Plastome data analysis of *Cucumis melo* subsp. *agrestis*

Mohammad Ajmal Ali[1,*,†], Khalid Mashay Al-Anazi[2,†], Waquar Akhtar Ansari[3], Joongku Lee[4,*]

[1]Department of Botany and Microbiology, College of Science, King Saud University, Riyadh-11451, Saudi Arabia; [2]Genetics Laboratory, Department of Zoology, College of Science, King Saud University, Riyadh-11451, Saudi Arabia; [3]Department of Botany, Banaras Hindu University, Varanasi-221005, India; [4]Department of Environment and Forest Resources, Chungnam National University, Daejeon 34134, Republic of Korea; *Corresponding authors; E-mail address: ajmalpdrc@gmail.com, alimohammad@ksu.edu.sa (M.A. Ali), joongku@cnu.ac.kr (J. Lee); †The first and second authors provided an equal contribution to this paper.

Author contacts: ajmalpdrc@gmail.com, alimohammad@ksu.edu.sa (Mohammad Ajmal Ali), Khalid Mashay Al-Anazi (kalanzi@ksu.edu.sa), Waquar Akhtar Ansari (waquar.ansari@gmail.com), Joongku Lee (joongku@cnu.ac.kr)

Declaration on Publication Ethics:
The author's state that they adhere with COPE guidelines on publishing ethics as described elsewhere at https://publicationethics.org/. The authors also undertake that they are not associated with any other third party (governmental or non-governmental agencies) linking with any form of unethical issues connecting to this publication. The authors also declare that they are not withholding any information that is misleading to the publisher in regard to this article.

Author responsibility:
The authors are responsible for the content of this article. The editorial and the publisher have taken reasonable steps to check the content of the article in accordance to publishing ethics with adequate peer reviews deposited at PUBLONS.

Declaration on official E-mail:
The corresponding author declares that official e-mail from their institution is not available for all authors

Abstract:
It is of interest to refine the taxonomic status of *C. melo* ssp. *agrestis* using its plastome data. The chloroplast size and GC% was found to be 1,56,016 bp and 36.92% respectively in *Cucumis melo* subsp. *agrestis*. The plastome of *C. melo* subsp. *agrestis* comprises of two inverted repeat (IR) regions of 25,797 bp each. It consisted of 133 genes with 88 protein-coding genes, 8 rRNA genes and 37 tRNA genes. Analysis of the *C. melo* ssp. *agrestis* plastome data will help breeders to improve the yield the crop.

Keywords: *Cucumis melo* ssp. *agrestis,* Cucurbitaceae, melon, plastome

Background:

The cucurbits are vegetable crops of the *Cucurbitaceae* family. This family carries total 98 genera and 1000 species. Muskmelon (*Cucumis melo* L.) is a member of the family Cucurbitaceae. The genus *Cucumis* possess large phenotypic diversity with *C. hystrix*, *C. callosus*, and *C. sativus* var. *hardwickii* [1]. A number of cucurbits including *Cucumis melo* subsp. *agrestis* are used in the indigenous system of medicine. Cucurbitacins in cucurbits possess renowned biological attributes [2]. Information on the reproductive biology of *Cucumis melo* subsp. *agrestis* is known [14]. Therefore, it is of interest to refine the taxonomic status of *C. melo* ssp. *agrestis* using next generation sequencing (NGS) plastome data for further application [3].

Figure 1: *Cucumis melo* subsp. *agrestis.* A) Habit in reproductive stage; B) Flowering; C) Fruit

Materials and Methods:

Plant material, DNA extraction, sequencing, assembly and annotation:

The fresh leaves sample of *Cucumis melo* subsp. *agrestis* (**Figure 1**) was collected from Nita Akaha, Bhagalpur, Bihar India, and fixed in 60-120 mesh size powder silica gel. The DNA extraction was performed using DNeasy Plant Mini Kit (QIAGEN) as per protocol. The *de novo* sequencing as a single end run of 51 bp was performed (# Illumina platform) at Macrogen, Republic of Korea. The Illumina Pipeline 1.3.2 was used. The FASTQC was used to filter the raw reads. The filtered high quality reads were assembled using SPAdes [4]. The assembled plastome data was annotated using GeSeq [5] as shown in **Figure 2**.
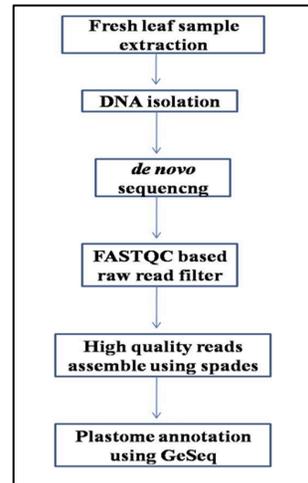

Figure 2: Methodology flowchart

Comparative analyses:

The plastome of a total number of 49 species of the family Cucurbitaceae was downloaded from **NCBI** (**Table 1**). The variation of plastome was analyzed. The percentage of GC and CDS number were plotted. The plastome of *C. melo* subsp. *agrestis* and *C. melo* were aligned with reference *C. sativus* using BRIG [6].

