



Long-span, mate-pair scaffolding and other methods for faster next-generation sequencing library creation

The NxSeq™ 40 kb Mate-Pair Cloning Kit facilitates the creation of scaffolds for *de novo* genome assembly. Supporting either Illumina or 454 sequencing, the kit produces long-span, mate-pair sequences with greater efficiency than existing protocols. In addition, NxSeq DNA Sample Prep Kits can be used to streamline workflow and speed up DNA library preparation for next-generation sequencing.

Large insert mate-pair reads have a major impact on the overall success of *de novo* genome assembly and the discovery of inherited and acquired structural variants. Molecular tools are required that bridge the gap between massively parallel short-read sequencing technologies (35–1,500 bases) and the large scaffolds (40 kb or larger) needed to accurately assemble complex repeat-rich genomes. The NxSeq™ 40 kb Mate-Pair Cloning Kit has been developed to facilitate genome assembly and gap closure. Results show that approximately 40 kilobases (kb) paired-end sequences can be obtained by either Illumina or 454 sequencing at an overall efficiency of >60%, which is significantly better than existing long-span, mate-pair systems. This Application Note will describe the unique pNGS FOS vector, which contains primer binding sites for next-generation sequencing (NGS) from Illumina or Roche 454 platforms, as well as the methodology used to create long-span, mate-pair libraries for *de novo* genome assembly. In addition, we describe the NxSeq DNA Sample Prep Kits, which feature a combined end-repair and A-tailing master mix to enable significantly faster DNA library preparation for NGS applications.

pNGS FOS vector

Lucigen's NxSeq 40 kb Mate-Pair Cloning system uses a unique fosmid vector design. Fosmid vectors are important tools for positional cloning, physical mapping and genomic sequencing. Lucigen's pNGS FOS vector minimizes transcription both into and out of the insert DNA, reducing the cloning bias found with conventional vectors, and is pre-cut and dephosphorylated, eliminating the need

Cheng-Cang Wu¹, Rosa Ye¹, Svetlana Jasinovica¹, Megan Wagner¹, Ronald Godiska¹, Amy Hin-Yan Tong², Si Lok², Amanda Krerowicz¹, Curtis Knox¹, David Mead¹ & Michael Lodes¹

¹Lucigen Corporation, Middleton, Wisconsin, USA. ²Genome Research Centre, University of Hong Kong, Hong Kong. Correspondence should be addressed to C.-C.W. (cww@lucigen.com).

for vector preparation. Most importantly, the pNGS vector contains primer binding sites for NGS from Illumina or Roche 454 platforms. A separate pNGS FOS vector with Ion Torrent primer binding sites will be introduced by the end of 2012.

The pNGS FOS vector contains a single blunt cloning site flanked by primer binding sites for NGS using Roche 454 or Illumina platforms. The vector also lacks the 4-base recognition sites GTAC (RsaI and CviQI) and CTAG (BfaI and FspBI), enabling paired-end or 'di-tag' sequencing of the insert (Fig. 1).

The NxSeq pNGS vector contains the following features: Illumina and 454 primer sites for NGS platforms, absence of the 4-base recognition sites GTAC and CTAG, single-copy and inducible medium-copy replication origins, transcriptional terminators to

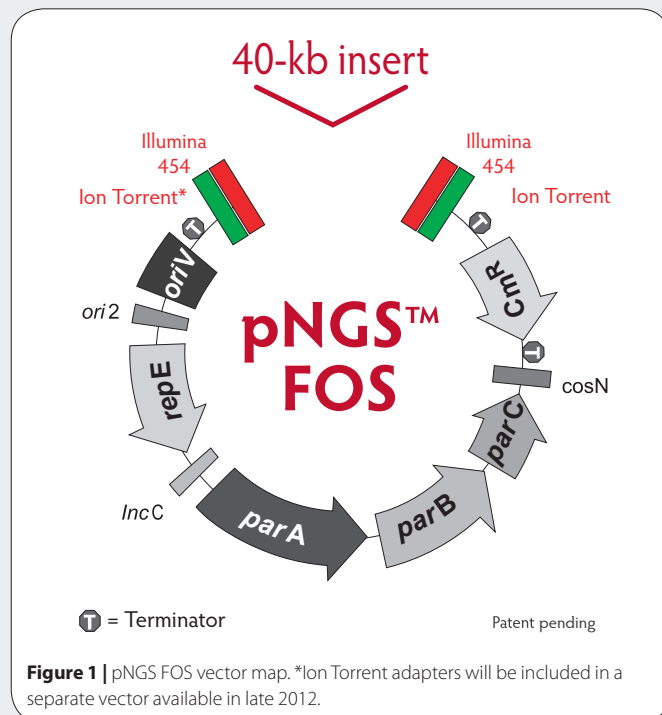


Figure 1 | pNGS FOS vector map. *Ion Torrent adapters will be included in a separate vector available in late 2012.

APPLICATION NOTES

Table 1 | Summary of Lucigen 40-kb long-span, mate-pair human genome library sequenced on Illumina platform.

