Novogene

The Utility of Short- and Long-read Sequencing in Genomics Research

Review of different studies that have used a combination of multiplex sequencing strategies, especially by introducing long-read sequencing to construct a more accurate and complete genome sequence of various species ranging from mammal, plant, to molluscs. The construction of complete genomes has laid an important foundation for exploring gene function and variation, trait expression, environmental adaptation, and evolution during climate changes.





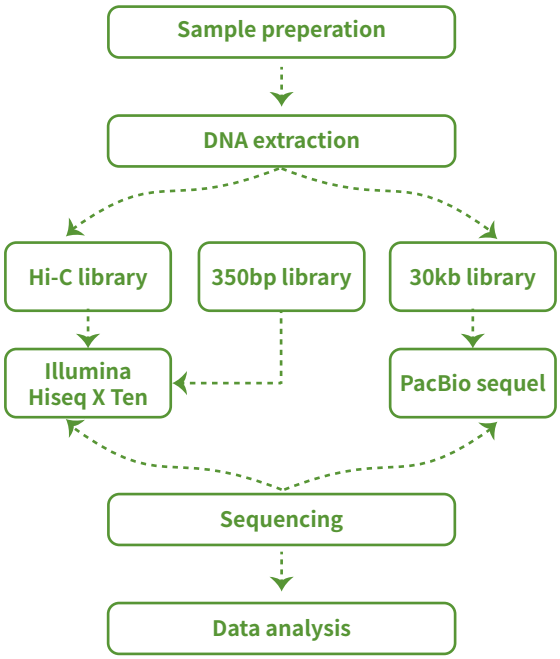



Novogene

Using Genomic Analyses to Examine The Adaptation and Evolution of A Living Fossil

The nautilus is commonly known as a living fossil due to its striking resemblance to ancestors who inhabited our seas half a million years ago. Until recently, genomic information on this species was lacking. Researchers used Illumina and PacBio sequencing technologies to produce the first chromosome-level genome assembly of *Nautilus pompiius*. [1]

1- First, genomic DNA was extracted from samples using the DNAsecure Plant Kit (Tiangen). This DNA was then used to construct libraries sequenced on the Illumina HiSeq X platform. One 30-kb size-selected SMRT bell library was constructed and then sequenced using the PacBio Sequel Platform. Illumina short read sequencing data was used to estimate the genome size. The long-reads from the PacBio Sequel Platform were used to assemble the contigs of the N. Pompiius genome using the WTDEG program before being polished using QLIVER. Short-reads from the Illumina sequencing were then used to correct any errors. Once the genome was assembled, the researchers used Hi-C technology to anchor the scaffolds into chromosomes.



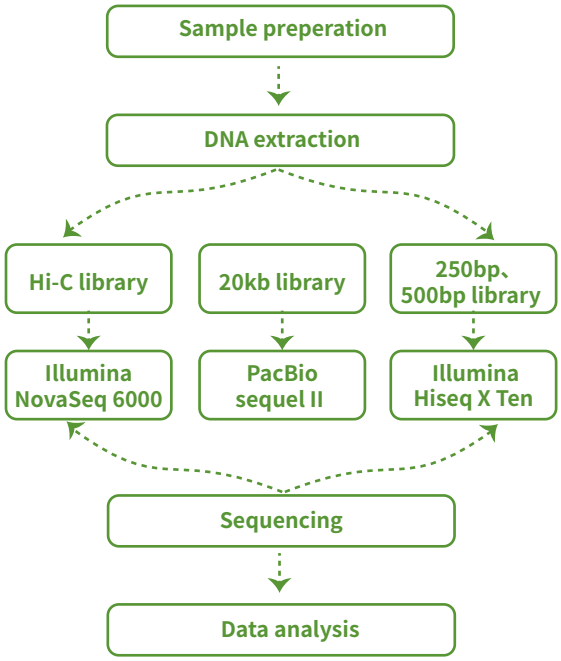
2- The combined use of SMRT sequencing on the PacBio Platform and Illumina sequencing enabled the researchers to generate 186 Gb of data for genome assembly. Once assembled, 15,096 gene models in the genome were identified, and 98% of the genes were functionally annotated. The sequencing enabled the researchers to identify gene families involved in specific adaptation and evolution events and provided a greater insight into the biology of N. Pompiius. For example, sequencing showed that the small size of the Nautilus genome is because of a reduction in repeat content and gene size. In addition, this study identified evolving conserved genes and recent transposable element insertions, which suggests that the evolution of this animal has been driven in a specific direction to help it to cope with fluctuations in environmental conditions. The use of long and short-read sequencing provides novel information that could be an important resource for future conversation genomic studies.

