

Table 1 Statistics of the Nanopore three-generation sequencing and genome assembly of *Artocarpus Nanchuanensis*.

Nanopore three-generation sequencing results		Statistics on assembly results	
SumBase	123,375,925,044	Contig length	769,440,982 bp
MaxLen	216,661 bp	Contig max	8,879,419 bp
SeqNum	7,057,335	Contig number	1,087
N50Len	19,177 bp	Contig N50	2,094,024 bp
N90Len	11,029 bp	Contig N90	402,757 bp

Note: SumBase means total base number of sequencing data; MaxLen means the longest reads length of sequencing data; SeqNum means the total read number of sequencing data; N50Len means the N50 length of sequencing data reads; N90Len means the N90 length of sequencing data reads; Contig length means the length of Contig in the middle of more than 1Kb of scaffolding; Contig Max means the length of the longest Contig; Contig Number means the Number of Contig in the middle of more than 1Kb scaffolding; Contig N50 means the length of Contig N50; Contig N90 means the length of Contig N90.

Table 2 Statistical table of Clean Data and genome comparison results.

Library	Mapping Type	Number	Ratio (%)
	Total Read Pairs	458,907,479	100
	Mapped Reads	728,487,984	79.37
	Unique Mapped Read Pairs	236,274,160	51.45

Table 3 Statistical table of Hi-C sequencing data types.

Library	Type	Number	Ratio (%)
	Unique Paired Alignments	236,274,160	100
	Valid Interaction Pairs	56,964,635	24.11
	Dangling End Pairs	138,663,715	58.69
	Re-ligation Pairs	15,169,250	6.42
	Self-cycle Pairs	975,544	0.41
	Dumped Pairs	24,501,016	10.37

Note: Unique Paired Alignments: The Unique Read Pairs alignment to the genome; Valid Interaction Pairs: the Valid Interaction Read Pairs; Dangling End Pairs: the Dangling End read pairs in the invalid data; Re-ligation Pairs: the re-ligation Read Pairs in the invalid data; Self-circle Ligation Pairs: Self-circle Ligation Read Pairs in the invalid data; Dumped Pairs: the dumped Read Pairs in the invalid data.

Table 4 The statistics table of Hi-C assembly data.

Group	Cluster Num	Cluster Len (bp)	Order Num	Order Len (bp)
LG01	46	26,514,107	24	24,676,255
LG02	41	26,638,661	16	24,489,134
LG03	30	24,254,703	16	23,044,270
LG04	34	22,404,888	13	20,644,200
LG05	33	21,646,681	16	20,177,649
LG06	35	29,133,579	18	27,822,153
LG07	69	32,924,820	27	29,467,719
LG08	45	29,858,101	20	27,605,363
LG09	77	29,556,483	29	25,185,028
LG10	45	22,896,788	20	20,243,522
LG11	67	25,833,105	20	21,750,724
LG12	37	24,385,337	15	22,370,729
LG13	47	23,481,896	24	21,098,278
LG14	46	29,162,015	19	26,857,340
LG15	61	28,431,484	30	25,341,045
LG16	32	21,965,556	16	20,879,538
LG17	41	25,915,114	19	24,032,910
LG18	49	34,941,454	27	32,502,827
LG19	54	29,520,137	21	25,685,935
LG20	50	32,513,478	18	29,815,261
LG21	50	28,639,915	21	25,613,043
LG22	42	27,392,871	24	25,873,084
LG23	42	28,655,389	16	26,447,344
LG24	52	27,753,222	24	25,148,606
LG25	46	23,720,417	16	21,152,151
LG26	63	33,995,937	28	30,220,329
LG27	58	28,458,315	24	25,577,647
LG28	44	25,907,258	22	23,985,053
Total (Ratio%)	1336 (97.95%)	766501711 (99.62%)	583 (43.64%)	697707137 (91.02%)

Note: the statistics do not include 100 Ns added by artificially connected pseudochromosomes.

Table 5 The genomic statistics after Hi-C assembly.

Scaffold number	Scaffold length (bp)	Scaffold N50 (bp)	Scaffold N90 (bp)	Scaffold max (bp)	Gap total length (bp)
809	769,496,482	25,150,906	20,179,149	32,505,427	55,500
Contig number	Contig length (bp)	Contig N50 (bp)	Contig N90 (bp)	Contig max (bp)	GC content (%)
1,364	769,440,982	1,778,064	200,000	8,646,128	32.34

Note: Scaffold represents the Scaffold generated after Hi-C connection; Contig represents the Contig after Hi-C error corrected. Scaffold number: the number of scaffolds above 1 Kb; Scaffold length (bp): the length of a Scaffold exceeding 1 Kb; Scaffold N50 (bp): length of Scaffold N50 above 1 Kb; Scaffold N90 (bp): length of Scaffold N90 above 1 Kb; Scaffold Max (bp): the length of the longest Scaffold exceeding 1 Kb; Gap total length (bp): the total length of Gap; Contig Number: the Number of Contig in the scaffold above 1 Kb; Contig length (bp): the length of Contig in the middle of the scaffold above 1 Kb; Contig N50 (bp): the length of Contig N50; Contig N90 (bp): the length of Contig N90; Contig Max (bp): the length of the longest Contig; GC content (%): the GC content percentage.

Table 6 The statistics of repeat sequences.

