DNBSEQ™ SERVICE OVERVIEW

Plant and Animal *De novo* Sequencing

Service Description

De novo sequencing refers to the sequencing of a novel genome without a reference sequence for alignment. The process of *de novo* genome sequencing involves the sequencing of small/large DNA fragments, assembling the reads into longer sequences (contigs) and finally ordering the contigs to obtain the entire genome sequence. We have an extensive experience in the *de novo* whole genome sequencing and assembly of more than 400 species genomes.

We offer a complete suite of technologies to support your *de novo* sequencing projects, along with expert assistance in the planning of optimal sequencing and bioinformatics options, to ensure your project is a success.

Sequencing Specification

Our plant and animal *de novo* services are executed with multiple squencing systems

Sample preparation and services

- Library preparations (DNBSEQ™/Illumina, Nanopore PromethION, PacBio Revio/Sequel II etc)
- Various sequencing modes
- Raw data, standard and customized data analysis
- Available data storage and bioinformatics applications

Sequencing quality standard

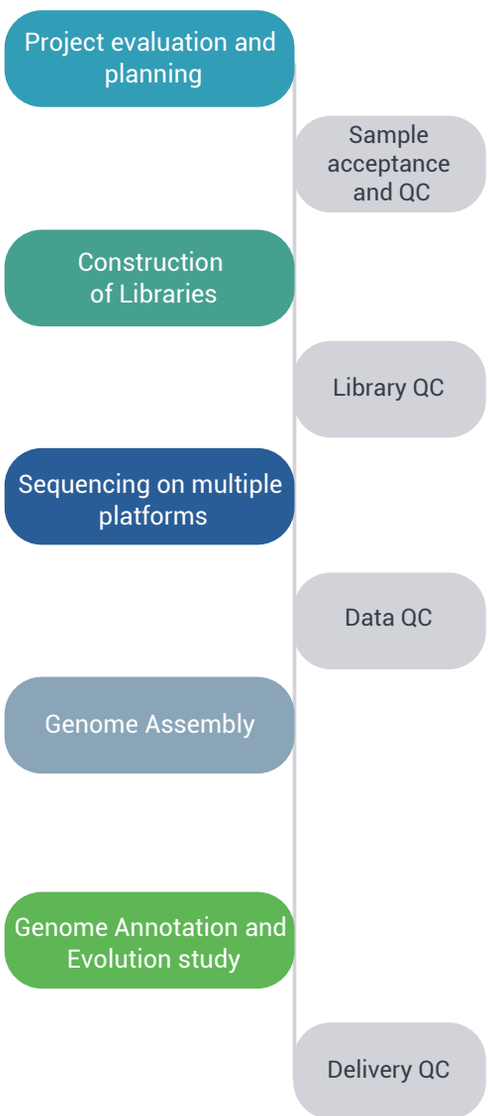
- Guaranteed ≥90% of DNBSEQ™ clean bases with quality score of Q20
- Guaranteed ≥100Gb Nanopore pass data with Q>7
- Guaranteed ≥80Gb PacBio Revio HiFi data except for some complex species

Turnaround Time

- 30 working days from sample QC acceptance to filtered data availability
- Case by case for the genome analysis