Table 1: The chloroplast genome of cucurbits retrieved from NCBI organellar genome for the comparative analyses.

S. No.	Taxon	GenBank
1	*Ampelosycios humblotii (Cogn.)* Jum. & H. Perrier	NC_046869.1
2	*Baijiania yunnanensis (A.M. Lu & Zhi Y. Zhang)* A.M. Lu & J.Q.Li	NC_046871.1
3	*Bryonia marmorata* E. Petit	NC_046865.1
4	*Cionosicys macranthus* Grayum & J.A. González	NC_046861.1
5	*Citrullus amarus* Schrad.	NC_035974.1
6	*Citrullus colocynthis (L.)* Schrad.	NC_035727.1
7	*Citrullus lanatus subsp. vulgaris (Schrad.)* Fursa	NC_032008.1
8	*Citrullus mucosospermus (Fursa)* Fursa	NC_033899.1
9	*Citrullus rehmii* De Winter	NC_035975.1
10	*Coccinia grandis (L..)* Voigt	NC_031834.1
11	*Corallocarpus boehmii (Cogn.)* C. Jeffrey	NC_046874.1
12	*Cucumis hystrix* Chakrav.	NC_023544.1
13	*Cucumis melo* L.	NC_015983.1
14	*Cucumis melo subsp. agrestis*	NC_048891.1
15	*Cucumis sativus* L.	NC_007144.1
16	*Cucurbita maxima* Duchesne	NC_036505.1
17	*Cucurbita moschata* Duchesne	NC_036506.1
18	*Cucurbita pepo* L.	NC_038229.1
19	*Cyclanthera pedata (L.)* Schrad.	NC_046860.1
20	*Cyclantheropsis parviflora (Cogn.)* Harms	NC_046870.1
21	*Gerrardanthus macrorhizus* Harv. ex Benth. & Hook.f.	NC_046873.1
22	*Gynostemma burmanicum* King ex Chakrav.	NC_036141.1
23	*Gynostemma cardiospermum* Cogn. ex Oliv.	NC_035959.1
24	*Gynostemma caulopterum* S.Z. He	NC_036135.1
25	*Gynostemma laxiflorum* C.Y. Wu & S.K. Chen	NC_036134.1
26	*Gynostemma longipes* C.Y. Wu	NC_036140.1
27	*Gynostemma pentagynum* Z.P. Wang	NC_036136.1
28	*Gynostemma pentaphyllum (Thunb.)* Makino	NC_029484.1
29	*Gynostemma pubescens (Gagnep.)* C.Y. Wu	NC_036142.1
30	*Hemsleya lijiangensis* A.M. Lu ex C.Y. Wu & Z.L. Chen	NC_039653.1
31	*Herpetospermum pedunculosum (Ser.)* C.B. Clarke	NC_046858.1
32	*Hodgsonia heteroclite (Roxb.)* Hook.f. & Thomson	NC_046857.1
33	*Hodgsonia macrocarpa (Blume)* Cogn.	NC_039628.1
34	*Indofevillea khasiana* Chatterjee	NC_046859.1

