

# 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 with the planning of optimal sequencing and bioinformatics options, to assure your project is a success.

### Sequencing Specification

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



#### Sample preparation and services

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

#### Sequencing quality standard

- Guaranteed  $\geq 90\%$  of DNBSEQ™ clean bases with quality score of Q20
- Guaranteed  $\geq 100\text{Gb}$  Nanopore pass data with Read length  $> 20\text{kb}$
- Guaranteed  $\geq 100\text{Gb}$  PacBio Sequel II CLR data with Read length  $N50 > 20\text{kb}$
- Guaranteed  $\geq 20\text{Gb}$  PacBio Sequel II CCS (HiFi library) data with accuracy greater than 99%

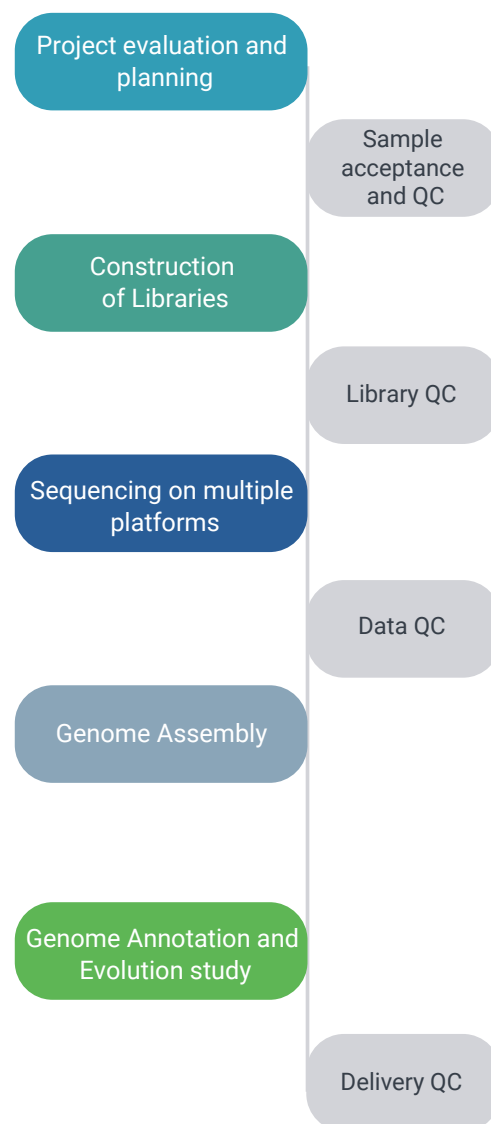


#### Turnaround Time For the species (genome size $\leq 5\text{Gb}$ ) :

- 40 working days from sample QC acceptance to filtered data availability;
- 40/70 working days for the bioinformatics of common/complex genome assembly;
- 30 working days for the bioinformatics of genome annotation;
- Case by case (for species with genome size  $> 5\text{Gb}$ ).

### 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 overall scientific objectives of the project.

Our plant and animal *de novo* sequencing service is usually performed using a combination of available platforms, including our own DNBSEQ™ technology NGS platform augmented with Nanopore PromethION, PacBio Sequel II, Hi-C platforms for sequencing, library preparation and mapping. In addition, we offer extensive bioinformatics data analysis options for genome assembly, annotation and evolution.

Platform Tools	Library type	Read Lengths	Recommended Sequencing Depth
DNBSEQ	350bp Library	PE100/PE150	≥60X
	stLFR Library	PE100	≥100X
	Hi-C Library	PE150	≥100X
Nanopore PromethION	20-50Kb Library	Read length ≥20Kb	≥100X
	Ultra long library (>50K)	Read length ≥50Kb	≥100X
PacBio Sequel II	15Kb CCS (HIFI) Library	Read length ≥15Kb	≥50X
	20-40K CLR library	Read length ≥20Kb	≥100X