Novogene

Chromosome-level Genome Assembly of The Arctic Fox (*Vulpes Lagopus*) Using PacBio Sequencing and Hi-C Technology

Understanding how genetic diversity helps animals adapt to extreme environments provides essential information for evolution and conservation studies. The Arctic fox (*Vulpes Lagopus*) is one such species that has adapted to survive life in the extreme cold, yet the molecular basis for this adaption has not been investigated. Peng et al. (2001) used PacBio sequencing and chromosome structure capture technique (Hi-C technique) to create the first assembly of the Artic fox genome to explore how their genetic diversity could contribute to their survival in extreme climatic conditions.[2]

1- Genomic DNA was extracted from the blood samples of female foxes. The Illumina HiSeq X Ten and PacBio Sequel II platforms were used to sequence enough data to assemble the genome. Errors were corrected using Quiver and long-read sequences before using Pilon and short reads to polish the genome. The sequencing reads were assembled to create a genome with a total length of 2.345 GB (contig N50 of 31.848, scaffold N50 of 131.537 Mb). This genome was used to predict 21,278 protein-coding genes, of which 99.14% were functionally annotated. Comparison with the genomes of other mammals showed that the Arctic fox is most closely related to *V. Vulpes*.



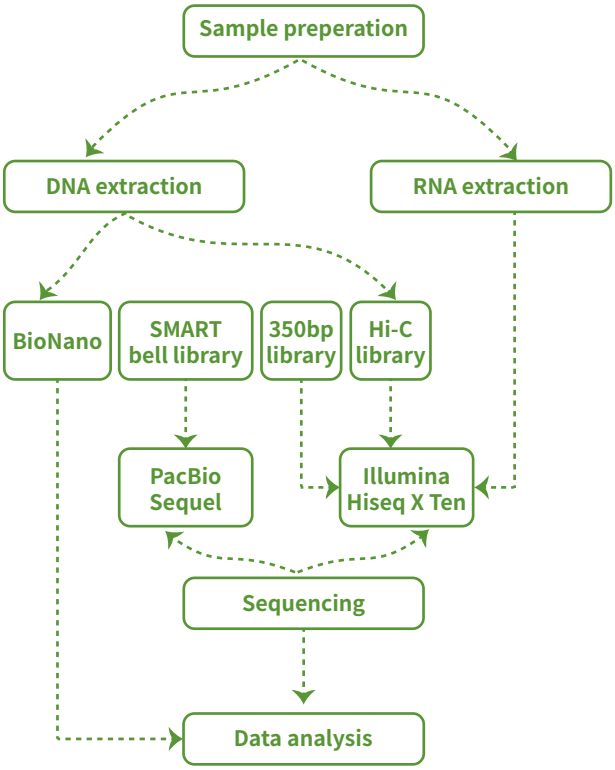
2- The results from the sequencing enabled the researchers to provide the first chromosome-level genome for a species from the Vulpes genus. This genome was then used for a comparative analysis that revealed novel information about the evolutionary history of the Arctic fox and the adaptations it uses to survive in extreme cold.

