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. BGI is a recognized leader in *de novo* Whole Genome Sequencing and has extensive experience from the de novo sequencing and assembly of more than 170 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

BGI 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 ≥90% of DNBSEQ™ clean bases with quality score of Q20
- Guaranteed ≥50Gb Nanopore pass data with polymerase length longer than 10kb
- Guaranteed ≥100Gb PacBio Sequel II CLR data with polymerase length longer than 10kb
- Guaranteed ≥20Gb PacBio Sequel II CCS (HiFi library) data with accuracy greater than 99%

Turnaround Time

- For the species (genome size ≤ 5Gb) :
- 70 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;
- For the species (genome size ＞ 5Gb) ， case by case;

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.

Sequencing Strategy

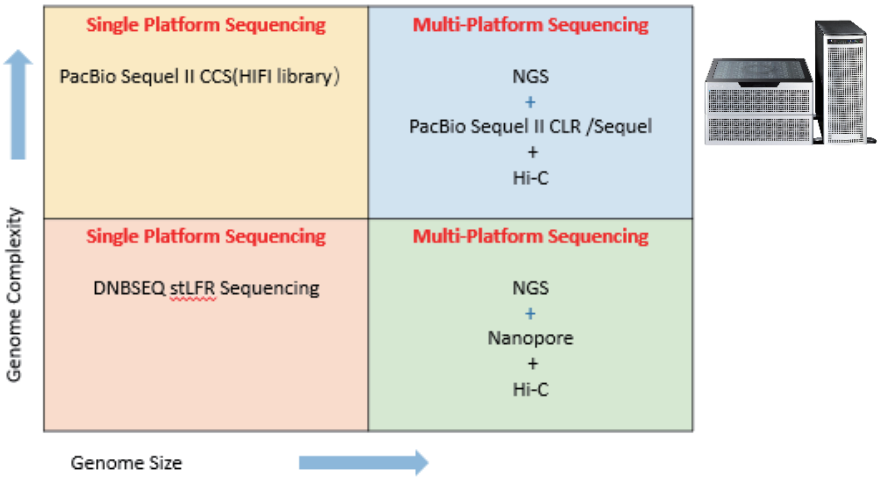
De novo sequencing usually requires a customized approach based on your subject species' genome size and complexity as well as overall scientific objectives of the project.
Our plant and animal *de novo* sequencing service is usually performed using a combination of available platforms, including our own DNBSEQ™ technology NGS platform augmented with Nanopore PromethION, PacBio Sequel II, Hi-C platforms for sequencing, library preparation and mapping. In addition, BGI offers extensive bioinformatics data analysis options for genome assembly, annotation and evolution.

Platform Tools	Library type	Sequencing /USE	Recommend sequencing depth
DNBSEQ	350bp normal Library	PE100/PE150	≥60X
	stLFR Library	PE150	≥100X
	Hi-C Library	PE150	≥100X
Nanopore PromethION	20Kb Library	Read length ≥10Kb	≥100X
PacBio Sequel II	20Kb CLR Library	Read length ≥10Kb	≥100X
	15Kb CCS（HIFI） Library	Read length ≥10Kb	≥50X

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



DNBSEQ Sequencing System

Combined with

- Nanopore PromethION
- PacBio Sequel II/ Sequel
- Hi-C

Data Analysis

Besides clean data output, BGI offers 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	1. Kmer estimation (Jellyfish + GenomeScope); 2. External pollution Analysis (BWA);
Genome Assembly	1. Reads correction; 2. Assembly; 3. Assembly result correction using long reads; 4. Assembly result correction using short reads; 5. BUSCO assessment ;
Gene Annotation	1. Repeat annotation; 2. Gene prediction; 3. Gene function Annotation;
Evolution	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; 7. Positive selection analysis
Auxiliary Assembly	Hi-C data auxiliary assembly
stLFR Assembly	Genome assembly using stLFR data

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). BGI also provide special sample extraction services to satisfy project requirements

Plant and Animal Genome De novo Sequencing(Genomic DNA)					
Platform	**Sample type**	**Mass**	**OD**	**Integrity (AGE)**	**Sample Purity**
Nanopore PromethION	20Kb library	≥10μg	OD260/280: 1.6-2.2 OD260/230: 1.6-2.2	The band shown on gel electrophoresis has little degradation, or of fragment size greater than 40kb.	No contamination with RNA, protein or salt ions; colorless and transparent; non-sticky
PacBio Sequel II	20Kb CLR library	≥10μg			
PacBio Sequel II	20Kb HIFI library	≥20μg			
PacBio Sequel	20Kb library	≥10μg			
DNBSEQ	stLFR library	≥500ng	–	The band shown on gel electrophoresis has little degradation, or of fragment size greater than 20kb.	
	350bp library	≥1μg			

