# Next Gen Sequencing:
## Library Prep Challenges and Solutions

Rob Brazas, Ph.D.
March, 2016

Lucigen®
www.lucigen.com

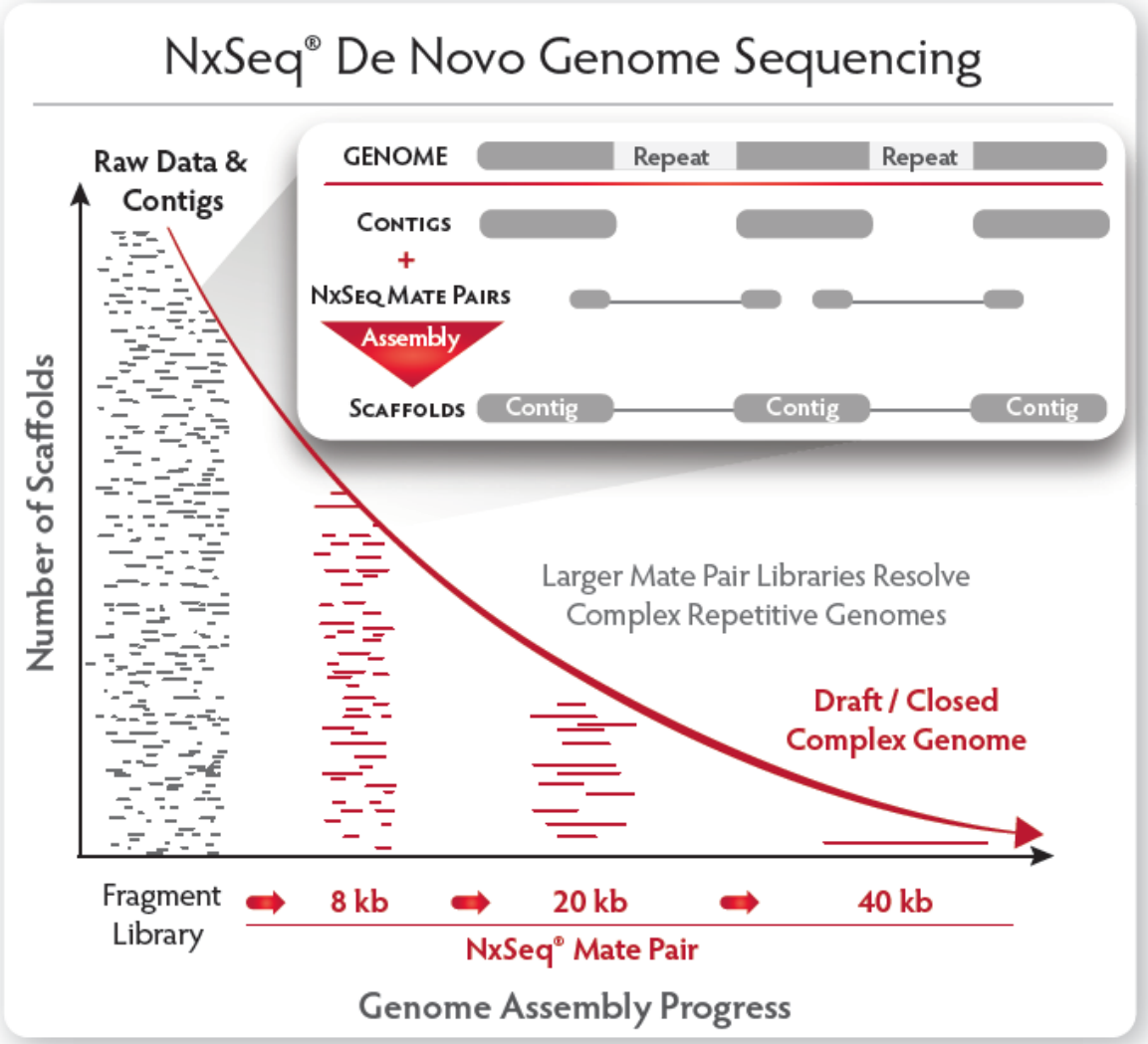
# Agenda

## Helping Solve Challenges in NGS DNA Library Prep

- Overview of de novo next gen sequencing
- Review of DNA fragment library construction
- Challenges of DNA fragment library prep
- NxSeq® AmpFREE Low DNA Library Kit: Solving the major challenges
- Overview of mate pair libraries
- Challenges of mate pair library construction
- NxSeq® Long Mate Pair Library Kit: Overcoming the major obstacles
- Summary

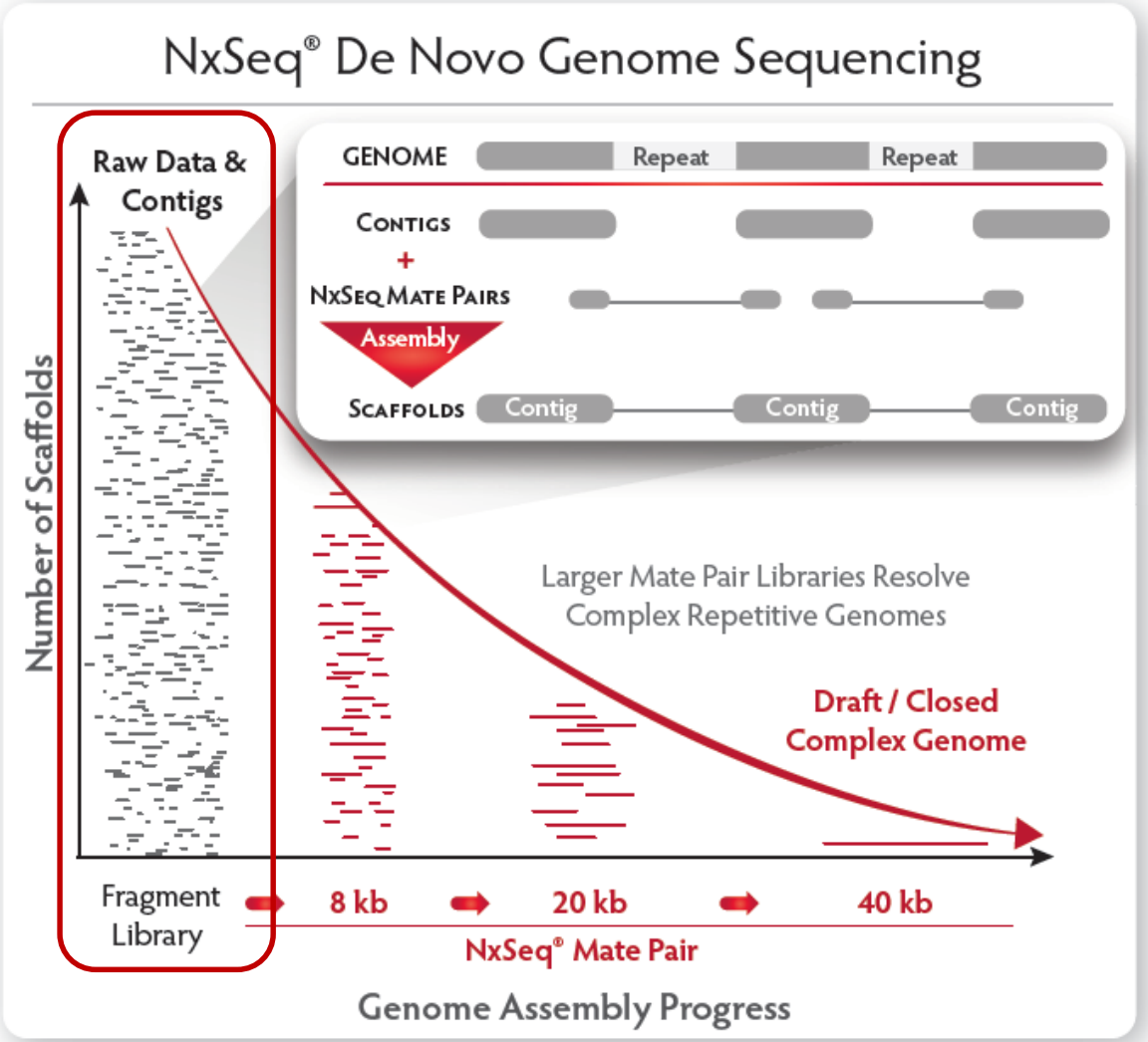
# De Novo Next Generation Sequencing (NGS)

