph shows the mean values obtained from 16 trials, and the error bars represent the standard deviations.
p values were calculated using the Wilcoxon test.

Conclusion

Adding a 40-base tail to the sequencing primer reduced the base call error in the reading start region and improved the start position of the base calling. In the verification using Std-seq, peaks and base callings were obtained from the 1st base downstream of the 3' end of the primer in all trials. It should be noted that the read base length is reduced by the size of the added tail. When adding a tail to the sequencing primer, please refer to this report and check the required analysis sections in advance.

Troubleshooting

- Noise is introduced in the sequence data if the tail of the sequencing primer anneals with the template DNA. Design the sequence of the tail so that it does not pair with the template DNA. In particular, make sure that the Tm value of the region where the tail anneals with the template DNA does not exceed the Tm value of the primer.
- If the tail anneals with the primer or the tail itself, or forms a higher-order structure, the reliability of the sequence data decreases. Carefully consider the tail and primer sequences when designing the primer with the tail. Some primer design programs have functions for eliminating self-pairing and higher-order structure formation[2]. Please consider using such programs.
- If the tail is too long, it becomes easier to form higher-order structures. It is recommended that the tail be generally no longer than approximately 40 bases.

Materials and Methods

The cycle sequencing reaction was performed using pTS1 (Nippon gene, Tables 1 to 3). The reaction product was purified by ethanol precipitation and dissolved in 10 μL of HiDi™ formamide (Thermo Fisher Scientific). Sequencing analysis was performed on the DS3000 (Table 4). The electropherogram was drawn by modifying the picture obtained using sangerseqR[3].

Table 1. Sequencing primers

Primer	Base sequence
T3 Primer	ATTAACCCTCACTAAAGGGA
T3 Primer with tail	CAAAACGTCGTGAGACAGTTTGTAGGGATAACAGGGTAATATTAACCCTCACTAAAGGGA

Table 2. Reaction conditions of cycle sequencing reaction

Temperature (°C)	Time (s)	Cycle
96	60	×1
96	10	×25
54	5	
60	240	
4	hold	

Table 3. Cycle sequencing reaction system

Reagent	Liquid volume (μL)
Template DNA (100 ng/μL)	2
BigDye™ Terminator v3.1 Ready Reaction Mix	4
10 μM Primer	0.64
BigDye™ Terminator v1.1 & v3.1 5×Sequencing Buffer	2
PCR grade water	11.36

Table 4. Electrophoresis conditions

Assay	Electrophoresis module	Electrophoresis		Sample injection	
		Time	Voltage	Time	Voltage
Fast-seq	AB_Seq_36_Fast	About 30 min	14 kV	4 s	1.2 kV
Std-seq	AB_Seq_36_Std	About 60 min	7.5 kV	4 s	1.2 kV

References
(1) Binladen, J., et al. 5'-Tailed sequencing primers improve sequencing quality of PCR products. BioTechniques. 42 (2), 174-176 (2007)
(2) Untergasser, A., et al. Primer3—new capabilities and interfaces. Nucleic Acis Research. 40 (15), e115 (2012) https://www.primer3plus.com/
(3) Hill, J., T., et al. Poly peak parser: Method and software for identification of unknown indels using sanger sequencing of polymerase chain reaction products. Developmental Dynamics. 243, 1632-1636 (2014)

Main specifications for the DS3000

Main unit specifications

Item	Details
Number of capillaries	4
Capillary length	36 cm
Sample format	8-tube strip × 4
Device control	Touch panel PC
Number of Dyes	6
Application	Sequencing analysis / Fragment analysis
Size	400 (W) × 600 (D) × 600 (H) mm
Weight	45 kg
Performance guarantee temperature	15 – 30°C
Performance guarantee humidity	20 – 80% RH (no condensation)
Power input	100 – 240 ±10% VAC, 50/60 Hz
Rated power	260 VA
Supported secondary analysis software	• Mutation Surveyor (SoftGenetics, LLC, sold separately) • GeneMarker (SoftGenetics, LLC, sold separately) • GeneMarker HID (SoftGenetics, LLC, sold separately)

Run module specifications

Run Module	Application	Polymer type	Contiguous Read Length*1 (bp, QV20 CRL)	Average run time (minutes)
Fast_Sequence36_Polymer7	Sequencing analysis	Polymer7	≥600	≤32
Standard_Sequence36_Polymer7	Sequencing analysis	Polymer7	≥700	≤60
BDx_Fast_Sequence36_Polymer7	BDx sequencing analysis	Polymer7	≥600	≤32
BDx_Standard_Sequence36_Polymer7	BDx sequencing analysis	Polymer7	≥700	≤60

Run Module	Application	Polymer type	Average run time (minutes)	Sizing precision*2 (bp, 50 – 400 bp)
Fragment_Analysis36_Polymer7	Fragment analysis	Polymer7	≤35	NA
Fragment_Analysis36_Polymer4	Fragment analysis	Polymer4	≤44	<0.16

*1 Contiguous Read Length (bp, QV20 CRL) is measured with BigDye™ Terminator v3.1. Sequencing Standard Kit (Thermo Fisher Scientific, sold separately)

*2 Sizing precision (bp, 50 - 400 bp) is measured with PowerPlex™ ESI17 Fast Allelic Ladder and WEN ILS 500 ESS. (Promega®, sold separately)

· BigDye™ is a registered trademark of Thermo Fisher Scientific Inc.

· Promega®, PowerPlex™ is a registered trademark of Promega Corporation.

Consumables specifications

Product name	Part number	Details	Maximum number of uses	Remarks
Capillary Cartridge 36 cm	613-0330	1 pcs	300 injections/unit	Storage temperature: 15 – 30°C
Buffer	613-0252	Anode Buffer × 2 cartridges Cathode Buffer × 2 cartridges	80 injections/unit	Storage temperature: 2 – 10°C
Polymer7	613-0251	4 cartridges	16 injections/unit	Storage temperature: 2 – 10°C
	613-0291	4 cartridges	24 injections/unit	Storage temperature: 2 – 10°C
Polymer4	613-0250	4 cartridges	16 injections/unit	
	613-0290	4 cartridges	24 injections/unit	
Septa for Cathode Buffer Cartridge	613-7231	10 pcs	Single-use	
Retainer for Cathode Buffer Cartridge	613-7233	4 pcs	–	
Septa for 8 well tubes	613-7230	24 pcs	Single-use	
Base and Retainer for 8 well tubes	613-7232	4 pcs	–	
Anode Electrode Assembly	613-7263	1 pc	–	

Specifications in this catalog are subject to change with or without notice,
as Hitachi High-Tech Corporation continues to develop the latest technologies and product for its customers.

CAUTION:For correct operation, follow the instruction manual when using the instrument.

NOTICE:The system is For Research Use Only, and is not intended for any animal or human therapeutic or diagnostic use.
Hitachi High-Tech does not guarantee the performance on the document with every possible sample under every possible condition.
Copyright (C) Hitachi High-Tech Corporation 2023 All rights reserved.

Hitachi High-Tech Corporation

Tokyo, Japan
www.hitachi-hightech.com/global/science/
Toranomon Hills Business Tower, 1-17-1 Toranomon, Minato-ku,
Tokyo 105-6409, Japan
customercenter.ev@hitachi-hightech.com