Project Workflow

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



Quality Data | Fast TAT | Cost Effective

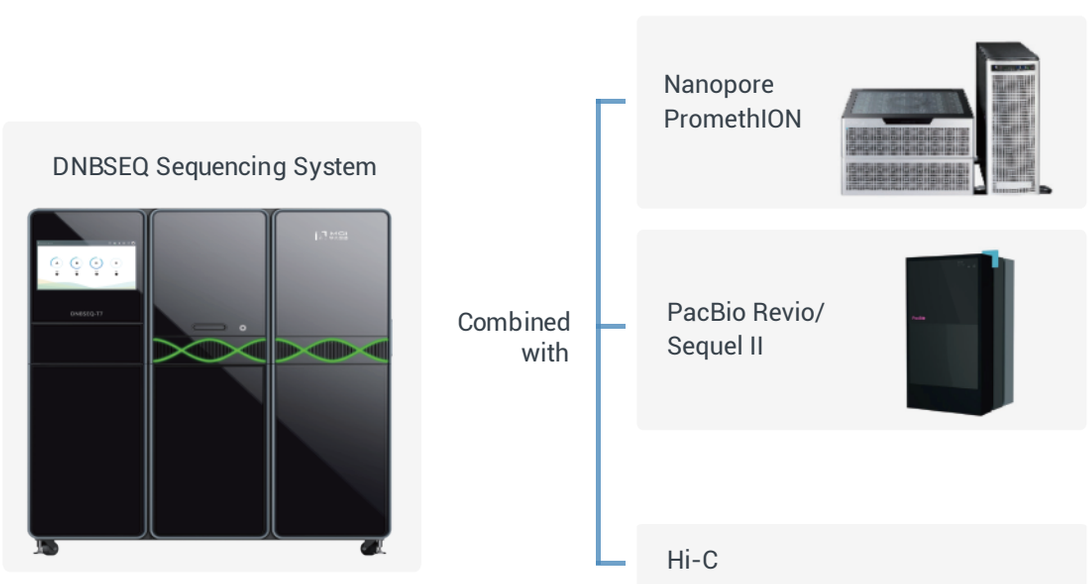
Sequencing Strategy

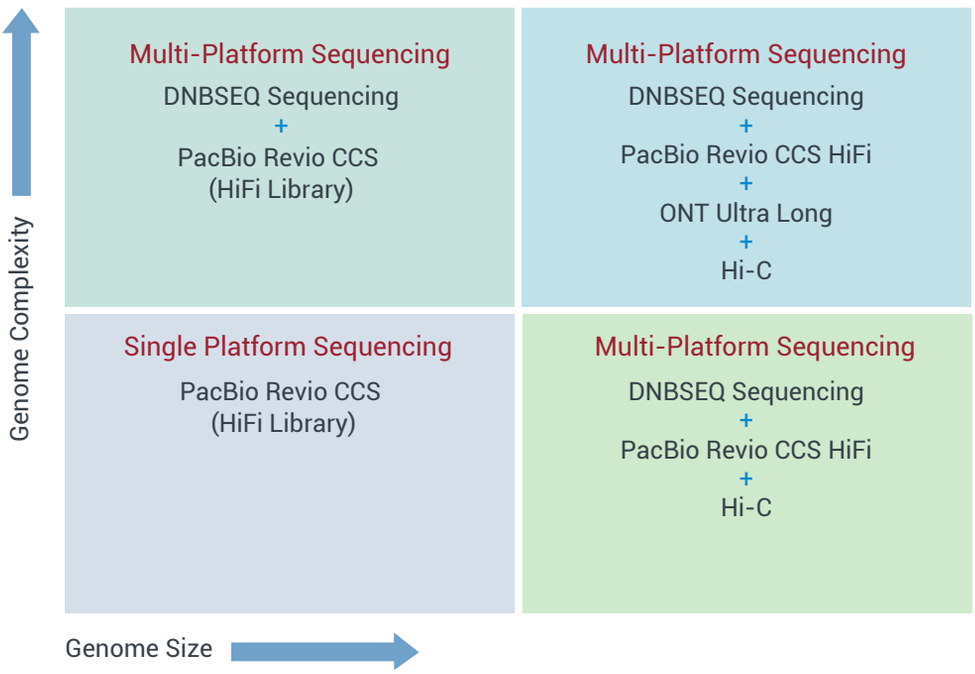
De novo sequencing usually requires a customized approach based on your subject species' genome size and complexity as well as the overall scientific objectives of the project.

Our plant and animal *de novo* sequencing services are usually performed using a combination of available platforms, including proprietary DNBSEQ™ NGS platforms augmented with Nanopore PromethION, PacBio Revio/Sequel II for sequencing, library preparation and mapping. In addition, we offer extensive bioinformatics data analysis options for genome assembly, annotation and evolution studies.

Platform Tools	Library Type	Sequencing /USE	Recommend Sequencing Depth
DNBSEQ™	DNBSEQ library	PE100/PE150	≥100X
	Hi-C Library	PE150	≥100X
Nanopore PromethION	20K-50Kb Library	Read length ≥10Kb	≥100X
	Ultra long library (>50Kb)	Read length ≥20Kb	≥100X
PacBio Revio/Sequel II	15K-20Kb CCS (HIFI) Library	Read length ≥10Kb	≥50X

Our Sequencing specialists will work with you to design the optimal strategies for your project, using platform combinations as appropriate for your project.





Data Analysis

Besides clean data output, we offer a range of standard and customized bioinformatics pipelines for your plant and animal *de novo* sequencing project.
Reports and output data files are delivered in industry standard file formats: BAM, .xls, .png. Raw FASTQ and FASTA data are available.

Genome Survey	1. Kmer estimation (Jellyfish + GenomeScope); 2. External pollution Analysis (BWA);
Genome Assembly (Pacbio HiFi data)	1. Assembly; 2. Assessment by short reads alignment; 3. BUSCO assessment;
Gene Annotation	1. Repeat annotation; 2. Gene prediction; 3. Gene function Annotation;
Evolution	Deliver published genome and allied species (less than 10 species) 1.Gene family identification (Animal TreeFam; Plant OrthoMCL;≤10 species); 2.Phylogenetic tree construction; 3.Estimation of divergence time; 4.Genome synteny analysis; 5.Whole genome duplication analysis; 6.Gene family expansion and contraction analysis;
Auxiliary Assembly	Hi-C data auxiliary assembly

Sample Requirements

We can process your DNA sample of plant and animal with the following general requirements (Actual sample requirements for each specific project will depend on the number and type of libraries to be constructed). We also provide special sample extraction services to satisfy custom project requirements.