## Multiple Steps to a Sequenced Genome



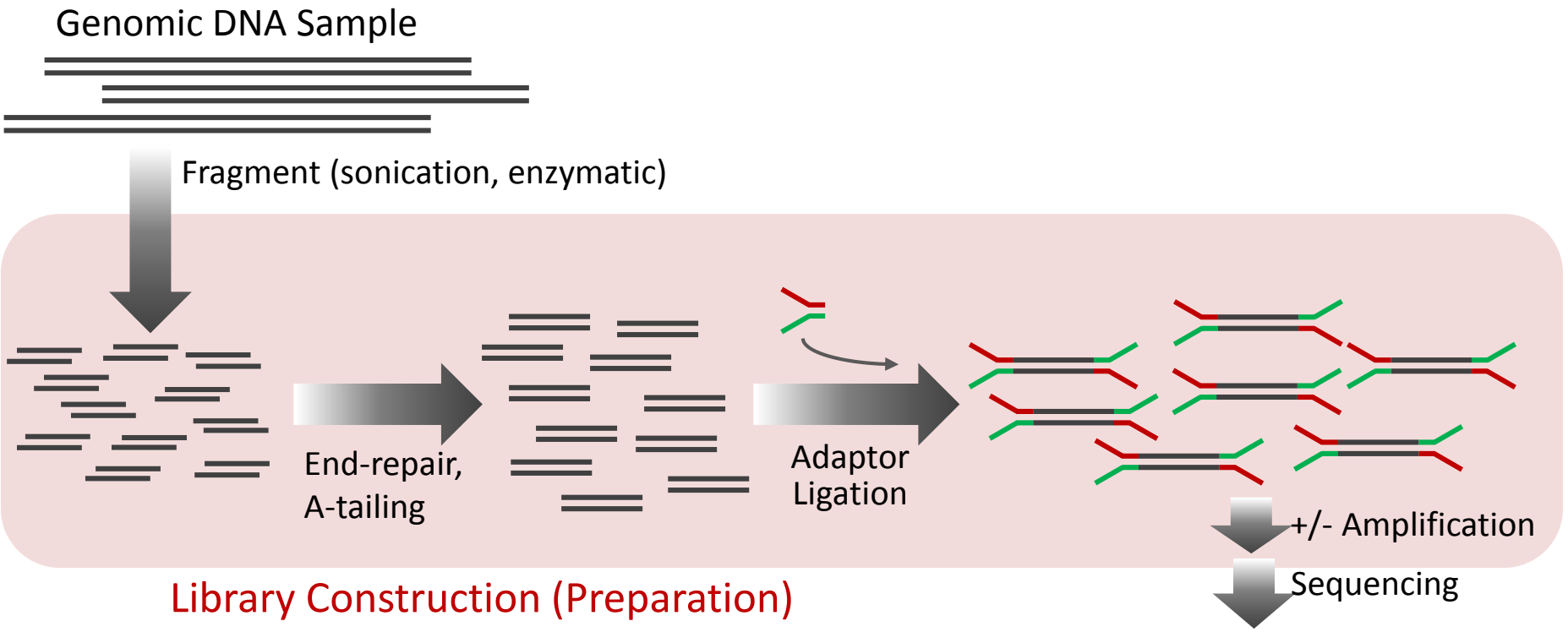
# Other Applications of Fragment Library Sequencing

- Resequencing
- Mutation/SNP detection
- ChIP-seq
- Targeted capture/ sequencing



# Fragment DNA Library Preparation

## A Critical Step in DNA Next Gen Sequencing



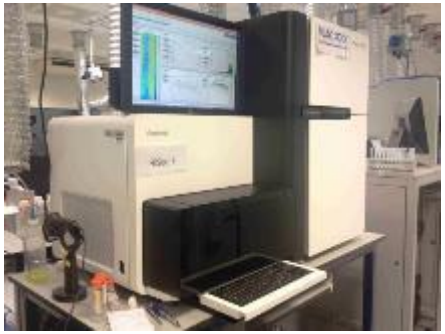

Library Construction (Preparation)

***Illumina***

HiSeq 2500, 3000, 4000

NextSeq 500

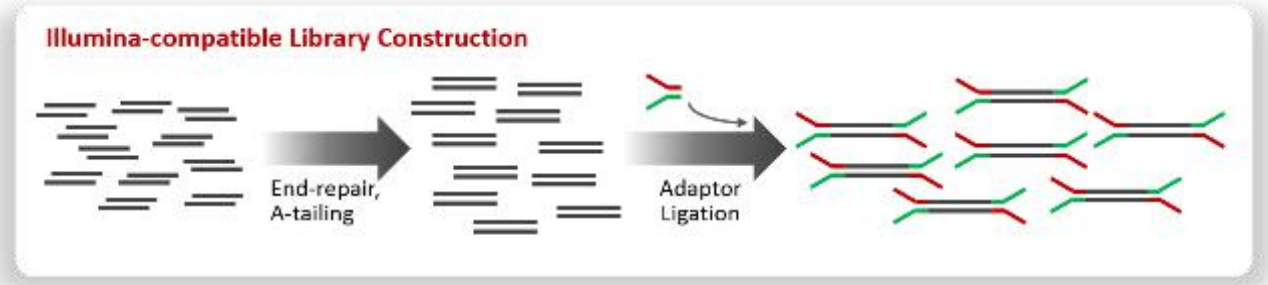
MiSeq



# Challenges of Fragment DNA Library Construction

## Input Amounts, Bias and Efficiency

- **Low amounts of starting DNA**
  - Many samples don't provide enough DNA to do a library prep without an amplification step
- **PCR introduces bias**
  - Some fragments amplify better while others amplify worse
  - Leads to uneven coverage across the genome or target regions
- **Inefficient library construction** (low percentage of fragments with correctly ligated adaptors)
  - Decreases amount of sequencing data obtained from each sample
  - Compounds when samples are multiplexed
  - Increases the number of chimeric fragments



# NxSeq® AmpFREE Low DNA Library Kit

## Minimal DNA Input, No Amplification Bias

- **Low DNA input with no amplification**
  - 75 ng (up to 1 µg) sheared DNA input, increases number of usable samples
  - No PCR = no amplification bias introduced
- **High efficiency reagents and protocol**
  - Optimized end-repair, A-tailing & ligation reactions to produce the highest efficiency libraries
  - Single tube protocol for all enzymatic steps
  - More complex libraries and better sequencing coverage
- **Fast and efficient**
  - DNA libraries in 2 hours = saves time and gets samples on the sequencer faster
- **Automation friendly**
  - Multi-channel 96-well plate protocol developed to increase throughput
  - Easy automation with single tube reaction format

Lucigen®

# Only 4 Key Components Simplify Usage

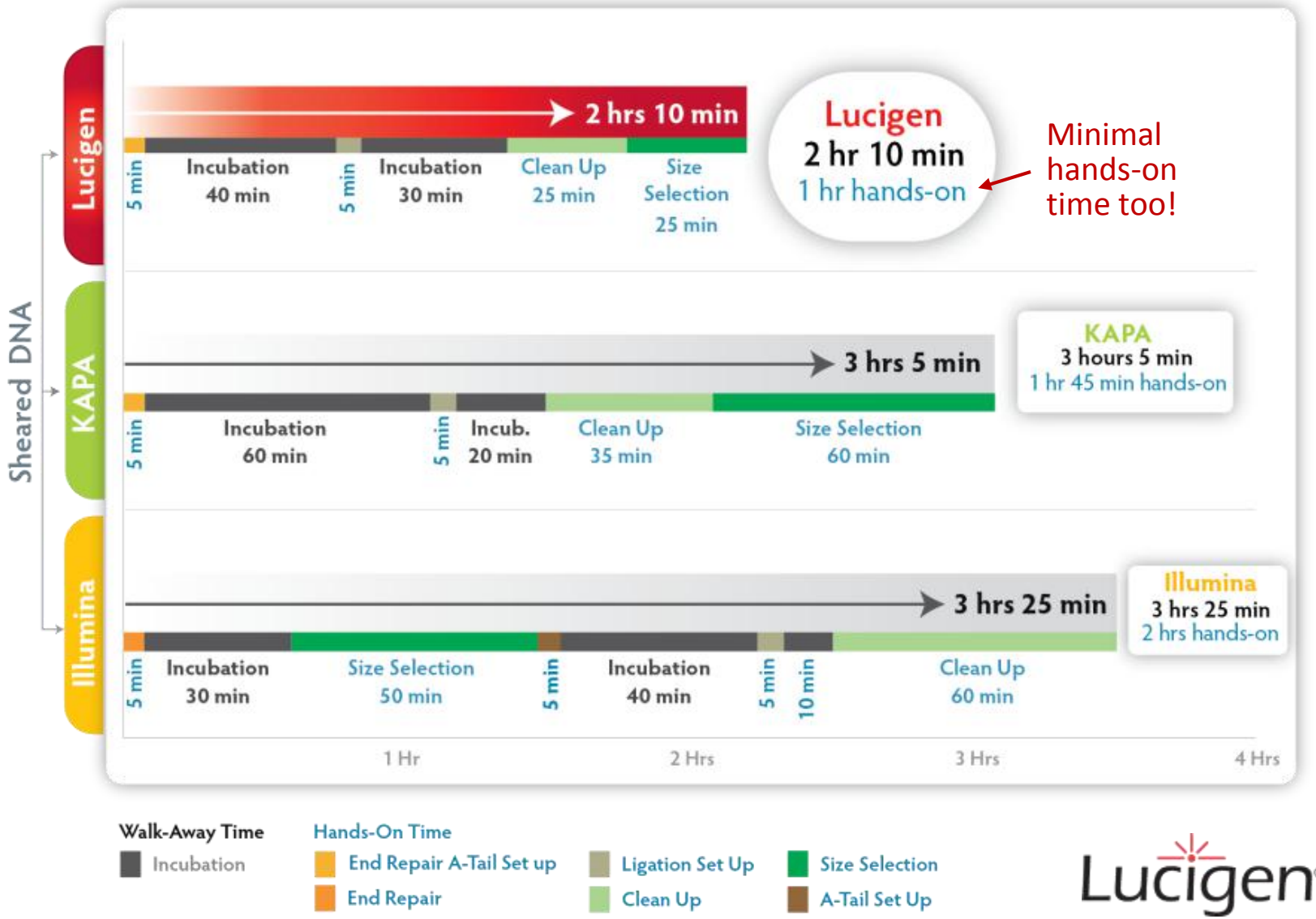
| Product | Lucigen Cat. No. | Size (rxn) |
|---|---|---|
| NxSeq® AmpFREE Low DNA Library Kit | 14000-1 | 12 |
| | 14000-2 | 48 |
| NxSeq® Adaptors, Box 1 | 14300-1 | 12 x 4 (12 adaptors, 4 rxn ea.) |
| NxSeq® Adaptors, Box 2 | 14400-1 | |

**NxSeq AmpFREE Low DNA Library Kit Components**:
Enzyme Mix, 2X Buffer, Ligase, Elution Buffer

*Adaptors must be purchased separately.

