

Baraminology of Cucurbitaceae Based on Chloroplast Genome Analysis

Matthew Cserhati, Concordia University, Irvine, California.

Abstract

The gourd family (Cucurbitaceae) is an economically important and diverse group of plants, with approximately 1000 species. Previous analysis of hybridization data and nuclear ITS regions in 70 species of 14 genera of the family Cucurbitaceae identified eight putative monobaramins. The present study expands upon this previous analysis by analyzing the sequence similarity of the chloroplast genome between 30 species from Cucurbitaceae. The clustering analysis showed six putative holobaramins: *Gynostemma*, *Gomphogyne+Hemsleya*, *Cucurbita*, *Cucumis*, *Citrullus*, *Coccinia grandis+Hodgsonia macrocarpa + Lagenaria siceraria + Momordica charantia + Siraitia grosvenorii + Trichosanthes kirilowii*. All six groups had a significant p-value except *Gomphogyne+Hemsleya*. Of these, *Gynostemma* and *Gomphogyne + Hemsleya* are two new putative holobaramins. The present study is the second of its kind, after an analysis of chloroplast genomes in Liliales. This study used chloroplast genome sequences supported by previous hybridization results. Future chloroplast genome studies could be performed on plant groups that have enough species with mitochondrial genomes.

Keywords: Gourd, chloroplast, genome, molecular baraminology, *Gynostemma*, *Gomphogyne*, *Hemsleya*

Introduction

Baraminology is the study of the created kinds, mentioned in the Genesis creation account. A kind, or holobaramin, is a reproductive community consisting of species that are continuous with one another and discontinuous with all other species. The goal of a baraminology study is to use various types of data (the Bible, hybridization data, DNA sequences, morphological, ecological, and behavioral data) to classify species into their appropriate kinds.

The gourd family (Cucurbitaceae) is an agriculturally important group of food crops that include cucumbers, gourds, squashes, melons, pumpkins, zucchinis, and luffas. The secondary metabolites found in gourds called glycoalkaloids are used in the pharmaceutical industry. The family is quite diverse and can be classified into 15 tribes, 95 genera, and around 1000 species (Chomicki, Schaefer, and Renner 2020).

Hybridization data exists for 131 unique species pairs of cucurbits, including data from experimental cross-pollination events, natural hybridization, and embryo/ovule cultures. Previous baraminology research based on hybridization found eight monobaramins: *Cucurbita*, *Citrullus*, *Cucumis*, *Lagenaria*, *Coccinia + Diplocyclos*, *Trichosanthes + Momordica*, *Luffa*, and *Bryonia* (Brophy, Gregory, and Townsend 2023). This study was further augmented by the sequence analysis of ITS regions of nuclear ribosomal genes for 21 cucurbits taken from a study by Jobst, King, and Hemleben (1998). The ITS region is a segment of the DNA situated between the small-subunit rRNA and the large-subunit rRNA and is used widely in phylogenetic studies.

However, the two ITS regions combined amount to approximately 450bp. Besides the ITS region, the whole chloroplast genome may be used to get more precise baraminic relationships, as it is much longer, approximately 155–160Kbp. Its sequence is also very conserved, easy to isolate, abundant in cells, and has low mutation rates (Kousar and Park 2023). More closely related species are more likely to have similar genome sizes and characteristics. Therefore, it makes sense to put two plant species into the same baramin based on a high chloroplast genome sequence similarity (Xiao-Ming et al. 2017).

This study would be the second of its kind, after the analysis of the chloroplast genomes of 163 species from the order Liliales (Cserhati 2023). The results from the present study can be compared with hybridization data from Brophy, Gregory, and Townsend (2023) and the previous results from the analysis of Liliales.

Materials and Methods

The chloroplast genome of 28 species from Cucurbitaceae as well as three species from the genus *Fagus* and three from the genus *Lycium*, as outgroups, were downloaded from the Chloroplast Genome Database (CpGDB) at <https://www.gndu.ac.in/CpGDB/index.aspx> (Singh et al. 2020). Two chloroplast genomes for *Gomphogyne cissiformis* var. *cissiformis* and var. *villosa* were downloaded from the NCBI database at <https://www.ncbi.nlm.nih.gov/nucleotide>. These 36 sequences were then aligned using the MAFFT software (Katoh, Rozewicki, and Yamada 2019; Kuraku et al. 2013) using the --retree 2 and --reorder parameters.

