

Research Article

Open Access

Sequencing of Chloroplast Genome of *Cucumis melo* L. var. *agrestis* Naud.

Ma Zongxin ✉

Academy of Agricultural Sciences of Fuyang, Fuyang, 236065, China

✉ Corresponding author email: mzxism965@163.comMolecular Plant Breeding, 2021, Vol.12, No.15 doi: [10.5376/mpb.2021.12.0015](https://doi.org/10.5376/mpb.2021.12.0015)

Received: 12 May, 2021

Accepted: 17 May, 2021

Published: 25 May, 2021

Copyright © 2021 Ma, This article was first published in Molecular Plant Breeding in Chinese, and here was authorized to translate and publish the paper in English under the terms of Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:Ma Z.X., 2021, Sequencing of chloroplast genome of *Cucumis melo* L. var. *agrestis* Naud., Molecular Plant Breeding, 12(15): 1-11 (doi: [10.5376/mpb.2021.12.0015](https://doi.org/10.5376/mpb.2021.12.0015))

Abstract *Cucumis melo* L. var. *agrestis* Naud. belong to the genus *Cucumis* in the Gourd family, it is rich in oil. It can extract high quality edible oil and provide raw materials for medicine and food industry. The chloroplast genome of *Cucumis melo* L. was sequenced for the first time. The results showed that the chloroplast genome of *Cucumis melo* L. was similar to that of common melon. This study can provide a theoretical reference for breeding.

Keywords *Cucumis melo* L. var. *agrestis* Naud.; Chloroplast genome; Sequencing

Field Muskmelon (*Cucumis melo* L. var. *agrestis* Naud.) belongs to the *Cucumis* of the family Cucurbitaceae, wild weeds. A number of scholars believe that the Field Muskmelon is the wild ancestor of Chinese oriental thin skin melon (Ma, 2020a). Seeds contain a lot of fat, which can be extracted from oil and eaten. Fruit rich in medicinal ingredients such as cucurbitacin, is a potential new type of wild plant (Ma, 2020b). At present, chloroplast genome sequencing fragments can be accurately assembled, but it's still difficult to assemble a larger complete genome, nuclear genome assembly is more difficult (Anton et al., 2012). Field Muskmelon is one of the most important cash crops in the world, Chloroplast genome sequencing has been successful, Field Muskmelon chloroplast genome contains 133 genes. The total GC content of the genome was 36.9% (Zhu et al., 2016). Genetic diversity of local muskmelon varieties in vietnam using chloroplast genome analysis (Phan et al., 2010). Genomic sequence polymorphisms were used to infer diversity and genetic differentiation of cultivated melon (Katsunori et al., 2013). And elaborated the melon chloroplast genome complete sequence and its systematic significance (Hong et al., 2020). So far, the study of chloroplast genome of Field Muskmelon has not been reported. In order to understand the composition of chloroplast genome and the relationship between genes and botanical traits, to provide theoretical and technical basis for breeding genetic improvement, The experiment material, Illumina high - throughput sequencing, The chloroplast genome sequencing of Field Muskmelon was completed.

1 Results and Analysis

1.1 Statistics and annotation of chloroplast genome sequencing data

Sequencing showed that the GC content of chloroplast genome was 37.45%, Q20 was 96.83% (Table 1). A total of 133 genes, tRNA37, rRNA8, mRNA88 (Table 2). Chloroplast gene function is mainly divided into photosynthetic synthesis, self-replication and other genes, there are also some genes with unknown function (Table 3).

Table 1 Sequencing data statistics

Sample ID	ReadSum	BaseSum	GC (%)	Q20(%)	Q30(%)
<i>Cucumis melo</i> L. var. <i>agrestis</i> Naud.	55968670	16790601000	37.45	96.83	91.75

Table 2 Statistics of chloroplast gene annotation information

ID	Gene	Trna	Rrna	Mrna	Pseudo
<i>Cucumis melo</i> L.var. <i>agrestis</i> Naud.	133	37	8	88	0

