Assessment of Nextera Long Mate-Pair Libraries: A Rapid, Low-Input Method for Mate-Pair Library Construction Yields Improved Assemblies

Cindi A. Hoover*[1], Kevin S. Eng[1], Hui Sun[1], Jeff Froula[1], and Feng Chen[1]
[1] DOE Joint Genome Institute, Walnut Creek, CA, USA

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

Introduction

Long mate-pair libraries are invaluable tools for genome assembly. However, traditional methods of long mate-pair library construction require large (20 µg) quantities of DNA and several days of hands-on time. Illumina's Nextera™ Long Mate-Pair (LMP) method is rapid and requires only 1 to 4 micrograms of input material. Here we present an initial assessment of the method for both gel-free and gel size-selected libraries using microbial, fungal, and plant samples. We observed uniform read coverage and high read uniqueness for Nextera™ LMP libraries. Assembly using ALLPATHS-LG generated low contig and scaffold numbers even with relatively low mate-pair coverage.

Methods Overview

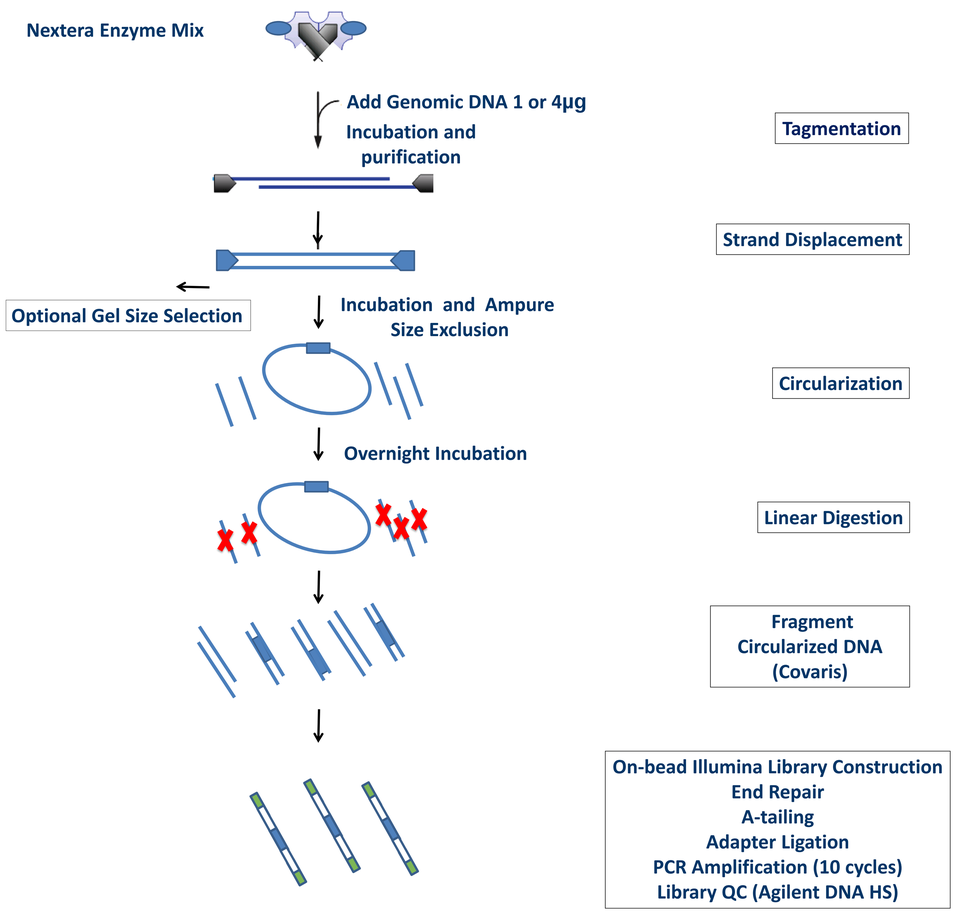

Figure 1. Nextera LMP Workflow

Organisms Tested

Species	%GC	Type
Phycomyces blakesleeanus	36%	Filamentous fungi
Spirochaeta smaragdinae	49%	Gram (-) microbe
Conexibacter woesei	73%	Gram (+) microbe
Cellumonas flavigena	74%	Gram (+) microbe
Suillus luteus	47%	Basidiomycete fungi
Sorghum bicolor	42%	Plant

