P02-021

Draft genome of medical plant

Cynanchum wilfordii, *C. auriculatum* and *Metaplexis japonica*

Sae Hyun Lee, Jiseok Kim, Eunbi Kim, Padmanaban Mohanan, Jee Young Park, Tae-Jin Yang*

Department of Plant Science, Plant Genomics and Breeding Institute, and Research Institute of Agriculture and Life Sciences,

College of Agriculture and Life Sciences, Seoul National University, Seoul, 08826, Republic of Korea

Background : *Cynanchum wilfordii* (Cw) and *C. auriculatum* (Ca) have been used as traditional medicine in Korea and China, respectively. *Metaplexis japonica* (Mj) also has been regarded as a nutritious tonic. In particular, Cw has recently been widely used as a raw material for health functional foods to alleviate menopausal symptoms. Cw, Ca and Mj are nearly allied species and have very similar phenotypes, but they show remarkable differences on traits in inter-species level, such as storage root development, size of aerial parts, and disease resistance. Although many studies on the pharmacological efficacy and metabolites were published, genome researches were limited to species authentication. Therefore, in this study, we tried to reveal the genome sequences of the three important medicinal plants.

Methods and Results : The hybrid assembly was conducted with Oxford Nanopore Sequencing Technology (ONT) long-read sequence, Illumina paired-end (550bp), and mate-pair sequencing (3k, 5k, 8k). Genome scaffolding was performed using the DNase Hi-C method, and high-quality draft genome sequences of three species were obtained. The genome size estimated through the k-mer analysis was about 250 Mbp, and the assembled sequences were about 180 Mbp. The BUSCO value and mapping rate of transcriptome read were both 99%. The Cw, Ca, Mj genome sequences had a quite high similarity. As a result of synteny analysis with *Coffea canephora*, a vividly conserved chromosome structure was found even though they were estimated to have a divergence event about 100 million years ago. Through the hidden paralogue sequence excavation with *Solanum lycopersicum* genome, these three apocynaceae species were predicted not to experience genome duplication events after gamma paleohexaploidization.

Conclusion : From the result of this research, high-quality draft genome sequences of three apocynaceae species were assembled, and they are expected to be useful for further genomic and metabolomic pathway study. This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government (MSIT) (No. 2020R1A2C3007885).

Materials

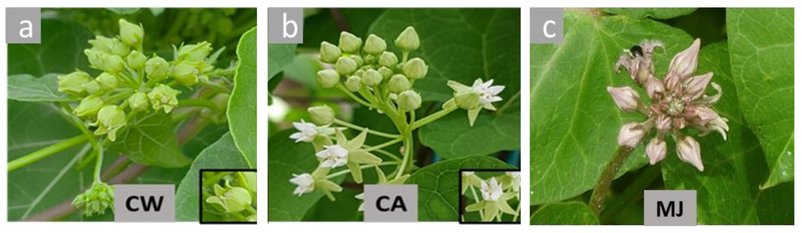


Figure 1. Three Apocynaceae species. a: *Cynanchum wilfordii*, b: *C. auriculatum*, c: *M. japonica*

