NRC Research Press

ARTICLE

Complete chloroplast genome sequencing and comparative analysis reveals changes to the chloroplast genome after allopolyploidization in *Cucumis*

Yufei Zhai, Xiaqing Yu, Junguo Zhou, Ji Li, Zhen Tian, Panqiao Wang, Ya Meng, Qinzheng Zhao, Qunfeng Lou, Shengli Du, and Jinfeng Chen

Abstract: Allopolyploids undergo "genomic shock" leading to significant genetic and epigenetic modifications. Previous studies have mainly focused on nuclear changes, while little is known about the inheritance and changes of organelle genome in allopolyploidization. The synthetic allotetraploid *Cucumis* ×*hytivus*, which is generated via hybridization between *C. hystrix* and *C. sativus*, is a useful model system for studying cytonuclear variation. Here, we report the chloroplast genome of allotetraploid *C.* ×*hytivus* and its diploid parents via sequencing and comparative analysis. The size of the obtained chloroplast genomes ranged from 154 673 to 155 760 bp, while their gene contents, gene orders, and GC contents were similar to each other. Comparative genome analysis supports chloroplast maternal inheritance. However, we identified 51 indels and 292 SNP genetic variants in the chloroplast genome of the allopolyploid *C.* ×*hytivus* relative to its female parent *C. hystrix*. Nine intergenic regions with rich variation were identified through comparative analysis of the chloroplast genomes within the subgenus *Cucumis*. The phylogenetic network based on the chloroplast genome sequences clarified the evolution and taxonomic position of the synthetic allotetraploid *C.* ×*hytivus*. The results of this study provide us with an insight into the changes of organelle genome after allopolyploidization, and a new understanding of the cytonuclear evolution.

Key words: allopolyploidy, chloroplast genome, SNP, phylogenetic analysis, *Cucumis*.

Résumé : Les allopolyploïdes subissent "choc génomique" menant à d'importants changements génomiques et épigénétiques. Les études antérieures ont principalement documenté les changements nucléaires, tandis que peu de choses sont connues de l'hérédité et des changements dans l'organisation du génome des organites lors de l'allopolyploïdisation. L'allopolyploïde synthétique *Cucumis* ×*hytivus*, lequel est généré par hybridation entre le *C. hystrix* et le *C. sativus*, est un modèle utile pour étudier la variation cytonucléaire. Dans ce travail, les auteurs examinent le génome chloroplastique de l'allotétraploïde *C.* ×*hytivus* et de ses parents diploïdes via le séquençage et des analyses comparées. La taille des génomes chloroplastiques obtenus variait entre 154 673 et 155 760 pb, tandis que le contenu génique, leur ordre ainsi que le contenu en GC étaient semblables. Une analyse comparée des génomes confirme la transmission maternelle du chloroplaste. Cependant, les auteurs ont identifié 51 indels et 292 variants SNP dans le génome chloroplastique de l'allopolyploïde *C.* ×*hytivus* par rapport à celui du parent femelle *C. hystrix*. Neuf régions intergéniques riches en variation ont été identifiées via analyse comparée des génomes chloroplastiques au sein du sous-genre *Cucumis*. Le réseau phylogénétique fondé sur les séquences de génomes chloroplastiques a clarifié l'évolution et la position taxonomique de l'allotétraploïde *C.* ×*hytivus*. Les résultats de cette étude jettent un éclairage sur les changements qui surviennent dans les génomes des organites après allopolyploïdisation et contribuent à une meilleure compréhension de l'évolution cytonucléaire. [Traduit par la Rédaction]

Mots-clés : allopolyploïdie, génome chloroplastiqu, SNP, analyse phylogénétique, *Cucumis*.

Received 3 August 2020. Accepted 20 December 2020.

Y. Zhai,* X. Yu,* J. Li, Z. Tian, P. Wang, Y. Meng, Q. Zhao, and Q. Lou. State Key Laboratory of Crop Genetics and Germplasm Enhancement, College of Horticulture, Nanjing Agricultural University, Nanjing, Jiangsu, China.
J. Zhou. College of Horticulture and Landscape, Henan Institute of Science and Technology, Xinxiang, Henan, China.
S. Du. State Key Laboratory of Vegetable Germplasm Innovation, Tianjin Kernel Cucumber Research Institute, Tianjin, China.
J. Chen. State Key Laboratory of Crop Genetics and Germplasm Enhancement, College of Horticulture, Nanjing Agricultural University, Nanjing, Jiangsu, China; State Key Laboratory of Vegetable Germplasm Innovation, Tianjin Kernel Cucumber Research Institute, Tianjin, China.
Corresponding authors: Jinfeng Chen (email: jfchen@njau.edu.cn); Junguo Zhou (email: junguo1020@163.com).
*These authors contributed equally to this work.

Introduction

Allopolyploids, formed by interspecific hybridization and whole-genome duplication (WGD), have played a key role in the evolution of plant species (Leitch and Bennett 1997; Soltis et al. 2009; Wendel 2000). Previous studies have indicated that all seed plants have undergone at least one round of genome doubling in their ancestry (Jiao et al. 2011). After the union of two different sets of genomes, the newly formed allopolyploids undergo a "genomic shock" (McClintock 1984), which can result in rapid inheritance, epigenetic, and genomic changes (Adams and Wendel 2005; Bottani et al. 2018; Buggs et al. 2012; Flagel and Wendel 2010; Paun et al. 2010; Yu et al. 2018).

The chloroplast genome is independent from the nuclear genome and exhibits semi-autonomous genetic character (Reyes-Prieto et al. 2007). Chloroplast genome has been commonly used for phylogenetic analyses and species identification because of its slow evolution and sequence conservation (Huang et al. 2016). During allopolyploidization, not only two different nuclear genomes collide but also chloroplast and mitochondrial genomes from different sources interact in the same cell. Given the complex coordination between the nuclear and organelle genomes (Taylor 1989), cytonuclear coevolution is an important aspect of allopolyploidization. So far, cytonuclear evolution has only been studied in relation to cytonuclear co-encoded protein complexes in a few allopolyploid plant systems (Gong et al. 2014; Wang et al. 2017). While nuclear–chloroplast incompatibility and altered cytonuclear interactions have been reported in interspecific hybrids and allopolyploid species (Sharbrough et al. 2017; Ferreira de Carvalho et al. 2019), the chloroplast genome evolution after allopolyploidization has rarely been systematically studied.

An interspecific cross was successfully made in *Cucumis* between a wild species of *Cucumis*, *C. hystrix* Chakr. (HH, $2n = 2x = 24$), and cultivated cucumber, *C. sativus* L. 'BeijingJietou' (CC, $2n = 2x = 14$) (Chen et al. 1997). The chromosome numbers of the *C. hystrix* × *C. sativus* F_1 interspecific hybrid (HC, $2n = 19$) were then doubled through somaclonal variation using embryo culture technique, and the synthesized allotetraploid species *C. ×hytivus* Chen and Kirkbride (HHCC, $2n = 4x = 38$) was obtained (Chen and Kirkbride 2000). The allotetraploid can self-pollinate and is cross-compatible with *C. sativus*, which provides a unique system to reveal the complicated processes in the formation and evolution of polyploid species. More importantly, it can serve as a genetic bridge to broaden the gene pool of cucumber by introgressing genes from wild relative species (Chen et al. 2003; Chen and Kirkbride 2000).

The genome of cucumber has three different inheritance modes, in which the chloroplast genome, mitochondrial genome, and nuclear genome are maternally, paternally, and biparentally inherited, respectively (Havey et al. 1998; Shen et al. 2015). For *C. ×hytivus*, Shen and colleagues concluded that mitochondrial DNA was paternally inherited, while chloroplast DNA was maternally inherited between species of *Cucumis* (Shen et al. 2013). In a previous study on the cytonuclear-encoded RuBisCO complex in *C. ×hytivus*, we found that the *rbcL* gene encoded by the chloroplast is consistent with its maternal *rbcL* sequence, and the nuclear-encoded *rbcS* gene inherits the copy type of both parents (Zhai et al. 2019). However, studies of individual genes or complexes are not sufficient to characterize the evolution of the whole genome. In this study, the chloroplast genomes of *C. ×hytivus* and its diploid parents were sequenced and comparatively analyzed to identify the genetic variations of chloroplast under the impact of allopolyploidization.

