Supporting information

1. Sequencing strategy

This project uses the Whole Genome Shotgun (WGS) strategy to construct different inserts. A Next-Generation Sequencing (NGS) DNA library was constructed and sequenced using the Illumina MiSeq- platform (2×250-bp paired end), which generated 4900736 reads. The data of each sample was statistically analyzed. The results are shown in Table S1. The total amount of sequence data comprises 1,224,164,919 bases

Table S1 Sequencing strategies and statistics of sequencing data

Sample	Library Name	Insert Size	Reads Number	Total Bases (bp)	N (%)	GC (%)	Q20 (%)	Q30 (%)	Sequencing Platform	Sequencing Mode
SMT-1	PE400	400 bp	4,900,736	1,224,164,919	0.0003	61.59	93.86	84.36	Illumina Miseq	Paired-end, 2×250 bp

Note: **N (%):** percentage of fuzzy bases. **Q20 (%):** The percentage of bases with base recognition accuracy above 99%; **Q30 (%):** Percentage of bases with base recognition accuracy above 99.9%.

2. Data quality

The sequencing data contains some low-quality reads ends, which will make a big impact on subsequent information analysis. Interference, in order to ensure the quality of subsequent information analysis, further filtering of the off-machine data is required. Data filtering standards, including the following points:

(1) Joint contamination removal, using adapter removal (ver. 2.1.7) (Mikkel S *et al*. 2016) to remove the low-quality ends;

(2) The quality-controlled reads were assembled using SOAPec (v2.0) (Luo *et al*. 2012), with a default set of k-mer sizes and options.

Table S2 Data quality filtering statistics

Sample	HQ Reads	HQ Reads %	HQ Data (bp)	HQ Data %	Coverage (×)
SMT-1	4,769,110	97.31	1,120,252,671	91.51	183

Note: HQ Reads%: high quality reads as a percentage of the machine reads

HQ Data (bp): high quality reads base number

HQ Data %: The percentage of high-quality sequence bases in the base

Coverage (×): Refers to the depth of sequencing, where the genome size refers to the length of the genomic sequence obtained by splicing.

3. Data statistics

The basic information of the reading frame is shown in Table S3. The amino acid sequence of all open reading frames can be found in "*.protein.faa".

Table.S3 Open reading frame prediction data statistics

Seq ID	Property	Value
SMT-1	ORF number	5,619
	ORF total length	5,345,943 bp
	ORF density	0.919 genes per kb
	Longest ORF length	18,933 bp
	ORF average length	951.4 bp
	Intergenetic region length	762,294 bp
	ORF / Genome (coding percentage)	87.50%
	Intergenetic length / Genome	12.50%
	GC content in ORF region	62.40%
	GC content in intergenetic region	55.80%

4. CRISPR

Using the CRISPR finder (http://crispr.i2bc.paris-saclay.fr/Server/) to predict DRs in the whole genome (forward repeat) and spacers (Bland *et al*. 2007). We obtained 3 CRISPR samples in the SMT-1 genome. The CRISPRs structure, statistical results also determined.

Table S4 CRISPR forecast results

Sample ID	Sequence Name	ID	CRISPRs type	Start	End	Number of Spacer	Length (bp)	Genome %
	contig11	**1**	Questionable	120,173	120,293	1	120	0.0020
SMT-1	Contig21	**1**	Confirmed	65,652	65,980	6	328	0.0054
	Contig33	**1**	Questionable	256	329	1	73	0.0012

5. **Protein coding gene function annotation**

The main purpose of functional annotation of protein-coding genes is to functionally resolve all protein-coding genes, thereby the species is analyzed horizontally. The main method for functional annotation of protein-coding genes is to encode all predicted protein-coding genes into various databases. In general, the longer the region of the two sequences is aligned, the higher the sequence identity of the two sequences. The annotation of the eggNOG of the protein-coding gene is done using the blast software, and the database used by blast is eggNOG (V4), the critical value is selected 1e-6; the function discriminant rule of eggNOG is: E-value < 1e-6. Protein coding gene function annotation information of SMT-1 were indicated in Table **S5**.

Table.S5 Protein coding gene function annotation

Sample	Annotation in Database	No. Of Genes	%
	NR	5,503	97.94%
	eggNOG	4,882	86.88%

SMT-1	KEGG	2,764	49.19%
	Swiss-Prot	3,966	70.58%
	GO	4,052	72.11%