Karyotyping

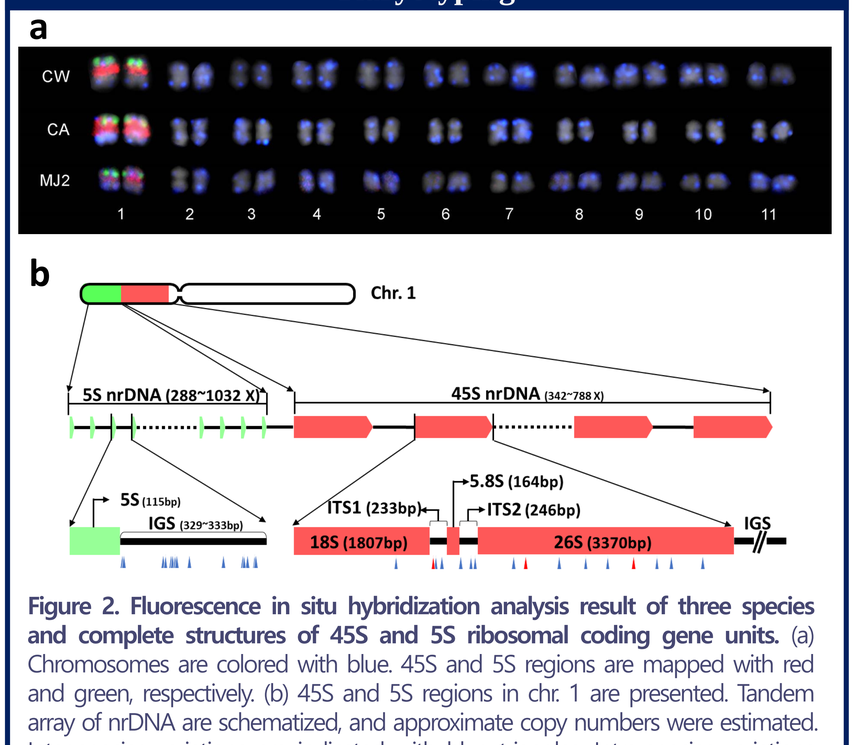


Figure 2. Fluorescence in situ hybridization analysis result of three species and complete structures of 45S and 5S ribosomal coding gene units. (a) Chromosomes are colored with blue. 45S and 5S regions are mapped with red and green, respectively. (b) 45S and 5S regions in chr. 1 are presented. Tandem array of nrDNA are schematized, and approximate copy numbers were estimated. Inter-species variations are indicated with blue triangles. Intra-species variations are marked with red triangles.

Genome size estimation (*k*-mer)

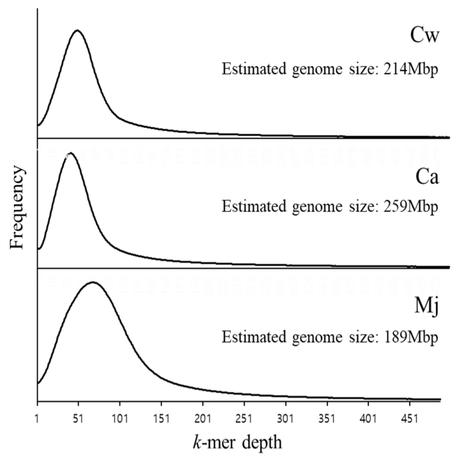
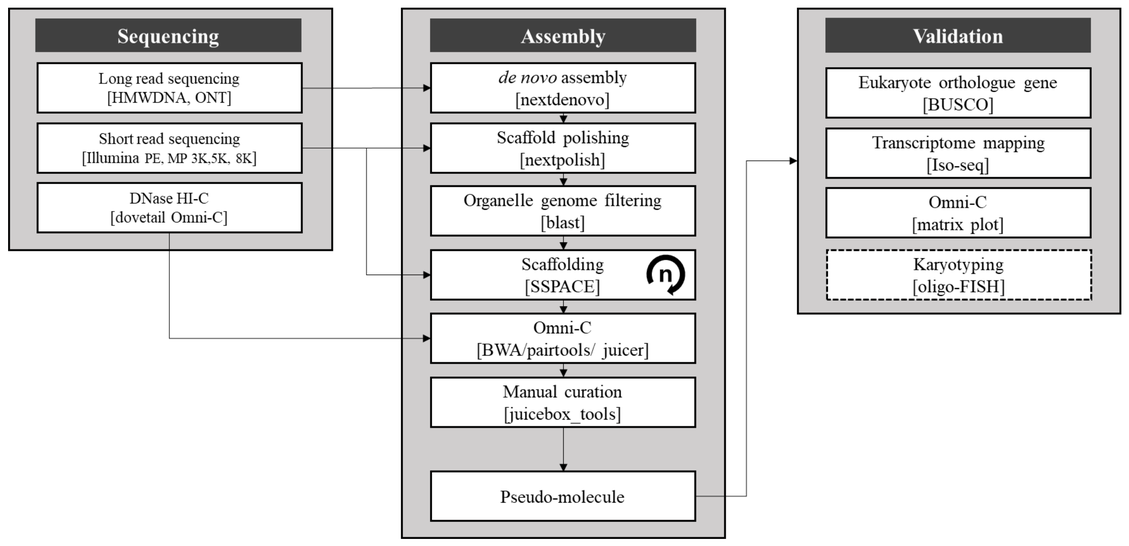


Figure 3. Result of *k*-mer analysis. *k*-mer analysis result showed clear single peak.