Materials and methods

Plant materials and chloroplast DNA isolation

Three species of *Cucumis* were used for this study, the cultivated cucumber *C. sativus* 'BeijingJietou' ($2n = 14$, genome CC), the self-cross plants (14th self-pollinated generation, S_{14}) of a synthesized new allotetraploid species *C. ×hytivus* ($2n = 38$, genome HHCC), and the wild species *C. hystrix* ($2n = 24$, genome HH). The two diploid parental plants used in this experiment are the inbred lines used for the interspecific cross. The individual plant used for sequencing was randomly selected from inbred lines.

About 10 g of fresh leaves were sampled from adult plants of three species of *C. hystrix*, *C. ×hytivus*, and *C. sativus*, respectively, for chloroplast DNA isolate. Total chloroplast DNA was isolated using improved sucrose gradient centrifugation method (Diekmann et al. 2008). The quality of chloroplast DNA was checked by monitoring the ratios of A260/A280 (DU800, Beckman Coulter, USA) and TBE polyacrylamide gel electrophoresis.

Library construction, genome sequencing, and assembly

The chloroplast DNA of each sample was randomly fragmented by sonication, and then the resulting DNA fragment was subjected to end-repair and phosphorylation using T4 DNA polymerase, Klenow DNA polymerase, and T4 PNK. After this, an 'A' base was inserted at the 3′ ends of the repaired DNA fragment, and then the Solexa adaptors at Illumina paired-end were attached to these DNA fragments to distinguish the different sequencing samples. Finally, the three paired-end libraries were sequenced using Illumina HiSeq™ 2000 according to the manufacturer's instructions (Illumina, San Diego, CA). In total, 1132, 2499, and 999 Mb paired-end clean reads (150 bp average read length) were obtained for *C. sativus*, *C. ×hytivus*, and *C. hystrix*, respectively.

Raw data were cleaned in several steps, including removing reads with unknown bases call (N) >10%, removing reads with 20 bp low-quality bases (≤Q20), removing adaptor contamination, and removing duplicated reads. The filtered reads were de novo assembled by SOAPdenovo v2.04 (Luo et al. 2015), and GapCloser v1.12 (Li et al. 2008) software was used to close gaps and finally remove the

redundant segment sequence to get the final assembly results. The chloroplast genome sequences obtained were used for the following analyses and are available in GenBank under the following accession numbers: MH424440 (*C. sativus* L. 'BeijingJietou'), NC_033871 (*C. ×hytivus*), and MH427087 (*C. hystrix*).

Chloroplast genome annotation

Following functional annotation using homologous alignment methods, the assembled sequences were compared with NR (Sayers et al. 2020) and Swiss-Prot (Magrane and Consortium 2011) databases using BLAST software (Altschul et al. 1990) to obtain functional annotation information for the encoded genes. Since each sequence alignment result may exceed one, to ensure its biological significance, the annotation retains an optimal match result as a comment for the gene. The chloroplast genome circular map was drawn using OGDraw (Lohse et al. 2007).

SNP and indel analysis

Chloroplast genome sequences were used to analyze the sequence polymorphism between allotetraploid *C. ×hytivus* and its diploid parents. Single nucleotide polymorphism (SNP) was detected from pairwise alignments using MUMmer alignment software (Kurtz et al. 2004). MUMmer was run on the complete chloroplast genomes to generate pairwise sequence alignments against the reference chloroplast genome using a sliding window to detect the potential SNP. A 100 bp sequence on each side of each potential SNP was extracted and then BLAT software was used to compare the extracted sequence with the assembly result to verify the SNP. If the length of the comparison was less than 101 bp, then it was considered an unreliable SNP and was removed. Finally, BLAST, TRF, and RepeatMask software were used to predict the repeat region of the reference sequence, and the SNPs in the repeat region were filtered to obtain reliable SNPs. Insertion and deletion (indel) analyses were carried out by alignment with the reference genome using LASTZ software (Harris 2007).

Repeat elements analysis

REPuter was used to identify repeat sequences, including four repeat types: palindromic, reverse, forward, and complement repeat (Kurtz et al. 2001), in which a minimum repeat size of 30 bp, 90% or greater sequence identity, and the Hamming distance of 3 were selected, respectively (Saski et al. 2005).

Comparative genome analysis of the chloroplast genome of species of subgenus *Cucumis*

The complete chloroplast genomes of seven species within the subgenus *Cucumis*, including allotetraploid *C. ×hytivus* and its diploid parents, *C. hystrix* and *C. sativus*, and four published *C. sativus* chloroplast genomes, which were downloaded from the NCBI Organelle Genome Resource database (Table S1[1]), were comparatively compared using the mVISTA program in a Shuffle-LAGAN mode (Frazer et al. 2004). Conserved regions on *C. ×hytivus* were identified through global alignment with the other six chloroplast genomes of species of subgenus *Cucumis*. DNA polymorphisms of nucleotide diversity (P_i) at the intraspecific level of *C. sativus* and the interspecific level of *C. sativus* and *C. hystrix* were calculated using DnaSP 5.10.01 (Librado and Rozas 2009).

Phylogenetic analysis

To illustrate the phylogenetic relationship of species of the family Cucurbitaceae, 13 chloroplast genomes were aligned using MAFFT v7 (Katoh and Standley 2013) and adjusted manually, including three newly assembled chloroplast genomes of species of subgenus *Cucumis* and 10 completed chloroplast genome sequences downloaded from the NCBI Organelle Genome Resource database (Table S1[1]). The aligned sequences were saved in PHYLIP and NEXUS format to generate a phylogenetic tree. The phylogenetic network was constructed with SplitsTree 5 (Huson and Bryant 2006) using the Neighbor-Net method.

SNP validation analysis

The polymerase chain reactions (PCRs) and direct sequencing were performed to verify the six SNPs identified in exons between the chloroplast genomes of *C. ×hytivus* and *C. hystrix*. Primers used are tabulated in Table S2[1]. The PCRs products were sequenced (TSINGKE, Beijing, China). We randomly selected five individuals of *C. ×hytivus* and five individuals of *C. hystrix* for PCR verification, and to exclude the possibility of PCR recombination and sequencing artifacts, we required that each sequence be detected thrice.

Results

Chloroplast genome structural features and gene content of *C. ×hytivus* and its parents

The assembled chloroplast genomes of three species of *Cucumis* ranged from 154 673 to 155 760 bp in length, with the allopolyploid *C. ×hytivus* being the largest and *C. hystrix* the smallest. The three chloroplast genomes of species of subgenus *Cucumis* showed a typical quadripartite structure of angiosperm chloroplast DNA consisting of a pair of inverted repeat (IR) regions (50 374–50 432 bp), a large single copy (LSC) region (86 247–87 004 bp), and a small single copy (SSC) region (17 996–17 324 bp). The total guanine-cytosine (GC) content of each genome was ~37% (Table 1; Fig. 1).

The three chloroplast genomes mapped to a circular molecule containing 113 unique genes and 17 duplicated genes, for a total of 130 genes; three chloroplast genomes were identical in gene arrangement and gene composition (Table 2; Fig. 1). Seven tRNA genes, four rRNA genes, and six protein-coding genes (*ndhB*, *rps7*, *rps12*, *rpl2*, *rpl23*, and *ycf2*) were completely duplicated in

[1]Supplementary data are available with the article at https://doi.org/10.1139/gen-2020-0134.

Table 1. Size comparison of three chloroplast genomic regions of *Cucumis*.

	C. hystrix	*C. ×hytivus*	*C. sativus*
Total chloroplast genome size (bp)	154 673	155 760	155 526
Large single copies in bp (% of genome)	86 247 (55.76%)	87 004 (55.86%)	86 878 (55.86%)
Inverted repeats in bp (% of genome)	50 430 (32.60%)	50 432 (32.38%)	50 374 (32.39%)
Small single copies in bp (% of genome)	17 996 (11.63%)	18 324 (11.76%)	18 274 (11.75%)
GC content (%)	37.00%	36.92%	36.94%
Protein-coding genes[a]	79 (85)	79 (85)	79 (85)
rRNA genes[a]	4 (8)	4 (8)	4 (8)
tRNA genes[a]	30 (37)	30 (37)	30 (37)

[a]First value excludes duplicates; value in parentheses includes them.

