

Sujin Kim¹, Seung-Chul Shin², Hyun Park² & Jong-Eun Lee¹
¹DNA Link, Inc. Seoul Korea / ²Korea Polar Research Institute, Incheon Korea

Abstract

Along with the advent of next-generation sequencing (NGS) techniques, it has become possible to sequence a microbial genome very quickly with high coverage. Recently, PacificBioscience developed single molecule real-time sequencing (SMRT) technology, 3rd generation sequencing platform, which provide much longer (average read length: 1.5Kb) reads without PCR amplification.

We did de novo sequencing of *Streptomyces* sp. using Illumina GAllx, Roche 454 and PacBio RS system and compared the data. The endosymbiotic bacteria *Streptomyces* sp. PAMC 26508 was isolated from Antarctic lichen *Psoroma* sp. that grows attached rocks on Barton Peninsula, King George Island, Antarctica (62, 13'S, 58, 47'W).

With 4 SMRT cells, we could get more than 15x coverage of corrected sequence data for de novo assembly. Comparing the performance of other sequencing platforms, PacBio platform could generate data in similar manner with general mid-level GC content organism.

In conclusion, PacBio RS system, SMRT technology, shows better performance with high GC content organisms and is expected to be the new tool to improve the *de novo* sequencing and assembly.

Summary

General Information

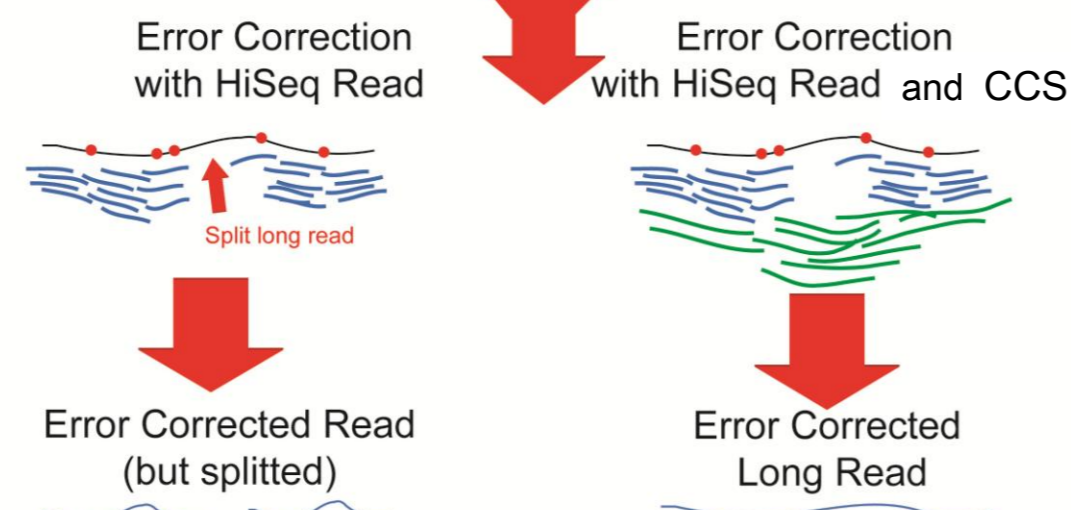
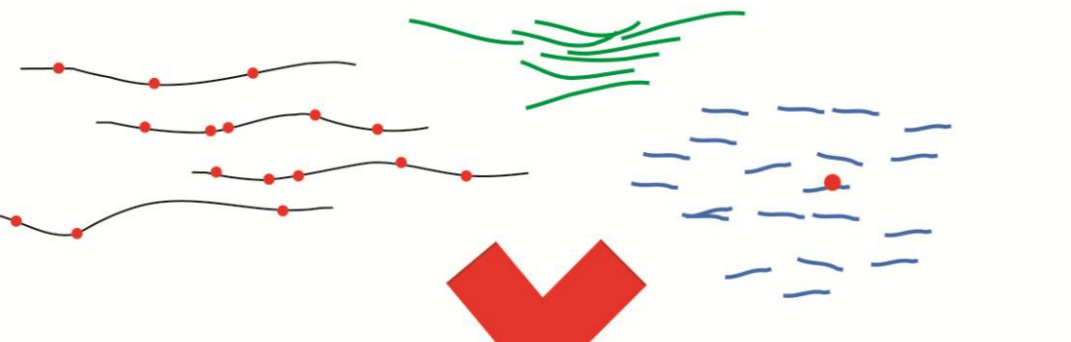
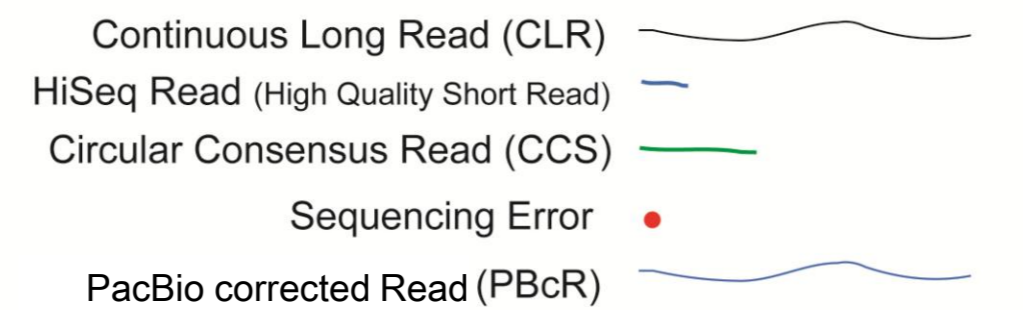
Strain No.	Scientific name	GC content (%)	Genome Size (Mb)	HiSeq Coverage	PBCr Coverage	GSFLX Coverage	CCS Coverage
PAMC 26508	<i>Streptomyces</i> sp.	70.09	7.6	100X	15X	6X	27X