Strategy for genome assembly



Result of *denovo* assembly

Table 1. Genome assembly with ONT & scaffolding

	CW	CA	MJ
Sequence No.	107	124	333
Total base	183,187,033	183,561,278	184,285,085
Maximum size	8,353,936	9,269,288	3,785,250
Minimum size	61,095	33,054	78,893
Average base	1,712,028	1,480,332	553,408
N50	5,186,031	4,789,127	1,072,055

Table 2. Transcript mapping result and BUSCO value

		CW	CA	MJ
Transcript mapping	Total transcript	26,430	28,718	26,674
	Mapped transcript	26,385	28,233	25,758
	Mapping rate	99.83%	98.31%	96.57%
BUSCO value	Eukaryotic	99.30%	99.60%	96.50%
	Eudicot	96.00%	96.10%	93.70%

Genome scaffolding

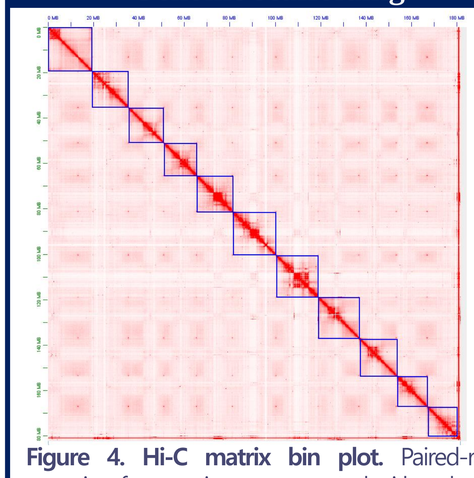


Figure 4. Hi-C matrix bin plot. Paired-read mapping frequencies are presented with red color. Chromosomes are marked with blue.

Information of draft genome seq.

Table 3. Length of each chr.

chr ID	Length (Kbp)
chr01	19,614
chr02	16,063
chr03	15,554
chr04	14,414
chr05	15,918
chr06	19,092
chr07	18,566
chr08	18,128
chr09	16,594
chr10	13,511
chr11	13,088

Table 4. Total length of CW genome

	Pseudo molecule	Chr0
Sequence No.	11	78
Total base	180,546,168	1,372,519
Maximum size	19,614,000	126,315
Minimum size	13,088,032	1,000
Average base	16,413,288	17,596
N50	16,594,229	27,243

Comparative genomic analysis of three apocynaceae species

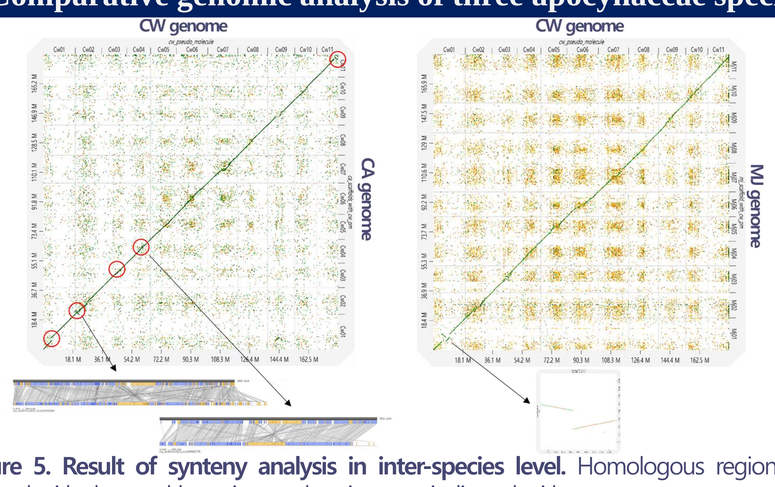


Figure 5. Result of synteny analysis in inter-species level. Homologous regions are marked with dots and large inverted regions are indicated with arrow.