The ‘dist.alignment’ command from the ‘seqinr’ package in R was used to create the sequence identity matrix using the ‘identity’ flag parameter. The dist.alignment command uses Fitch matrixes, which contain the square-root of the pairwise distances (Fitch 1966). The ‘heatmap’ function in R was used to create the heatmap of the 34 species seen in fig. 1. Species clustering was performed using the ‘ward.D’ method, which implements Ward’s agglomerative clustering method (Ward 1963).

The ‘upgma’ method was used to create a UPGMA-based baraminic tree for fig. 2. Bootstrapping was done using the ‘boot.phylo’ method in R for 100 bootstrap replicates. The ward.D2 clustering method was used. See Cserhati (2023) for a more detailed description of methods utilized in this study.

Version 4.3.1 of R was used. The New King James Version of the Bible was used in the biblical analysis. All supplementary files and figures are available on Zenodo at <https://zenodo.org/records/10372288>.

Biblical Analysis

Mentions of cucurbits in the Bible are scant. Isaiah 1:8 mentions cucumbers: “So the daughter of Zion is left as a booth in a vineyard, as a hut in a garden of cucumbers, as a besieged city.” 2 Kings 4:39 mentions gourds: “So one went out into the field to gather herbs, and found a wild vine, and gathered from it a lapful of wild gourds, and came and sliced them into the pot of stew, though they did not know what they were.” Jonah 4:6–10 mentions a plant called קִיקָיוֹן (*kikayon*), which some allege to be either a gourd, or the castor-oil plant, but which is disputed. The word for gourd in 2 Kings 4:39 is פַּקוּחָה (*paquah*), which is different from the word in Jonah 4.

Lastly, Numbers 11:5 mentions two members of the family Cucurbitaceae: “We remember the fish which we ate freely in Egypt, the cucumbers, the melons, the leeks, the onions, and the garlic.” Since cucumbers and melons are listed separately, this might be an indication that they possibly belong