Table 3 Classification and statistics of chloroplast gene function

Classification	Genome	Gene name
Photosynthesis	Subunits of photosystem I	psaA, psaB, psaC, psaI, psaJ
	Subunits of photosystem II	psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ
	Subunits of NADH dehydrogenase	ndhA *, ndhB *(2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK
	Subunits of cytochrome b/f complex	petA, petB *, petD *, petG, petL, petN
	Subunits of ATP synthase	atpA, atpB, atpE, atpF *, atpH, atpI
	Large subunit of rubisco	rbcL
	Subunits photochlorophyllide reductase	-
Self-replication	Proteins of large ribosomal subunit	rpl14, rpl16*, rpl2*(2), rpl20, rpl22, rpl23(2), rpl32, rpl33, rpl36
	Proteins of small ribosomal subunit	rps11, rps12**(2), rps14, rps15, rps16*, rps18, rps19, rps2, rps3, rps4, rps7(2), rps8
	Subunits of RNA polymerase	rpoA, rpoB, rpoC1*, rpoC2
	Ribosomal RNAs	rrn16(2), rrn23(2), rrn4.5(2), rrn5(2)
	Transfer RNAs	trnA-UGC *(2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnG-UCC *, trnH-GUG, trnI-CAU(2), trnI-GAU *(2), trnK-UUU *, trnL-CAA(2), trnL-UAA *, trnL-UAG, trnM-CAU, trnN-GUU(2), trnP-UGG, trnQ-UUG, trnR-ACG(2), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC(2), trnV-UAC *, trnW-CCA, trnY-GUA, trnM-CAU
Other genes	Maturase	matK
	Protease	clpP **
	Envelope membrane protein	cemA
	Acetyl-CoA carboxylase	accD
	c-type cytochrome synthesis gene	ccsA
	Translation initiation factor	infA
	Other	-
Genes of unknown function	Conserved hypothetical chloroplast ORF	ycf1(2), ycf15(2), ycf2(2), ycf3**, ycf4

Note: Gene *: contains one intron; Gene **: contains two introns; # Gene: pseudogenes; and genes with a copy number greater than 1 copy number in parentheses

1.2 Chloroplast genome map

The chloroplast genome of Field Muskmelon is covalently closed double-stranded cyclic molecule (Figure 1), Total 156 017 bp, Includes 1 pair of reverse repeat (IR) zones (25 797 bp), 1 large single copy (LSC) area (86 336 bp) and 1 small single copy (SSR) area (18 087 bp), genome-wide GC content of 37.45%. Six ATP synthase genes and 11 NADH dehydrogenase genes.

1.3 RSCU codon preference

Codon preference indicates that, TGG separate corresponding Trp codes (Figure 2), Met correspond to GTG, ATT, ATG and ATC codes, respectively. both Arg, Leu and Scr correspond to 6 codon codes (Figure 2). The results of codon preference analysis showed that the UAA and GCU values were higher (Table 4), while the GCG, UGC, UAG and UGA values are lower (Table 4).

Table 4 RSCU codon preference

Amino acid	Symbol	Codon	Quantity No.	Preference RSCU
*	Ter	UAA	56	1.909 2
*	Ter	UAG	15	0.511 5
*	Ter	UGA	17	0.579 6
A	Ala	GCA	398	1.136 4
A	Ala	GCC	223	0.636 8
A	Ala	GCG	148	0.422 4
A	Ala	GCU	632	1.804 4
C	Cys	UGC	89	0.566 8
C	Cys	UGU	225	1.433 2

1.4 Disperse repeats

The results showed that there were more forward and palindrome repeats, but less reverse repeats and no complementary repeats (Figure 3).