Fig. 1. Chloroplast genome map of *Cucumis hystrix*, *C. ×hytivus*, and *C. sativus*. Genes on the outside of the large circle are transcribed counterclockwise and those inside are transcribed clockwise. The genes are color-coded based on their function. Dashed area represents the GC composition of the chloroplast genome.

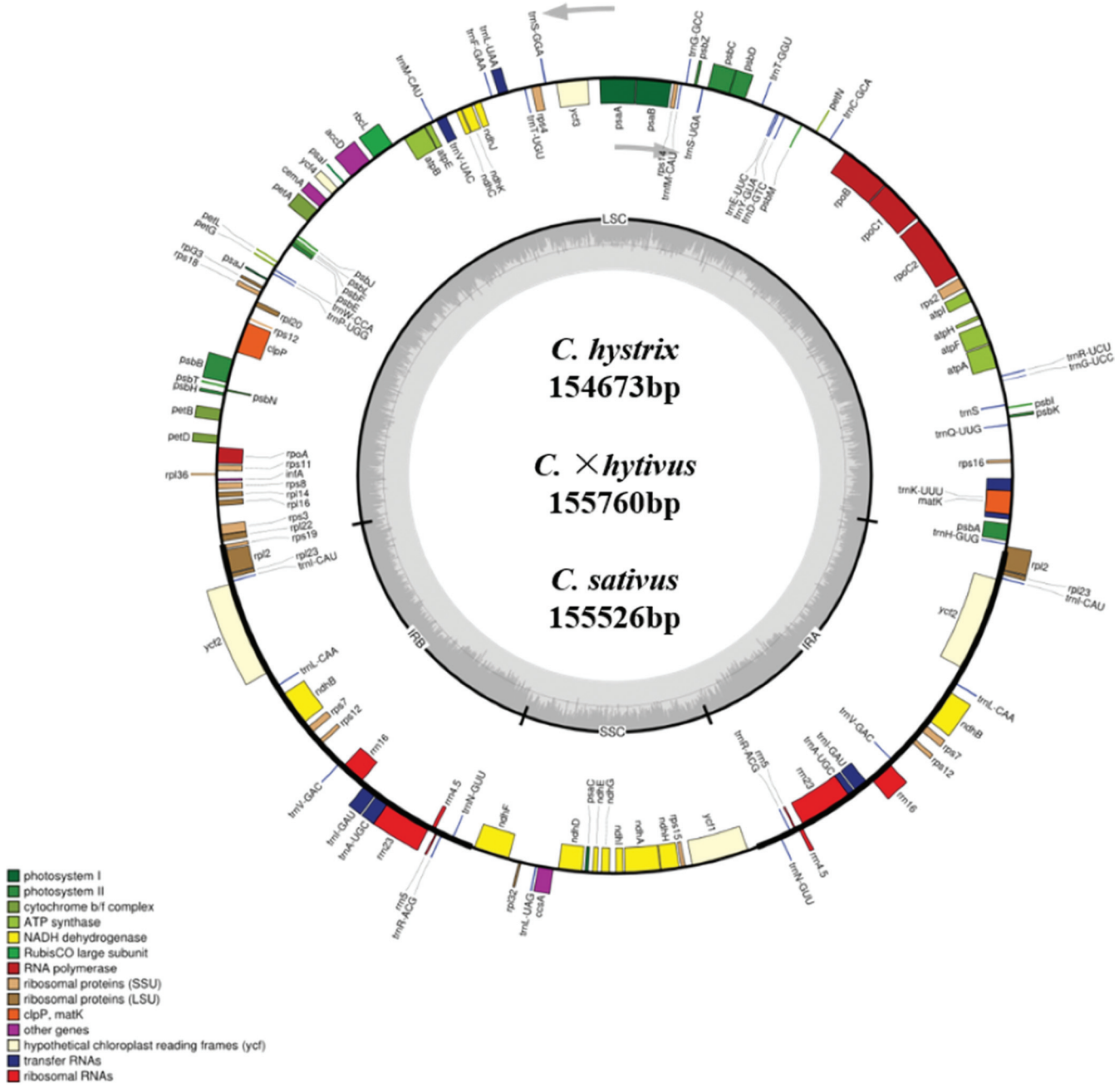


Table 2. Genes identified in the chloroplast genomes of *Cucumis hystrix*, *C. ×hytivus*, *and C. sativus*.

Category for genes	Group of gene	Name of gene
Photosynthesis-related genes	Photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*
	Photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*
	Cytochrome b/f compelx	*petA*, **petB*, **petD*, *petG*, *petL*, *petN*
	ATP synthase	*atpA*, *atpB*, *atpE*, **atpF*, *atpH*, *atpI*
	Cytochrome c synthesis	*ccsA*
	Assembly/stability of photosystem I	**ycf3*, *ycf4*
	NADPH dehydrogenase	**ndhA*, **ndhB* (2×), *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*
	Rubisco	*rbcL*
Transcription- and translation-related genes	Transcription	*rpoA*, *rpoB*, **rpoC1*, *rpoC2*
	Ribosomal proteins	*rps2*, *rps3*, *rps4*, *rps7* (2×), *rps8*, *rps11*, **rps12* (2×), *rps14*, *rps15*, **rps16*, *rps18*, *rps19*, **rpl2* (2×), *rpl14*, **rpl16*, *rpl20*, *rpl22*, *rpl23* (2×), *rpl32*, *rpl33*, *rpl36*
Other genes	RNA processing	*matK*
	Translational initiation factor	*infA*
	Carbon metabolism	*cemA*
	Fatty acid synthesis	*accD*
	Proteolysis	**clpP*
Genes of unknown function	Conserved reading frames	*ycf1*, *ycf2* (2×)
RNA genes	Ribosomal RNA	*rrn5* (2×), *rrn4.5* (2×), *rrn16* (2×), *rrn23* (2×)
	Transfer RNA	**trnA-UGC* (2×), *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnfM-CAU*, **trnG-UCC*, *trnG-GCC*, *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*

Note: Intron-containing genes are marked by an asterisks (*); Duplicated genes are marked by (2×) behind genes (genes present in the IR regions).

Table 3. SNP and Indel analysis between the chloroplast genomes of *Cucumis hystrix*, *C. ×hytivus*, and *C. sativus*.