Novogene

Genomic Insights Into The Origin, Domestication, and Diversification of *Brassica Juncea*

Brassica juncea is an important agricultural species worldwide, more commonly known as mustard. Four subspecies have been identified. The classification of these subspecies depends on the use of the plant and includes *uncea* (seed mustard), integrifolia (leaf mustard), *napiformis* (root mustard), and *tumida* (stem mustard) .[3]Despite its importance as an agricultural species, not a lot is known about the origin, domestication, and diversification of *Brassica juncea*.

1- The researchers used techniques in population genomics to answer some fundamental questions about the origin and domestication of *Brassica juncea*. First, they generated a chromosome-scale de novo scale assembly of the genome of one variant of Brassica juncea, commonly known as Sichuan Yellow. PacBio long reads were combined with Illumina short-read sequencing, BioNano optical mapping, and Hi-C chromatin data. The PacBio reads were assembled using FALCON, and contig correction was carried out using Illumina reads. This enabled the researchers to generate a V.1 assembly.



2- Following the assembly of the genome of Sichuan Yellow, the researchers re-sequenced 480 *B.juncea* accessions from 38 countries. From this, they identified approximately 4.52 million SNPs and 0.97 million insertion-deletion polymorphisms. Sequencing revealed that major genetic variation exists between the different accessions. This information enabled the researchers to reconstruct both the evolutionary and domestication history of B. Juncea. Together, the results demonstrate that B. Juncea originated in West Asia. The evolution of new forms occurred through spontaneous gene mutations and introgressions as they spread east across the world.

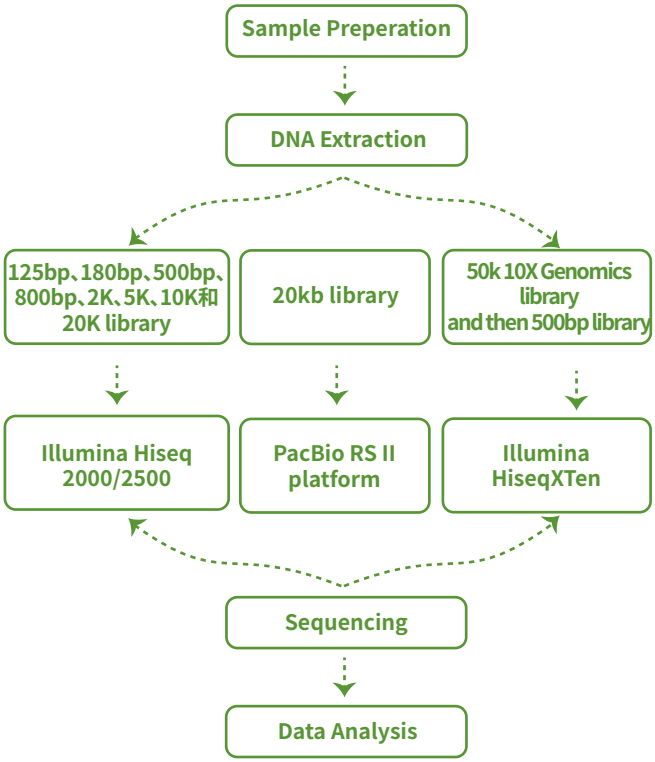
Novogene

The Apostasia Genome and The Evolution of Orchids

Orchids plants (Orchidaceae) are rich in biodiversity, representing 10% of flowering plant species. They are capable of inhabiting a large number of different habits throughout the world which makes them interesting to look at from an evolutionary perspective. Recently, researchers published an article that looked at the evolution of orchids in nature by using next-generation sequencing to obtain a high quality genome for the orchid *Apostasia schenzhenica*. This genome was then used as a reference to infer the genome structure and content of other orchid species with the aim of providing an insight into the origins and evolutionary history of these plants and their genomes. [4]

1- To first obtain the genome for A. *schenzhenica*, sequencing libraries were constructed using a library construction kit provided by Illumina, and short-read sequencing was carried out with sequencing sizes of 180 bp to 20 kb. 80.02 Gb raw reads were generated by Illumina HiSeq 2000 platform. cDNA libraries were also constructed using the Illumina HiSeq 2000 platform, generating 100 based pair paired-end reads. A preliminary assembly of the genome was then carried out using ALLPATHS-LG software.