Lucigen®

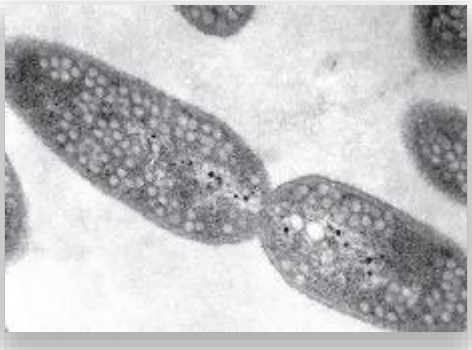
# Fastest Protocol – Only ~2 Hours Total




Lucigen®

# Multiple Genomes Analyzed to Cover Complexity and GC Content

| Genome Characteristic | Human | *Staphylococcus aureus* | *Rhodobacter sphaeroides* | *E. coli* K12 |
|---|---|---|---|---|
| **Percent GC** | 45% | 24% | 68% | 50% |
| **Size** | > 3 Gb | 2.8 Mb | 4.6 Mb | 4.6 Mb |



Lucigen®

# Sufficient Library DNA Generated from Only 75 ng of Sheared Input DNA

* MiSeq requires input of library DNA at 2 nM in 5 µL volume

| | Total Library Yield in 20 µL | Library Concentration | | Volume of Library required to create 5 µL of a 2nM stock |
|---|---|---|---|---|
| Library -Input | ng | ng/µL | nM | |
| Human - 75 ng | 28.6 | 1.43 | 3.61 | 2.77 µL |
| Staph – 75 ng | 29.8 | 1.49 | 4.33 | 2.31 µL |
| Rhodo – 75 ng | 23.2 | 1.16 | 2.83 | 3.53 µL |

**Results:**

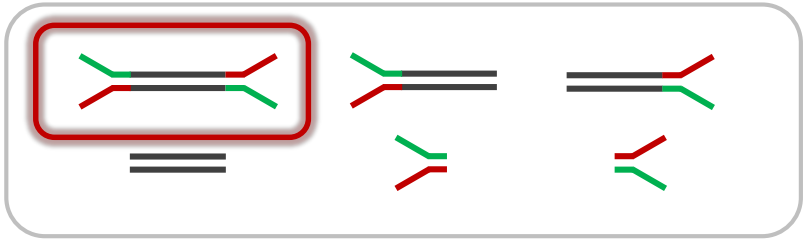
- Final library in 20 µL
- On average, generated >1.7X more concentrated library than needed for a MiSeq run
- Sufficient yield from each library at a high enough concentration for multiple sequencing runs (≥5 runs per library)

Lucigen®

# Measuring Library Efficiency by qPCR

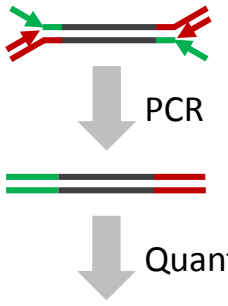
## A Simplified View of the Protocol

What percentage of library is correct?



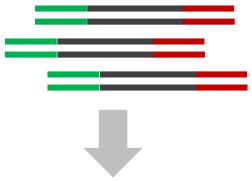
### qPCR Quantitation Method

1. Make or buy an amplified library to use as a standard.



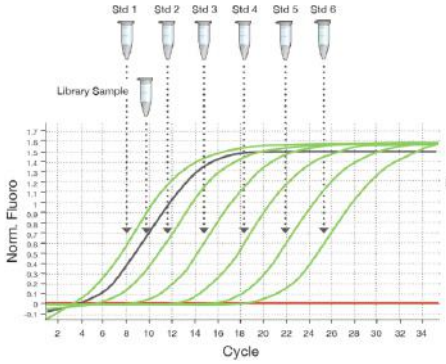

*e.g.* 100 ng/µL

2. Serially dilute library to make standards or use kit standards

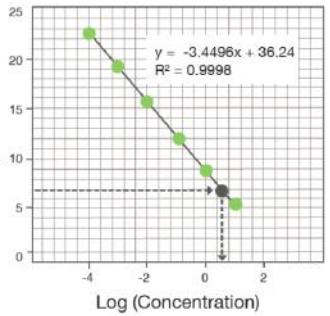


*e.g.*
Std 1: 10 ng/µL
Std 2: 1 ng/µL
Std 3: 0.1 ng/µL
Std 4: 0.01 ng/µL
Std 5: 0.001 ng/µL
Std 6: 0.0001 ng/µL

3. Do qPCR with standards and unknown library




4. Divide qPCR quantitation by fluorescence quantitation to get library efficiency (% correct)



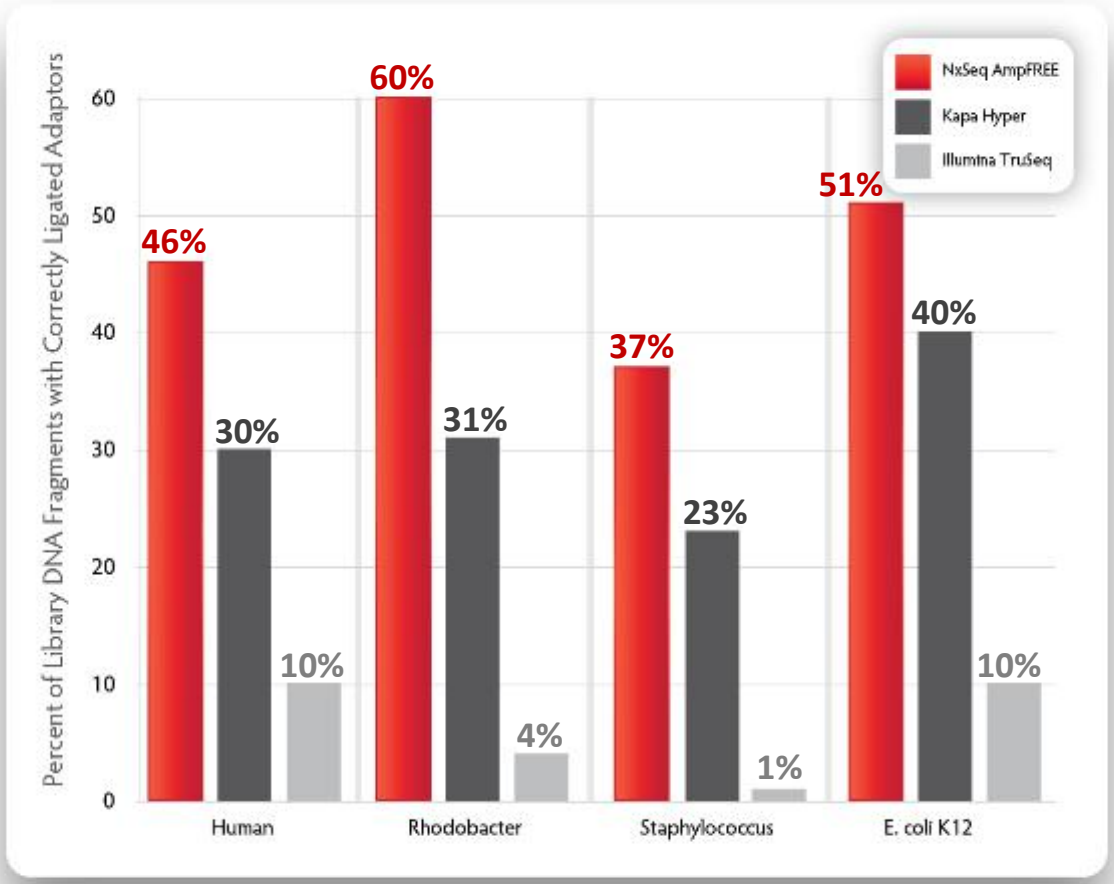

qPCR =1.5 ng/µL

Fluorescence =10 ng/µL

Library efficiency =15%

# Highly Efficient Library Construction

## Produces More Complex Libraries




**Methods**:

- Duplicate fragment DNA libraries were prepared with gDNA from each organism using the indicated kits according to manufacturer's recommended protocols and DNA input amounts:
  - NxSeq® AmpFREE, 75 ng
  - Kapa Hyper Prep, 250 ng
  - Illumina TruSeq PCR-free, 1 µg
- Adaptor ligation (library) efficiency was measured by qPCR using the Kapa Library Quantification Kit (Complete, ROX Low, #KK4873) and the results averaged.