Type	Number	Length	Rate (%)
ClassI	715,880	354,430,815	46.06

Type	Number	Length	Rate (%)
ClassI/DIRS	26,959	23,880,069	3.10
ClassI/LARD	238,559	75,579,627	9.82
ClassI/LINE	12,327	3,488,322	0.45
ClassI/LTR/Copia	210,964	147,520,139	19.17
ClassI/LTR/Gypsy	209,038	129,742,980	16.86
ClassI/LTR/Unknown	3,210	597,880	0.08
ClassI/PLE	1,322	456,964	0.06
ClassI/SINE	8,910	2,272,141	0.30
ClassI/TRIM	4,219	1,617,309	0.21
ClassI/Unknown	372	66,025	0.01
ClassII	103,050	33,806,173	4.39
ClassII/Crypton	17	1,003	0.00
ClassII/Helitron	22,320	5,961,597	0.77
ClassII/MITE	1,183	204,923	0.03
ClassII/Maverick	1,392	1,762,270	0.23
ClassII/TIR	66,336	25,334,820	3.29
ClassII/Unknown	11,802	879,933	0.11
PotentialHostGene	30,328	8,227,684	1.07
SSR	10,909	2,628,783	0.34
Unknown	187,171	55,237,782	7.18
Total	1,047,338	422,782,698	54.94

Note: Type: repeat sequence Type; Number: The Number of repeats that obtained; Length: the total Length of the predicted repeating sequence; Rate (%): the proportion of repetitive sequences in the total genome.

Table 7 The gene prediction results statistics.

Method	Software	Species	Gene number
Ab initio	Genscan	-	42,670
	Augustus	-	47,491
	GlimmerHMM	-	74,635
	GeneID	-	70,861

	SNAP	-	88,742
Homology-based	GeMoMa	Arabidopsis_thaliana	33,745
		Morus_notabilis	41,079
		Prunus_avium	37,127
		Rosa_chinensis	37,535
RNAseq	TransDecoder	-	84,498
	GeneMarkS-T	-	52,056
	PASA	-	48,695
Integration	EVM	-	41,636

Table 8 The statistical table of genetic information.

GeneNum	Genelen	AveGenlen	ExonLen	AveExonLen
41,636	158,114,419	3,797.54	62,835,343	1,509.16
ExonNum	AveExonNum	CDSLen	AveCDSlen	CDSNum
233,559	5.61	50,445,441	1,211.58	226,727
AveCDSNum	IntronLen	AveIntronLen	IntronNum	AveIntronnum
5.45	95,279,076	2,288.38	191,923	4.61

Note: Gene Num: the number of predicted gene; GeneLen (bp): the total number of gene; AveGeneLen(bp): the average length of gene; ExonLen (bp): the total number of exons; AveExonLen (bp): the average of the total exon length of each gene; ExonNum: the number of exons; AveExonNum: the average number of exons per gene; CDSLen (bp): total length of CDS ; AveCDSLen (bp): average CDS length; IntronLen (bp): total intron length; AveIntronLen (bp): the total intron mean length for each gene; IntronNum: the number of introns; AveIntronnum: the introns average number in per gene.

Table 9 The gene function annotated statistics.

Annotation database	Annotated number	Percentage (%)
GO_Annotation	21,275	51.10
KEGG_Annotation	13,903	33.39
KOG_Annotation	21,567	51.80
TrEMBL_Annotation	39,571	95.04

Annotation database	Annotated number	Percentage (%)
nr_Annotation	39,553	95.00
All_Annotated	39,596	95.10

Note: Annotation database: Functional annotation database; Annotated number: the number of genes Annotated to the corresponding database; Percentage (%): The percentage of genes annotated to the total number of databases.

Table 10 The statistical results of non-coding RNA.

RNA classification	Number	Family
miRNA	138	24
rRNA	409	4
tRNA	512	24

Note: RNA classification: The classification of RNA; Number: The predicted RNA number; Family: The RNA family number.

Table 11 Statistical classification of gene families.

Name	Total gene	Cluster num	Total family	Unifamily
A.thaliana	27,369	23,106	12,753	726
A.trichopoda	16,986	15,058	11,147	254
P.trichocarpa	41,335	33,270	14,725	950
A.chinensis	39,040	25,888	12,648	1,327
V.vinifera	26,346	19,238	12,682	665
M.notabilis	26,965	20,423	14,794	524
T.cacao	21,432	20,070	13,810	176
A. Nanchuanensis	41,636	33,925	15,436	512

Note: Name: the name of Species; Total gene: the number of total gene; Cluster num: the number of genes that involved in family classification; Total family number: the number of gene families that can be divided; Uni family: the number of unique gene families.

Table 12 The annotation of protein gene family.

GeneFamily	Pfam	Function
GF_12673	PF00646.28	F-box domain
GF_10548	PF00031.16	Cystatin domain
GF_8	PF00069.20	Protein kinase domain
GF_13176	PF13639.1	Ring finger domain

Note: Gene family : the gene family cluster; Pfam: the ID of protein family alignment to the Pfam database ; Function : the function of the protein family that can be alignmented.

Table 13 The rapidly evolving genes selected by CodeML.

GeneID	P-value	Sites
EVM0035972.1	0.05	298,G,0.993**
EVM0031735.1	0.06	74,E,0.984*
EVM0026117.1	0.35	68,K,0.997**
EVM0015119.1	0.00	232,E,0.990**

Note: Gene ID mean the ID of gene, ω_0 mean ka/ks for the studied Species, ω_1 mean the average ka/ks for other species, ω_2 mean ka/ks for the whole evolutionary tree.