Comparative analysis with diploid coffee genome (*Coffea canephora*)

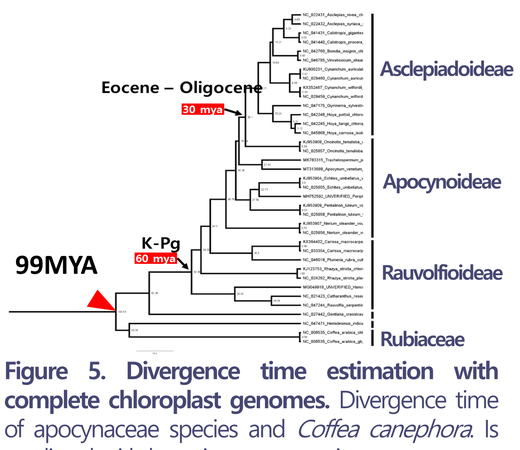


Figure 5. Divergence time estimation with complete chloroplast genomes. Divergence time of apocynaceae species and *Coffea canephora*. Is predicted with bayesian tree anaysis.

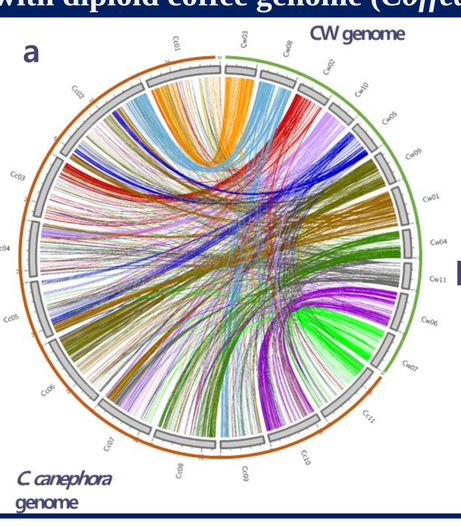


Figure 6. Conserved chromosome structures between *Coffea canephora* and Cw genomes. (a) Genome sequences of Cw and *Coffea canephora* are marked with green and brown colors, respectively. Syntenic regions are linked with colored lines. (b) High similar regions are indicated with signals.

Absent of the duplication event evidence in Cw genome

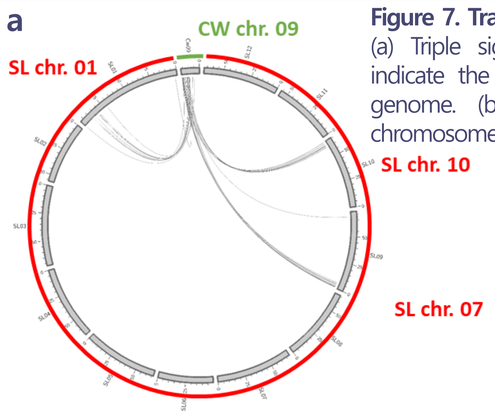


Figure 7. Traces of the Cw genome conservation. (a) Triple signals from *S. lycopersicum* genome indicate the absent of WGD, WGT event in CW genome. (b) Signals that splited into several chromosomes were labeled with colored lines.

History of the genome duplication events

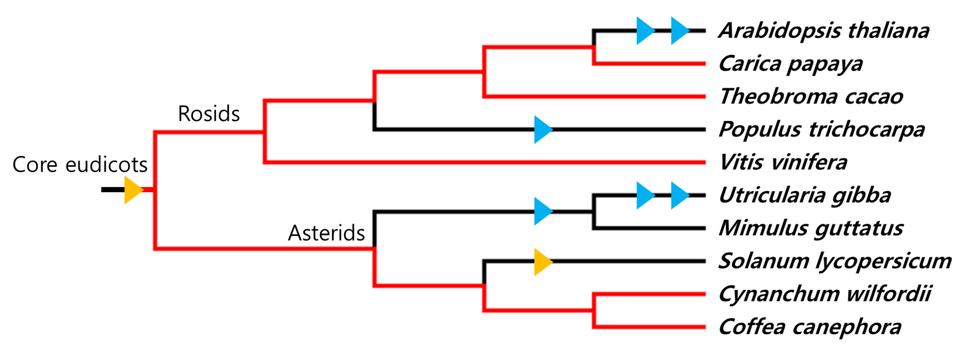


Figure 8. Schematic diagram of genome duplication history. Whole genome duplication and triplication events are labeled with blue and yellow triangles, respectively. The species without duplication events are marked with red color.

*Corresponding authors: tjyang@snu.ac.kr