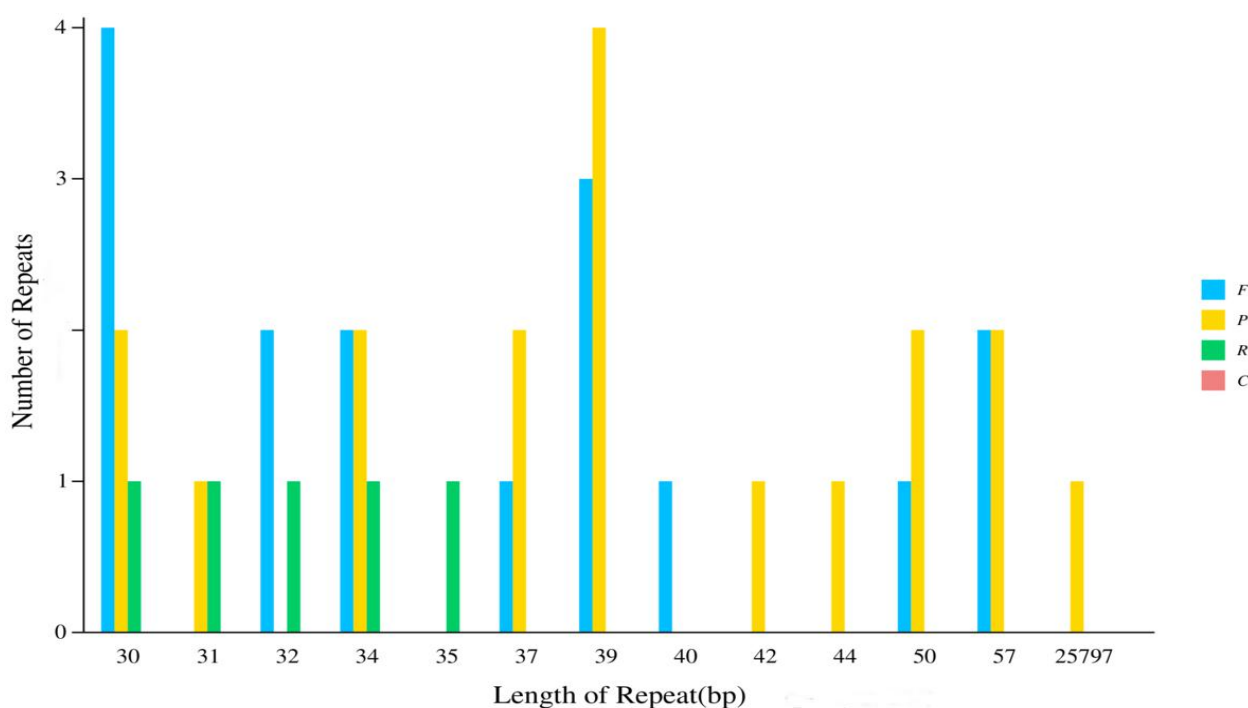


Figure 3 Statistics of scattered repeated sequences

1.5 CpSSR Results

191 single nucleotide repeats in the chloroplast genome of SSR (Table 5; Table 6), 66.55 per cent; SSR 11 dinucleotides, A total of 71 dinucleotide SSR, 24.74 per cent; Four, five and six nucleotides SSR less, were 9, 3 and 2, respectively (Table 5). A, T nucleotide repeats are the largest in single nucleotide repeats, fewer repeats C, G single nucleotide (Figure 4; Table 7).

Table 5 SSR Statistics of the analysis

Unit size	Number of SSRs	Percent (%)
1	191	66.55
2	11	3.83
3	71	24.74
4	9	3.14
5	3	1.05
6	2	0.70

Table 6 Results of SSR analysis

ID	SSR nr.	SSR type	SSR	size	start	end
1	1	p2	(TA)7	14	132	145
1	2	p1	(T)9	9	1 524	1 532
1	3	c	(A)8taagt (A)8	21	1 675	1 695
1	4	p1	(T)9	9	2 170	2 178
1	5	p1	(A)8	8	2 837	2 844
1	6	p3	(TTC)3	9	3 081	3 089
1	7	p1	(A)12	12	3 220	3 231
1	8	p3	(AAT)3	9	3 541	3 549
1	9	p1	(T)8	8	3 954	3 961