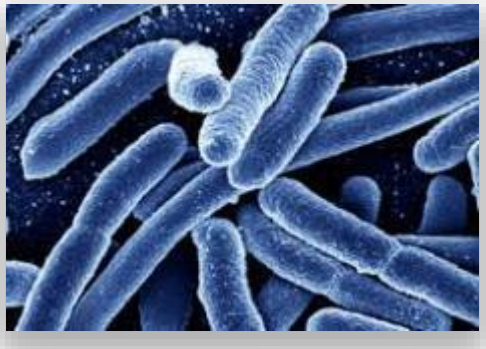
Lucigen®

# What Does Higher Efficiency Really Mean?

## More Sequencing Reads Per Library

| | Number of Reads Per Sample (Multiplexed MiSeq Run) | |
|---|---|---|
| **Library Kit** | ***Staphylococcus aureus*** | ***E. coli* K12** |
| **NxSeq® AmpFREE** | **5,649,946** | **4,305,882** |
| **Kapa Hyper Prep** | 4,838,726 (-15%) | 1,647,452 (-62%) |
| **Illumina TruSeq DNA PCR-Free** | 38,768 (-99%) | 1,543,558 (-64%) |

Note: Added same molar amount (based on fluorescence quantitation) of each library to a multiplexed MiSeq run



# Higher Library Efficiency Improves Data from Challenging FFPE Samples

| Library Kit | Sample Type | Input Amount | Total Reads | Mapped Reads (repeat masked) |
|---|---|---|---|---|
| **NxSeq® AmpFREE** | Normal gDNA | 75 ng | 2,163,636 | 900,338 (41.6%) |
| | FFPE gDNA | 75 ng | 1,767,818 | 688,074 (38.9%) |
| | FFPE gDNA | 150 ng | 1,706,714 | 656,658 (38.5%) |
| **Kapa Hyper Prep** | Normal gDNA | 250 ng | 1,567,276 (-28%) | 650,296 (41.5%) |
| | FFPE gDNA | 250 ng | 1,270,870 (-28%) | 487,872 (38.4%) |

- Samples: **Biochain normal gDNA** (Cat. #: D1234142-S02) and **matched FFPE human kidney tissue** (Cat #: T2234142-S02)
- Extracted DNA from FFPE tissues using **Qiagen AllPrep DNA/RNA FFPE Kit**
- Sheared DNA samples to about 250 bp and made libraries in parallel following recommended protocols for each kit
- Added same molar amount (based on fluorescence quantitation) of each library to a multiplexed **MiSeq run** (7 libraries in total, some data not shown)
- **2 x 150 bp** sequencing chemistry

# MiSeq Human Genome Sequencing Example
## >95% of Reads Map to the Genome

| Human DNA Input Library | |
|---|---|
| Genome size, GC | ~3 Gb  45% GC |
| Raw reads | 3,131,114 |
| Mapped reads | 2,979,237 (**95.15%**) |
| Read length | 148.9 bp |
| Total bases | 443,767,447 |
| Genome fraction | 0.11 |
| Avg. coverage | 0.15X |

- Used 75 ng sheared input DNA for fragment library prep
- 2 x 150 bp sequencing chemistry

Lucigen®

# Other MiSeq Genome Sequencing Examples
## Highly Mappable Reads

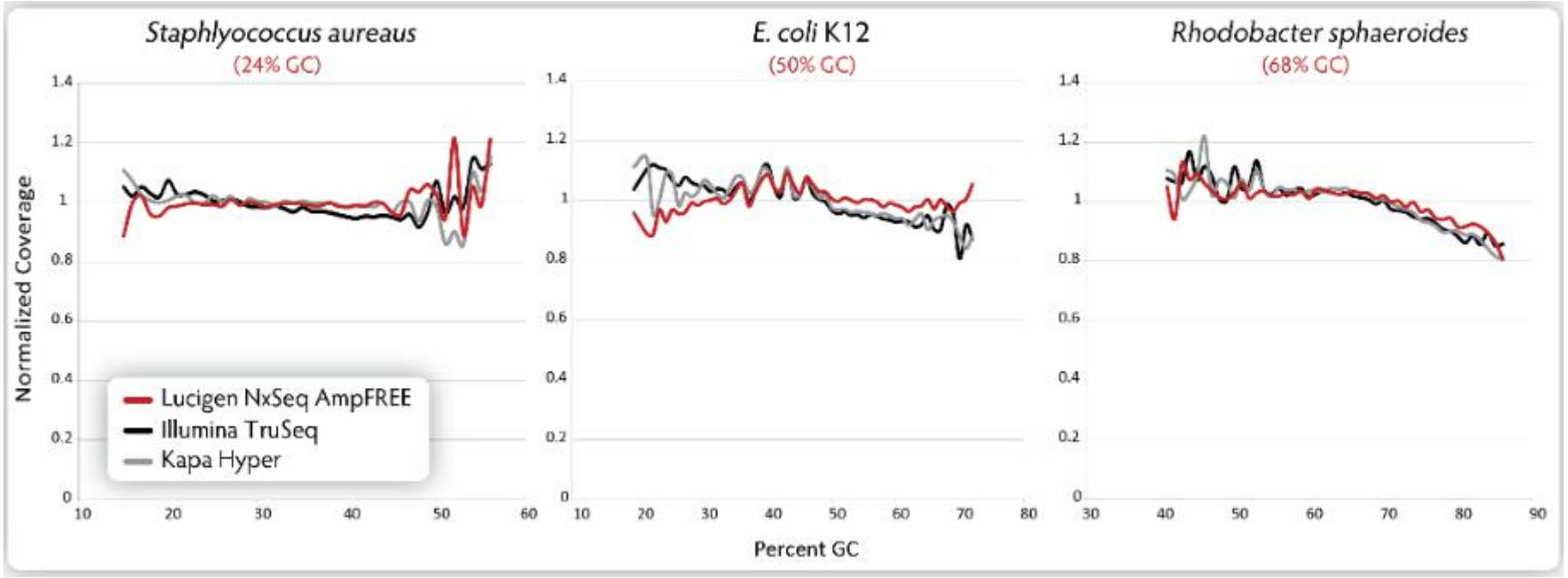
| | *Staphylococcus* | *Rhodobacter* |
|---|---|---|
| Genome size, GC | 2,821,361 24%GC | 4,602,977 68%GC |
| Raw reads | 1,260,836 | 3,900,174 |
| Mapped reads | 1,174,111 (**93.12%**) | 3,613,165 (**92.64%**) |
| Read length | 148.8 bp | 149.6 bp |
| Total bases | 174,694,261 | 540,403,552 |
| Genome fraction | 0.97 | 1.00 |
| Avg. coverage | 62X | 117X |

- Used 75 ng sheared input DNA for fragment library prep
- 2 x 150 bp sequencing chemistry

Lucigen®

# Minimal/Similar Bias to Other Kits Promotes Even Coverage Across Different Genomes

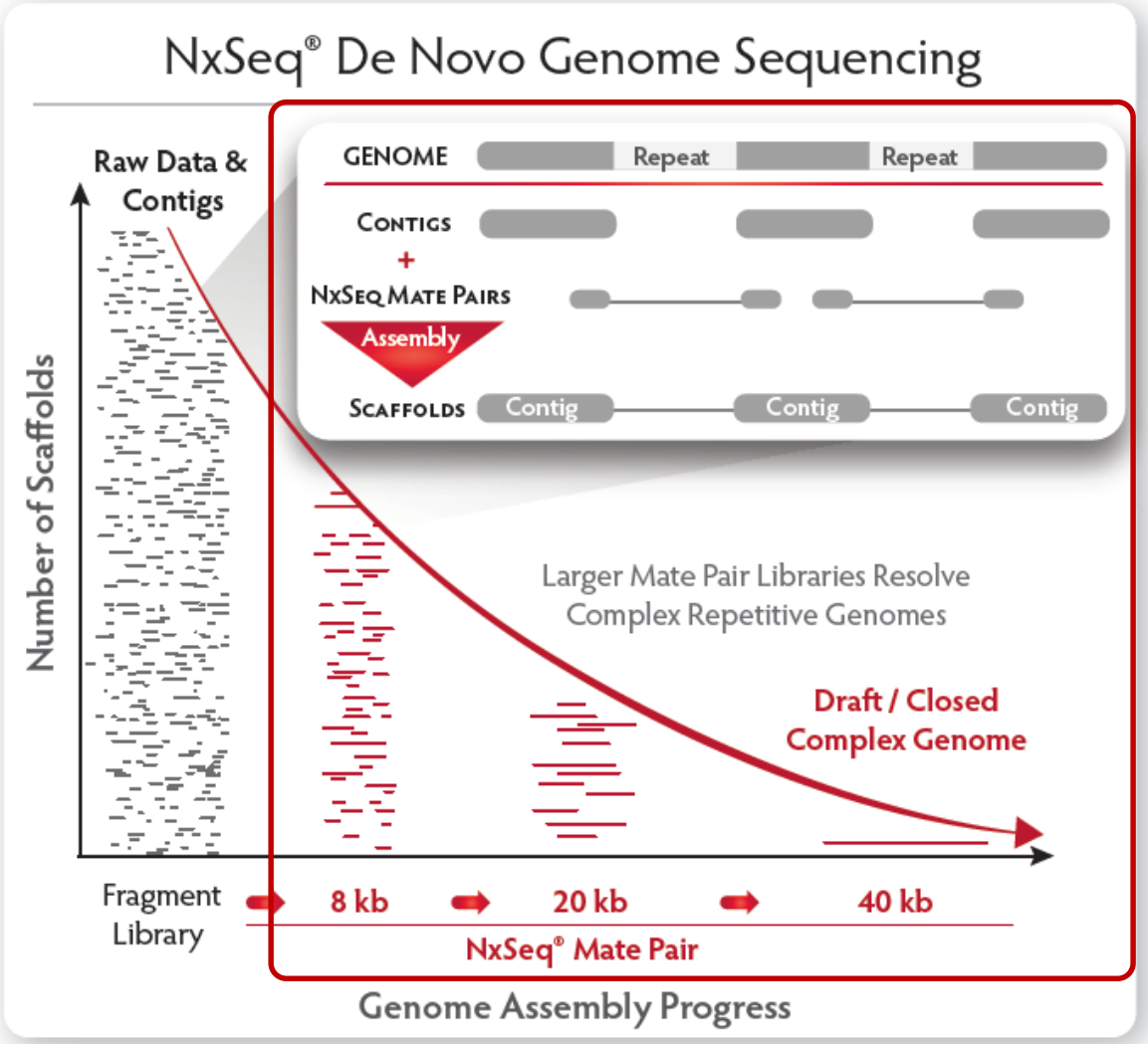
Occurs due to low number of genomic regions at this percent GC




$$\text{Normalized Coverage} = \frac{\text{Average coverage of all windows with X\% GC content}}{\text{Overall average coverage}}$$