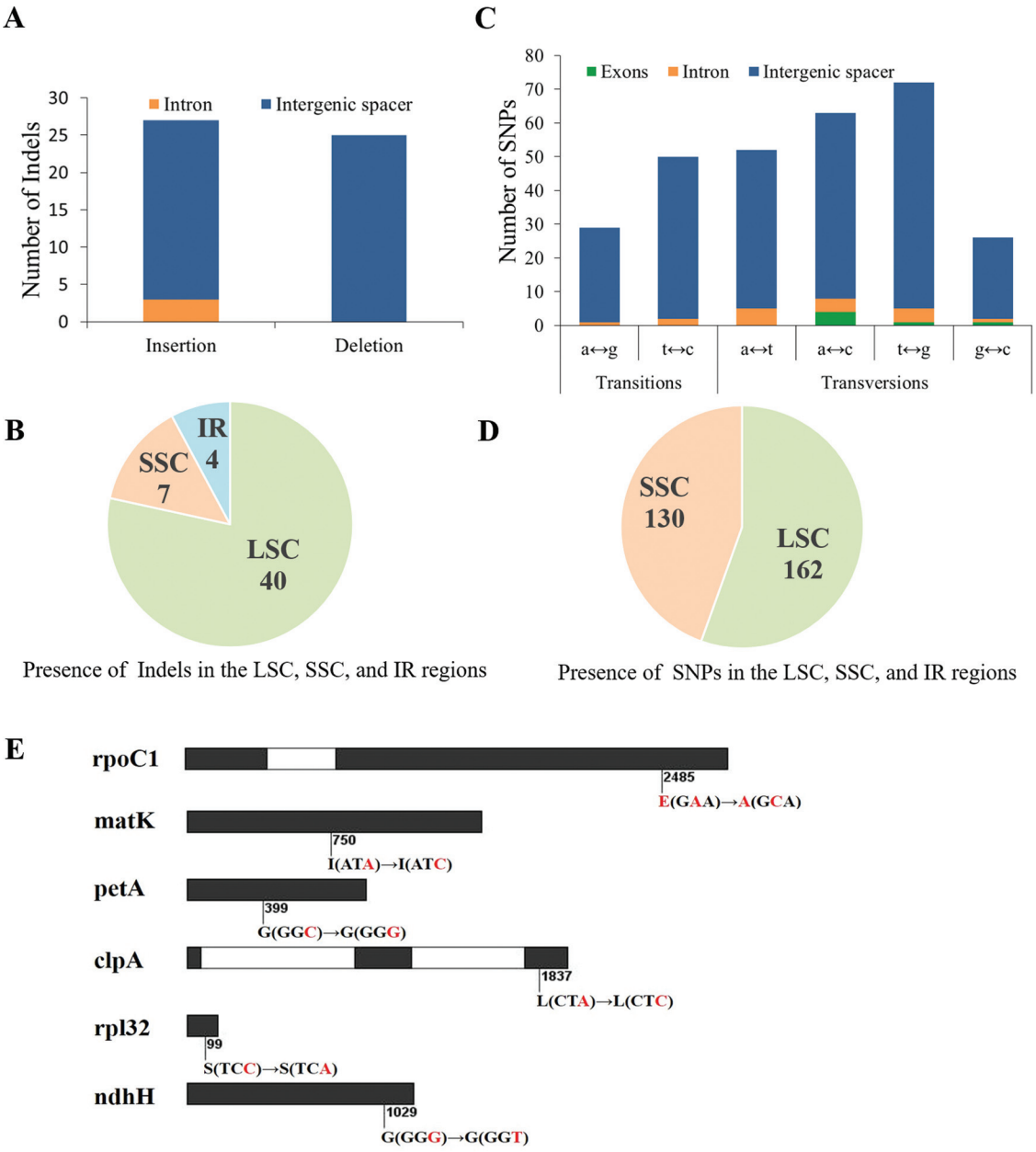
SNP	Synonymous	Nonsynonymous	Total CDS SNP	Intergenic	Total SNP
C. hystrix vs. *C. ×hytivus*	5	1	6	286	292
C. sativus vs. *C. ×hytivus*	93	106	200	661	861
Indel	**Insertion**	**Deletion**	**CDS with Indel**	**Intergenic**	**Total Indel**
C. hystrix vs. *C. ×hytivus*	27	24	3	48	51
C. sativus vs. *C. ×hytivus*	74	48	2	120	122

the IR regions. One copy of *ycf1* gene was a pseudogene because the normal copy of *ycf1* was incompletely duplicative at the IRa and SSC boundary (Fig. 1). Eighteen genes contained introns among the 113 unique genes, 12 of which were from protein-coding genes and six were from tRNAs. The *rps12* gene was encoded as *trans*-spliced with a single 5′ end in the LSC region and a repeated 3′ end in both IR regions, which is generally found in many other land plants (Hildebrand et al. 1988; Schmitz-Linneweber et al. 2006).

Analysis of the chloroplast genomes of *C. ×hytivus* and its parents: SNP, indels, and repeat elements

Sequence alignment between allotetraploid *C. ×hytivus* and its diploid parents revealed a total of 292 SNPs and 51 indels in *C. hystrix* versus *C. ×hytivus*, and 861 SNPs and 122 indels in *C. sativus* versus *C. ×hytivus* (Table 3). This result supports maternal chloroplast inheritance, that is, the chloroplast of *C. ×hytivus* is derived from its maternal ancestor, *C. hystrix*, and it also revealed the genetic variation of chloroplast genome after allopolyploidization in *Cucumis*. Further, DNA polymorphism of chloroplast genome at the intra- and interspecific levels was calculated using DnaSP 5.10.01 (Librado and Rozas 2009). The results showed that the nucleotide diversity (Pi) was 0.00531 between *C. sativus* and *C. hystrix*, which was higher than that in *C. sativus* (Pi = 0.00001 ∼ 0.00024). While the nucleotide diversity between *C. ×hytivus* and *C. hystrix* (Pi = 0.00211) was between the interspecific level (*C. hystrix* vs. *C. sativus*) and the intraspecific level (*C. sativus*), indicating increased molecular diversity of the chloroplast genome after allopolyploidization.

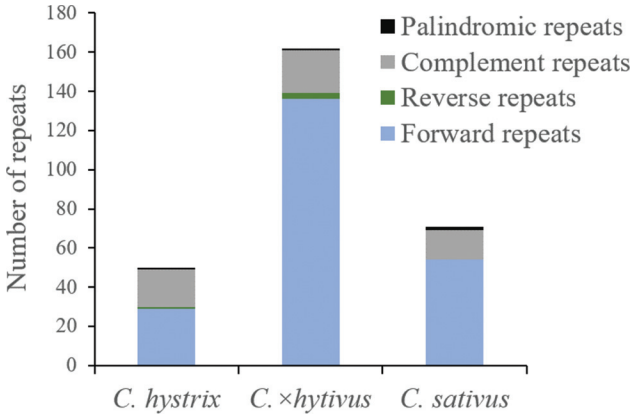
Fig. 2. Distribution, type, and presence of indels and SNPs identified in the chloroplast genomes of *Cucmis ×hytivus* relative to *C. hystrix*. (A) The distribution of indels in the chloroplast genomes. (B) Presence of indels in the large single copy (LSC), small single copy (SSC), and inverted repeat (IR) regions. (C) The distribution and type of SNPs in the chloroplast genomes. (D) Presence of SNPs in the LSC, SSC, and IR regions. (E) Schematic representations of the six substitution events identified located in exons regions, among which one induced nonsynonymous (rpoC1) and five induced synonymous (*matK*, *petA*, *clpP*, *rpl32*, and *ndhH*).



Of the 51 indels identified in *C. ×hytivus* relative to *C. hystrix* chloroplast genomes, 27 were insertions and 24 were deletions (Fig. 2A; Table S3[1]). These indels were all present in non-coding regions, 48 in intergenic spacer regions, and 3 in the intron regions of *ycf3* and *rpoC1*. Most of them (40/51) were located in the LSC region, and a few were located in the SSC region (7/51) and the IR region (4/51) (Fig. 2B). The majority of SNPs (269/292) were located in intergenic spacer regions, while only 17 and six SNPs were located in introns and exons, respectively (Fig. 2C). Of the total 292 SNPs, 79 and 213 were transitional (Ts) and transversional (Tv) changes, respectively, showing a lower Ts/Tv bias (≈0.37) (Fig. 2C), and 162 of them were located in the LSC region, and 130 were located in the SSC region (Fig. 2D; Table S4[1]).

The six SNPs in exons were verified by PCR and direct sequencing, five *C. ×hytivus* individuals and five *C. hystrix* individuals were randomly selected for PCR validation, and the results confirmed the effectiveness of

Fig. 3. Repeated elements analysis in the chloroplast genome of three species of *Cucumis*.



sequencing (Fig. 2E). Five out of six SNPs induced synonymous substitution (*matK, petA, clpP, rpl32*, and *ndhH*), while only one induced nonsynonymous substitution (*rpoC1*, Glu to Ala) (Fig. 2E).

A total of 50, 162, and 71 repeats were detected in the chloroplast genomes of *C. hystrix*, *C. ×hytivus*, and *C. sativus*, respectively, using REPuter, including forward, reverse, complement, and palindromic repeats (Table S5[1]). Forward repeats were the most common, followed by complement repeats. Only one palindromic repeat in *C. hystrix* and *C. ×hytivus*, two in *C. sativus*, and one and three reverse repeats were identified in *C. hystrix* and *C. ×hytivus*, but none were identified in *C. sativus* (Fig. 3). In these chloroplast genomes, a minority of repeats were found in coding regions, while the majority were located in noncoding regions (Table S5[1]).