Plant and Animal Genome *De novo* Sequencing (Genomic DNA)						
Platform	Sample type	Mass	Concentration	OD	Integrity (AGE)	Sample Purity
DNBSEQ	350bp library	≥1 μg	12.5 ng/ul	-	Main band≥20 kb.	No contamination with RNA, protein or salt ions; colorless and transparent; non-sticky
Nanopore PromethION	Normal library	≥9 μg	90 ng/ul	OD260/280: 1.8-2.2 OD260/230: 1.8-2.2	No degradation or little degradation with main band≥40 kb, with smear no smaller than 20Kb	
Nanopore PromethION	Ultralong library	≥16 μg	153 ng/ul	OD260/280: 1.6-2.5 OD260/230: 1.6-2.5	No degradation or little degradation with main band≥60kb, with smear no smaller than 25Kb	
PacBio	15-20 Kb HIFI library	≥15 μg	80 ng/ul	OD260/280: 1.6-2.2 OD260/230: 1.6-2.5	No degradation or little degradation with main band≥40kb.	

Examples of *de novo* projects executed by us recently

Species	Heterozygosity	Genome Size	Sequencing Platform	Sequencing reads N50	ContigN50
Plant	0.89%	850M	Nanopore	22KB	7.3MB
Plant	0.20%	2.4G	Nanopore	30KB	23.7MB
Plant	0.80%	400M	Nanopore	30KB	17.3MB
Plant	1%	1.1G	Nanopore	23KB	6.5MB
Plant	1.10%	10G	Nanopore	30KB	1.6MB
Plant	0.38%	550M	Nanopore	21KB	10MB
Animal	0.30%	3G	Pacbio CLR	22KB	27.8MB
Plant	0.30%	4G	Pacbio CLR	20KB	6.6MB
Plant	0.10%	1.1G	Pacbio CLR	20KB	17.0MB
Plant	0.40%	1.5G	Pacbio CLR	25KB	1.5MB
Autopolyploid plant	3%	3G	Pacbio HIFI	17KB	4.8MB
Plant	0.90%	650M	Pacbio HIFI	15KB	8.6MB
Plant	0.80%	2.4G	Pacbio HIFI	17KB	48.4MB
Animal	1.30%	1.4G	Pacbio HIFI	18KB	7MB

A case study of water lily genome

Title: The water lily genome and the early evolution of flowering plants

Species:

Nymphaea colorata

Strategy:

1. 124X (49.8 Gb) PacBio long reads data for gap filling and scaffolding.
2. 254X (346Mb PE150 Reads) for scaffolding

A high-quality water lily genome sequence was assembled by sequencing with a hybrid strategy. The genome was assembled into 1,429 contigs (with a contig N50 of 2.1 Mb) and total length of 409 Mb with 804 scaffolds, 770 of which were anchored onto 14 pseudo-chromosomes.
The phylogenomic analyses support Amborellales and Nymphaeales as successive sister lineages to all other extant angiosperms.

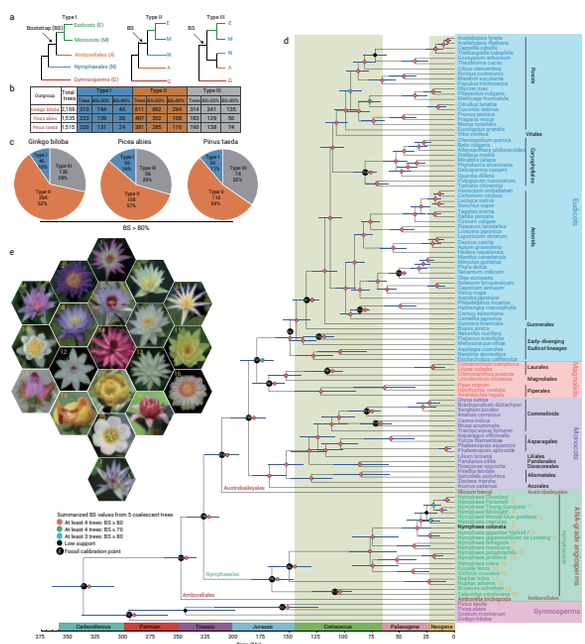


Fig 1. | Phylogenomic relationships of angiosperms.

The N. colorata genome and 19 other water lily transcriptomes reveal a Nymphaealean whole-genome duplication event, which is shared by Nymphaeaceae and possibly Cabombaceae.