2- Once the genome was assembled, the researchers were able to use new technologies, such as 10X Genomics and Pacbio, to significantly improve the assembly level of the genome for A. *schenzhenica* as well as the genomes of two other orchids; *Dendrobium catenatum* and *Xiaolanyu Phalaenopsis*. This information was then used in conjunction with the transcriptomes of a variety of orchids and other plants to construct a phylogenetic tree was constructed and the study shows that this subfamily possesses certain shared genes with other subfamilies, such as the Phalaenopsis and Dendrobium, suggesting a whole-genome duplication (WGD) amongst all the extended orchids.



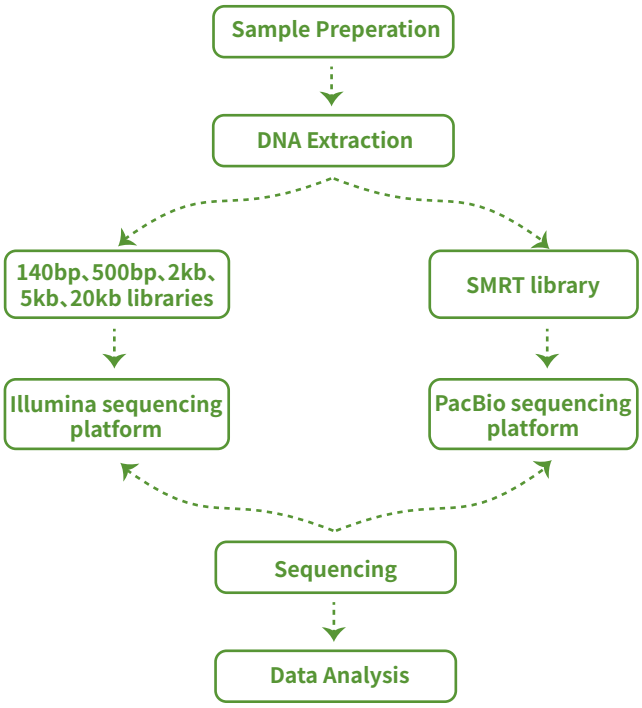
3- This study demonstrates the effectiveness of 10X Genomics and Pacbio third-generation sequencing in improving the assembly level of the genomes which has enabled the researchers to provide a complete analysis of the evolutionary history of orchid plants.

Novogene

Genomic Analyses of Primitive, Wild and Cultivated Citrus Provide Insights Into Asexual Reproduction

Apomixis, a type of asexual reproduction, is observed in modern citrus plants and is where the offspring produced is a maternal clone. In the following article, researchers sought to find the locus for reproduction-associated genes and polyembryony, the presence of multiple (up to 30 and more) embryos within one seed. [5]

1- A total of four citrus genomes were used for annotation and de novo assembly; the *C. grandis, Atalantia buxifolia* (primitive citrus), *C. ichangensis* (papeda, or 'wild citrus') and C. medica (citron), using Pacbio single molecule sequencing in combination with Illumina data. The advancement of sequencing technology allows researchers to use inserts ranging from small (140 to 500 bp) to long ones, reaching even 20 kb. These sequenced reads were assembled using the algorithm from the SOAPdenovo package. This use of shotgun sequencing allowed the researchers to obtain a sweet orange genome that was 361 Kb in size.