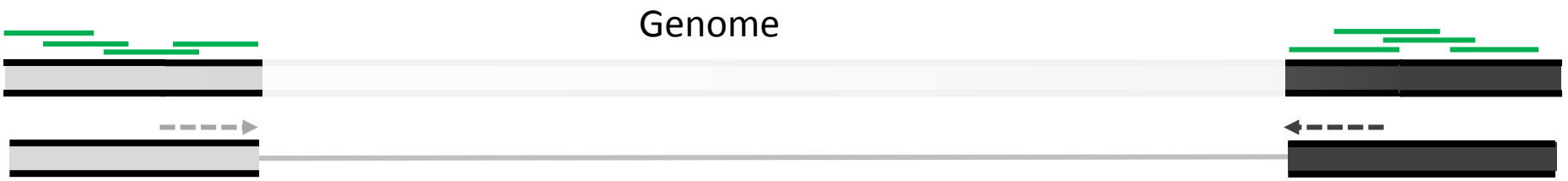
# De Novo Next Generation Sequencing (NGS)

## Multiple Steps to a Sequenced Genome

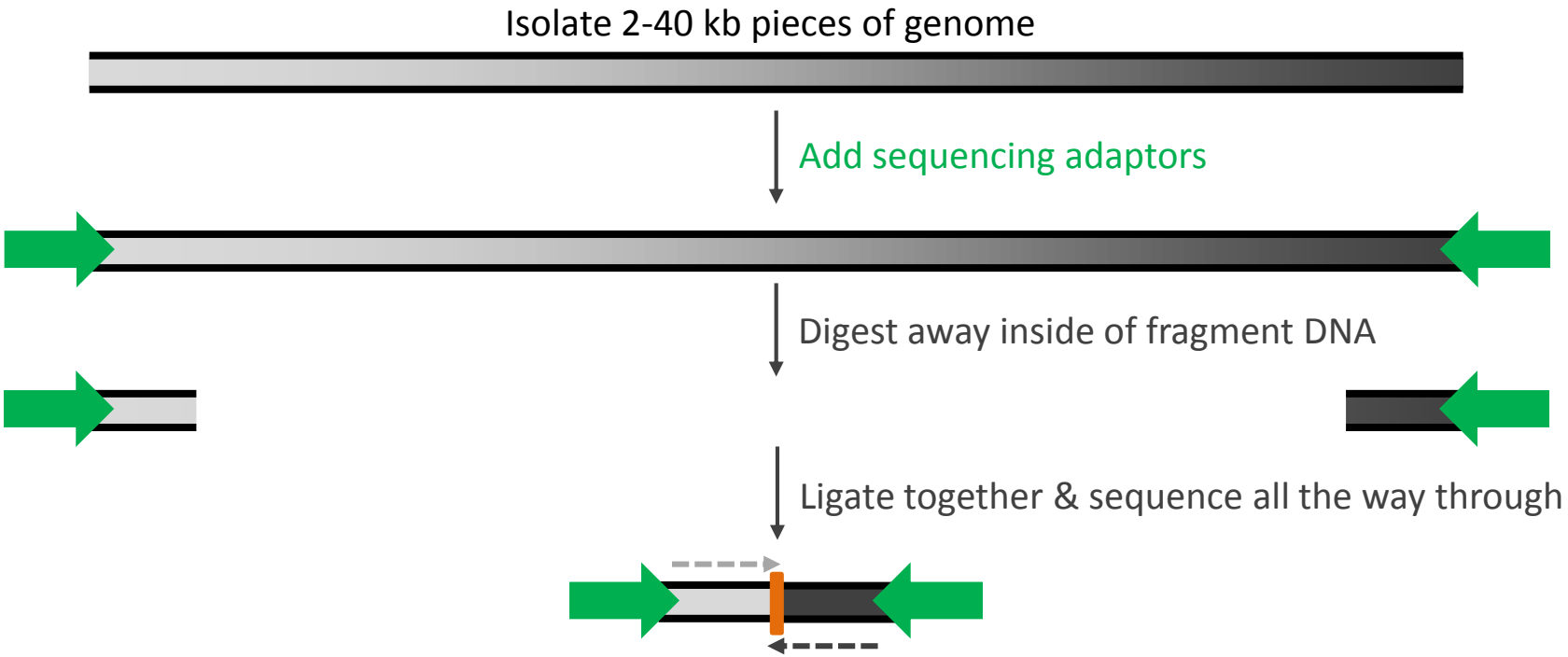


# Mate Pair Library Construction

## Bringing the Ends of Long Fragments Together

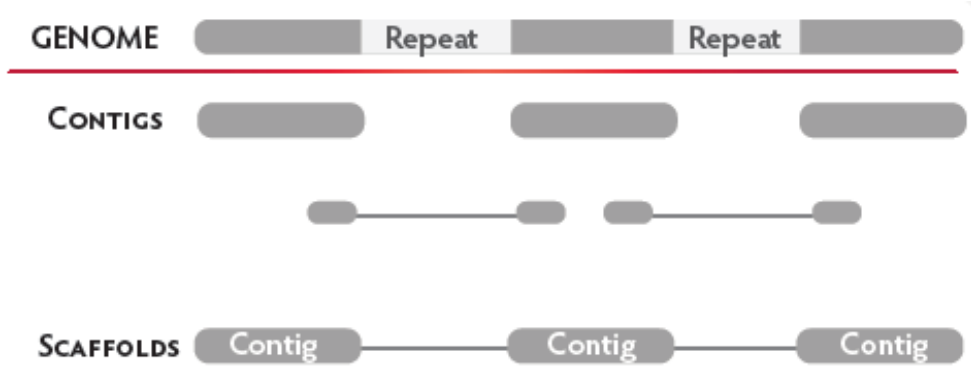



### Mate Pair Sequencing Concept



# Other Applications of Mate Pair Library Sequencing

## De Novo Genome Sequencing




## Mapping Insertion Sites

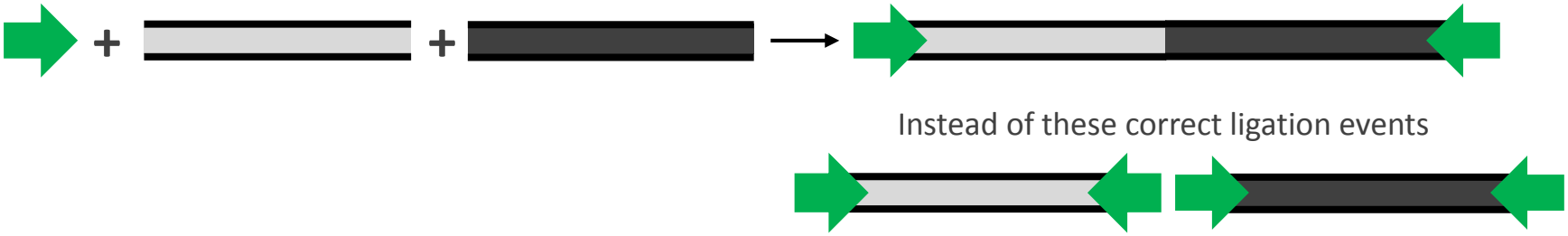
- Viruses
- Transgenes


## Analyzing Structural Variation

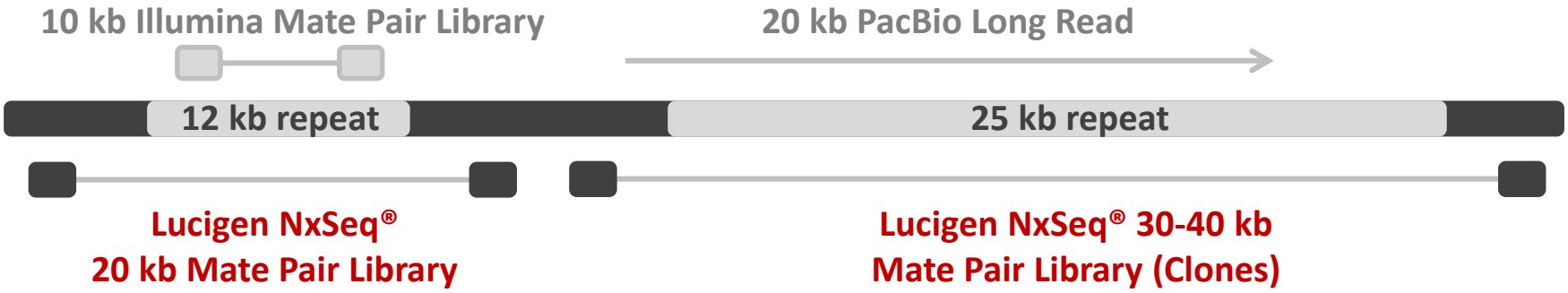
# Challenges of Mate Pair Sequencing

## Chimeras that Produce False Mate Pairs

- **False mate pairs and inefficient library construction**