Comparative analysis of the chloroplast genomes of species of subgenus *Cucumis*

Conservative areas and rich variation regions in chloroplast genome are useful for species identification and understanding of the functions of particular DNA area or the mechanisms for cpDNA evolution (Cartwright 2005; Shaw et al. 2007). mVISTA (Frazer et al. 2004) was used to study the chloroplast genome sequence variations in the species of subgenus *Cucumis*, *C. ×hytivus*, *C. hystrix*, *C. sativus* 'BeijingJietou', *C. sativus* 'Chipper', *C. sativus* 'Baekmibaekdadagi', *C. sativus* 'GY14', and *C. sativus* var. Hardwickii, of which *C. ×hytivus* was set as a reference. Broadly, the result shows that the IR region was higher conserved than the LSC and SSC regions, and the coding region was found to be higher conserved than the non-coding regions (Fig. 4). Conserved regions on *C. ×hytivus* were identified through global alignment with the other six chloroplast genomes of subgenus *Cucumis* (Table S6[1]). The intergenic regions of *rps16-trnQ*, *rpl32-trnL*, *trnK-rps16*, *trnT-psbD*, *ndhF-rpl32*, *psbE-petL*, *petA-psbJ*, *rbcl-accd*, and *rpl16-rps3* were highly variable in subgenus *Cucumis*, seven of which were included in the 13 hotspots reported in the genomes of several plants (Shaw et al. 2007), the other two (*rbcl-accd* and *rpl16-rps3*) areas displayed specificity to subgenus *Cucumis*, which can be used for molecular studies at low taxonomic levels.

Phylogenetic analysis of chloroplast genome sequences within the family Cucurbitaceae

The phylogenetic network was built in SplitsTree5 (Huson and Bryant 2006) using the Neighbor-Net method based on the chloroplast genome sequences of 13 taxa. As shown in Fig. 5, C. ×*hytivus* and *C. hystrix*, along with *C. sativus*, are members of the subgenus *Cucumis*. The phylogenetic relationship of the chloroplast genome in other species of Cucurbitaceae is consistent with their botanical classifications (Fig. 5). Splits in the phylogenetic network are a result of reticulate events such as hybridization, horizontal gene transfer, or recombination (Huson and Bryant 2006); the cyclic split of *C. ×hytivus*, *C. hystrix*, and *C. sativus* may correspond to the interspecies hybridization event in *C. hystrix* and *C. sativus*.

Discussion

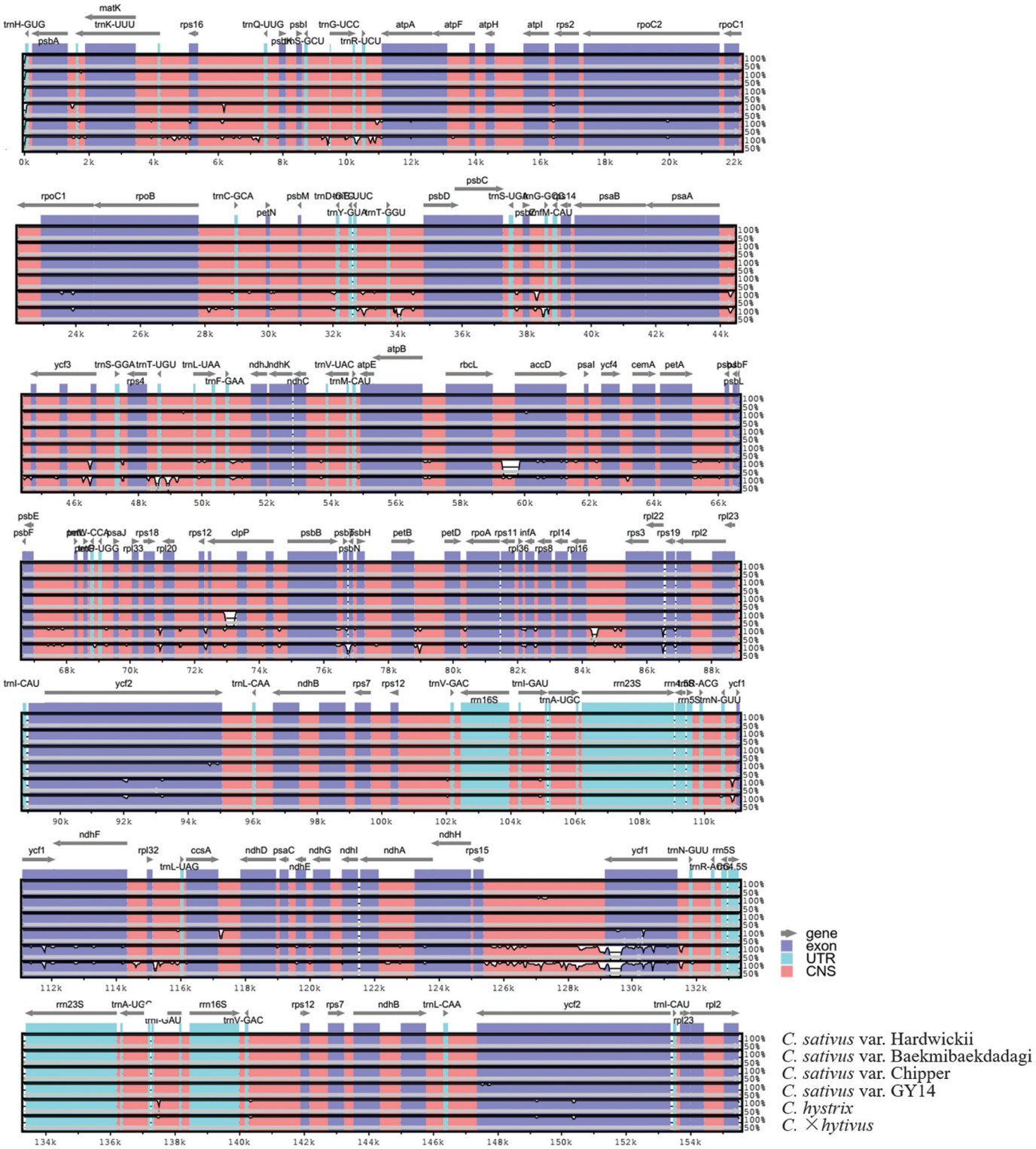
In this study, we used a synthesized *Cucumis* allopolyploid as a model system to explore chloroplast genome variation. For the first time, complete chloroplast genome sequencing and sequence alignment demonstrated chloroplast maternal inheritance and revealed the genetic variation sites of chloroplast genome in *Cucumis* allotetraploid. Our research may contribute to reveal the evolutionary features of allopolyploid organelles genome.

Chloroplast genome inheritance after allopolyploidization

The genome of cucumber has three different inheritance modes, in which the chloroplast genome, mitochondrial genome, and nuclear genome are maternally, paternally, and biparentally inherited, respectively (Havey et al. 1998; Shen et al. 2015). Shen and colleagues found that some chloroplast DNA fragments in the allotetraploid were maternally inherited (Shen et al. 2013). Whole chloroplast genome sequencing and sequence alignment in this study supports maternal inheritance of chloroplast in *C. ×hytivus* (Table 3). The allotetraploid chloroplast genome of *Cucumis* is 155 760 bp in length, which shows a trend toward increased genome size, and is larger than its male and female parents (155 526 and 154 673 bp), likely owing to differences of repeated elements and insertions and deletions in intergenic regions (Fig. 3; Tables S3, S5[1]). The allotetraploid chloroplast DNA is highly similar to its parents and also similar to previously published chloroplast genomes of species within Cucurbitaceae in structure and gene order (Kim et al. 2006; Plader et al. 2007; Rodriguez-Moreno et al. 2011), suggesting a high level of genetic conservative in Cucurbitaceae.