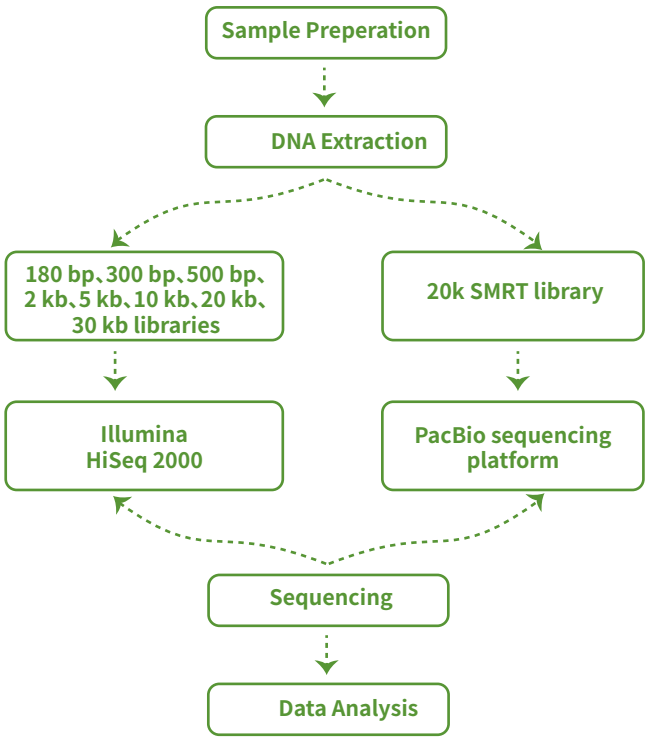
2- In addition, the researchers resequenced more than 100 citrus plants and used the information to analyze the population evolution of the citrus plant. They found that there was a string selection signal of 220Kb on chromosome 9. This information was used to map the relevant sites of the citrus polyembryonic reproduction using BSA trait positioning on a section of 80Kb.From this they found that CitRWP is most closely related to polyembryonic reproduction. Further, the researchers found that when the transposon was inserted into promoter region of the CitRWP gene it resulted in high expression in the ovules of polyembryonic varieties.

Novogene

Scallop Genome Reveals Molecular Adaptations to Semi-sessile Life and Neurotoxins

Bivalve mollusks are descendants of a long existing lineage that dates back to the Cambrian period. These species have survived several mass extinctions events, currently living on different environments; ranging from world oceans to freshwater zones. For this study, researchers were interested in understanding how the scallop *Chlamys farreri* adapted to its living environment by examining various levels of regulation using genomics, transcriptomics, proteomics, and morphology. [6]

1- Understanding the mutations that scallops have undergone to obtain the phenotypes that allow them to inhabit specific environments is crucial and provides a better picture of their evolutionary history. The genome of a 2-year-old *Chlamys farreri*, known as the Chinese scallop, was sequenced using the Illumina HiSeq 2000 and the Pacbio sequencing platform, using both short (180, 300 and 500 bp) and long inserts (2, 5, 10, 20 and 30 kb) DNA libraries. A modified SOAPdenovo method was used for the genome assembly. The protein-coding genes was annotated using the SwissProt, TrEMBL, InterPro, GO (gene ontology), and KEGG (Kyoto Encyclopedia of Genes and Genomes) databases. Single Nucleotide Polymorphisms (SNPs) were characterized through alignment to the assembled genome using the BWA software.



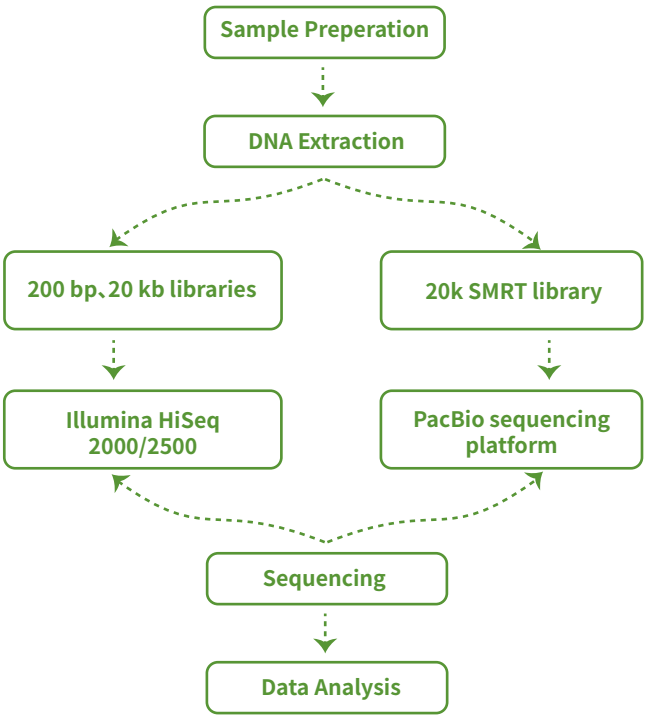
2- This study made use of a wide range of sequencing techniques, covering genomics, transcriptomics, proteomics and morphology-related research methods. This enabled the researchers to provide a new understanding of the evolutionary origin of some of the adaptive traits of these mollusks as well as how their semi-adherent life is regulated.