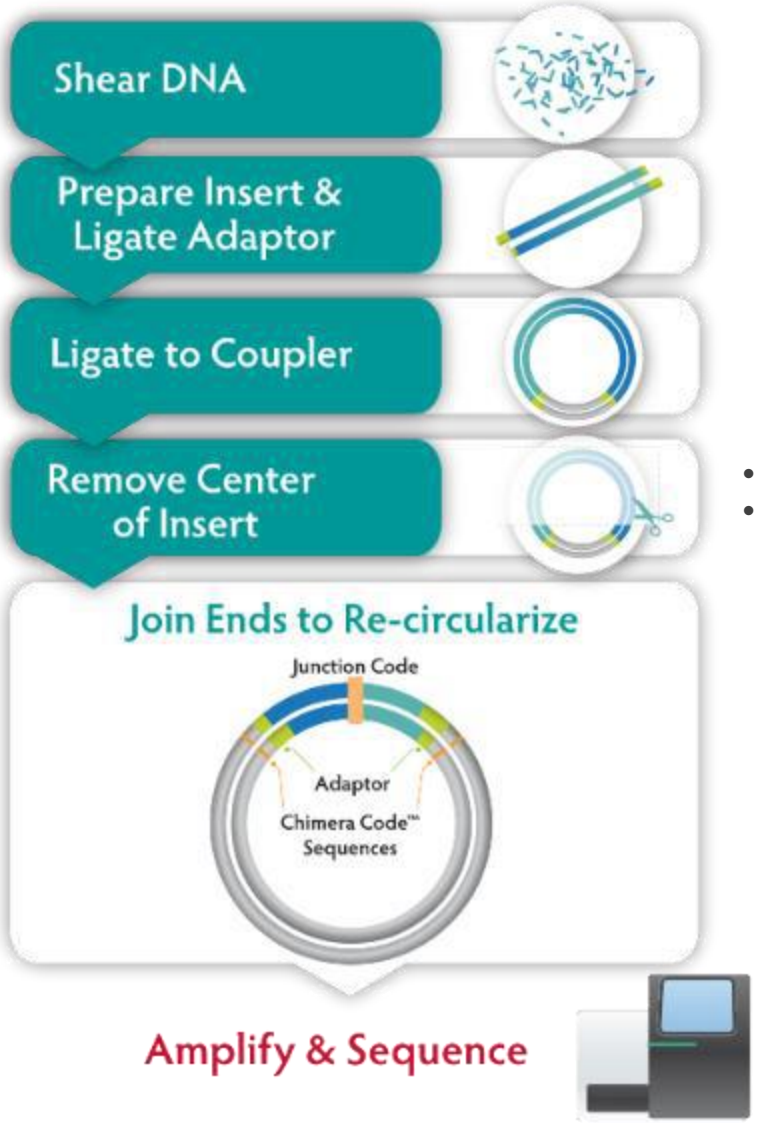



- **Inability to make large mate pair libraries to span large repeats**



# NxSeq® Long Mate Pair Library Kits

## Optimized Protocols to Produce Correct Mate Pairs

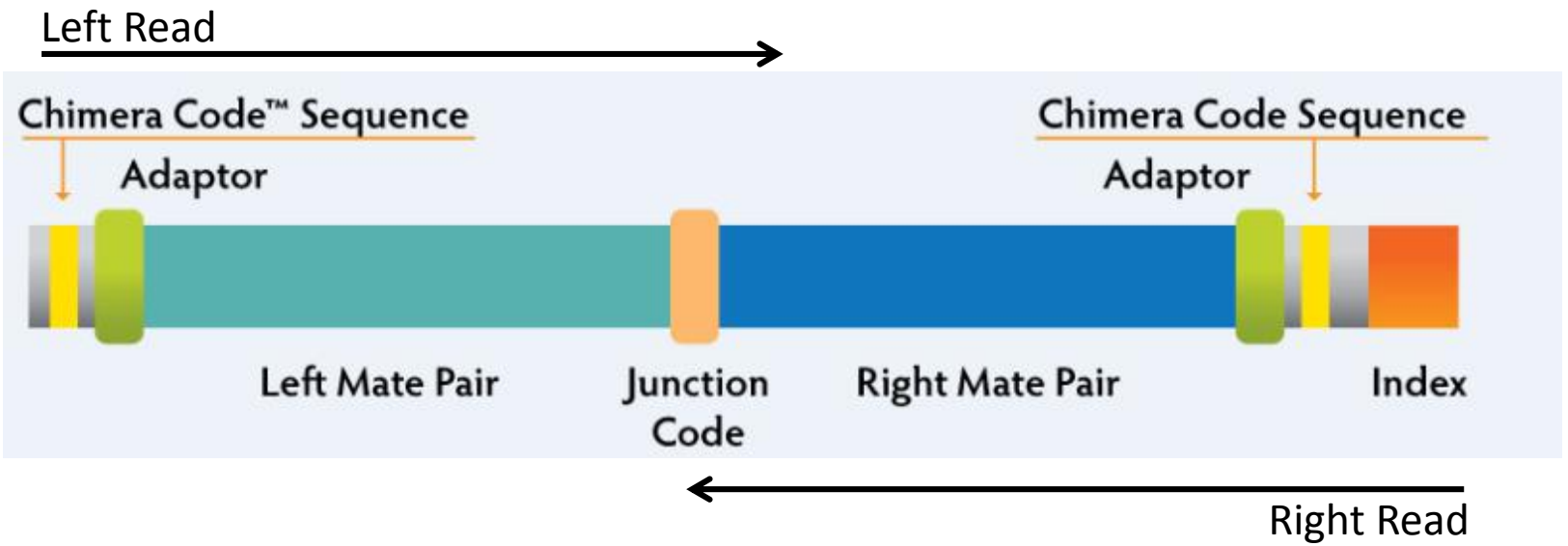



- Enzyme based
- No cutting in coupler

Lucigen®

# NxSeq® Long Mate Pair Library Kits

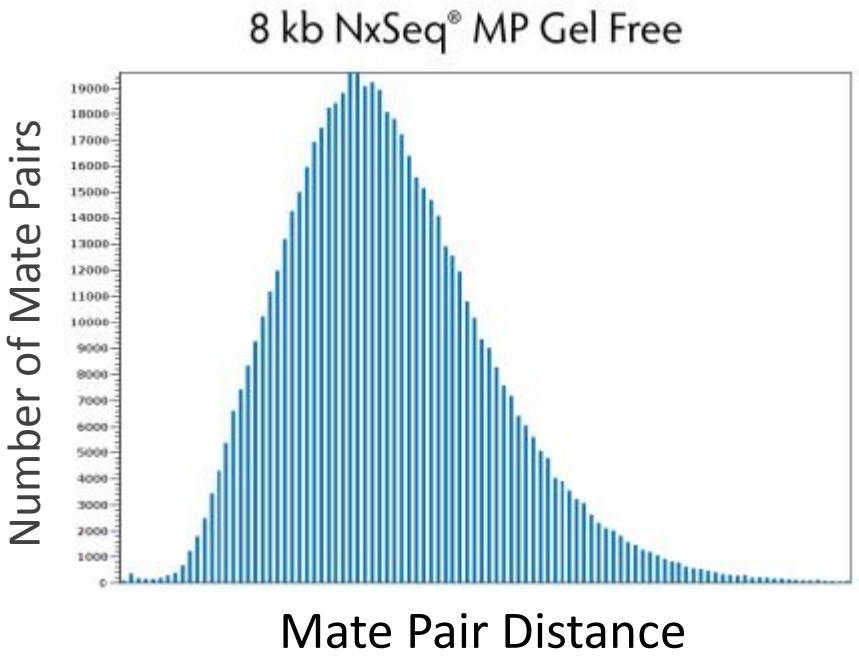
## ≤20 kb Mate Pair Libraries

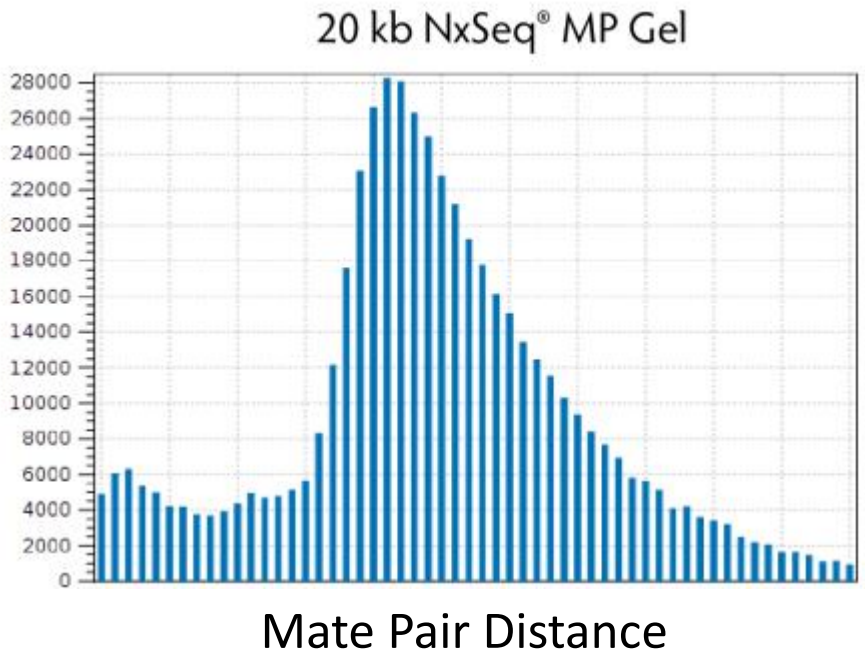



- Mate pair efficiency >95%
- Encrypted Chimera Code™ detection eliminates false mates
- Junction Code™ sequence identifies left & right mate pairs
- User-defined mate pair libraries up to 20 kb
- Multiplexing, cross contamination control via index/barcodes