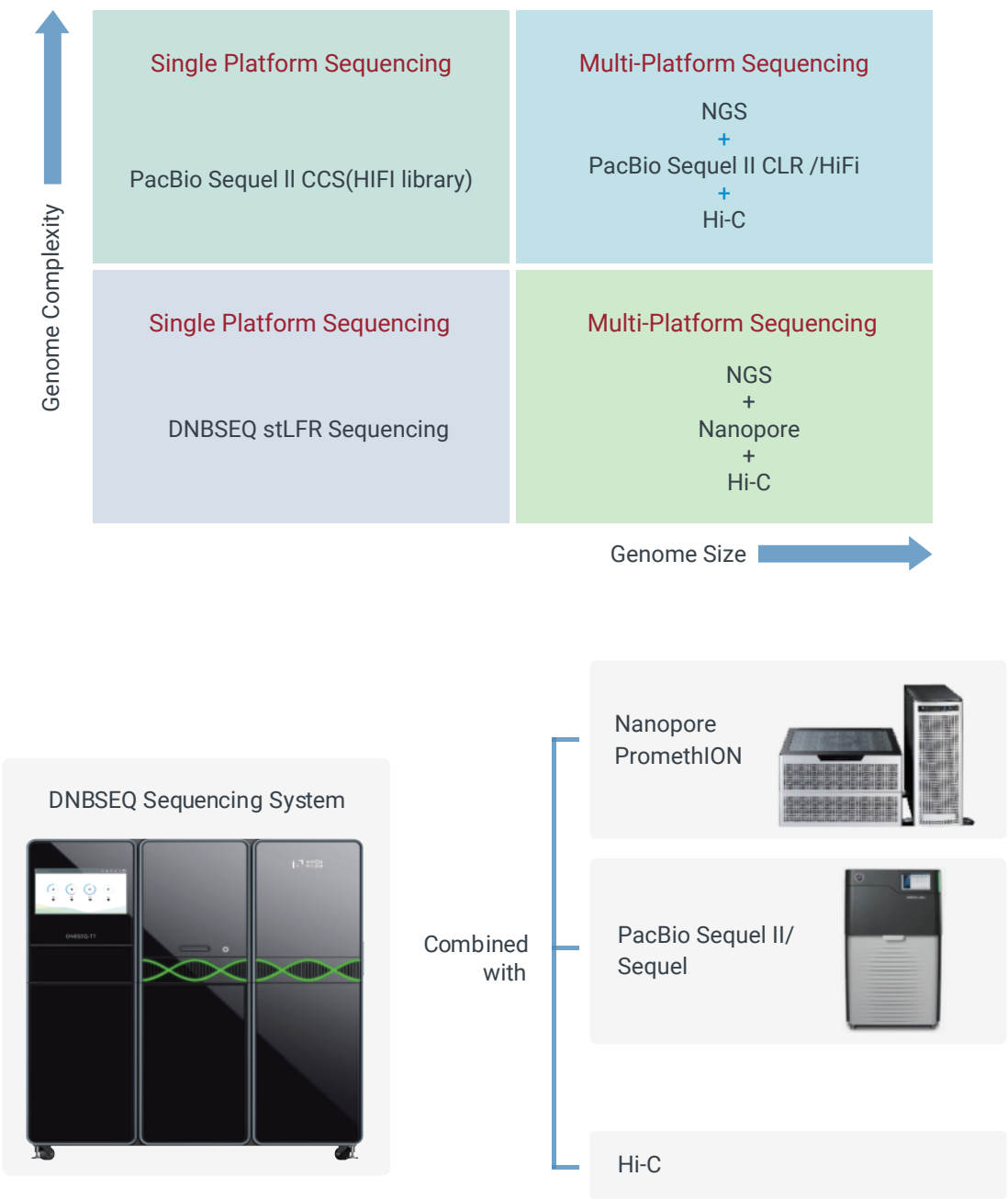
1. Packaging strategy for common genome:

- 1) Nanopore PromethION 100X+ DNBSEQ 60X;
- 2) PacBio Sequel II CLR 100X+ DNBSEQ 60X;
- 3) StLFR 100X;

2. Packaging strategy for highly heterozygous genome:

- 1) Nanopore PromethION 150X + DNBSEQ 60X;
- 2) PacBio CLR 150X+ DNBSEQ 60X
- 3) PacBio HIFI 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 is available.