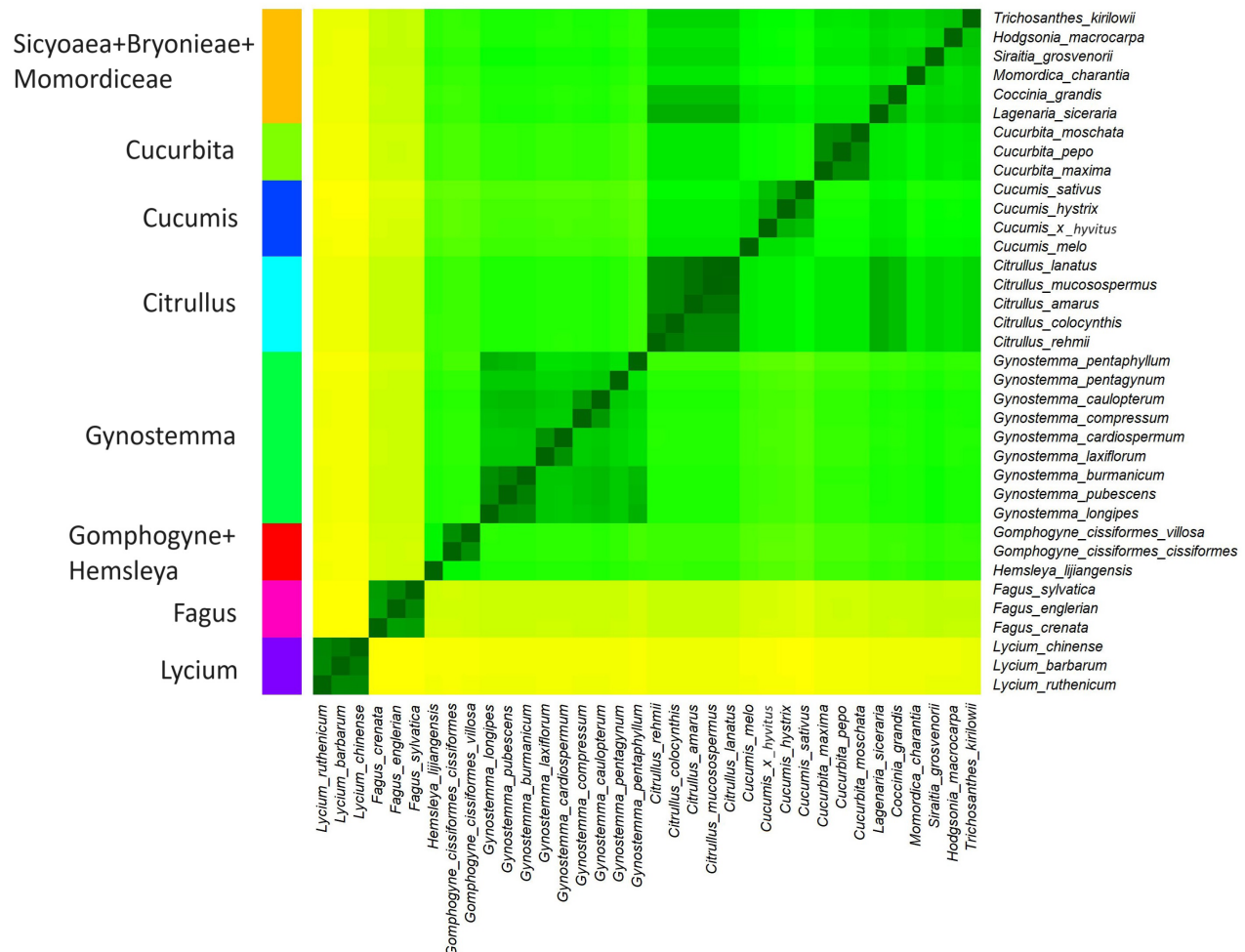


Fig. 1. Baraminic heatmap of 30 species from Cucurbitaceae, three species of *Fagus*, and three species of *Lycium*. Darker, greener colors represent species with high chloroplast genome sequence similarity. These species likely belong to the same monobaramin. Lighter, yellow colors represent species with lower sequence similarity, and which may belong to different baramins. Eight putative monobaramins were found: orange: Sicyoaea + Bryonieae + Momordiceae, light green: *Cucurbita*, blue: *Cucumis*, cyan: *Citrullus*, green: *Gynostemma*, red: *Gomphogyne*+*Hemsleya*, magenta: *Fagus*, purple: *Lycium*.

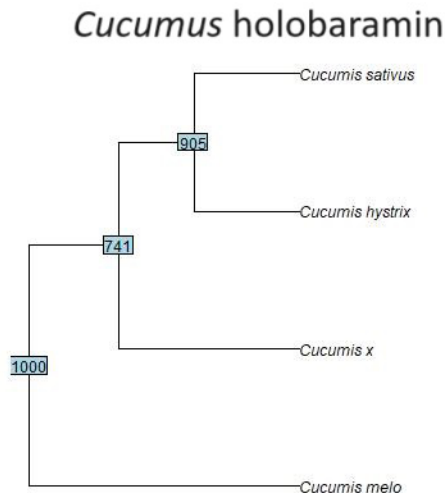


Fig. 2. Baraminic tree of the four *Cucumis* species analyzed in this study. *Cucumis melo* may be an earlier form in this monobaramin than the other species. The tree was constructed using the WPGMA algorithm.

to separate baramins, but this is not decisive. Cucumbers and many melons belong to the genus *Cucumis*. However, ‘melon’ is a common term and may cover other species outside *Cucumis*. Since the genus is generally held to be the lower bound of the baramin, some researchers believe that cucumbers and many of the melons are members of the same baramin (Brophy et al. 2023).