# Choose the Size of Your Mate Pair Library

## 2 kb to 20 kb Options Available

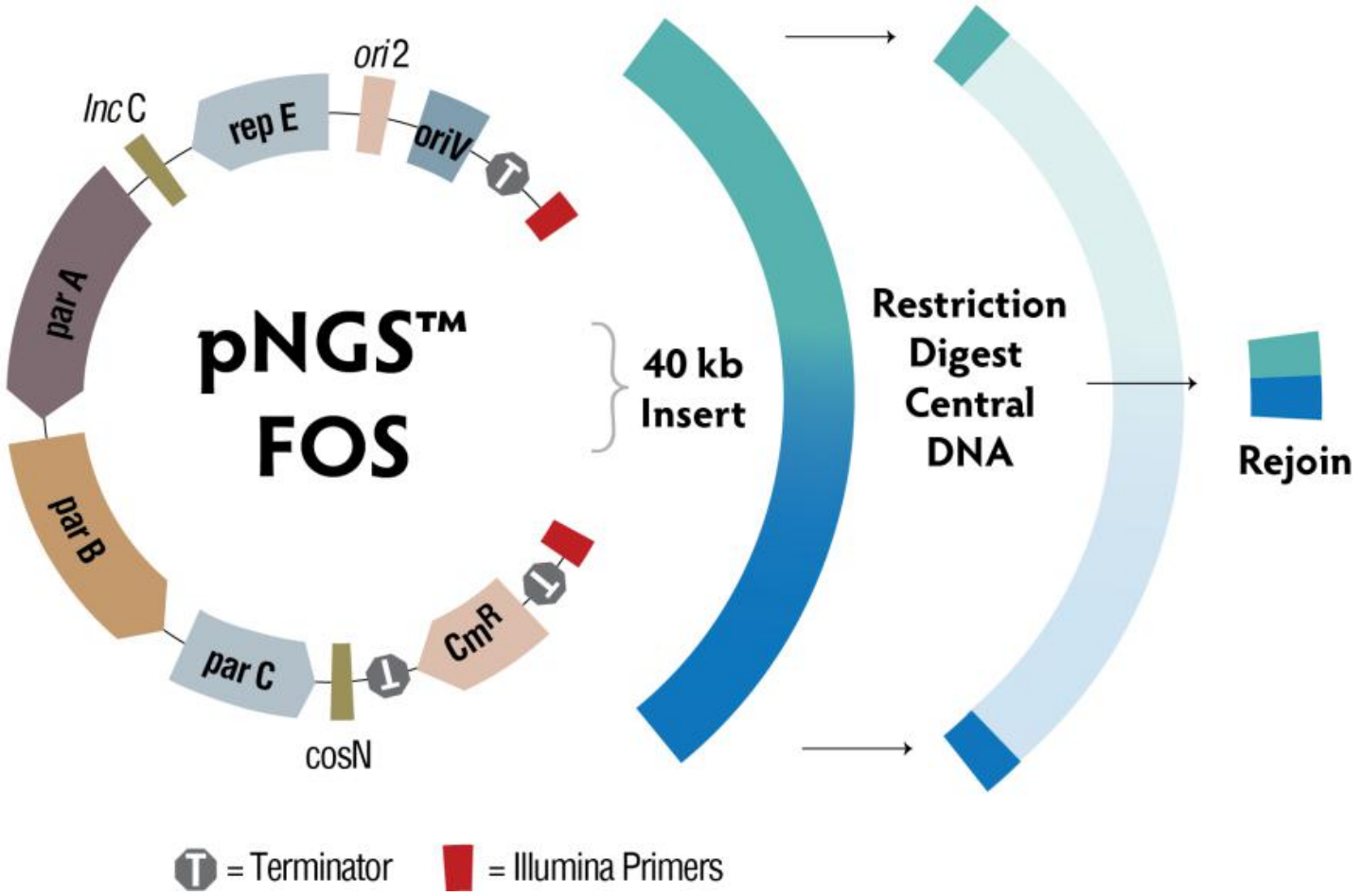



Lucigen®

# NxSeq® 40 kb Mate Pair Cloning Kit

## Optimized to Produce Correct Mate Pairs




US Patent 8329400

Lucigen®

# Comparison to Nextera Mate Pair Technology

## More True Mates with NxSeq® Mate Pair Technology

| Characteristic | Lucigen NxSeq® Mate Pair Technology | Illumina Nextera Mate Pair Technology |
|---|---|---|
| **Library Size Supported** | User-defined, up to 40 kb | 2-10 kb |
| **Mate Pair Efficiency** | >95% true mates | ~5-10% true mates |
| **Chimeras** | Largely prevented, most detected | Numerous, no detection |
| **Microbial Genomes Finished** | 7 out of 7 | 0 out of 9* |

*Peer J. 2015 Jun 2;3:e996. doi: 10.7717/peerj.996. eCollection 2015.
NxRepair: error correction in *de novo* sequence assembly using Nextera mate pairs.

Lucigen®

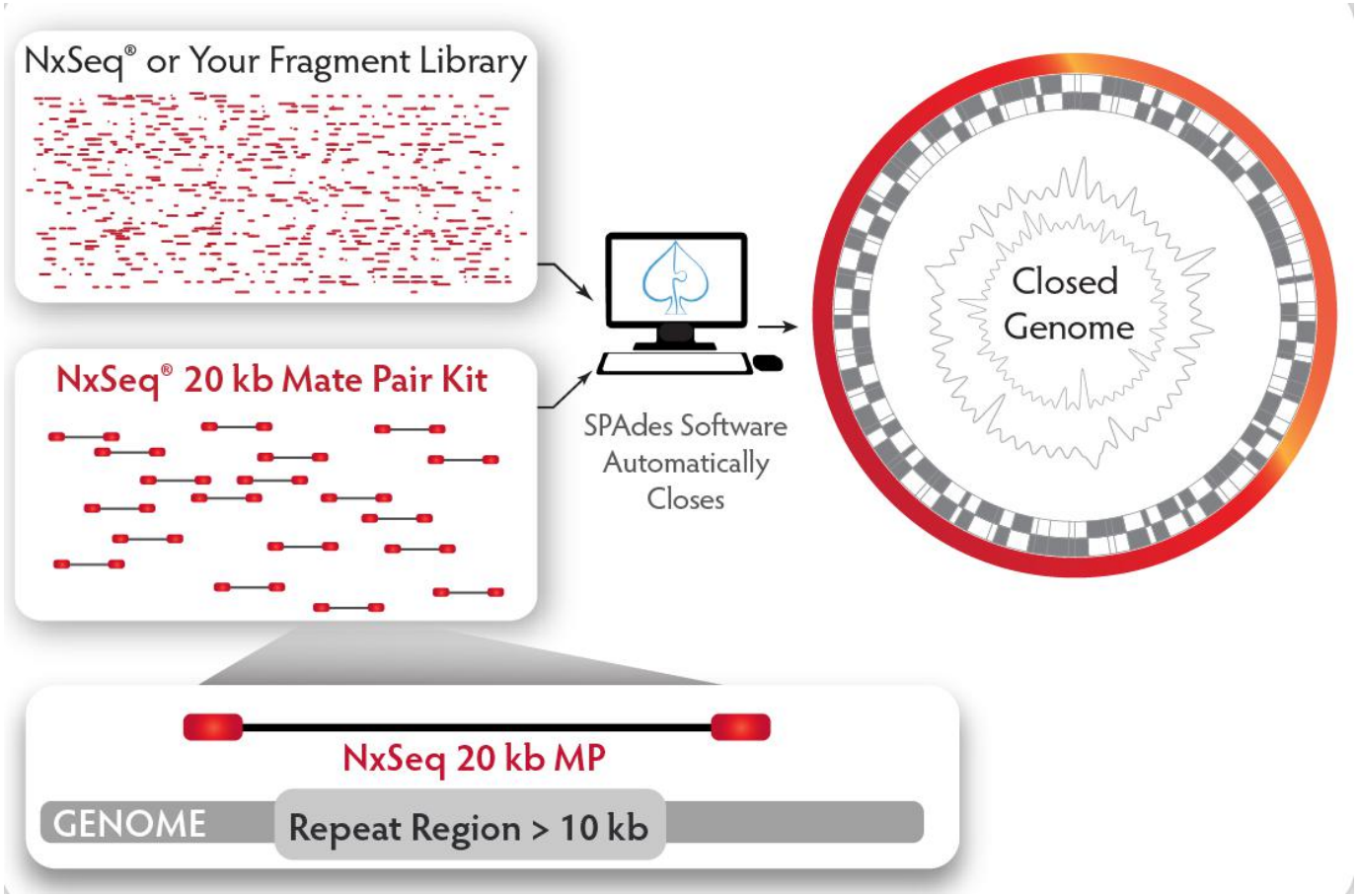
# Seven Difficult-to-Close Microbial Genomes