|		
Lagenaria siceraria (Molina) Standl.		NC_036808.1
Linnaeosicyos amara (L.) H. Schaef. & Kocyan		NC_046863.1
Momordica charantia L.		NC_036807.1
Momordica sessilifolia Cogn.		NC_046872.1
Nothoalsomitra suberosa (F.M. Bailey) I. Telford		NC_046876.1
Sechium edule (Jacq.) Sw.		NC_051498.1
Siraitia grosvenorii (Swingle) C. Jeffrey ex A.M. Lu & Zhi Y. Zhang		NC_043881.1
Trichosanthes baviensis Gagnep.		NC_046864.1
Trichosanthes homophylla Hayata		NC_046868.1
Trichosanthes kirilowii Maxim.		NC_041088.1
Trichosanthes lobata Roxb.		NC_046885.1
Trichosanthes Pilosa Lour.		NC_046884.1
Trichosanthes tricuspidata Lour.		NC_046866.1
Trichosanthes truncata C.B. Clarke		NC_046875.1
Trichosanthes tubiflora (Wight & Arn.) H.J. de Boer		NC_046867.1
Trichosanthes wallichiana (Ser.) Wight		NC_046882.1
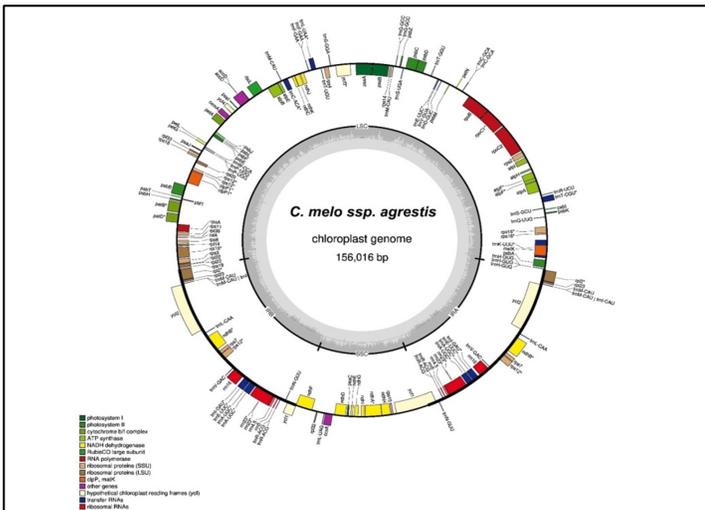
Figure 3: The plastome genome map of *Cucumis melo* subsp. *agrestis*.
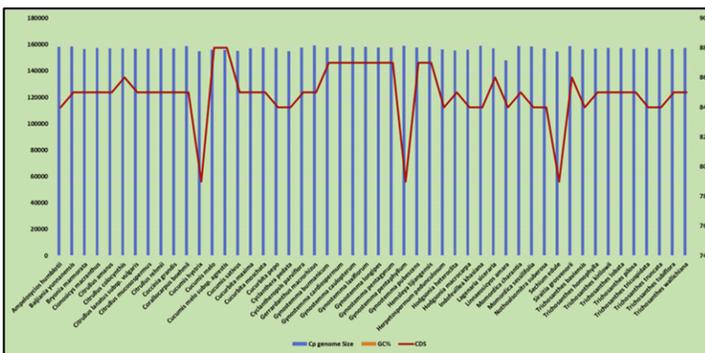

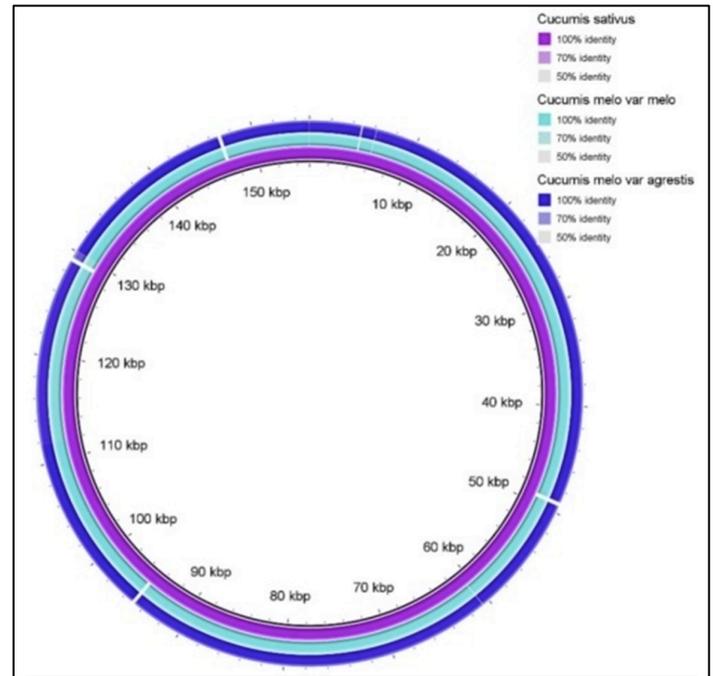**Figure 4:** Pattern of cp genome size, GC% and number of CDS across the cucurbits

Figure 5: The plastome of *Cucumis melo* subsp. *agrestis* and *Cucumis melo* aligned with reference *C. sativus*.