Results and Discussion

The 36 species of Cucurbitaceae, *Fagus*, and *Lycium* were clustered into eight groups, all but one with significant p-values (fig. 1). The elbow plot in Supplementary fig. 1 (available on Zenodo at <https://zenodo.org/records/10372288>) appears to show a drastic drop at seven clusters. However, the silhouette plot is inconsistent and appears to show that the optimal number of clusters is two (Supplementary fig. 2). The Hopkins statistic has a value of 0.881 which indicates very good clustering. The two outgroups, *Fagus* and *Lycium* form groups distinct from the cucurbit species. The sequence identity matrix, the putative clusters, and the descriptive statistics for each of the eight putative groups found in this study are available in Supplementary File 1.

The first cluster is made up of nine species, from the genus *Gynostemma*. The mean sequence similarity within this group is 0.829. This group was not analyzed by Brophy, Gregory, and Townsend (2023), so they may indeed be a separate holobaramin within Cucurbitaceae.

Hemsleya lijiangensis is very similar to *H. zhejiangensis*, which both cluster closely to the genus *Gynostemma* in a chloroplast genome tree in Li et al. (2019). However, in another chloroplast study including two subspecies of *Gomphogyne cissiformis*

(var. *cissiformis* and var. *villosa*), *H. lijiangensis* clustered together with the two *Gomphogyne* species (Zhang et al. 2018). The question then is, does *Hemsleya* + *Gomphogyne* form a monobaramin within *Gynostemma*; or are they their own holobaramin?

Based on the heatmap, *Hemsleya* + *Gomphogyne* form a separate small putative holobaramin, with a mean sequence similarity of 0.85. The tribe Gomphogyneae consists of mainly dioecious perennial lianas that grow in tree canopies but also annual creepers (Schafer and Renner 2011). However, the p-value for this group is 0.107, so further analysis is needed to confirm the baraminic status of these two genera.

Three putative holobaramins were also discovered corresponding to the genera *Cucurbita* (*Cucurbita pepo*, *Cucurbita moschata*, and *Cucurbita maxima*), *Cucumis* (*Cucumis x hytivus*, *Cucumis hystrix*, *Cucumis melo*, and *Cucumis sativus*), and *Citrullus* (*Citrullus rehmii*, *Citrullus amarus*, *Citrullus colocynthis*, *Citrullus mucosospermus*, *Citrullus lanatus*). These three groups were also identified via hybridization data by Brophy et al. (2023).

Zhou et al. (2023) found that the largest chloroplast genome sequence divergence calculated for seven species of *Citrullus* was 0.00641 (between *C. naudinianus* and *C. lanatus* subsp. *vulgaris*). The chloroplast genome size for these seven species is 157,005±102bp, suggests that the genomes of the species in this baramin have remained fairly stable throughout their history.

The four species of *Cucumis* were also found to form a holobaramin, the baraminic tree of which can be seen in fig. 2. The genome size ranges from 154,673 to 155,670bp in several species of *Cucumis* studied by Zhai et al. (2021), which is less than 1Kbp. Exons and inverted repeat (IR) regions were also found to be conserved in the chloroplast genome of *Cucumis x hytivus* (Zhai et al. 2021). It may be that *C. melo* is an earlier form in the putative *Cucumis* holobaramin. The mean chloroplast genome sequence similarity value between *C. melo* and *C. x hytivus*, *C. hystrix*, and *C. sativus* is only 82.8±0.8%, whereas among the latter three *Cucumis* species it is 89.7±2.9%. Analyses of chromosomal numbers and ITS regions indicate that *C. melo* (2n=24) is ancestral to other species (Sebastian et al., 2010; Garcia-Mas, Monforte, and Arús 2004), (which have a chromosome configuration of 2n=14). Therefore *C. melo* seems to be an earlier form of the *Cucumis*.

Interestingly, although 66 members of the genus *Cucumis* have a monoecious (hermaphrodite) mating system; and annual lifestyle, whereas other species have a dioecious mating system (male and female reproductive organs on separate plants) and

perennial habit (Sebastian et al. 2010). Apparently, there may be a more or less simple genetic switch that lets species alternate between monoecious and dioecious mating systems within a baramin. Similarly, Wood and Cavanaugh (2001) found that various species of the subtribe Flaveriinae use the C_3 , C_3 - C_4 and C_4 photosynthetic pathways. Both monoecious and dioecious mating systems and the C_3 , C_3 - C_4 and C_4 photosynthetic pathways appear to be specially designed morphological and molecular structures that allow plants to plastically adapt to new environments.