| Microbial Genome | Size Mb | %GC | #Repeats / Types | Max Repeat Size / Highest Copy # | Fragment Library Contigs |
|---|---|---|---|---|---|
| *Escherichia coli* | 4.6 | 50.8 | 75 / 21 | 5.3 kb / 11 | 78 |
| *Thermus aquaticus* | 2.3 | 68.1 | 44 / 14 | 3.5 kb / 8 | 65 |
| *Staphylococcus aureus** | 2.8 | 32.8 | 99 / 27 | 16.2 kb / 9 | 15 |
| *Streptomyces spp. A115* | 8.7 | 71.0 | 26 / 10 | 37.9 kb / 8 | 60 |
| *Nonomurea spp. F4* | 10.3 | 70.7 | 96 / 34 | 5.7 kb / 8 | 123 |
| *Bacillus amyloliquefaciens AP183* | 4.0 | 46.5 | 24 / 7 | 2.7 kb / 6 | 24 |
| *Aeromonas hydrophila S14-451* | 5.1 | 60.7 | 32 / 10 | 5.8 kb / 8 | 13 |

*Staphylococcus aureus**: failed to close with PacBio

Lucigen®

# Fragment Library + Long Mate Pair Library and SPAdes 3.5, Make Closing Microbial Genomes Easier




http://bioinf.spbau.ru/spades

J Comput Biol. 2012 May; 19(5): 455–477. doi: 10.1089/cmb.2012.0021

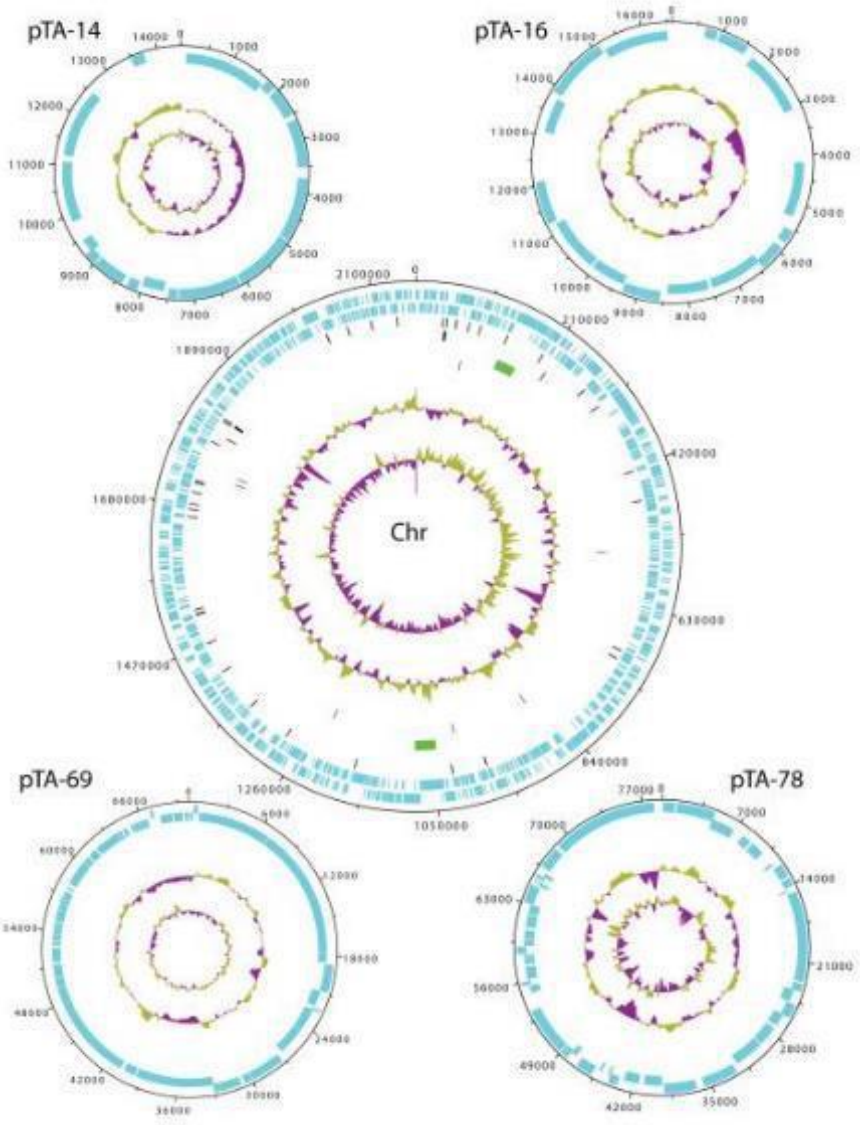
Lucigen®

# Fragment Library + Long Mate Pair Library and SPAdes 3.5, Make Closing Microbial Genomes Easier

| Microbial Genome | Fragment Library Contigs | SPAdes Scaffolds | Manual Curation Scaffolds | Finishing Requirements |
|---|---|---|---|---|
| *Escherichia coli* | 78 | 1 | 1 | Manual curation |
| *Thermus aquaticus* | 65 | 4 Chr + 5 Plasmids | 1 Chr + 4 Plasmids | 18 PCRs + Sanger |
| *Staphylococcus aureus* | 15 | 1 | 1 | Manual curation |
| *Streptomyces spp. A115* | 60 | 2 | 1 | 4 PCRs + Sanger |
| *Nonomurea spp. F4* | 123 | 1 | 1 | 1 PCR + Sanger |
| *Bacillus amyloliquefaciens AP183* | 24 | 2 | 1 | Manual curation |
| *Aeromonas hydrophila S14-451* | 13 | 2 | 1 | Manual curation |

Lucigen®

# Finished and Closed *Thermus aquaticus* Genome with the NxSeq® Long Mate Library




- 1 big chromosome
- 4 megaplasmids

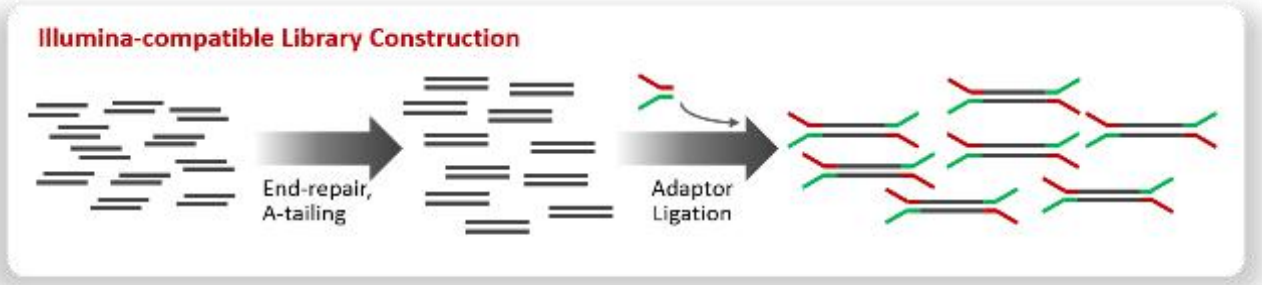
Lucigen®

# Summary: NxSeq® AmpFREE Library Kit

## Better Library Efficiency = More Data

**NxSeq® AmpFREE Low DNA Library Kit**

- Highest efficiency PCR-free kit enables use of very small amounts of input DNA (75 ng) without the need for amplification and produces more high quality sequencing data from each library
- Rapid protocol saves valuable time
- Lower price decreases costs thus extending budgets



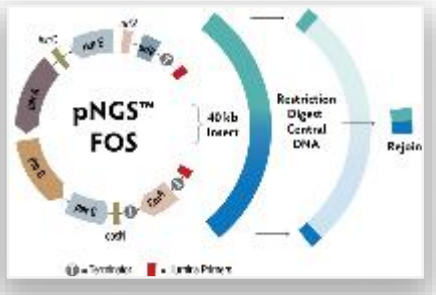

Lucigen®

# Summary: NxSeq® Mate Pair Kits

## More True Mates = Easier Data Analysis

**NxSeq® Long Mate Pair Library Kit and the 40 kb Mate Pair Cloning Kit**

- Generate long read information on short-read Illumina sequencers
- >95% true mate pair efficiency improves assembly efficiency and accuracy
- Mate pair libraries are for more than just genome assembly and closure
  - Transgene/viral insertion and structural variation mapping


Lucigen®

# Questions? www.lucigen.com

**Lucigen Tech Support**
techserv@lucigen.com
1 (608) 831-9011
8 am – 5 pm central time

**Contact me.**
**Rob Brazas, Ph.D.**
NGS Product Manager
rbrazas@lucigen.com

Thank You!

Lucigen®