Assembly Summary

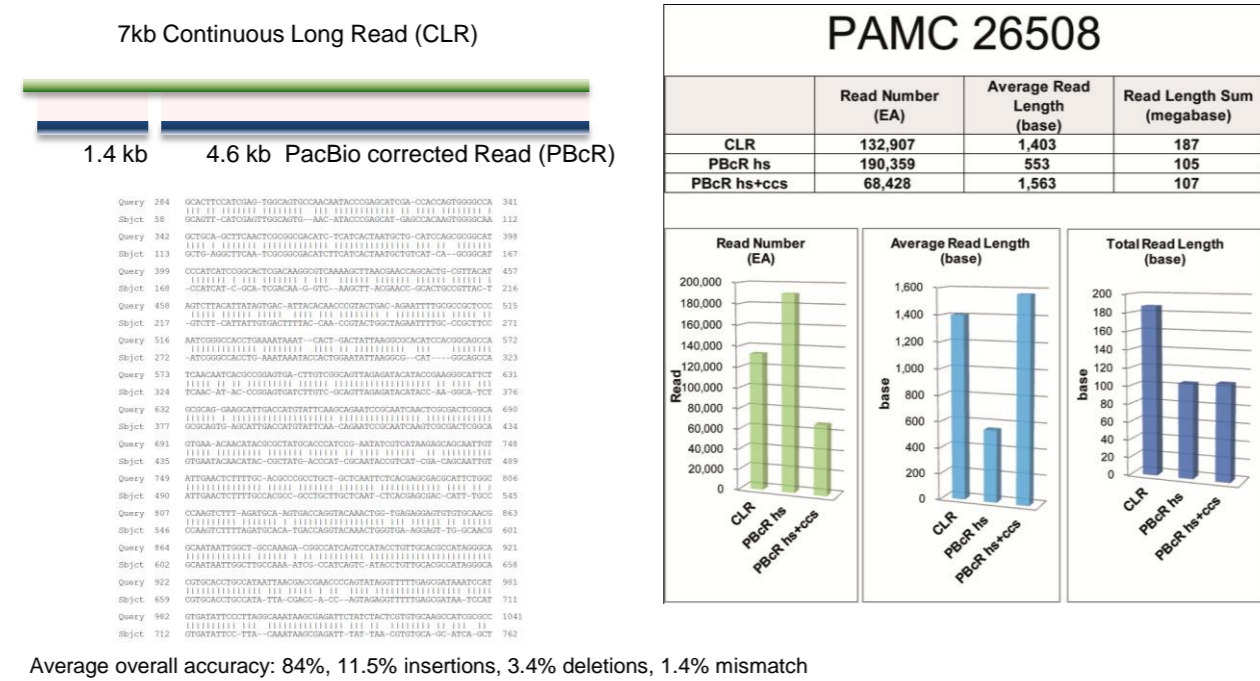
Strain No.	PAMC 26508	
	Scientific name	<i>Streptomyces</i> sp.
Contigs (EA)	HiSeq	185
	PBCr hs	78
	PBCr hs+ccs	26
Max Contig Bases (bp)	HiSeq	221,220
	PBCr hs	644,602
	PBCr hs+ccs	1,915,364
N50 Contig Bases (bp)	HiSeq	68,326
	PBCr hs	166,072
	PBCr hs+ccs	1,268,506
Total Big Contigs (10k <)	HiSeq	141
	PBCr hs	66
	PBCr hs+ccs	11
Big Contig Length	HiSeq	7,372,844
	PBCr hs	7,585,025
	PBCr hs+ccs	7,637,132

Error correction

Error Correction and Assembly



Assemble Error Corrected Reads with Celera Assembler



Assembly

PAMC26508 : *Streptomyces* sp.