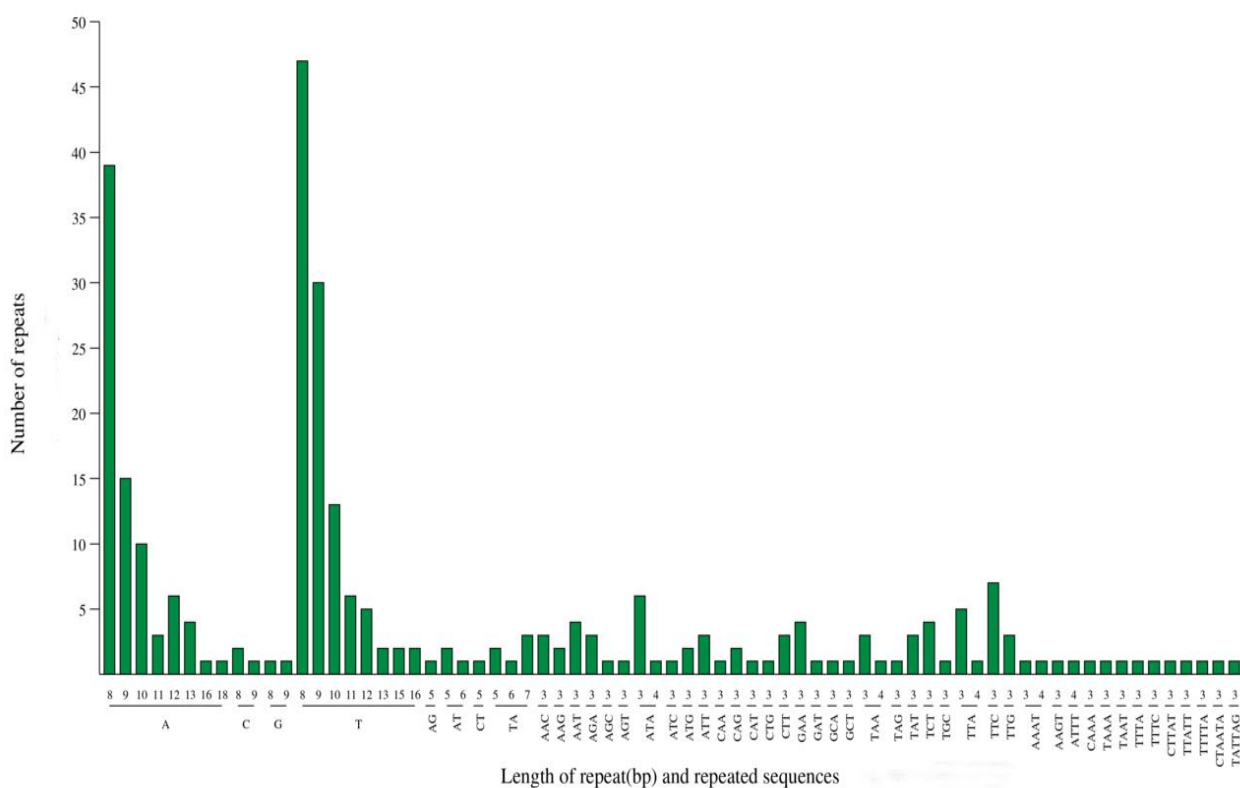


Figure 4 Quantitative statistics of various SSR types

Table 7 Result of SSR primer design

Gene ID	FPr1(5'-3')	RPr1(5'-3')	PSize	Start	End
1_1	AATCCACTGCCTTGATCCAC	GCATGAACGTAATGCTCACAA	259	88	346
1_2	GGACTCCCAAGCACATGAAT	TCCGACTAGTTCCGGGTTTC	280	1 375	1 654
1_3	TCATTAATATGGGTTGCCCG	GAAAAATGCAAGCACGGTTT	114	1 610	1 723
1_4	AGAACCTGTTCTTGCTCCGT	GATCGATTTCGTCGGATTTTG	234	2 025	2 258
1_5	GGAAAAATGGATTTCGTATTCACA	GTTCAAATCCCTTCGCTCCTG	190	2 767	2 956

1.6 KaKs experimental results

The ratio of non-synonymous mutation rate (Ka) to synonymous mutation rate (Ks) indicates the choice. The ratio is greater than 1, which indicates that the positive selection effect is less than 1, which indicates that there is purification selection effect. KaKs value is between 0.4 and 0.8, indicating purification selection (Table 8).

Table 8 Analysis and Statistics of chloroplast gene KaKs

Gene	Method	Ka/Ks	P-Value
rps12vs MF536699_rps12-2	MLWL	0.864801	0.565321
petD vs MF536699_petD	MLWL	0.6753	0.159274
rps12-2vs MF536699_rps12-2	MLWL	0.864801	0.565321
ycf1-2vs MF536699_ycf1-2	MLWL	0.456595	0.0830734

1.7 Analysis pi nucleic acid diversity

The higher pi values in the ssc and IR regions also mean that the region has a large degree of variation and a small value in the lsc interval, indicating a small degree of variation (Figure 5).