Results & Discussion:

The plastome of *C. melo* subsp. *agrestis* comprise (**Figure 3**) of 1,56,016 bp circular DNA carrying 25,797 bp each of two inverted repeat (IR) regions, divided by large single-copy (LSC) regions of 86,335 bp and small single-copy (SSC) regions of 18,088 bp, respectively. It contained a GC% of 36.92, a total number of 133 genes, including 88 protein-coding genes, 37 tRNA genes, and eight rRNA genes. This data is consistent with *Cucumis hystrix* (size 1,55,031 bp, GC% 36.98%, and 79 CDS), *Cucumis melo* (size 1,56,017 bp, GC% 36.92%, and 88 CDS), *Cucumis sativus* (size 1,55,293 bp, GC% 37.07%, and 85 CDS). The plastome size varies from 1,59,232 bp in *Gerrardanthus macrorhizus* to 1,47,874 bp in *Linnaeosicyos amara*, the differences of GC% was recorded in the range of 36 to 37%, *Ampelosycios humblotii* was recorded with minimum 79 CDS, while maximum 88 CDS was recorded in *Trichosanthes wallichiana* (**Figure 4**). The plastome of *C. melo* subsp. *agrestis* and *C. melo* aligned at reference *C. sativus* revealed 100% similarity (**Figure 5**). The *agrestis* group melons are monoecious; fruits, typically light-green, round, elliptic or oval, smooth surface with dark-greens spots and with numerous small seeds. The phylogenetic analyses suggested the *C. melo* comprise of various cultivar-groups [**1**]. The

genus *Cucumis* comprise of many economically significant crops, for example cucumber and melon with several important landraces [**7**]. The phylogenetic analyses of plastome data of *Cucumis* species also show proximity of *C. melo* subsp. *agrestis* with *C. melo* [**8**]. Variation at phenotypic level between wild and cultivated species is evident e.g. tomato [**9**], sunflower [**10**], rice [**11**], *Cucumis* [**12-13**], pepper [**9**], wheat [**10**], maize [**10**], attributed to genetic loci. The comparative transcriptomics showed the similarity at genetic level between the cultivated cucumber (*C. sativus*) and its wild relative (*C. hystrix*), the changes in transcription levels which may include alteration in stress tolerance to different abiotic stresses including salinity, heat, cold [**7**] resistance are continuous during the domestication process [**9-11**].

Conclusion:
We document the plastome data analysis
of *Cucumis melo* subsp. *agrestis* to glean insights on crop breeding.

Conflict of interest:
The authors declare that there are no competing interests.

References:

[1] Kocyan A *et al. Mol Phylogenet Evol.* 2007 **44:**553. [PMID: 17321763].

[2] Salehi B *et al*. *Molecules*. 2019 **24:**1854. [PMID: 31091784].
[3] Daniell H *et al. Genome Biol.* 2016 **17:**1. [PMID: 27339192].
[4] Bankevich A *et al. J Comput Biol*. 2012 **19:** 455. [PMID: 22506599].
[5] Tillich M, *et al, Nucleic Acids Res.* 2017 **45:** W6. [PMID: 28486635].
[6] Alikhan NF *et al. BMC genomics*. 2011 **12:**1. [PMID: 21824423].
[7] Abdel-Salam EM *et al*. *Plants*. 2020 **9:**63. [PMID: 31947725].
[8] Perl-Treves R *et al. Theor Appl Genet.*1985 **71:**430. [PMID: 24247448].
[9] Paran I & Van Der Knaap E. *J. Exp. Bot.* 2007 **58:**3841. [PMID: 18037678].
[10] Doebley JF *et al*. *Cell*. 2006 **127(7):**1309. [PMID: 17190597].
[11] Izawa T *et al*. 2009 **12(2):**185. [PMID: 19185529].
[12] Perl-Treves R & Galun E. *Theor Appl Genet*. 1985 **71:**417. [PMID: 24247447].
[13] Perl-Treves R *et al. Theor Appl Genet.*1985 **71:**430. [PMID: 24247448].
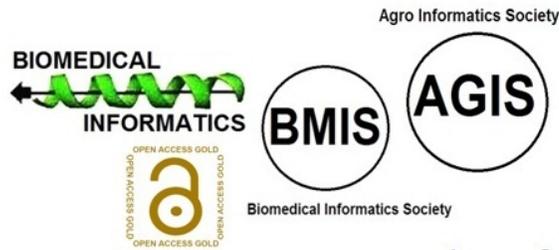[14] Kouonon LC e*t al. Ann Bot.* 2009 **104:**1129. [PMID: 19671577]

Articles published in BIOINFORMATION are open for relevant post publication comments and criticisms, which will be published immediately linking to the original article for FREE of cost without open access charges. Comments should be concise, coherent and critical in less than 1000 words.

BIOMEDICAL INFORMATICS

BMIS
Biomedical Informatics Society

Agro Informatics Society
AGIS

since 2005

BIOINFORMATION
Discovery at the interface of physical and biological sciences

indexed in

PMC

Pub Med

EBSCO

INDEXED IN
EMERGING
SOURCES
CITATION
(Web of Science)
CLARIVATE ANALYTICS

WEB OF SCIENCE

Web of
Science
Group

doi®

Crossref

ResearchGate

R G

publons