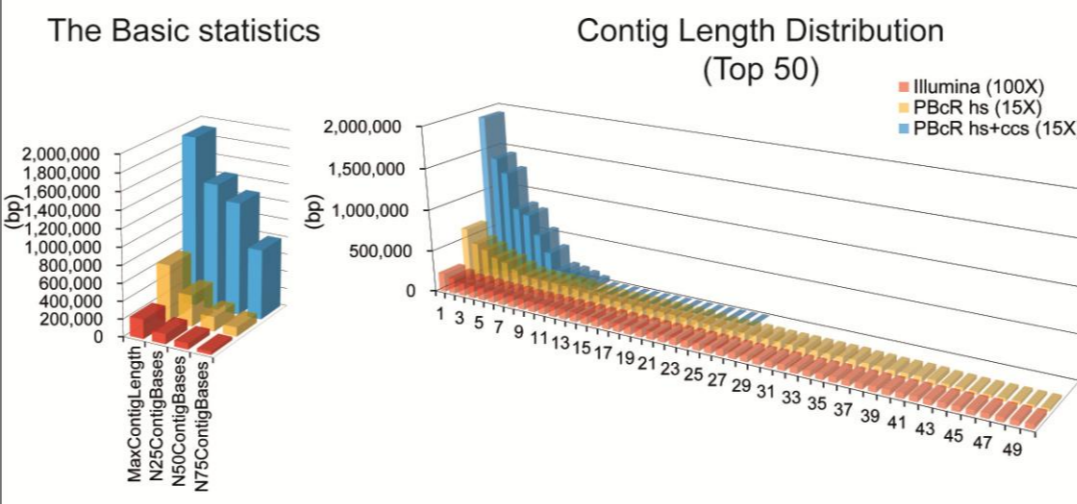
GC content : 70.90%
Genome Size : 7.6Mb

Assembler : Celera Assembler
Sequencing technologies :
Illumina Genome Analyzer : 500bp paired-end library
Pacbio RS: Continuous Long Read and Circular Consensus Read

Assembly results

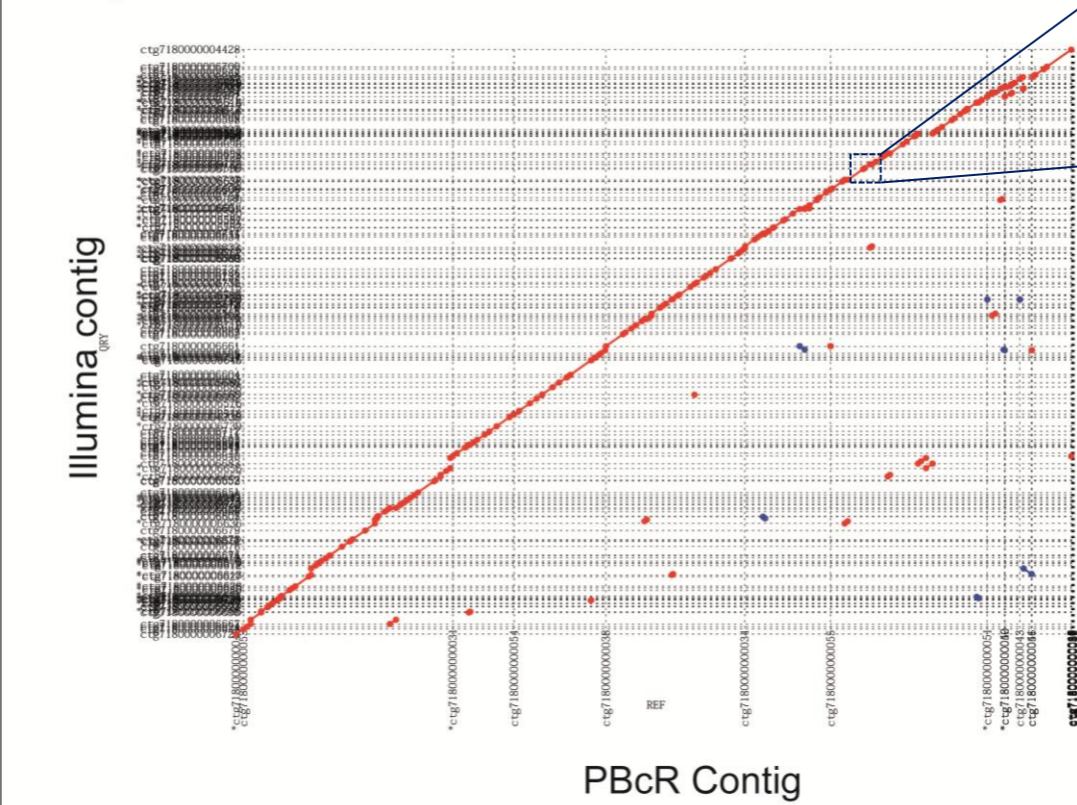
	Illumina (100X)	PBCr hs (15X)	PBCr hs+ccs (15X)
Scaffolds			
TotalScaffolds (EA)	80	78	26
TotalContigsInScaffolds (EA)	185	78	26
TotalBasesInScaffolds (bp)	7,595,626	7,630,446	7,669,881
MaxBasesInScaffolds (bp)	449,494	644,602	1,915,364
N50ScaffoldBases (bp)	202,976	166,072	1,268,506
Contigs			
TotalContigsInScaffolds (EA)	185	78	26
MaxContig.Length (bp)	221,220	644,602	1,915,364
N50ContigBases (bp)	68,326	166,072	1,268,506
TotalBigContigs (10k<)	141	66	11
BigContig.Length (bp)	7,372,844	7,585,025	7,637,132

The improvement of de novo assemblies with Pacbio RS



Assembly comparison

Program : mummer 3.0



SMRT Sequencing Advantage