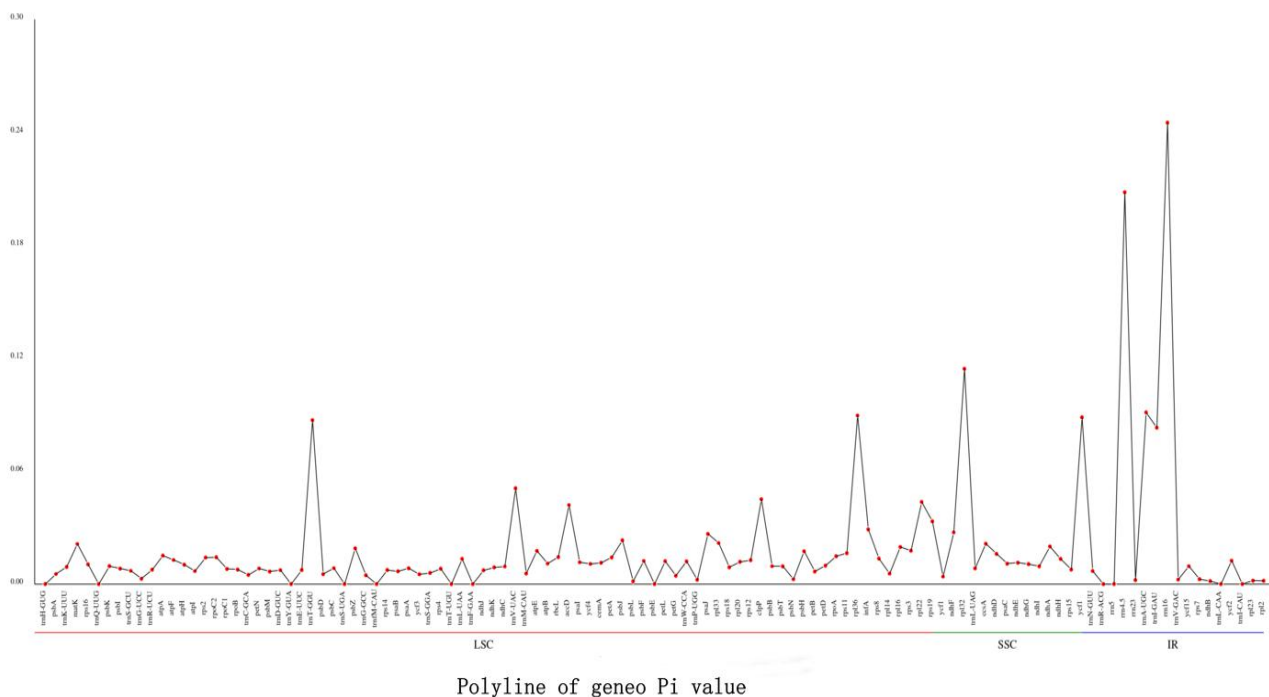


Figure 5 Polyline of geneo Pi value

1.8 IRscope test results

The chloroplast genome is a circular structure, IR has four boundaries with LSC and SSC, namely LSC-IRb, IRb-SSC, SSC-IRa and IRa-LSC. During the evolution of the genome, IR boundaries expand and shrink. Bringing certain genes into IR or single copy regions, Use the SVG module in the Perl to visualize the boundary information (Figure 6). There's a slight difference between several melon species, the length of the genome varies slightly, slightly different lengths between the four boundaries, only subtle differences, perhaps this subtle difference, which determines the different species of melon. The traits were different (Figure 6).

1.9 Comparison of chloroplast structures

Chloroplast structure is a ring-shaped structure. The results are similar by comparing the melon varieties of the proximal genus. Except for a few sites, the other approximations were higher (Figure 7).

1.10 Evolution tree clustering results

The results of cluster analysis showed that common melon, melon and snake melon were one kind, while Field Muskmelon was closely related to common melon, melon and snake melon, but far from bitter gourd, pumpkin and loofah (Figure 8).

