DNBSEQ™ SERVICE OVERVIEW
Plant and Animal Whole Genome Re-Sequencing

Service Description

Plant and animal whole genome re-sequencing (WGRS) involves sequencing the entire genome of a plant or animal and comparing the sequence to that of a known reference genome. Re-sequencing of the plant and animal genome will identify genetic variations such as SNPs and Indels and discover other genetic changes of the sequenced species. It has been used for the identification of functional genes and markers of important traits to facilitate molecular breeding and to improve agricultural production and conservation.

Highlights of DNBSEQ Technology

· Even coverage of reads
· Much less duplication
· True PCR-Free
· Index hopping free

Sequencing Service Specification

Our Plant and Animal Whole Genome Re-Sequencing services are executed with the DNBSEQ sequencing technology.

Sample Preparation and Services

· Library preparation
· 100 bp and 150 bp paired-end sequencing available
· Raw data, standard and customized data analysis
· Available data storage and bioinformatics applications

Seuquencing Quality Standard

· Guaranteed ≥90% of clean bases with quality score of Q20
· Standard sequencing coverage of 10-30X is recommended for the study of individuals and 5-10X for population studies

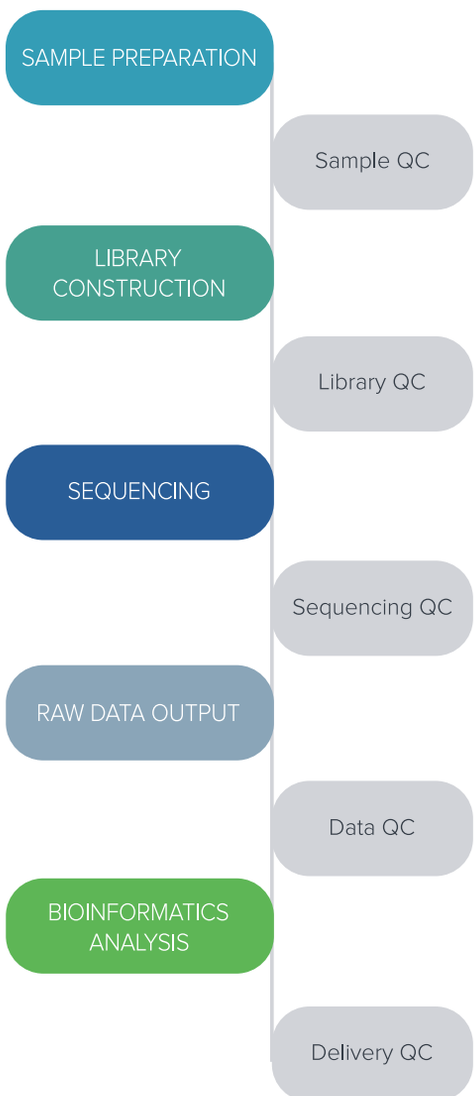
Turnaround Time

· Typical 15-30 working days from sample QC acceptance to filtered raw data availability
· Expedited services are available.Contact our specialist for details

Quality Data | Fast TAT | Cost Effective

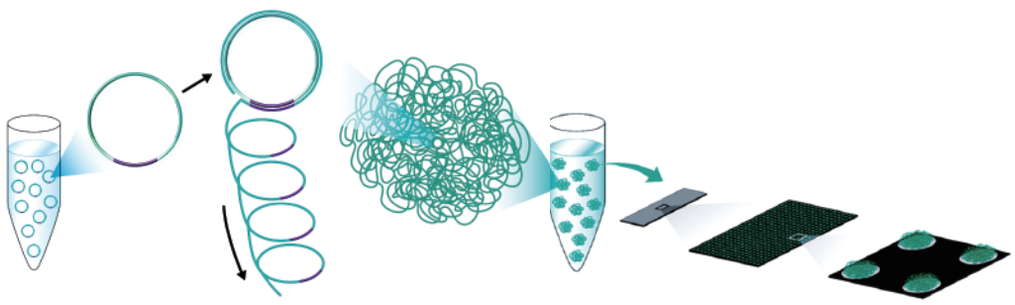
Project Workflow

We care for your samples from the start through to the result reporting. Highly experienced laboratory professionals follow strict quality procedures to ensure the integrity of your results.