Novogene

Penaeid Shrimp Genome Provides Insights Into Benthic Adaptation and Frequent Molting

Amongst all the crustaceans, the family *Penaeidea* represents one of the largest groups out there, more specifically, the penaeid shrimp. This shrimp constitutes one of the largest portions of global shrimp catch (~1.3 million tonnes), so it is very important economically. In 2019, researchers deciphered a high quality prawn genome map which has provided essential theoretical information that can be used in crustacean research as well as shrimp genome breeding and molecular improvement .[75]

1- The researchers used second- and third- generation sequencing techniques combined with bioinformatics to establish the whole genome of the Pacific white shrimp *Litopenaeus vannamei*. The de novo sequencing was carried out using the HiSeq 2500 platform and assembly of the whole genome covered ~1.66 Gb with 25,596 protein-coding genes. Genes related to the visual system, nerve signal conduction and the ecdysone signal pathway (a steroid hormone) were identified. The whole *L. vannamei* genome size was measured to be 2.45 Gb with flow cytometry, using the extracted DNA of a single male adult.



2- The researchers found that a large number of species-specific genes in the shrimp genome may be closely related to the evolution of the shrimp family. In addition, a large number of tandem repeat genes in the shrimp genome were also found.

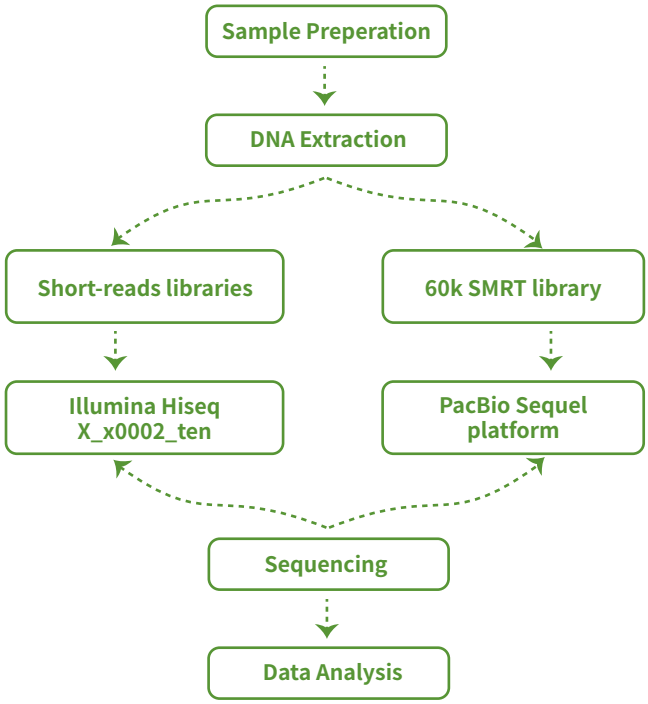
Novogene

Genome of Tripterygium Wilfordii And Identification of Cytochrome P450 Involved in Triptolide Biosynthesis

The *Tripterygium wilfordii*, colloquially known as the *thunder god vine*, has been traditionally for its supposed effects on rheumatoid arthritis, psoriasis and as contraceptive in men. However, it's also been demonstrated that it possesses hepatotoxic components. [8]

1- The researchers sequenced the genome of T. *wildfordii* using Pacbio and 10X Genomics. The total length of the final assembly was 348.38 Mb. The final assembly was further refined using high throughput chromosome conformation capture (Hi-C) data to mount the sequence onto 23 chromosomes. Further, the researchers were able to annotate 28, 321 protein-coding genes. A phylogenetic tree was then constructed to estimate the point of divergence of this plant from other 15 plant species. The most possible date was 87.1 million years ago. The researchers also identified 97 CYP genes, although only 13 were functional.