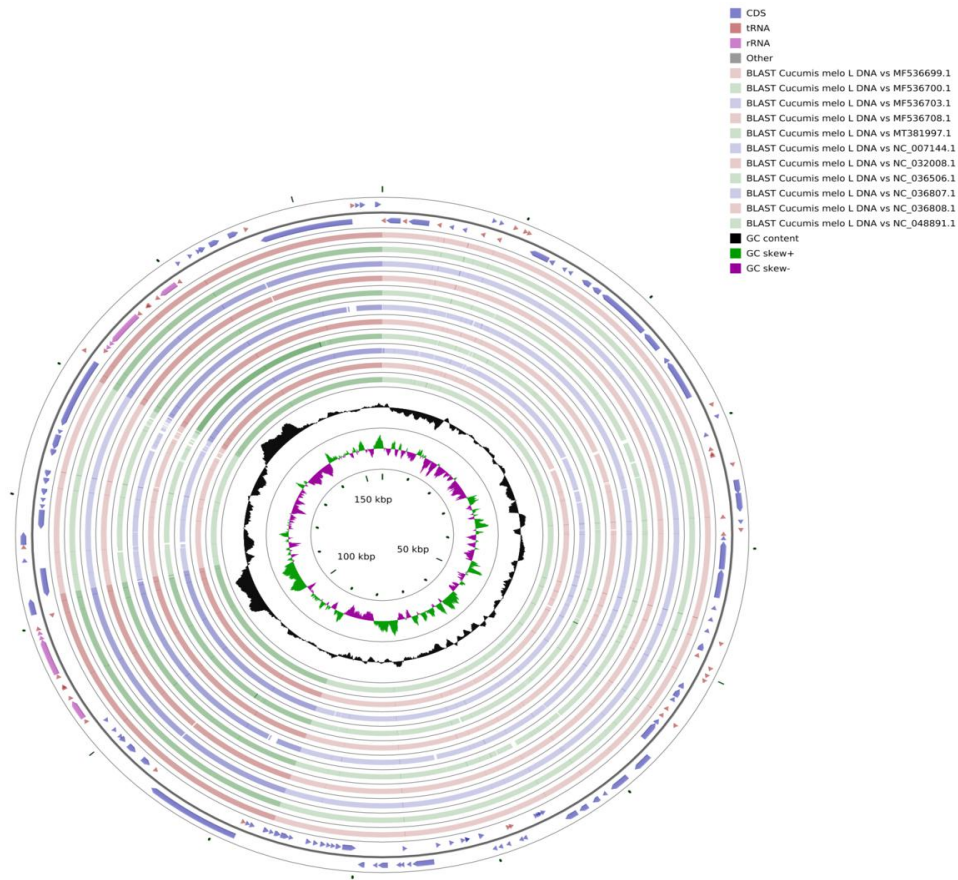


Figure 7 Comparison of Chloroplast Structure

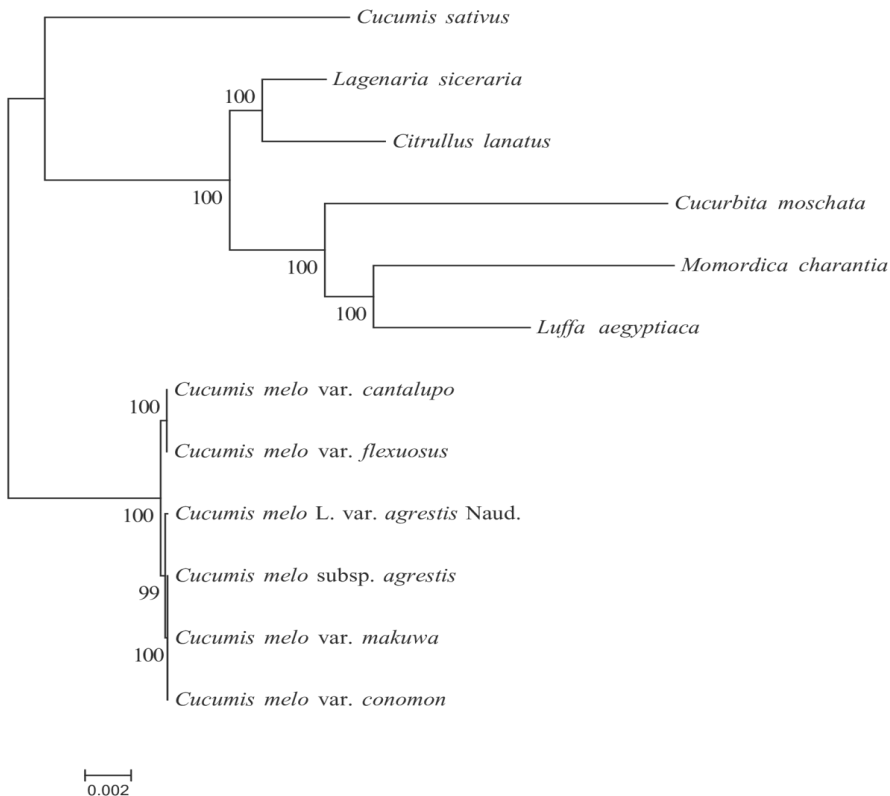


Figure 8 Evolutionary tree of chloroplast system

1.11 Homologous and collinearity analysis of chloroplast sequences

using the Mauve (<http://darlinglab.org/mauve>) software default parameters for genome alignment, sequence homology indicates that equine pomelo has homologous similarity with existing known melon, vietnamese melon, cucumber and other crops, except for a few loci, most of the sequences are approximate. Especially with Yue melon, common melon is more similar. far away from balsam pear and loofah (Figure 9).