DNBSEQ Sequencing Technology

DNBSEQ is an innovative high-throughput sequencing solution. The system is powered by combinatorial Probe-Anchor Synthesis (cPAS), linear isothermal Rolling-Circle Replication and DNA Nanoballs (DNB™) technology, followed by high-resolution digital imaging.



The combination of linear amplification and DNB technology reduces the error rate while enhancing the signal. The size of the DNB is controlled in such a way that only one DNB is bound per active site on the DNBSEQ flow cell. This densely patterned array technology provides optimal sequencing accuracy and increases flow cell utilization.

Data Analysis

Besides clean data output, we offer a range of standard and customized bioinformatics pipelines for your whole genome re-sequencing project.

Reports and output data files are delivered in industry standard file formats: FASTQ, BAM, VCF, .xls, .png

Standard Analysis

- Data Filtering
- Alignment
- SNP/InDel/SV/CNV calling, annotation and circus figure

Advanced Analysis

- Population-genetic analysis
- GWAS analysis

Customized Analysis

Further customization of bioinformatics analysis to suit your unique project is available.

Please contact technical representative for details.

Sample Requirements

We can process your gDNA samples from a variety of species, with the following general requirements:

DNA Sample	Library type	Mass	Concentration	Integrity (AGE)	Sample Purity
Regular Samples	PCR	≥200 ng (Recommend ≥400 ng)	≥8 ng/μL	The band shown on gel electrophoresis has little degradation, or of fragment size greater than 20 kb.	No contamination with RNA, protein or salt ions; colorless and transparent; non-sticky
	PCR-free	≥1 μg (Recommend ≥2 μg)	≥12.5 ng/μL		

Data Performance

1 soybean sample and 1 mouse sample were used to validate the DNBSEQ sequencing technology. DNBSEQ sequencing datasets from 2 technical replicates of each species were generated and compared to datasets from the N platform. Libraries were prepared according to the manufacturer's protocols.

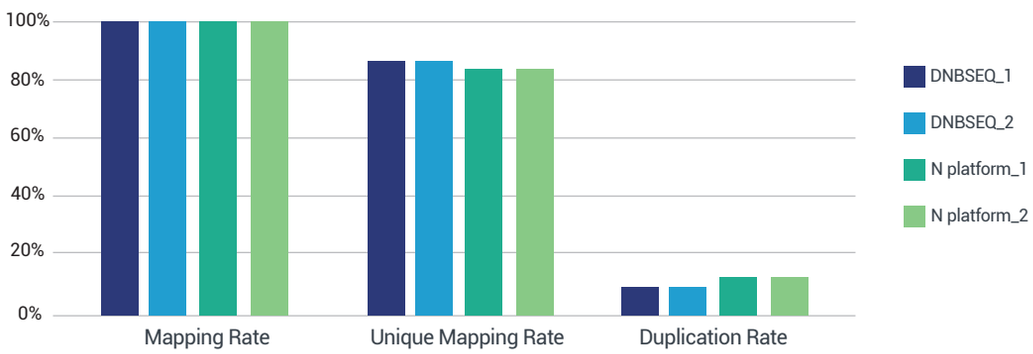
The DNBSEQ sequencing technology generated high quality PE100 data that is comparable to that from the N platform (Table 1). At similar sequencing output between different platforms of each species, over 95% of bases score above Q20 while the genome mapping rate and the unique mapping rate of DNBSEQ data and N platform data are at the comparative level. Notably, the clean data rates of all DNBSEQ replicates are higher than those of N platform replicates, generating more filtered bases for the following analysis. Moreover, the duplication rates of DNBSEQ reads of both species are constantly lower than those of N platform reads. Since the duplicated reads are usually skipped for downstream variant analysis, the DNBSEQ will generate more valid data for variant analysis than the N platform at the same data output.

▼ Table 1. Data quality for soybean and mouse samples (averages from 2 technical replicates).

