Dear Customer,

Thank you for your interest in NRGene™ DeNovoMAGIC™ and DeNovoMAX services.

We look forward to establishing our collaboration and provide you with a fast, accurate, and cost-effective genome assembly. Our genome assemblies are well known in the scientific community as well as in the commercial market for their high standards and quality.

Optimal genome assembly rely on high quality gDNA extraction and a well-designed library preparation and sequencing. In order to make sure you will get the most satisfying results, kindly adhere to the following standard instructions and specifications.

Please feel free to address us with any inquiry with the bellow email address,

Thank you for the collaboration,

NRGene Genomic Analysis Services team.

GAS\_support@nrgene.com

1. **Sequencing data specifications**

It is recommended to use NRGene’s certified sequence providers to prepare the sequencing data. NRGene has invested many efforts to provide you with the best available services with most affordable prices.

Nonetheless, whether you choose to prepare the sequencing data in-house or with the aid of a 3rd party contractor, the following criteria and instructions shall be met in order to allow NRGene to provide the best assembly results

1. **Coverage Calculation**

Coverage calculation depend on the zygosity of the organism being sequenced:

1. **Homozygous genome** – coverage should be calculated considering the haploid size of the genome[[1]](#footnote-1).
2. **Heterozygous genome** - coverage should be calculated considering the **double** haploid size of the genome.
3. **Sequencing Data Specifications**

The following table shows the libraries that are required to be constructed and the sequencing machines that can be used. Minimal coverage required from the library is an estimate that should be calculated considering the genome size[[2]](#footnote-2).

**DeNovoMAGIC**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| # | Library type | Insert size | Sequencing Instrument[[3]](#footnote-3) | Read length | Minimal Coverage |
| 1 | PCR-free PE Shotgun library(fragments size selection) | 450bp-460bp | HiSeq 2500, NovaSeq6000 | PE250bp | X60 |
| 2 | PCR-free PE Shotgun library(TruSeq™ v.2) | 700bp | HiSeq 4000, HiSeqX (700bp)[[4]](#footnote-4), NovaSeq6000 | PE150-160bp | X30 |
| 3 | MP (Nextera MP Gel Plus) | 2-4kbp | HiSeq 3000/4000, HiSeqX, NovaSeq6000 | PE150-160bp | X30 |
| 4 | MP (Nextera MP Gel Plus) | 5-7kbp | HiSeq 3000/4000, HiSeqX, NovaSeq6000 | PE150-160bp | X30 |
| 5 | MP (Nextera MP Gel Plus) | 8-10kbp | HiSeq 3000/4000, HiSeqX, NovaSeq6000 | PE150-160bp | X30 |
| 6 | Chromium (10X Genomics) | N/A | HiSeq 4000, HiSeqX, NovaSeq6000 | PE150bp | >X30 |
| 7 | PacBio SMRTbell | ~10Kbp | PacBio Sequell  | N/A | X30 |

**DeNovoMAX**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| # | Library type | Insert size | Sequencing Instrument | Read length | Minimal Coverage |
| 1 | PCR-free PE Shotgun library(fragments size selection) | 450bp-460bp | NovaSeq6000 | PE250bp | X45 |
| 3 | MP (Nextera MP Gel Plus) | 2-4kbp | NovaSeq6000 | PE150-160bp | X25 |
| 6 | Chromium (10X Genomics) | N/A | NovaSeq6000 | PE150bp | X25 |
| 7 | PacBio SMRTbell | ~10Kbp | PacBio Sequell  | N/A | X30 |

**Notes:**

1. In case of large genomes, it is recommended to prepare multiple libraries sets from each fragment size (1 illumina libraries set per each 5Gbp of the genome and 1 PacBio SMRTbell library per each 3Gbp), this should diversify the sequenced fragments and decrease the duplication level.
2. For the same reason, it is advisable to use the largest recommended amount of gDNA in the protocol of the MP libraries preparation.

**All output fastq files and relevant output analysis files should be referred to NRGene for QC and approval.**

1. **Quality Measurements**

The following table shows the minimal quality thresholds required from the sequencing data produced[[5]](#footnote-5).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| # | Library type | Quality(any applicable software) | Overlapping[[6]](#footnote-6) | Nextera junction adapter percentage[[7]](#footnote-7) |
| 1 | PCR-free PE library (450bp-460bp) | Q30>80% | 1. 85-90% of read pairs are merged.
2. Distribution Up to 2σ (std)
3. Average size- 450-460bp
 | - |
| 2 | PCR-free PE library (700bp) | Q30>80% | - | - |
| 3 | MP (Nextera MP Gel Plus)[[8]](#footnote-8) | Q30>80% | - | >20% and <75% |
| 4 | Chromium (10X) | Q30>80% | - | - |

***Please note:*** *Should you encounter any difficulties, please contact NRGene’s Professional Services Team; the team will do its utmost to assist you with its own expertise.* (e-mail: gas\_support@nrgene.com).

1. Haploid genome size= flow-cytometry 1c and/or amount of DNA in a gamete cell; Genome size estimates should be rounded up to avoid missing coverage in the sequencing calculations. [↑](#footnote-ref-1)
2. For example: a heterozygous genome of 3Gbp haploid genome size requires 360Gbp (2\*3\*60) of data from the PCR-free PE library with 450bp insert size, sequenced on HiSeq 2500/NovaSeq6000, read length 250bp. Coverage may be extended where data quality does not suffice. [↑](#footnote-ref-2)
3. Since May 2019- NRGene will prioritize sequencing on NovaSeq6000 and gradually phase-out other Machines. [↑](#footnote-ref-3)
4. Multiplexing of 700bp library with MP libraries on HiSeq4000 or HiSeqX is not advisable- the 700bp should run on a dedicated lane. [↑](#footnote-ref-4)
5. Other quality criteria may apply, NRGene will conduct additional QC procedures to ensure data compatibility. [↑](#footnote-ref-5)
6. For example - calculated using the following software run:

pear-0.9.5-bin-64/pear -f <R1.fastq> -r R2.fastq -o <output.fastq> -m 500 -n 250 -t 150 -v 10 -q 10 &

The resulting output fastq should be run through FASTQC software for evaluation. Merged reads should be 85% - 90% of the reads pairs in the input files R1/R2.fastq [↑](#footnote-ref-6)
7. Search of exact match of first 19bp out of 38bp sequence of the adapter, reads are assumed to be PE150-160bp. [↑](#footnote-ref-7)
8. MP libraries should be prepared from HMW extracted DNA. The required protocol to be used is Nextera Mate Pair Gel-Plus. It is recommended to have 2 cycles of tagmentation to increase efficiency of library preparation. The quality of the outcome is crucial in order to achieve long and accurate scaffolds. [↑](#footnote-ref-8)