Stage	No. of reads	%
Pipeline	8,683,854	
Filtered reads	6,506,126	
Reads mapped	5,173,778	79.2
Reads with pair sequences	4,685,996	90.6
Long-span, mate pairs ^a	3,018,576	64.4

^aMapped on same chromosome, ~40 kb apart; uniquely mapped.

stabilize recombinant clones, transcription- and translation-free cloning for unstable DNAs, bacteriophage lambda *cos* site for lambda packaging or terminase cleavage, *loxP* site for Cre-recombinase recognition, rare-cutting restriction sites on either side of the insert, and a chloramphenicol resistance gene.

Unique workflow

The lack of specific 4-cutter restriction sites within the pNGS-FOS vector allows significant workflow enhancements over other methods for long-span, mate-pair libraries. Fosmid packaging ensures an initial insert size of approximately 40 kb, and the inclusion of Illumina and 454 primer sites removes the need for adapter ligation. The basic steps to construct a 40-kb mate-pair NGS library include (i) shear and size-select for 40-kb fragments, ligate to the pNGS FOS vector and perform lambda packaging; (ii) Pool 10^3 – 10^6 clones and purify fosmid DNA; (iii) cut with a 4-base restriction enzyme (RsaI, CviQI, BfaI or FspBI); (iv) reconstitute the restriction site to create a known mate-pair junction; and (v) amplify with either Roche 454 or Illumina primers and proceed to sequencing (Fig. 2).

High-efficiency 40-kb paired-end sequencing

To demonstrate the efficiency of cloning and sequencing with the pNGS FOS vector, we performed large-scale, long-span, mate-pair sequencing of a human cell line (GM15510; Coriell) using Illumina technology. Sixty-four percent of filtered reads accurately mapped to the genome (Table 1). This efficiency, many-fold higher than that of existing systems, will allow the accurate assembly of genomes with NGS platforms for the first time.

A faster method for DNA library prep

In addition to enabling the creation of long-span, mate-pair libraries, Lucigen has also recently released the NxSeq DNA Sample Prep Kits for Illumina and Roche 454 platforms. These kits significantly decrease hands-on and total time to finished libraries by combining end-repair and A-tailing of DNA fragments in one master-mix step. Buffers are optimized to allow direct addition of ligase and adapters to the A-tailed fragments without the need for cleanup steps in between. The result is a kit that can go from sheared DNA to final size selection in less than 2 h, compared to 4 h for the most commonly used library prep kits. Because of the reduced number of steps, hands-on time is reduced by 60–70% compared with existing technologies. Validation studies confirm that coverage and data quality are equivalent to those of platform manufacturers' kits. For more information, please visit <http://www.lucigen.com/ngs/>.

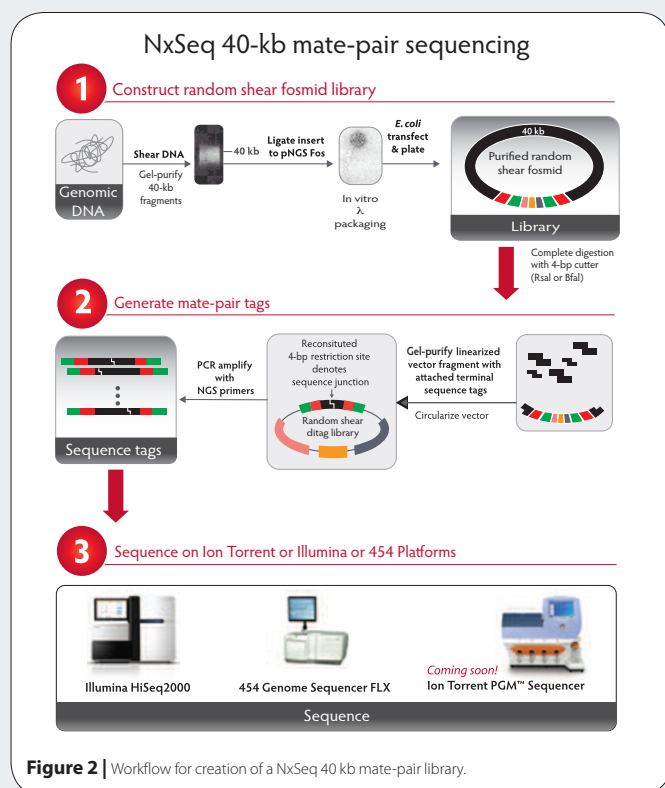
Summary

As the number of *de novo* genomes targeted for sequencing rapidly increases, tools that facilitate the creation of scaffolds used to assemble those genomes are becoming essential. The NxSeq 40 kb Mate-Pair Cloning Kit enables production of long-span, mate-pair libraries with greater control and less bias than previous methods. In addition, standard NGS library construction can be a notable bottleneck in the overall NGS process. The NxSeq DNA Sample Prep Kits reduce the overall time needed for library creation while also offering the shortest hands-on processing time. These new tools will improve efficiency and sequence assembly when adopted by NGS labs.

ACKNOWLEDGMENTS

This work was supported by a Small Business Innovation Research grant from the US National Human Genome Research Institute and by a Wisconsin State loan.

This article was submitted to *Nature Methods* by a commercial organization and has not been peer reviewed. *Nature Methods* takes no responsibility for the accuracy or otherwise of the information provided.

**Figure 2** | Workflow for creation of a NxSeq 40 kb mate-pair library.