The last group is made up of six species from various genera: *Coccinia grandis*, *Hodgsonia macrocarpa*, *Lagenaria siceraria*, *Momordica charantia*, *Siraitia grosvenorii*, and *Trichosanthes kirilowii*. This group has a mean sequence similarity of 0.847, which is the lowest of all eight groups, but has a p-value of 5.59E-20. Brophy, Gregory, and Townsend (2023) separate *Lagenaria* and *Coccinia* + *Diplocyclos* into their own groups, although both groups are represented by only two species in their study. However, Brophy, Gregory, and Townsend (2023) place *Trichosanthes*, *Luffa*, *Echinopepon*, *Cyclanthera*, *Sicyos*, *Bryonia*, *Ecballium*, and *Momordica* into one single putative monobaramin. They did this based on the genetic distances between 21 cucurbit species using the ITS region analyzed by Jobst, King, and Hemleben (1998).

For example, based on Brophy, Gregory, and Townsend's (2023) hybridization data, *Momordica charantia* and *Trichosanthes anguina* hybridize directly and *M. charantia* hybridizes with the same third species as *T. cucumerina* and *T. lobata*, indicating that they all belong to the same holobaramin. The genetic distance between *M. charantia* and *T. cucumerina* based on ITS regions is 0.149 (Jobst, King, and Hemleben 1998; see also Supplementary File 2). By inference, any species pair with distances less than 0.142 may possibly also be able to hybridize. This means that the 20 species in the previously mentioned eight genera might be united into one single holobaramin. This would also support the six species in the last group being assigned to their own holobaramin.

The chloroplast genome map for 26 cucurbits and the outlier species *Lycium ruthenicum* can be seen in fig. 3. The gene order appears largely conserved between all 26 species, with only minor differences. This is evidence that all cucurbits could be united into a single holobaramin (Brophy, Gregory, and Townsend 2023). However, this is not decisive. The gene order in the mitochondrial genome is also almost the same among the primates, yet we know that humans belong to a separate baramin compared to apes. The chloroplast gene order also

seems to be conserved in non-cucurbit plant species as well, such as *Nicotiana*, *Arabidopsis*, *Panax*, and *Oenothera* (Plader et al. 2007). Each element in the chloroplast genome on the genome map is listed in Supplementary File 3.

Summary and Conclusion

This is the second chloroplast genome-based baraminology study performed to date after the one on Liliales (Cserhati 2023). The present study benefited from the extra hybridization data analysis done by Brophy, Gregory, and Townsend (2023). The present study confirmed the holobaraminic status of *Cucurbita*, *Cucumis*, *Citrullus*, and *Coccinia grandis*, *Hodgsonia macrocarpa*, *Lagenaria siceraria*, *Momordica charantia*, *Siraitia grosvenorii*, and *Trichosanthes kirilowii*. Two further putative holobaramins were predicted, namely *Gynostemma* and *Hemsleya* + *Gomphogyne*, although the latter needs further analysis.

Brophy, Gregory, and Townsend (2023) consider Cucurbitaceae to be a single holobaramin. Their finding is based on the inclusion of species considered to possibly be in the same baramin based only on genetic similarity of hybridizing species. This genetic distance was determined by the sequence similarity of the ITS region. However, according to Tippery and Les (2008), genetic distances mean the reliability of such data deteriorates above the genus level. As such, the holobaraminic status of Cucurbitaceae is only tentative. While this may be a valid assessment on the part of Brophy, Gregory, and Townsend, the Silhouette value analysis shows that the maximum Silhouette value is 0.6 for eight clusters as opposed to 0.52 for four clusters (Cucurbitaceae, *Gynostemma*, *Fagus*, *Lycium*). The Silhouette values can be seen in Supplementary File 2. This consideration indicates that eight holobaramins may be a better division of the data as opposed to four.

This study demonstrated how chloroplast genomes can be used to classify species into holobaramins. Since chloroplast genomes are very easy to determine, many more chloroplast genome-based baraminology studies can be carried out in the future. Results from mitochondrial and chloroplast genome studies can be compared as supportive evidence for baraminic classification alongside hybridization and other datasets.