2- One of these CYP genes is a catabolic enzyme called cytochrome PYP728B70. It is responsible for catalyze oxidation reactions that result in triptolide biosynthesis. This product possesses antitumoral effects. The researchers used overexpression technology to verify that this CYP gene is involved in triptolide biosynthesis and demonstrated that it does play a role in increasing the production of triptolide.



Conclusion

These studies demonstrate how effective the combined use of short- and long-read sequencing is as a tool for genomic studies across a wide range of organisms. Companies such as Illumina and PacBio are at the vanguard of producing state-of-the-art sequencing technology with satisfactory results. With the development of sequencing technology and upgrading platforms, increasing production capacity with higher accuracy and reducing sequencing cost will continue to become the tread. Meanwhile, more accurate and HiFi long reads sequencing technology will be more widely applied to the study of genome structure and function, which will greatly solve the difficulties of analyzing repetitive regions and large structural variants.

References

- Huang, Z., Huang, W., Liu, X., Han, Z., Liu, G., Boamah, G. A., ... & Ke, C. (2022). Genomic insights into the adaptation and evolution of the nautilus, an ancient but evolving “living fossil”. Molecular Ecology Resources, 22(1), 15-27. https://doi.org/10.1111/1755-0998.13439
- Peng, Y., Li, H., Liu, Z., Zhang, C., Li, K., Gong, Y., ... & Li, X. (2021). Chromosome‐level genome assembly of the Arctic fox (Vulpes lagopus) using PacBio sequencing and Hi‐C technology. Molecular Ecology Resources, 21(6), 2093-2108. https://onlinelibrary.wiley.com/doi/abs/10.1111/1755-0998.13397
- Kang, L., Qian, L., Zheng, M., Chen, L., Chen, H., Yang, L., ... & Liu, Z. (2021). Genomic insights into the origin, domestica tion, and diversification of Brassica juncea. Nature genetics, 53(9), 1392- 1402. https://doi.org/10.1038/s41588-021-00922-y
- Zhang, GQ., Liu, KW., Li, Z. et al. The Apostasia genome and the evolution of orchids. Nature 549, 379–383 (2017). https://doi.org/10.1038/s41467-017-01927-0
- Wang, X., Xu, Y., Zhang, S. et al. Genomic analyses of primitive, wild, and cultivated citrus provide insights into asexual reproduction. Nat Genet 49, 765–772 (2017). https://doi.org/10.1038/ng.3839
- Li, Y., Sun, X., Hu, X. et al. Scallop genome reveals molecular adaptations to semi-sessile life and neurotoxins. Nat Commun 8, 1721 (2017). https://doi.org/10.1038/s41467-017-01927-0
- Zhang, X., Yuan, J., Sun, Y. et al. Penaeid shrimp genome provides insights into benthic adaptation and frequent molting. Nat Commun 10, 356 (2019). https://doi.org/10.1038/s41467-018-08197-4
- Tu, L., Su, P., Zhang, Z. et al. Genome of Tripterygium wilfordii and identification of cytochrome P450 involved in triptolide biosynthesis. Nat Commun 11, 971 (2020). https://doi.org/10.1038/s41467-020-14776-1

Novogene Co., Ltd.

en.novogene.com

Novogene Global | Novogene AMEA | Novogene America | Novogene Europe

inquiry_us@novogene.com | marketing@novogene-europe.com | marketing_amea@novogeneait.sg

Follow us on LinkedIn