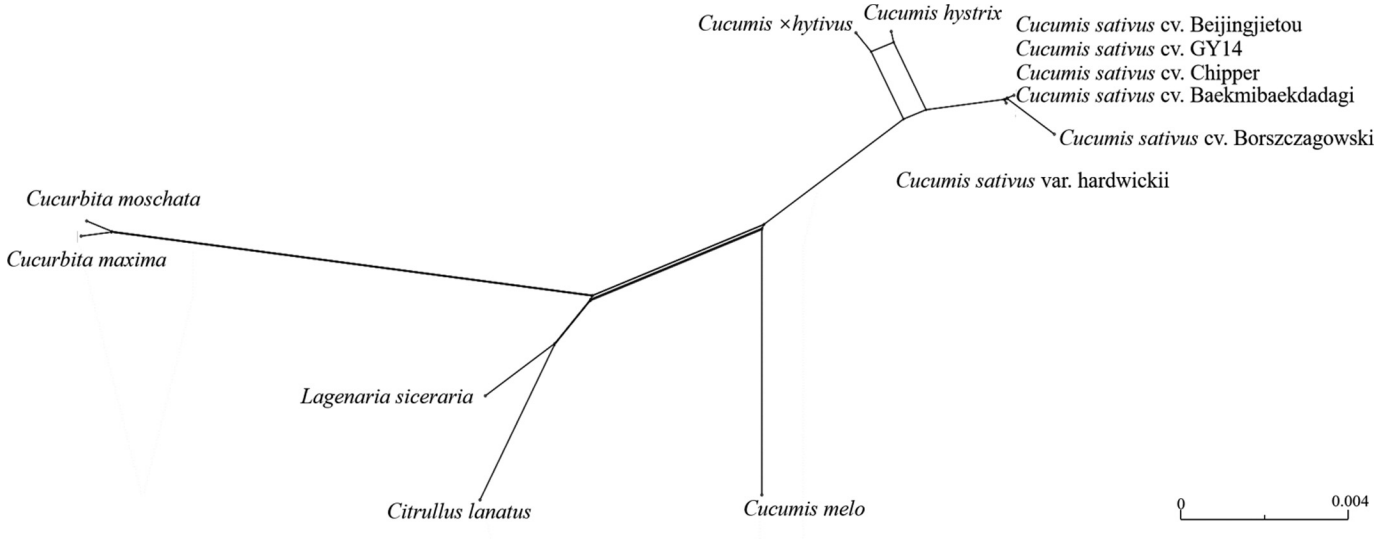
Previous investigation of reciprocal crosses showed that when *C. hystrix* was used as the female parent, the diploid hybrid ($2n$ = 19, HC) and allotetraploid ($2n$ = 38,

Fig. 4. Visualization of alignment of the genome of *Cucumis ×hytivus* and other published chloroplast genomes in the cucumber subgenus using VISTA. Gray arrows indicate the position and direction of each gene. Blue and red areas indicate genic and intergenic regions, respectively. *Y*-axis indicates the range of identity (50%–100%).



HHCC) had more fruit sets, whereas when the cucumber was used as the female parent, both diploid hybrid (2*n* = 19, CH) and allotetraploid (2*n* = 38, CCHH) were highly sterile with very low fruit sets (Chen et al. 2002*b*). That explains why no genetic anti-cross validation was included in this experiment. Allopolyploid plants often encounter many genetic and epigenetic variations, which leads to various phenotypic variation. *Cucumis ×hytivus* showed complex

Fig. 5. Phylogenetic relationship among the species of Cucurbitaceae based on the complete chloroplast genome sequences. The scale bar indicates the distance of the edges. The planar graph was constructed with Splits-Tree 5.



phenotypic variation (e.g., flowering time, fruit shape, and yellow-green leaf color), while some of the traits of *C. ×hytivus* plants are similar to the female parent *C. hystrix*, such as multiple-branching habit, ovate fruit, and densely brown hairs (Chen et al. 2002*a*; Yu et al. 2015). These matroclinous phenotypes of *C. ×hytivus* presumably could be contributed by the maternal inheritance of chloroplast and (or) nuclear–cytoplasmic interactions during the process of allopolyploidization.

The effect of allopolyploidization on chloroplast genomes

Almost all seed plants have experienced polyploidy in their evolutionary history (Chen et al. 2002*b*). The early evolution of polyploid genome is the epitome of the evolution of the plant genome. Through the comparison of genomic data, the study of biological evolution brought by WGD has been a hot topic in the study of plant polyploid genome. Duplicated nuclear genes after allopolyploidy may take different evolutionary pathways, including gene loss, pseudogenization, de novo functionalization or subfunctionalization, and redistribution of duplicate copy expression between tissues or developmental stages (Grover et al. 2012). Whereas the uniparentally inherited organelle genome lacks systematic study, some previous studies used chloroplast genes instead of the whole chloroplast genome to track nuclear–cytoplasmic evolution (Gong et al. 2012, 2014; Wang et al. 2017; Zhai et al. 2019). Organelle genome sequence analysis of synthetic allopolyploid systems can lay the foundation for further investigation of cytonuclear evolution.

Nucleotide substitution and indel (insertions and deletions) events are the major driving forces of gene and genome evolution, and whether insertion, deletion, or point mutation is not random (Kelchner 2000; Massouh et al. 2016; Morton 1995; Yamane et al. 2006). The results of the validation of SNP and sequencing analysis are consistent, which indicates that the mutation detected in *C. ×hytivus* (S_{14}) is reliable. Relative to *C. hystrix*, all indels and 97.9% of the SNPs in *C. ×hytivus* are located in noncoding region, all the SNPs are located in LSC and SSC regions, and of the 51 indel events, only 7% (4 indels) are located in IR regions, which is much less than 32.38% of the IR in genome. These results demonstrat that exons and IR regions of *C. ×hytivus* chloroplast genomes were relatively conservative, but its introns and intergenic spacer regions were more polymorphic in the process of evolution. The contraction and expansion events of IR regions are common in evolutionary history (Wang et al. 2008), and it plays an important role in stabilizing chloroplast genome structure (Marechal and Brisson 2010). The base substitutions, known as SNPs, include transitions (Ts; A↔G, T↔C) and transversions (Tv; A↔C, A↔T, C↔G, G↔T). In theory, the Ts/Tv ratio should be 0.5, but due to the genome content, the genetic characteristics of codons, and the corresponding pattern of codon replacements, this value is usually higher than that in practice (Morton 2003). Transitions (A↔G, T↔C) generally occur more frequently than transversions (A↔C, A↔T, C↔G, G↔T) among spontaneous mutations because of the different molecular structures of pyrimidines (C, T) versus purines (A, G) (Wakeley 1996). The relatively higher ratios of Ts/Tv (1.0, 0.76, and 0.7328) have been found in the evolution of *Morus mongolica*, *Haloxylon*, and rice chloroplast genome, respectively (Kong and Yang 2016; Tong et al. 2016). In our case, the ratios in the whole chloroplast genomes, intron, and intergenic spacer regions of *C. ×hytivus* relative to *C. hystrix* were 0.37, 0.27, and 0.39, respectively. In the coding region, the value reached zero with no transition and six transversions, indicating that transversions are more frequent in the process of the chloroplast genome evolution along with *Cucumis*

interspecies hybridization and polyploidy differing from other species. It is concluded that interspecific hybridization has an effect on chloroplast genomes, and this change may be due to exposure to nuclear genomic shocks as well as altered genomic doses, the relevant results can be further explored in more allopolyploids. This is new knowledge to us since chloroplast has been considered conservative in genetics and evolution. The relationship between evolutionary mechanisms that caused the abnormal types of mutations and complex nuclear–cytoplasmic interactions and incompatibilities in allopolyploid needs further research.

The evolution of species of *Cucumis*

The family Cucurbitaceae contains many important cultivated species, including popular vegetables such as cucumber, pumpkin, melon, and watermelon. Whole chloroplast genomes of 13 cucurbitaceous plants were used to analyze their evolutionary relationship. Our results correspond to earlier studies on the relationship between cucumber, melon, watermelon, and pumpkin (Sebastian et al. 2010) (Fig. 5). The evolutionary relationship between two important vegetable crops in the genus *Cucumis*, cucumber (*C. sativus* L. $2n = 14$) and melon (*C. melo* L. $2n = 24$), is controversial, with the following opposite hypotheses being proposed: a fragmentation hypothesis from $n = 7$ to $n = 12$ and a fusion hypothesis from $n = 12$ to $n = 7$. *Cucumis hystrix* is a wild species of *Cucumis* grouped into the same subgenus as *C. sativus*, where it has a chromosome number of $2n = 2x = 24$. The results of phylogenetic network analysis based on chloroplast genomic sequence in this study and other molecular phylogenetic studies suggest that $n = 12$ is ancestral in the genus *Cucumis* (Ghebretinsae et al. 2007; Sebastian et al. 2010). More importantly, we demonstrated increased molecular divergence of the newly synthetic allotetraploid *C.* ×*hytivus* at the chloroplast genome level, which may be related to allopolyploidization.