Figure 9 Chloroplast sequence homology

2 Discussion and Conclusions

Referring to the genome sequencing of Field Muskmelon, we obtained the complete chloroplast sequencing and genome map of the plant. Conomon melon (156 017 bp) was found between the equinox and the same species, melon (*Cucumis melo* var. *Makuwa*) (156 016 bp), melon subspecies (*Cucumis melo* subsp. *Agrestis*) (156 016 bp) close. Homologous and collinear analysis of chloroplast sequences showed that equine melon was consistent with melon and common melon, but different from watermelon and pumpkin in tRNA, few genes. The results of cluster analysis showed that the Field Muskmelon was closest to the common melon, and the melon was closest to the common melon, snake melon and Yue melon. The number of introns in Field Muskmelon was 21, which was basically consistent with that in common melon (Wang, 2017). Among the cucurbitaceae plants, the Field Muskmelon and common melon have the same origin, which is consistent with the previous research that the Field Muskmelon is the wild ancestor of Chinese thin skin melon (Ma, 2020a). The chloroplast genome is relatively conserved and can better explain the evolution and kinship of species. and this study shows that the genetic and taxonomic results of the Field Muskmelon are consistent with those of previous studies (Ma, 2020a).

Melon has a GC content of 36.9. Cucumber is also 36.9%(Zhu et al., 2016), and the GC content of Field Muskmelon is 37.45%. and the LSC content is 55.34%; SSC 11.59%, The IRs content of 33.1% was basically the same as that of melon (Zhu et al., 2016).

3 Materials and Methods

3.1 Plant materials

Fresh leaves of the Field Muskmelon plant. The collection site is Fuyang Academy of Agricultural Sciences Science and Technology Park.

3.2 Bioinformatics analysis

The Clean Data were assembled according to the chloroplast genome sequence of the reference species, the results of chloroplast sequence assembly were obtained, the results of chloroplast sequence assembly were annotated with gene structure, the chloroplast genome map was made, and the basic contents such as chloroplast genome SSR were analyzed. KaKs analysis, phylogenetic analysis and other advanced analysis content. Use fastp (version 0.20.0, <https://github.com/OpenGene/fastp>) software to filter raw data. The filtering criteria are as follows:

- (1) Sequencing connectors and primer sequences in the Reads were removed
- (2) Filter out reads whose average mass is less than Q5
- (3) Filter out reads with N number greater than 5

Using OGDRAW (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) to make chloroplast genome maps. Because of the characteristics of codon biology, the codon usage rate of different species is very different. This inequality in the use of synonymous codons is called codon preference (Relative Synonymous Codon Usage, RSCU). This preference is considered to be a combination of natural selection, species mutation and genetic drift. The calculation methods are as follows: (the number of codon encoding one of the amino acids / the number of all codon encoding the amino acid) / (1/the type of codon encoding the amino acid), and (the actual frequency of codon / the theoretical frequency of the codon). Calculate using a self-written Perl script.

Scattered in repeated sequences using vmatch v2.3.0 (<http://www.vmatch.de/>) software combined with Perl scripts to identify repeat sequences. the parameters are set as: minimum length (minimum length)=30 bp, hamming distance (hamming distance), and the identification forms are four: forward (Forward), palindrome (Palindromic), reverse (Reverse), complementary (Complement).

SSR markers on the chloroplast genome are called cpSSR markers. Using MISA v1.0 software to carry out cpSSR analysis, the parameters 1-8(single base repeat 8 times and above) were 2-5,3-3,4-3,5-3,6-3.

Pi (nucleic acid diversity) can reveal the variation of nucleic acid sequences of different species, and regions with high variation can provide potential molecular markers for population genetics. The homologous gene sequences of different species were compared globally by mafft software (auto mode), and the pi values of each gene were calculated by dnap5.