Genome Survey	<ol style="list-style-type: none"> <li>1. Kmer estimation (Jellyfish + GenomeScope);</li> <li>2. External pollution Analysis (BWA);</li> </ol>
Genome Assembly (Pacbio HiFi data)	<ol style="list-style-type: none"> <li>1. Assembly;</li> <li>2. Assessment by short reads alignment;</li> <li>3. BUSCO assessment;</li> </ol>
Genome Assembly (Pacbio CLR /Nanopore data+NGS)	<ol style="list-style-type: none"> <li>1. Reads correction;</li> <li>2. Assembly;</li> <li>3. Assembly result correction using long reads;</li> <li>4. Assembly result correction using short reads;</li> <li>5. BUSCO assessment</li> </ol>
stLFR Assembly	Genome assembly by stLFR data
Gene Annotation	<ol style="list-style-type: none"> <li>1. Repeat annotation;</li> <li>2. Gene structure annotation;</li> <li>3. Gene function Annotation;</li> <li>4. Transcrip factors (plant)</li> </ol>
Evolution	Deliver published genome and allied species (less than 10 species) <ol style="list-style-type: none"> <li>1. Gene family identification (Animal TreeFam; Plant OrthoMCL; <math>\leq 10</math> species);</li> <li>2. Phylogenetic tree construction;</li> <li>3. Estimation of divergence time;</li> <li>4. Genome syteny analysis;</li> <li>5. Whole genome duplication analysis;</li> <li>6. long fragments duplication (Animal LASTZ);</li> <li>7. Gene family expansion and contraction analysis</li> </ol>
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 <i>de novo</i> Sequencing (Genomic DNA)						
Platform	Sample type	Mass	Concentration	OD	Integrity (AGE)	Sample Purity
Nanopore PromethION	20-40Kb library	≥9µg	90 ng/µl	OD260/280: 1.8-2.2	No degradation or little degradation with main band ≥50Kb	No contamination with RNA, protein or salt ions; non-sticky
	Ultra-long library(>40k)	≥10µg	150 ng/µl	OD260/230: 1.8-2.2		
PacBio Sequel II	15-20Kb HIFI library	≥15µg	80 ng/µl	OD260/280: 1.6-2.5	No degradation or little degradation with main band ≥30Kb	
	20-40kb CLR library	≥7µg	70 ng/µl	OD260/230: 1.6-2.5	No degradation or little degradation with main band ≥40Kb	
DNBSEQ	350bp library	≥1µg	12.5 ng/µl	-	No degradation or little degradation with main band ≥20Kb	
	StLFR library	≥500ng	1 ng/µl	-	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.