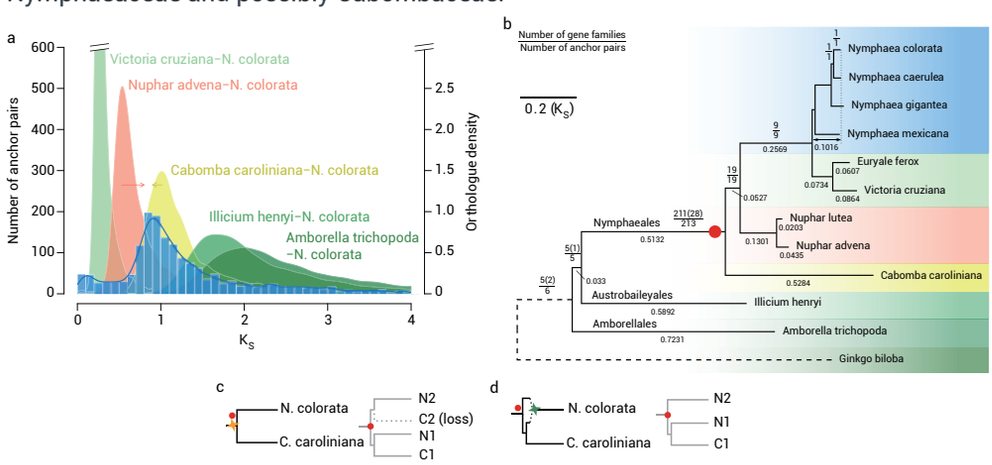


Fig. 2 | A Nymphaealean WGD shared by Nymphaeaceae and possibly Cabombaceae.

Among the genes retained from this whole-genome duplication are homologues of genes that regulate flowering transition and flower development.

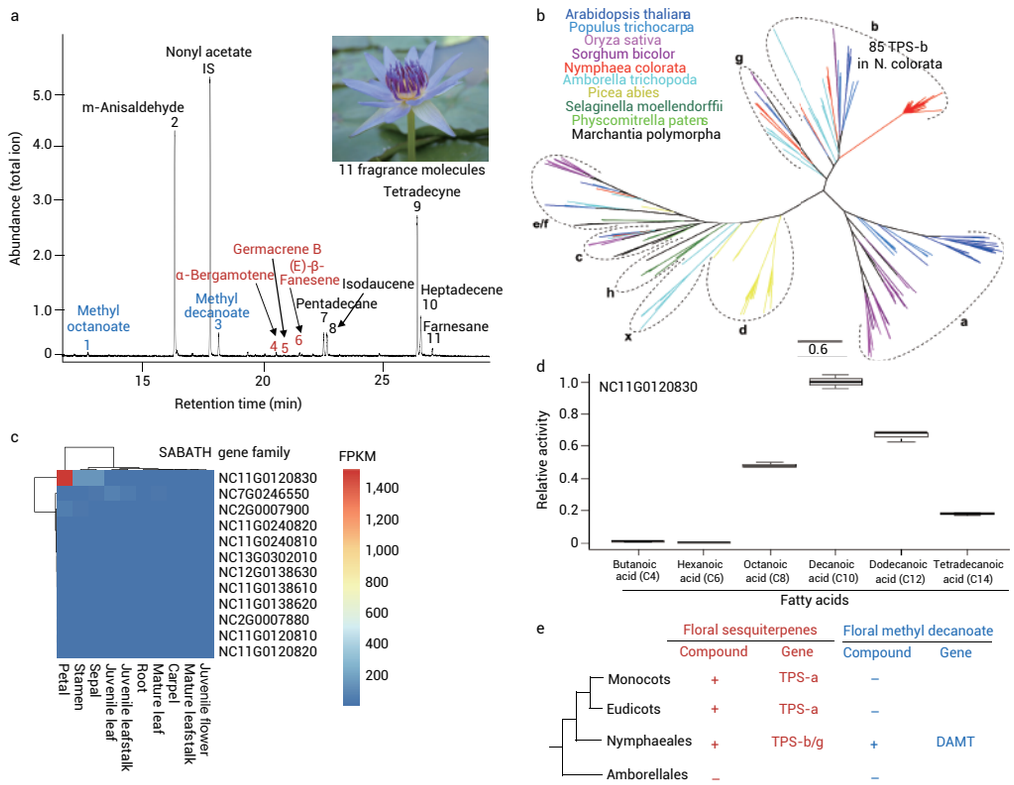


Fig. 4 | Floral scent and biosynthesis in N. colorata.

Reference:

Zhang L, Chen F, Zhang X, et al. The water lily genome and the early evolution of flowering plants[J]. *Nature*, 2020, 577(7788): 79-84.

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