Using software CGVIEW (http://stothard.afns.ualberta.ca/cgview_server/) default parameters, comparative analysis of chloroplast genome structure was performed for near-source species.

The whole genome is used for evolutionary tree analysis by default, the ring sequence is set the same starting point, and the interspecies sequence is compared by MAFFT software (v7.427,--auto mode). The compared data are trimmed with trimAl (v1.4.rev15), then the RAxML v8.2.10(<https://cme.h-its.org/exelixis/software.html>) software is used to select GTRGAMMA model, rapid Bootstrap analysis, bootstrap=1 The maximum likelihood evolution tree is constructed.

3.3 Genome assembly

To reduce the complexity of the next sequence assembly, bowtie2 (v2.2.4, <http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>) was used very-sensitive-local align the database and sequence. The assembly core module uses SPAdes [1]v3.10.1(<http://cab.spbu.ru/software/spades/>) software to

assemble chloroplast genome. Three aspects of genome quality control are performed to ensure the accuracy of assembly results:

- (1) reads specific genomes, statistics of genome coverage, size of inserted fragments, etc;
- (2) Genome alignment reference sequences to see collinear analysis of genome conservation and rearrangement.
- (3) Genome alignment of reference sequence structure information to compare the differences between the two.

A reference sequence *Cucumis_melo*: MT240857.1. is used for this project.

Author's contributions

Ma Zongxin is responsible for project design, data processing and gene sequencing, and completes the writing and revision of the paper. I have read and approved the final manuscript.

Acknowledgements

This research is supported by key R & D projects (1704 f0704067) of Anhui Science and Technology Department.

References

- Anton B., Sergey N., and Dmitry A., 2012, A new genome assembly algorithm and its applications to single-cell sequencing, *J Comput Biol.*,19(5):455-477
<https://doi.org/10.1089/cmb.2012.0021>
PMid:22506599 PMCID:PMC3342519
- Hong C., Wei P.K., Zhang M.M., and Hou D., 2020, The complete chloroplast genome of *Cucumis melo* L.'Shengkaihua'(Cucurbitaceae) and its physiognomic implications, *Mitochondrial DNA Part B*, 5(2):1253-1254
<https://doi.org/10.1080/23802359.2020.1731364>
- Katsunori T., Yukari A., Kenji F., Tatsuya Y., Yasheng A., Hidetaka N., Chun L.L., Hiromichi Y., YoIchiro S., and Kenji K., 2013, Diversification and genetic differentiation of cultivated melon inferred from sequence polymorphism in the chloroplast genome, *Breeding Science*, 63(2):183-196
<https://doi.org/10.1270/jsbbs.63.183>
PMid:23853513 PMCID:PMC3688380
- Ma Z.X., 2020a, Brief history of production of *Cucumis melo* L. var. *agrestis* Naud., *Nongcun Jingji Yu Keji* (Rural Economy and Science-Technology), 31(7):76-77
- Ma Z.X., 2020b, Investigation on new food resources of a new wild plant, *Nongye Yu Jishu* (Agriculture and Technology), 40(1):32-34
- Phan T.P.N., Yukari A., Tran T.M.H., Katsunori T., Yasheng A., Tatsuya Y., Hidetaka N., Long C.L., and Kenji K., 2010, Genetic diversity in Vietnamese melon landraces revealed by the analyses of morphological traits and nuclear and cytoplasmic molecular markers, *Breeding Science*, 60(3):255-266
<https://doi.org/10.1270/jsbbs.60.255>
- Wang H.G., Bai Y.L., Xu H.J., Liu S.H., Zhang W.J., Liu Q., Zhang X.M., Qiao M.Q., and Wang Y., 2017, Construction and stability of cucumber chloroplast expression vector, *Nankai Daxue Xuebao (Acta Scientificarum Naturalium Universitatis Nankaiensis)*, 50(5):61-66
- Zhu Q.L., Gao P., Liu S., Amanullah S., and Luan F.S., 2016, Comparative analysis of single nucleotide polymorphisms in the nuclear, chloroplast, and mitochondrial genomes in identification of phylogenetic association among seven melon (*Cucumis melo* L.) cultural., 66(5):711-719
<https://doi.org/10.1270/jsbbs.16066>
PMid:28163587 PMCID:PMC5282756