Table 1. Initial testing organisms and their GC-content



Results

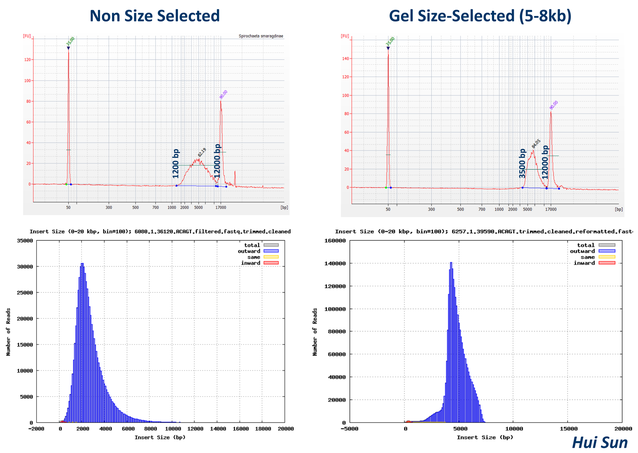

Figure 2. Size distribution of gel-free and gel size-selected libraries

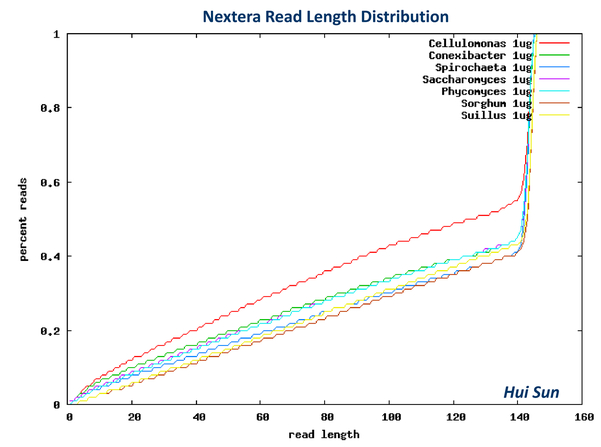

Figure 3. Uniform line indicates that Nextera transposon is inserted randomly across entire read length distribution.

Species	% Mapped Reads Non size-selected	% Mapped Reads Gel size-selected
Phycomyces blakesleeanus	96%	96%
Spirochaeta smaragdinae	97%	98%
Conexibacter woesei	88%	NA
Cellumonas flavigena	88%	94%
Suillus luteus	71%	NA
Sorghum bicolor	94%	NA

Table 2. Nextera LMP yields high percentage of mapped reads.

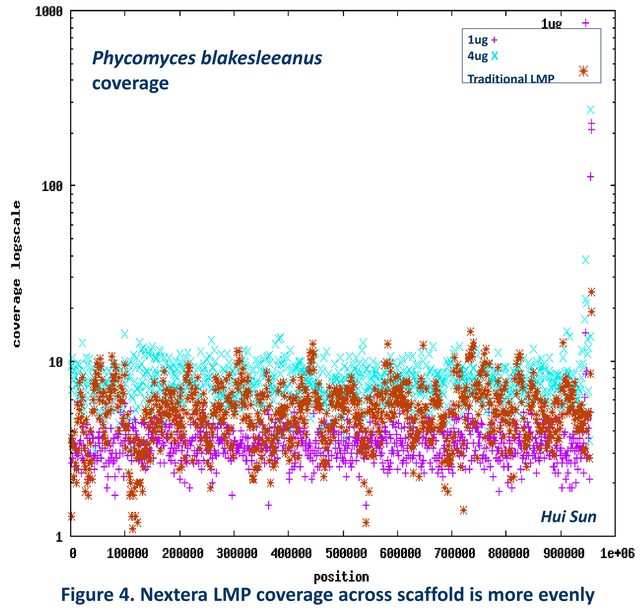

Figure 4. Nextera LMP coverage across scaffold is more evenly distributed compared to the traditional LMP library.

Organism & Assembly Type	Scaffolds	Contigs	Scaffold L50	Contig L50
Conexibacter woesei Frag+Traditional LMP	1	7	6355 Kb	1190 Kb
Conexibacter woesei Frag+ Nextera LMP	1	8	6328 Kb	744 Kb
Cellumonas flavigena Frag+Traditional LMP	8	48	4060 Kb	188 Kb
Cellumonas flavigena Frag+ Nextera LMP	4	27	3493 Kb	408 Kb
Suillus luteus Fragment only	1944	2113	57.6 Kb	51.3 Kb
Suillus luteus Frag+ Nextera LMP	397	1477	240 Kb	54.6 Kb

Table 3. ALLPATHS-LG assemblies were improved with the inclusion of Nextera LMP data compared to traditional LMP data.

Summary

- User-friendly protocol with short hands-on time
- Low template requirement compared to traditional long-mate pair methods (1µg/4 µg)
- Read uniqueness is high for Nextera LMP libraries
- Nextera LMP libraries have uniform read coverage
- Insert size doesn't seem to have significant impact on contig N50
- ALLPATHS-LG generated low contig and scaffold numbers for microbes, even with low coverage
- Addition of Nextera LMP data generally improved assembly results

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231