This might prove to be more difficult than may be anticipated. This is because mitochondrial genomes range from 200 to 2,400Kbp in size (rivaling the genome size of some bacteria) in plants (Ward, Anderson, and Bendich 1981). Some of these plant mitochondrial genomes are also multichromosomal. For example, the mitochondrial genome size of *Citrullus lunatus* (watermelon) is 379Kbp, that of

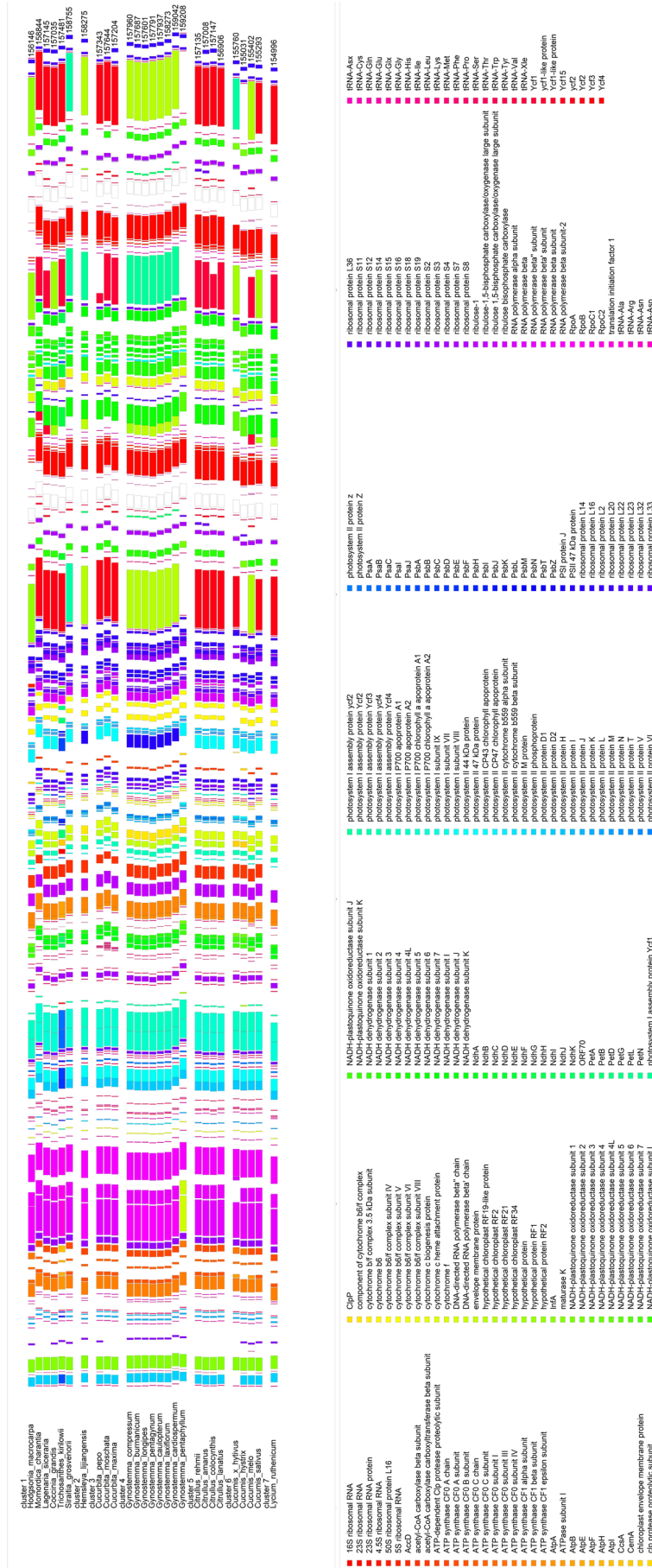


Fig. 3. Chloroplast genome map of several species of the six putative Cucurbitaceae holobaramins and one species from *Lycium* as a reference.