Examples of de novo projects executed by BGI

Species	Sequencing		Mass		
	Sequencing strategy	Read length	Assembled length (Gb)	Scaffold N50 (Mb)	Contig N50 (kb)
Human	10X Genomics Chromium library 48X	PE150	2.93	7.27	45.68
Marine Mammal	NGS libraries (200bp-10kb) 21X + 10X genomics Chromium library	PE150 & PE50	2.58	37.03	116.6
Wheat	BAC+ NGS + PacBio + BioNano + 10X genomics	PE150, PE250; PacBio CLR & BioNano NLRS	4.79	3.67	344
Tea	NGS libraries (170-40kb) 436X + PacBio	PE150, PE250 & PE50; PacBio CLR	3.14	1.39	67.07

A case study of water lily genome

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

Project Partners:

1. Fujian Agriculture and Forestry University, China
2. Nanjing Agricultural University, China
3. Ghent University, Belgium
4. BGI-Shenzhen, China.

Species: Nymphaea colorata

Strategy:

124X (49.8 Gb) PacBio long reads data for gap filling and scaffolding.

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.

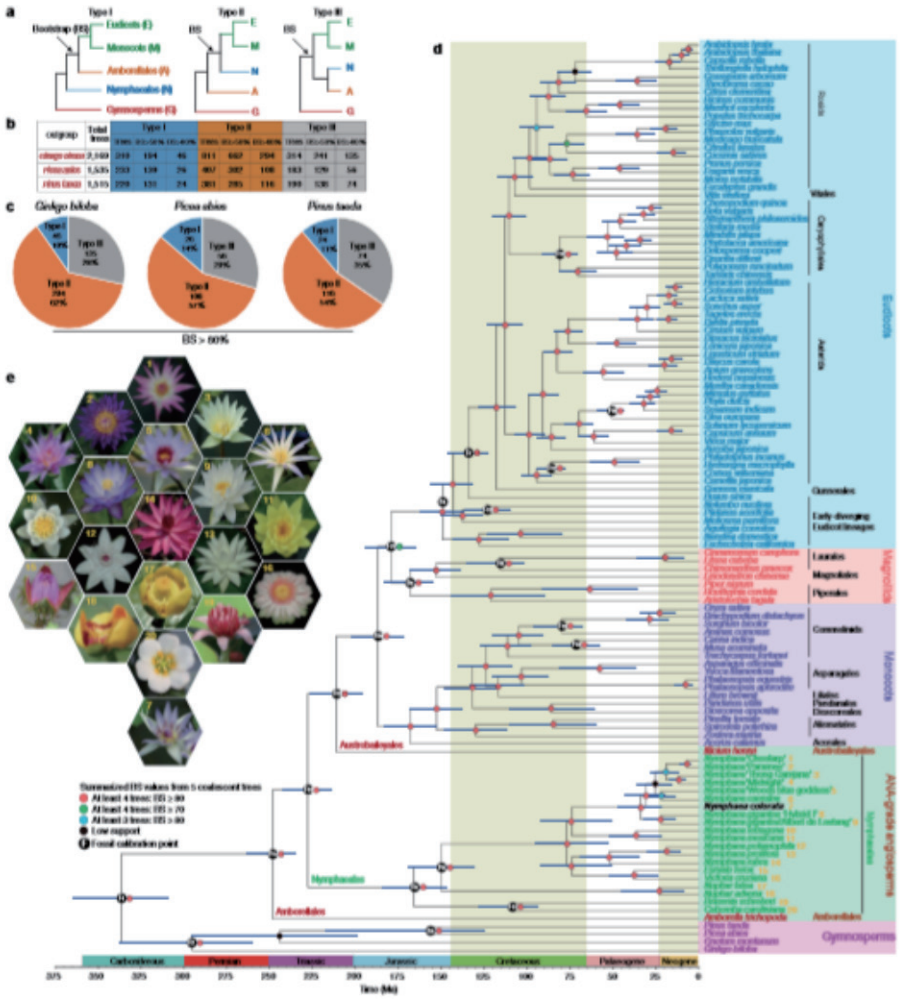


Fig 1. | Phylogenomic relationships of angiosperms.

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.

