

## **DNA Link, Inc** www.dnalink.com

# Genome sequencing of endosymbiotic bacteria Streptomyces sp. from Antarctic lichen using Single Molecule Real-time sequencing(SMRT) technology

### Sujin Kim<sup>1</sup>, Seung-Chul Shin<sup>2</sup>, Hyun Park<sup>2</sup> & Jong-Eun Lee<sup>1</sup> <sup>1</sup>DNA Link, Inc. Seoul Korea / <sup>2</sup>Korea 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 GAIIx, 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 on 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 Infomation									
Strain No.	Scientific name	GC content (%)	Genome Size (Mb)	Hiseq (Coverage)	PBcR (Coverage)	GSFLX (Coverage)	CCS (Coverage)		
PAMC 26508	Streptomyces sp.	70.09	7.6	100X	15X	6X	27X		
Assemb	oly Summar	у							
	Strain	No.			PAMC 265	08			

Scientific n	Streptomyces sp.		
	Hiseq	185	
Contigs (EA)	PBcR hs	78	
	PBcR hs+ccs	26	
	Hiseq	221,220	
Max Contig Bases Bases (bp)	PBcR hs	644,602	
	PBcR hs+ccs	1,915,364	
	Hiseq	68,326	
N50 Contig Bases Bases (bp)	PBcR hs	166,072	
	PBcR hs+ccs	1,268,506	
	Hiseq	141	
Total Big Contigs (10kb < )	PBcR hs	66	
	PBcR hs+ccs	11	
	Hiseq	7,372,844	
Big Contig Length	PBcR hs	7,585,025	
	PBcR hs+ccs	7,637,132	



7kl	o C	continuous Long Read (CLR)	PAMO			
				Read Number (EA)		
1 1 1 1	-	4.6 kb DeeDie corrected D	and (DDaD)	CLR	132,907	
1.4 KD		4.6 KD PacBio corrected Read (PBCR)		PBcR hs	190,359	
				PBcR hs+ccs	68,428	
Query	284	GCACTTCCATCGAG-TGGCAGTGCCAACAATACCCGAGCATCGA-CCACCAGTGGGGCC2	341			
Sbjct	58	GCAGTT-CATCGAGTTGGCAGTGAAC-ATACCCGAGCAT-GAGCCACAAGTGGGGGCAJ	112			
Query	342	GCTGCA-GCTTCAACTCGCGGCGACATC-TCATCACTAATGCTG-CATCCAGCGCGGCAT	398	Read Number	Avera	
Sbjet	113	GCTG-AGGCTTCAA-TCGCGGCGACATCTTCATCACTAATGCTGTCAT-CAGCGGCA1	167	(EA)		
Query	399	CCCATCATCCSGCACTOGACAAGGCGTCAAAAGCTTAACGAACCAGCACTG-CGTTACAT	457	200.000	1 600	
Sbjet	168	-CCATCAT-C-GCA-TCGACAA-G-GTCAAGCTT-ACGAACC-GCACTGCCGTTAC-1	216	180.000	1,000	
Query	458	AGTUTTACATTATAGTGAC-ATTACACAACCOTACTGAC-AGAATTTTGCGCCGCTCO	515	100,000	1,400	
Duerv	516	AATOORGCCACCTGRAAAATAAATCACT-GACTATTAAGGCGCACATCCACGGCAGCC	572	160,000	1 200	
Sbict	272	-ATCOGGCCACCTG-AAATAAATACCACTGGAATATTAAGGCGCATGGCAGCCP	323	140,000		
Query	573	TCARCAATCACGCCGGAGTGA-CTTGTCGGCAGTTAGAGATACATACCGAAGGGCATTCI	631	p <sup>120,000</sup>	1,000	
Sbjct	324	TCAAC-AT-AC-CCGGAGTGATCTTGTC-GCAGTTAGAGATACATACC-AA-GGCA-TCT	376	§100,000		
Query	632	GCGCRG-GAAGCATTGACCATGTATTCAAGCAGAATCCGCAATCAACTOGCGACTCGGC	690	80,000	ä	
Sbjct	377	GCGCAGTG-AGCATTGACCATGTATTCAA-CAGAATCCGCAATCAAGTCGCGACTCGGC	434	60,000	600	
Query	691	GTGAA-ACAACATACGCOCTATGCACCCATCCG-AATATCGTCATAAGAGCAGCAATTG	748	40.000	400	
Sbjct	435	GTGAATACAACATAC-CGCTATG-ACCCAT-CGCAATACCGTCAT-CGA-CAGCAATTGI	489	20.000	000	
Query	749	ATTGAACTCTTTTGC-ACGCCCGCCTGCT-GCTCAATTCTCACGAGCGACGCATTCTGGC	806	20,000	200	
Sbjct	490	ATTGAACTCTTTTGCCACGCC-GCCTGCTTGCTCAAT-CTCACGAGCGAC-CATT-TGCC	545	4		
Query	807	CCAMGECTTT-AGATGCA-AGTGACCAGGTACAAACTGG-TGAGAGGAGTGTGTGCAACC	863	Chi a the cos	~	
Sbjct	546	CCARGTCTTTTAGATGCACA-TGACCAGGTACAAACTGGGTGA-AGGAGT-TG-GCAACC	601	oBCI IST		
Query	864	GCANTAATTGGCT-GCCAAAGA-CGGCCATCAGTCCATACCTGTTGCACGCCATAGGGCP	921	act		
spjet	602	WARLING INCOMPANY AND A CONTRACT AND	030	Q~		
Query	922	COTOLOCIONALITACIACIANICICUANTINOGITITTOACCATAATCCAT	201			
Overv	482	CTORES CONTRACTOR CONT	1041			
Chief	712	ITTELLET THE THEFT ACCOUNTS AND A CONTRACT OF A CONTRACT O	762			