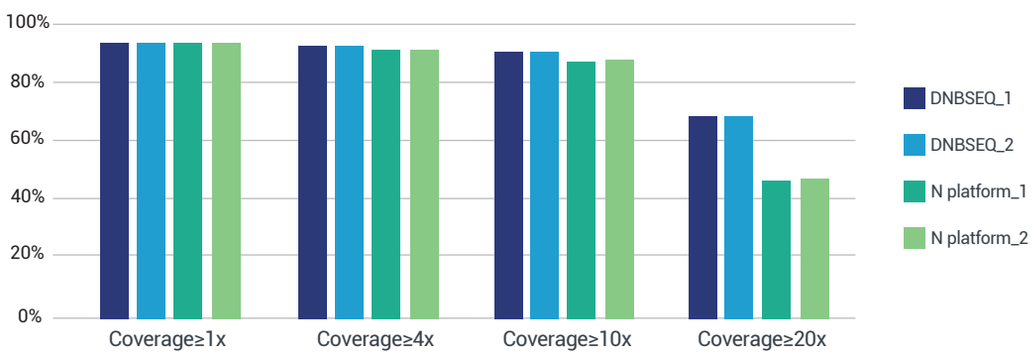
Species	Soybean		Mouse	
Sequencer	DNBSEQ	N platform	DNBSEQ	N platform
Raw data amount (Gb)	33.12	30.56	70.09	70.33
Clean data amount (Gb)	31.20	26.69	65.50	57.40
Clean rate (%)	94.19	87.34	93.45	81.62
Clean read Q20 (%)	96.59	97.81	96.16	97.66
Mapping rate (%)	99.53	99.77	99.76	99.92
Unique mapping rate (%)	82.90	82.61	86.57	83.97
Duplication rate (%)	3.64	6.42	7.85	11.21
Mismatch rate (%)	0.95	0.68	0.68	0.39
Average sequencing depth	29	26	22	18
Coverage (%)	95.48	94.59	93.51	93.26
Coverage at least 4X (%)	94.15	92.78	92.93	92.36
Coverage at least 10X (%)	92.05	89.32	91.00	87.72
Coverage at least 20X (%)	81.20	67.11	66.05	44.72

Mouse sample data performance

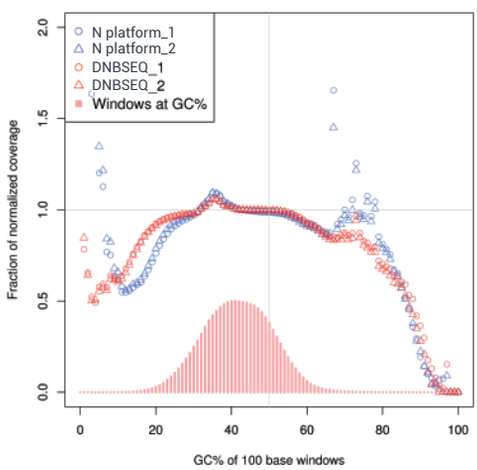
1 mouse liver DNA sample was sequenced by both DNBSEQ and N platform. 2 technical replicates for DNBSEQ and 2 for N platform are included in the experiment. Both DNBSEQ and N platform data show high mapping rates, while less bias is found in the read distribution of DNBSEQ datasets at low GC content (Fig. 1, 2 and 3). Moreover, DNBSEQ duplication rates are significantly lower (Fig. 1), due to less PCR bias from DNBSEQ's rolling circle replication, resulting in more uniform distribution of reads for variant analysis.



▲ Figure 1. Mapping rate and duplication rate of 2 DNBSEQ datasets and 2 N platform datasets with comparable data output of 70 Gb.

