Title
Assessment, optimization and Applications of 454 FLX Titanium Sequencing Systems

Authors
Singh, Kanwar
Zhao, Zhiying Jean
Zvenigorodsky, Natasha
et al.

Publication Date
2009-03-24

Assessment, Optimization and Applications of 454 FLX Titanium Sequencing Systems

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

Kanwar Singh*, Zhiying Jean Zhao*, Natasha Zvenigorodsky*, Jeff Froula*, Len A. Pennacchio and Feng Chen

ABSTRACT

Next generation DNA sequencing provides new opportunities to efficiently accomplish a variety of genomic tasks such as the de novo assembly of genomes. The newly released 454 FLX Titanium in October 2008 advanced the previous 454 FLX Standard system with nearly double the sequence reads and read lengths. A thorough assessment and optimization of 454 FLX Titanium sequencing system was done by the Technology Development Group at Joint Genome Institute. Currently applications like de novo whole genome shotgun assembly, transcriptome profiling, pyrotag sequencing and sequence capture technology are being optimized on the new 454 FLX Titanium systems.

ASSESSMENT AND OPTIMIZATION OF 454 FLX TITANIUM SYSTEM

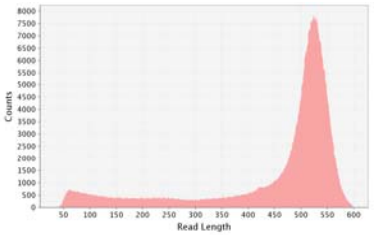
454 FLX Standard *vs.* 454 FLX Titanium

	454 FLX TITANIUM	454 FLX STANDARD
Throughput/Run	400-600 millions	100 millions
Run Time (hrs)	10	7.5
Read Length	400bp	250bp
Accuracy (Q20)	99% at 400 bases	>99.5% at 250 bases
Reads/Run	> 1 million	400K

Record Run at JGI

Region	cpb	Enrichment efficiency	# Total Bases	%Key Pass	%Pass Filter	%Dots	%Mixed	%TSQ	Avg Length
I	0.75	8.0	346,354,570	99	73	3	3	21	408.63
II	0.75	10.0	312,940,739	99	63	5	7	24	412.87
			659,295,309	**99**	**68**	**4**	**5**	**23**	**410.75**

Example of read length distribution of 687,205 reads with average read length of 452 bases



Observed considerable bias in DNA read coverage from high GC samples

As part of the initial assessment of the new 454 FLX Titanium platform, we sequenced six finished genomes of various GC content and aligned the resulting reads back to each corresponding reference genome. We did not observe any GC coverage bias in the two low GC content genomes, *Methanococcus voltae* (28.6% GC) and *Thermoanaerobacter ethanolicus* (34.5% GC), or two moderate GC content genomes, *Pyrobacterium islandicum* (49.6% GC) and *Escherichia coli DH10B* (50.8% GC) (data not shown). However, there was a noticeable effect on coverage for organisms containing high GC content including *Sanguibacter Keddieii* (71.9% GC) and *Brachybacterium faecium* (72.1% GC).

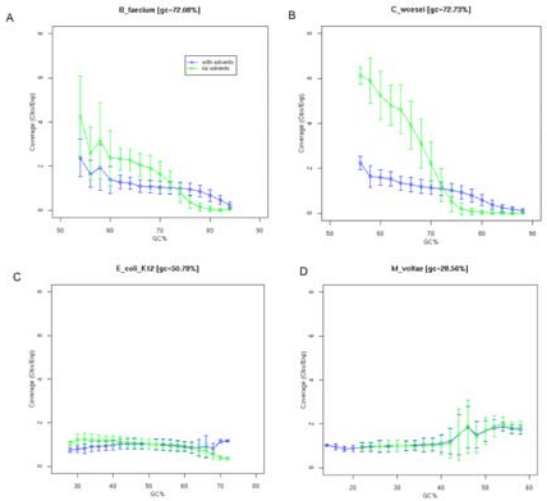
Comparative coverage analysis of 454 FLX Standard and 454 FLX Titanium sequencing