Author contributions

J.C. and J.Z. conceived of the study. J.L. and Q.L. helped designed the experiments. Y.Z. and X.Y. performed the experiment and drafted the manuscript. Z.T., P.W., and S.D. assisted in data analysis. Q.Z. and Y.M. helped with the synthesis of materials. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This research was partially supported by the National Natural Science Foundation of China (NO. 31902006), the Fundamental Research Funds for the Central Universities (KJQN202029), the Natural Science Foundation of Jiangsu Province, China (BK20180536), the National Key Research and Development Program of China (2016YFD0100204-25), and National Key Research and Development Program of China (2016YFD0101705-5). A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

Adams, K.L., and Wendel, J.F. 2005. Polyploidy and genome evolution in plants. Curr. Opin. Plant Biol. **8**(2): 135–141. doi:10.1016/j.pbi.2005.01.001. PMID:15752992.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic Local Alignment Search Tool. J. Mol. Biol. **215**(3): 403–410. doi:10.1016/S0022-2836(05)80360-2. PMID:2231712.

Bottani, S., Zabet, N.R., Wendel, J.F., and Veitia, R.A. 2018. Gene expression dominance in allopolyploids: hypotheses and models. Trends Plant Sci. **23**(5): 393–402. doi:10.1016/j.tplants.2018.01.002. PMID:29433919.

Buggs, R.J.A., Chamala, S., Wu, W., Tate, J.A., Schnable, P.S., Soltis, D.E., et al. 2012. Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. Curr. Biol. **22**(3): 248–252. doi:10.1016/j.cub.2011.12.027. PMID:22264605.

Cartwright, R.A. 2005. DNA assembly with gaps (Dawg): simulating sequence evolution. Bioinformatics, **21**(3): 31–38. doi:10.1093/bioinformatics/bth471. PMID:15333453.

Chen, J.F., and Kirkbride, J.H. 2000. A new synthetic species of *Cucumis* (Cucurbitaceae) from interspecific hybridization and chromosome doubling. Brittonia, **52**(4): 315–319. doi:10.2307/2666583.

Chen, J.F., Staub, J.E., Tashiro, Y., Isshiki, S., and Miyazaki, S. 1997. Successful interspecific hybridization between *Cucumis sativus* L. and *C. hystrix* Chakr. Euphytica, **96**(3): 413–419. doi:10.1023/A:1003017702385.

Chen, J.F., Staub, J., Adelberg, J., Lewis, S., and Kunkle, B. 2002*a*. Synthesis and preliminary characterization of a new species (amphidiploid) in *Cucumis*. Euphytica, **123**(3): 315–322. doi:10.1023/A:1015095430624.

Chen, J.F., Zhuang, F.Y., Lou, Q.F., Xu, Y.B., Qian, C.T., Ren, G., and Lou, X.D. 2002*b*. Studies on reciprocal differences in interspecific hybridization in *Cucumis*. Acta Horticulturae Sinica, **29**: 483–485. doi:10.1006/jfls.2001.0409.

Chen, J., Staub, J., Qian, C., Jiang, J., Luo, X., and Zhuang, F. 2003. Reproduction and cytogenetic characterization of interspecific hybrids derived from *Cucumis hystrix* Chakr. × *Cucumis sativus* L. Theor. Appl. Genet. **106**(4): 688–695. doi:10.1007/s00122-002-1118-7. PMID:12595999.

Diekmann, K., Hodkinson, T.R., Fricke, E., and Barth, S. 2008. An optimized chloroplast DNA extraction protocol for grasses (Poaceae) proves suitable for whole plastid genome sequencing and SNP detection. PLoS One, **3**(7): e2813. doi:10.1371/journal.pone.0002813. PMID:18665252.

Ferreira de Carvalho, J., Lucas, J., Deniot, G., Falentin, C., Filangi, O., Gilet, M., et al. 2019. Cytonuclear interactions remain stable during allopolyploid evolution despite repeated whole-genome duplications in *Brassica*. Plant J. **98**: 434–447. doi:10.1111/tpj.14228. PMID:30604905.

Flagel, L.E., and Wendel, J.F. 2010. Evolutionary rate variation, genomic dominance and duplicate gene expression evolution during allotetraploid cotton speciation. New Phytol. **186**(1): 184–193. doi:10.1111/j.1469-8137.2009.03107.x. PMID:20002320.

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and Dubchak, I. 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Res. **32**: W273–W279. doi:10.1093/nar/gkh458. PMID:15215394.

Ghebretinsae, A.G., Thulin, M., and Barber, J.C. 2007. Relationships of cucumbers and melons unraveled: Molecular phylogenetics of *Cucumis* and related genera (Benincaseae, Cucurbitaceae). Am. J. Bot. **94**(7): 1256–1266. doi:10.3732/ajb.94.7.1256. PMID:21636491.

Gong, L., Salmon, A., Yoo, M.J., Grupp, K.K., Wang, Z.N., Paterson, A.H., and Wendel, J.F. 2012. The cytonuclear dimension of allopolyploid evolution: an example from cotton using Rubisco. Mol. Biol. Evol. **29**(10): 3023–3036. doi:10.1093/molbev/mss110. PMID:22490824.

Gong, L., Olson, M., and Wendel, J.F. 2014. Cytonuclear evolution of Rubisco in four allopolyploid lineages. Mol. Biol. Evol. **31**(10): 2624–2636. doi:10.1093/molbev/msu207. PMID:25015644.

Grover, C., Gallagher, J., Szadkowski, E., Yoo, M., Flagel, L., and Wendel, J.J.N.P. 2012. Homoeolog expression bias and expression level dominance in allopolyploids. New Phytol. **196**(4): 966–971. doi:10.1111/j.1469-8137.2012.04365.x. PMID:23033870.

Harris, R.S. 2007. Improved pairwise alignment of genomic DNA. PhD thesis, The Pennsylvania State University.

Havey, M.J., McCreight, J.D., Rhodes, B., and Taurick, G. 1998. Differential transmission of the *Cucumis* organellar genomes. Theor. Appl. Genet. **97**(1-2): 122–128. doi:10.1007/s001220050875.

Hildebrand, M., Hallick, R.B., Passavant, C.W., and Bourque, D.P. 1988. Trans-splicing in chloroplasts: the *rps12* loci of *Nicotiana tabacum*. Proc. Natl. Acad. Sci. U.S.A. **85**: 372–376. doi:10.1073/pnas.85.2.372. PMID:3422433.

Huang, Y., Li, X., Yang, Z., Yang, C., Yang, J., and Ji, Y. 2016. Analysis of complete chloroplast genome sequences improves phylogenetic resolution in *Paris* (Melanthiaceae). Front. Plant Sci. **7**: 1797. doi:10.3389/fpls.2016.01797. PMID:27965698.

Huson, D., and Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. **23**: 254–267. doi:10.1093/molbev/msj030. PMID:16221896.

Jiao, Y.N., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., et al. 2011. Ancestral polyploidy in seed plants and angiosperms. Nature, **473**(7345): 97–100. doi:10.1038/nature09916. PMID:21478875.

Katoh, K., and Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. **30**(4): 772–780. doi:10.1093/molbev/mst010. PMID:23329690.

Kelchner, S.A. 2000. The evolution of non-coding chloroplast DNA and its application in plant systematics. Ann. Mo. Bot. Gard. **87**(4): 499–527. doi:10.2307/2666142.

Kim, J.S., Jung, J.D., Lee, J.A., Park, H.W., Oh, K.H., Jeong, W.J., et al. 2006. Complete sequence and organization of the cucumber (*Cucumis sativus* L. cv. Baekmibaekdadagi) chloroplast genome. Plant Cell Rep. **25**(4): 334–340. doi:10.1007/s00299-005-0097-y. PMID:16362300.

Kong, W.Q., and Yang, J. H. 2016. The complete chloroplast genome sequence of *Morus mongolica* and a comparative analysis within the Fabidae clade. Curr. Genet. **62**(1): 165–172. doi:10.1007/s00294-015-0507-9. PMID:26205390.

Kurtz, S., Choudhuri, J.V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. **29**(22): 4633–4642. doi:10.1093/nar/29.22.4633. PMID:11713313.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. 2004. Versatile and open software for comparing large genomes. Genome Biol. **5**(2): R12. doi:10.1186/gb-2004-5-2-r12. PMID:14759262.