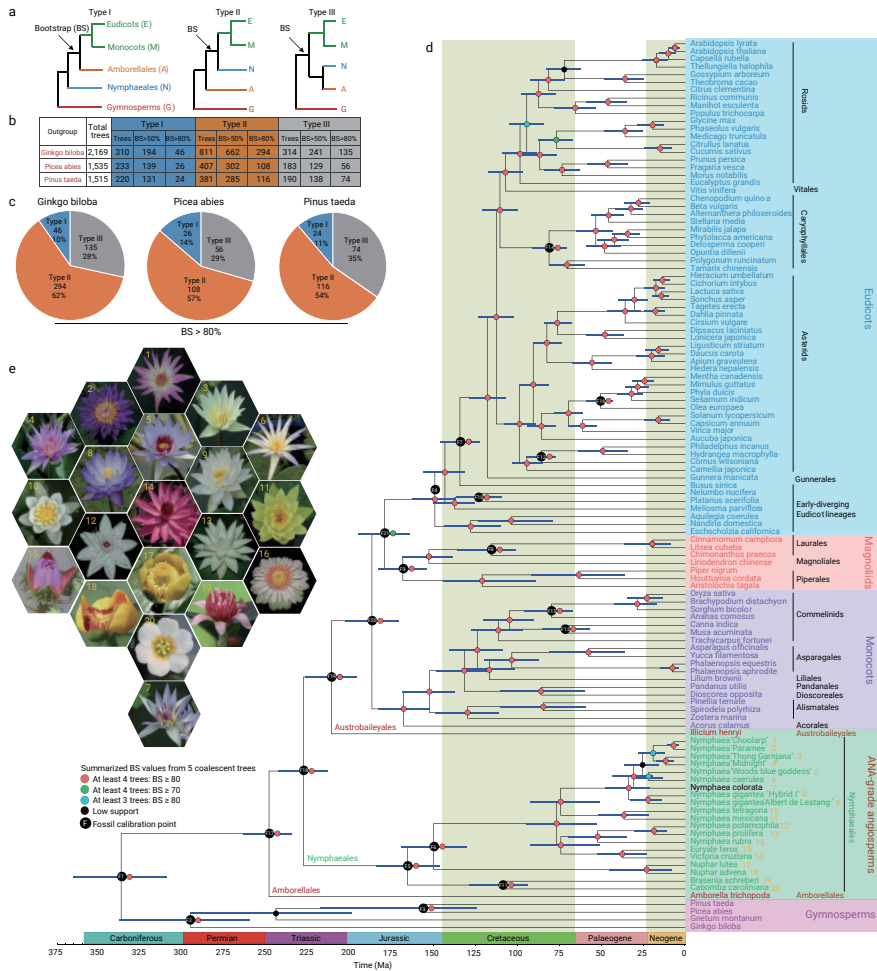


Fig 1. | Phylogenomic relationships of angiosperms.

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.

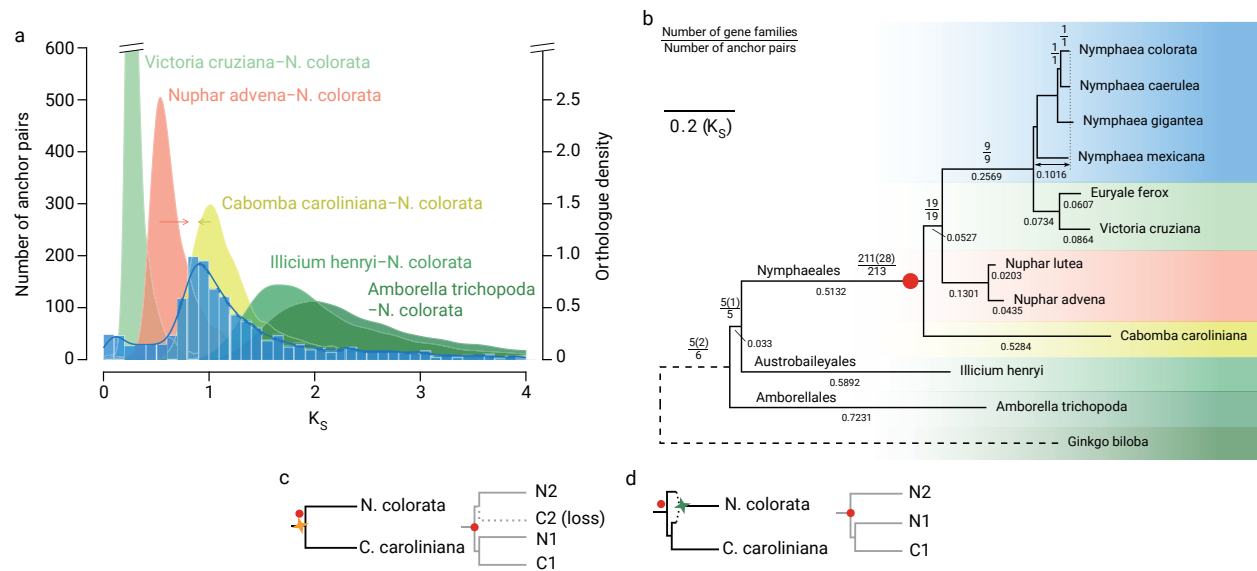


Fig. 2 | A Nymphaealean WGD shared by Nymphaeaceae and possibly Cabombaceae.