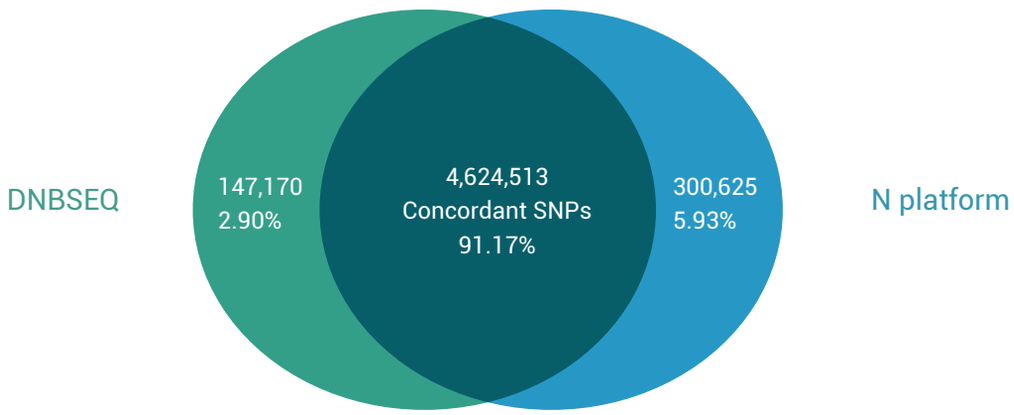


▲ Figure 2. Sequencing coverage of 2 DNBSEQ datasets and 2 N platform datasets with comparable data output of 70 Gb.



▲ Figure 3. Normalized read coverage by GC content (duplication reads excluded). It shows a less biased distributionof DNBSEQ reads at low GC content when compared to N platform read distribution.

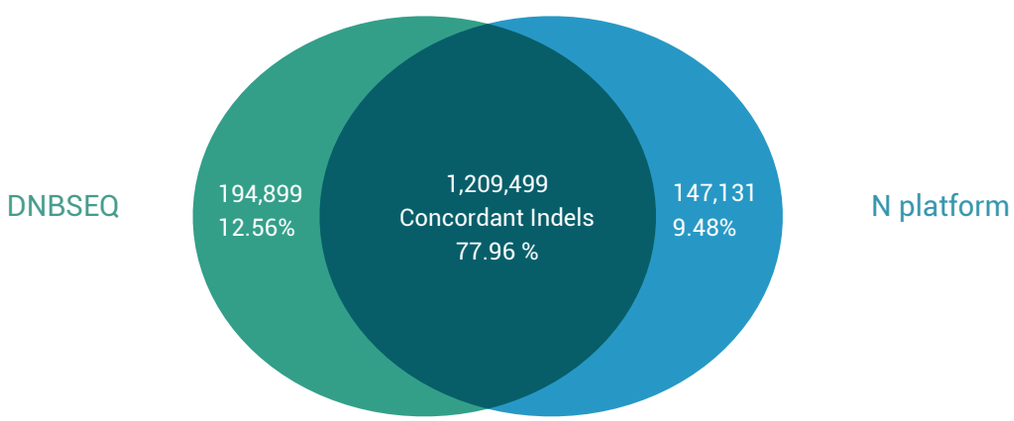
The DNBSEQ demonstrated equivalent variant calling reproducibility to N platform, with 93.63% and 76.64% for SNP and Indel calling, respectively. The concordant rates of SNP and Indel calling between platforms were 91.17% and 77.96%, respectively (Fig. 4 and Fig. 5).



DNBSEQ	
DNBSEQ_1	4,639,014
DNBSEQ_2	4,639,471
DNBSEQ total SNPs	4,791,741
Reproducibility	93.63%

N platform	
N platform_1	4,772,466
N platform_2	4,785,071
N platform total SNPs	4,946,249
Reproducibility	93.23%

▲ Figure 4. The SNP reproducibility and cross-platform consistency statistics from 2 DNBSEQ replicates and 2 N platform replicates.



DNBSEQ	
DNBSEQ_1	1,299,372
DNBSEQ_2	1,307,108
DNBSEQ total Indels	1,477,897
Reproducibility	76.64%

N platform	
N platform_1	1,250,891
N platform_2	1,269,891
N platform total Indels	1,414,150
Reproducibility	78.25%

▲ Figure 5. The Indel reproducibility and cross-platform consistency statistics from 2 DNBSEQ replicates and 2 N platform replicates.

To Learn More

If you have any questions or would like to discuss how our services can help you with your research, please don't hesitate to contact us at P_contact@innomics.com. We look forward to hearing from you!

Innomics Inc.
750 N. Pastoria Ave
Sunnyvale, CA 94085
USA

www.innomics.com