Cucurbita pepo (field pumpkin) is 983Kbp, and that of *Cucumis melo* (musk melon) is 2,740Kbp, the longest mitochondrial genome in existence (Rodríguez-Moreno et al. 2011). The determination of baraminic status of species in the cucumber family may be complicated due to different inheritance patterns between the nucleus, the chloroplast and the mitochondrion. In *Cucumis* species, for example, mitochondrial DNA is paternally inherited, whereas chloroplast DNA is maternally inherited (Shen et al. 2015; Zhai et al. 2021).

References

- Brophy, Timothy R., Jack R. Gregory, and Brigitte Townsend. 2023. "Hybridization and Genetic Distances Suggest One Large Monobaramin in the Gourd Family (Cucurbitales: Cucurbitaceae)." In *Proceedings of the Ninth International Conference on Creationism*. Edited by J. H. Whitmore, 657–658. Cedarville, Ohio: Cedarville University International Conference on Creationism.
- Chomicki, Guillaume, Hanno Schaefer, and Susanne S. Renner. 2020. "Origin and Domestication of Cucurbitaceae Crops: Insights from Phylogenies, Genomics and Archaeology." *The New Phytologist* 226, no. 5 (June): 1240–1255.
- Cserhati, Matthew. 2023. "Chloroplast Genome-based Baraminology Study of Liliales." *Creation Research Society Quarterly* 60: 84–96.
- Fitch, Walter M. 1966. "An Improved Method of Testing for Evolutionary Homology." *Journal of Molecular Biology* 16, no. 1 (March): 9–16.
- García-Mas, J., A. J. Monforte, and P. Arús. 2004. "Phylogenetic Relationships Among *Cucumis* Species Based on the Ribosomal Internal Transcribed Spacer Sequence and Microsatellite Markers." *Plant Systematics and Evolution* 248 (20 July): 191–203.
- Jobst, Jürgen, Klaus King, and Vera Hemleben. 1998. "Molecular Evolution of the Internal Transcribed Spacers (ITS1 and ITS2) and Phylogenetic Relationships among Species of the Family Cucurbitaceae." *Molecular Phylogenetics and Evolution* 9: 204–219.
- Katoh, Kazutaka, John Rozewicki, and Kazunori D. Yamada. 2019. "MAFFT Online Service: Multiple Sequence Alignment, Interactive Sequence Choice and Visualization." *Briefings in Bioinformatics* 20, no. 4 (July 19): 1160–1166.
- Kousar, Muniba, and Joonho Park. 2023. "Comparative Analysis of the Chloroplast Genome of *Sicyos angulatus* with Other Seven Species of Cucurbitaceae Family." *Genes* 14, no. 9 (8 September): 1776.
- Kuraku, Shigehiro, Christian M. Zmasek, Osamu Nishimura, and Kazutaka Katoh. 2013. "aLeaves Facilitates On-Demand Exploration of Metazoan Gene Family Trees on MAFFT Sequence Alignment Server With Enhanced Interactivity." *Nucleic Acids Research* 41, no. W1 (1 July): W22–W28.
- Li, Yanrong, Jinghui Li, Lingli Fang, and Ming Jiang. 2019. "Characterization of the Complete Chloroplast Genome Sequence of *Hemsleya zhejiangensis* (Cucurbitaceae), a Rare and Endangered Wild Plant Species in Zhejiang Province, China." *Mitochondrial DNA, Part B, Resources* 4, no. 2 (November): 4098–4099.
- Plader, Wojciech, Yasushi Yukawa, Mashiro Sugiura, and Stefan Malepszy. 2007. "The Complete Structure of the Cucumber (*Cucumis sativus* L.) Chloroplast Genome: Its Composition and Comparative Analysis." *Cellular and Molecular Biology Letters* 12, no. 4 (July 3): 584–594.
- Rodríguez-Moreno, Luis, Víctor M. González, Andrej Benjak, M. Carmen Martí, Pere Puigdomènech, Migueal A. Aranda, and Jordi Garcia-Mas. 2011. "Determination of the Melon Chloroplast and Mitochondrial Genome Sequences Reveals that the Largest Reported Mitochondrial Genome in Plants Contains a Significant Amount of DNA Having a Nuclear Origin." *BMC Genomics* 12 (August 20): 424.