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

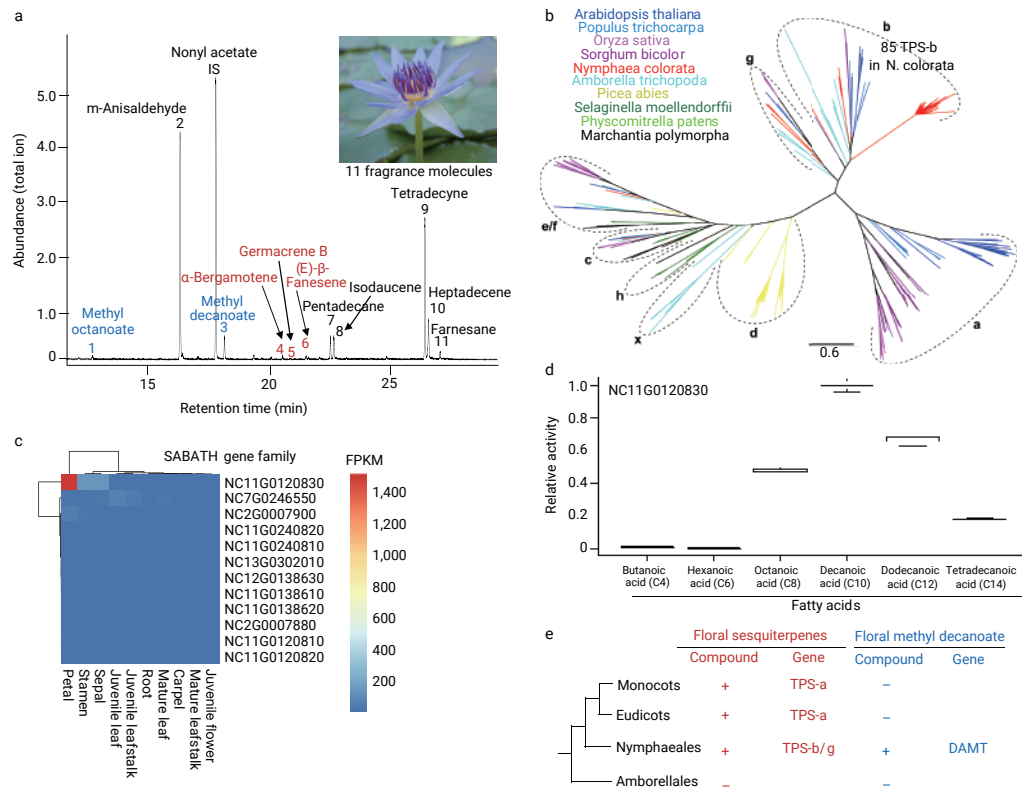


Fig. 4 | Floral scent and biosynthesis in *N. colorata*.

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 we can meet your specific needs or for expert advice on experiment design, from sample to bioinformatics, please don't hesitate to contact us at [P\\_contact@innomics.com](mailto:P_contact@innomics.com)

### **Innomics Inc.**

2904 Orchard Parkway  
San Jose, CA 95134  
USA

For Research Use Only. Not for use in diagnostic procedures (except as specifically noted).

Copyright© Innomics Inc. 2023. All trademarks are the property of Innomics Inc. or their respective owners. This material contains information on products targeted to a wide range of audiences and could contain product details or information otherwise not accessible or valid in your country. Please be aware that we do not take any responsibility for accessing such information, which may not comply with any legal process, regulation, registration, or usage in the country of your origin. Please get in touch with a representative for regional availability. The company reserves the right of final interpretation.