Genome	Genome Size (Mb)	GC of Genome (%)	454 System	Total Reads	Read Length (bp)	Reads Aligned (%)	Fold Coverage (X)	GC of Gaps (%)	Bases Not Covered (%)
Brachybacterium faecium	3.6	72.1	FLX	382,553	218	95.9	22.6	73.6	0.4
			Tiantium	498,553	374	97.5	50.7	78.5	6.2
Sanguibacter Keddieii	4.3	71.9	FLX	507,659	229	99.6	27.2	67.1	0.3
			Tiantium	600,493	365	75.6	39.2	77.5	6.2

Combination of 2.5% DMSO and 0.125M Betaine dramatically improves the coverage of high GC regions

emPCR Co-solvent	Enrichment Efficiency (%)	Total Reads	Read Length (bp)	Fold Coverage (X)	Reads Aligned (%)	GC of Gaps (%)	Bases Not Covered (%)
454 Titanium without solvent	7.1	296,114	373	30.1	98.0	77.2	16.5
5.0% DMSO	6.9	319,448	393	34.6	99.0	79.3	2.1
0.125M Betaine	4.8	325,190	404	36.1	98.7	78.8	4.3
2.5% DMSO + 0.125M Betaine	7.3	316,165	385	33.6	98.9	74.0	0.2

Assessment of 2.5% DMSO and 0.125M Betaine on multiple genomes



The co-solvents have a strong effect for the two high (>70%) genomes (A) and (B) but not on the mid to low GC genomes, (C) and (D). The percent GC (x-axis) was derived by a sliding window of 800bp along the genome. The y-axis reflects the ratio of observed to expected coverage at each percent GC bin.

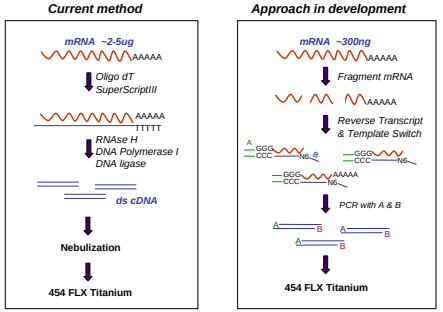
ACKNOWLEDGEMENTS

454 Life Sciences for early access to the 454 FLX Titanium system.
JGI collaborators for providing samples.
Transcriptome Sequencing: **Erika Lindquist, Mei Wang, and Cindy Choi**
PyroTag Sequencing: **Phil Hugenholtz and Anna Engelbrektson**
NimbleGen Capturing: **Crystal Wright and Anna Lipzen**

APPLICATIONS OF 454 FLX TITANIUM SEQUENCING SYSTEM

Overview Of Transcriptome Sequencing Using 454 FLX Titanium

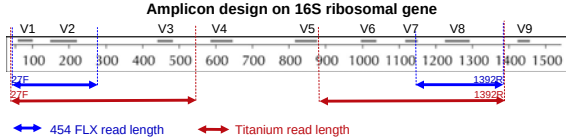
We have implemented transcriptome sequencing using the 454 FLX Titanium platform. The development of *direct RNA-seq* should give less biased coverage of the transcripts.



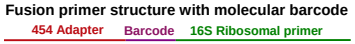
PyroTag sequencing

PyroTag sequencing is a powerful tool used to detect microbial diversity in environmental samples. Ribosomal hypervariable regions, used to distinguish between species may be sequenced together when they are within the target sequencing read length. We are exploiting the longer read lengths of the FLX Titanium platform to increase flexibility in the analysis of these samples.
The use of molecular bar codes allows for the sequencing of multiple microbial communities in one experiment. The high throughput of the Titanium platform allows us to multiplex several microbial communities in one experiment.

Amplicon design on 16S ribosomal gene

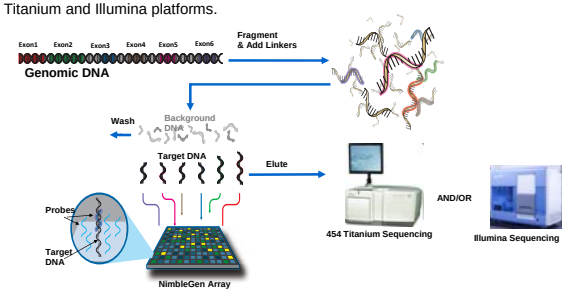


Fusion primer structure with molecular barcode



Sequence capture technology

We have optimized the Nimblegen Captured and Enriched DNA to work with 454 FLX Titanium and Illumina platforms.



* Equally Contributed to Work

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.