- Schaefer, Hanno, and Susanne S. Renner. 2011. "Phylogenetic Relationships in the Order Cucurbitales and a New Classification of the Gourd Family (Cucurbitaceae)." *Taxon* 60, no. 1 (1 February): 122–138.
- Sebastian Patrizia, Hanno Schaefer, Ian R. H. Telford, and Susanne S. Renner. 2010. "Cucumber (*Cucumis sativus*) and melon (*C. melo*) have Numerous Wild Relatives in Asia and Australia, and the Sister Species of Melon is From Australia." *Proceedings of the National Academy of Sciences USA* 107, no. 32 (August 10): 14269–14273.
- Shen Jia, Juan Zhao, Grzegorz Bartoszewski, Stefan Malepszy, Michael Havey, and Jinfeng Chen. 2015. "Persistence and Protection of Mitochondrial DNA in the Generative Cell of Cucumber is Consistent with its Paternal Transmission." *Plant and Cell Physiology* 56, no. 11 (November): 2271–2282.
- Singh, Bhupinder Pal, Ajay Kumar, Harpreet Kaur, Harpreet Singh, and Avinash Kaur Nagpal. 2020. "CpGDB: A Comprehensive Database of Chloroplast Genomes." *Bioinformatics* 16, no. 2 (February 29): 171–175.
- Tippery, Nicholas P., and Donald H. Les. 2008. "Phylogenetic Analysis of the Internal Transcribed Spacer (ITS) Region in Menyanthaceae using Predicted Secondary Structure." *Molecular Phylogenetics and Evolution* 49, no. 2 (November): 526–537.
- Ward, Joseph H. Jr. 1963. "Hierarchical Grouping to Optimize an Objective Function." *Journal of the American Statistical Association* 58, no. 301: 236–244.
- Ward, Bernard L., Robert S. Anderson, and Arnold J. Bendich. 1981. "The Mitochondrial Genome is Large and Variable in a Family of Plants (Cucurbitaceae)." *Cell* 25, no. 3: 793–803.
- Wood, Todd Charles, and David P. Cavanaugh. 2001. "A Baraminological Analysis of Subtribe Flaveriinae (Asteraceae: Helenieae) and the Origin of Biological Complexity." *Origins* 52: 7–27.
- Xiao-Ming, Zheng, Junrui Wang, Feng Li, Liu Sha, Pang Hongbo, Qi Lan, Li Jing, Sun Yan, Qiao Weihua, Zhang Lifang, Cheng Yunlian, and Yang Qingwen. 2017. "Inferring the Evolutionary Mechanism of the Chloroplast Genome Size by Comparing Whole-Chloroplast Genome Sequences in Seed Plants." *Scientific Reports* 7, no. 1 (8 May): 1555.
- Zhai, Yufei, Xiaqing Yu, Junguo Zhou, Ji Li, Zhen Tian, Panqiao Wang, Ya Meng, Qinzhen Zhao, Qunfeng Lou, Shengli Du, and Jinfeng Chen. 2021. "Complete Chloroplast Genome Sequencing and Comparative Analysis Reveals Changes to the Chloroplast Genome after Allopolyploidization in *Cucumis*." *Genome* 64, no. 6 (June): 627–638.
- Zhang, Xiao, Tao Zhou, Jia Yang, Jingjing Sun, Miaomiao Ju, Yuemei Zhao, and Guifang Zhao. 2018. "Comparative Analyses of Chloroplast Genomes of Cucurbitaceae

Species: Lights into Selective Pressures and Phylogenetic Relationships.” *Molecules* 23, no.9 (28 August): 2165.

Zhou, Cong, Putao Wang, Qun Zeng, Rongbin Zeng, Wei Hu, Lei Sun, Shi Liu, Feishi Luan, and Qianglong Zhu. 2023. “Comparative Chloroplast Genome Analysis of Seven Extant *Citrullus* Species Insight into Genetic Variation, Phylogenetic Relationships, and Selective Pressure.” *Scientific Reports* 13, no.1 (25 April): 6779.

Supplementary Materials

All supplementary files and figures are available on Zenodo at <https://zenodo.org/records/10372288>.