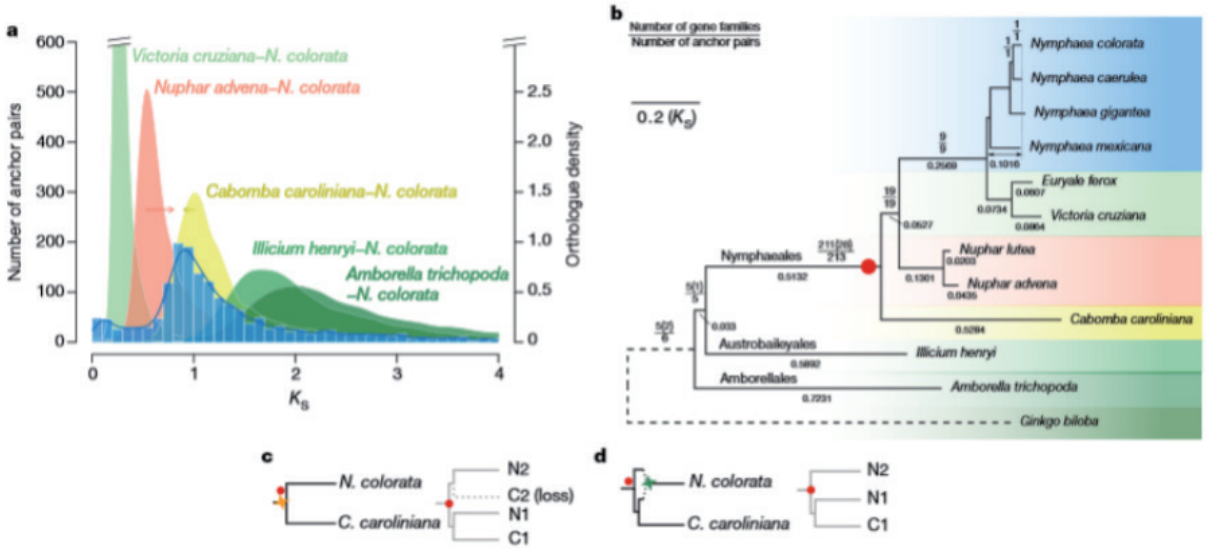


Fig. 2 | A Nymphaealean WGD shared by Nymphaeaceae and possibly Cabombaceae.

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

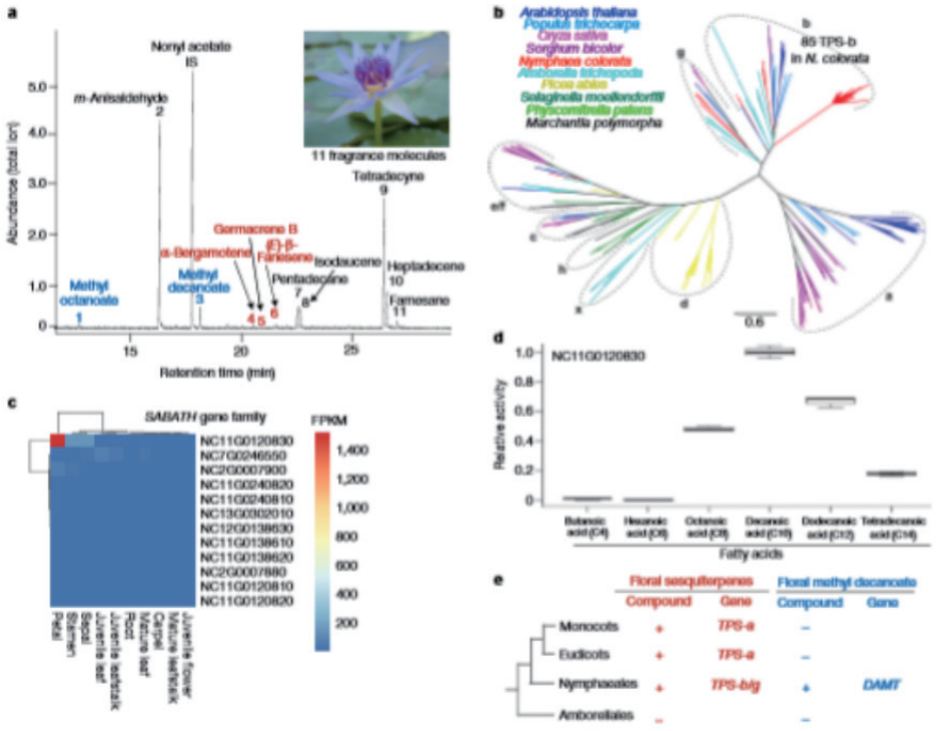


Fig. 4 | Floral scent and biosynthesis in N. colorata.

Request for Information or Quotation

Contact your BGI account representative for the most affordable rates in the industry and to discuss how we can meet your specific project requirements or for expert advice on experiment design, from sample to bioinformatics.

info@bgi.com
www.bgi.com

International Head Offices

BGI Americas
One Broadway, 14th Floor
Cambridge, MA 02142,
USA
Tel:+1 617 500-2741

BGI Europe
Ole Maaløes Vej 3,
DK-2200 Copenhagen N,
Denmark
Tel:+45 7026 0806

BGI Asia
16 Dai Fu Street,
Tai Po Industrial Estate,
New Territories, Hong Kong
Tel:+852 36103510

BGI Australia
L6, CBCRC, 300 Herston
Road, Herston, Brisbane,
Queensland 4006, Australia
Tel: +61 (07) 3362 0475

BGI Genomics BGI_Genomics

Published June 2020.

All Services and Solutions are for research use only.

We Sequence, You Discover