Leitch, I.J., and Bennett, M.D. 1997. Polyploidy in angiosperms. Trends Plant Sci. **2**(12): 470–476. doi:10.1016/S1360-1385(97)01154-0.

Li, R.Q., Li, Y.R., Kristiansen, K., and Wang, J. 2008. SOAP: short oligonucleotide alignment program. Bioinformatics, **24**(5): 713–714. doi:10.1093/bioinformatics/btn025. PMID:18227114.

Librado, P., and Rozas, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics, **25**(11): 1451–1452. doi:10.1093/bioinformatics/btp187. PMID:19346325.

Lohse, M., Drechsel, O., and Bock, R. 2007. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr. Genet. **52**(5–6): 267–274. doi:10.1007/s00294-007-0161-y. PMID:17957369.

Luo, R.B., Liu, B.H., Xie, Y.L., Li, Z.Y., Huang, W.H., and Yuan, J.Y., et al. 2015. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience, **4**(1): 18–2012. vol. doi:10.1186/s13742-015-0061-x, doi:10.1186/s13742-015-0069-2. PMID:25897398.

Magrane, M., and Consortium, U. 2011. UniProt Knowledgebase: a hub of integrated protein data. Database, **2011**: bar009. doi:10.1093/database/bar009. PMID:21447597.

Marechal, A., and Brisson, N. 2010. Recombination and the maintenance of plant organelle genome stability. New Phytol. **186**(2): 299–317. doi:10.1111/j.1469-8137.2010.03195.x. PMID:20180912.

Massouh, A., Schubert, J., Yaneva-Roder, L., Ulbricht-Jones, E.S., Zupok, A., Johnson, M.T.J., et al. 2016. Spontaneous chloroplast mutants mostly occur by replication slippage and show a biased pattern in the plastome of *Oenothera*. Plant Cell. **28**(4): 911–929. doi:10.1105/tpc.15.00879. PMID:27053421.

McClintock, B. 1984. The significance of responses of the genome to challenge. Science, **226**: 792–801. doi:10.1126/science.15739260.

Morton, B.R. 1995. Neighboring base composition and transversion transition bias in a comparison of rice and maize chloroplast noncoding regions. Proc. Natl. Acad. Sci. U.S.A. **92**(21): 9717–9721. doi:10.1073/pnas.92.21.9717. PMID:7568204.

Morton, B.R. 2003. The role of context-dependent mutations in generating compositional and codon usage bias in grass chloroplast DNA. J. Mol. Evol. **56**(5): 616–629. doi:10.1007/s00239-002-2430-1. PMID:12698298.

Paun, O., Bateman, R.M., Fay, M.F., Hedren, M., Civeyrel, L., and Chase, M.W. 2010. Stable epigenetic effects impact adaptation in allopolyploid orchids (Dactylorhiza: Orchidaceae). Mol. Biol. Evol. **27**(11): 2465–2473. doi:10.1093/molbev/msq150. PMID:20551043.

Plader, W., Yukawa, Y., Sugiura, M., and Malepszy, S. 2007. The complete structure of the cucumber (*Cucumis sativus* L.) chloroplast genome: Its composition and comparative analysis. Cell. Mol. Biol. Lett. **12**(4): 584–594. doi:10.2478/s11658-007-0029-7. PMID:17607527.

Reyes-Prieto, A., Weber, A.P.M., and Bhattacharya, D. 2007. The origin and establishment of the plastid in algae and plants. Annu. Rev. Genet. **41**: 147–168. doi:10.1146/annurev.genet.41.110306.130134. PMID:17600460.

Rodriguez-Moreno, L., Gonzalez, V.M., Benjak, A., Marti, M.C., Puigdomenech, P., Aranda, M.A., and Garcia-Mas, J. 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**: 424. doi:10.1186/1471-2164-12-424. PMID:21854637.

Saski, C., Lee, S.B., Daniell, H., Wood, T.C., Tomkins, J., Kim, H.G., and Jansen, R.K. 2005. Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. Plant Mol. Biol. **59**(2): 309–322. doi:10.1007/s11103-005-8882-0. PMID:16247559.

Sayers, E.W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K.D., and Karsch-Mizrachi, I. 2020. GenBank. Nucleic Acids Res. **48**(D1): D84–D86. doi:10.1093/nar/gkz956. PMID:31665464.

Schmitz-Linneweber, C., Williams-Carrier, R.E., Williams-Voelker, P.M., Kroeger, T.S., Vichas, A., and Barkan, A. 2006. A pentatricopeptide repeat protein facilitates the *trans*-splicing of the maize chloroplast *rps12* pre-mRNA. Plant Cell, **18**(10): 2650–2663. doi:10.1105/tpc.106.046110. PMID:17041147.

Sebastian, P., Schaefer, H., Telford, I.R.H., and Renner, S.S. 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. Proc. Natl. Acad Sci. U.S.A. **107**(32): 14269–14273. doi:10.1073/pnas.1005338107. PMID:20656934.

Sharbrough, J., Conover, J.L., Tate, J.A., Wendel, J.F., and Sloan, D.B. 2017. Cytonuclear responses to genome doubling. Am. J. Bot. **104**(9): 1277–1280. doi:10.3732/ajb.1700293. PMID:29885242.

Shaw, J., Lickey, E.B., Schilling, E.E., and Small, R.L. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. Am. J. Bot. **94**(3): 275–288. doi:10.3732/ajb.94.3.275. PMID:21636401.

Shen, J., Kere, M.G., and Chen, J.F. 2013. Mitochondrial genome is paternally inherited in *Cucumis* allotetraploid (*C. ×hytivus*) derived by interspecific hybridization. Sci. Hortic. **155**: 39–42. doi:10.1016/j.scienta.2013.03.009.

Shen, J., Zhao, J., Bartoszewski, G., Malepszy, S., Havey, M., and Chen, J. 2015. Persistence and protection of mitochondrial DNA in the generative cell of cucumber is consistent with its paternal transmission. Plant Cell Physiol. **56**(11): 2271–2282. doi:10.1093/pcp/pcv140. PMID:26412781.

Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C.F., et al. 2009. Polyploidy and angiosperm diversification. Am. J. Bot. **96**(1): 336–348. doi:10.3732/ajb.0800079. PMID:21628192.

Taylor, W.C. 1989. Regulatory interactions between nuclear and plastid genomes. Annu. Rev. Plant Physiol. **40**: 211–233. doi:10.1146/annurev.pp.40.060189.001235.

Tong, W., Kim, T.S., and Park, Y.J. 2016. Rice chloroplast genome variation architecture and phylogenetic dissection in diverse *Oryza* species assessed by whole-genome resequencing. Rice, **9**(1): 57. doi:10.1186/s12284-016-0129-y. PMID:27757948.

Wakeley, J. 1996. The excess of transitions among nucleotide substitutions new methods of estimating transition bias underscore its significance. Trends Ecol. Evol. **11**: 158–162. doi:10.1016/0169-5347(96)10009-4. PMID:21237791.

Wang, R.J., Cheng, C.L., Chang, C.C., Wu, C.L., Su, T.M., and Chaw, S.M. 2008. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. BMC Evol. Biol. **8**: 36. doi:10.1186/1471-2148-8-36. PMID:18237435.

Wang, X., Dong, Q., Li, X., Yuliang, A., Yu, Y., Li, N., et al. 2017. Cytonuclear variation of Rubisco in synthesized rice hybrids and allotetraploids. Plant Genome, **10**(3): plantgenome2017.05.0041. doi:10.3835/plantgenome2017.05.0041.

Wendel, J.F. 2000. Genome evolution in polyploids. Plant Mol. Biol. **42**(1): 225–249. doi:10.1023/A:1006392424384. PMID:10688139.

Yamane, K., Yano, K., and Kawahara, T. 2006. Pattern and rate of indel evolution inferred from whole chloroplast intergenic regions in sugarcane, maize and rice. DNA Res. **13**(5): 197–204. doi:10.1093/dnares/dsl012. PMID:17110395.

Yu, X.Q., Hyldgaard, B., Rosenqvist, E., Ottosen, C.O., and Chen, J. 2015. Interspecific hybridization in *Cucumis* leads to the divergence of phenotypes in response to low light and extended photoperiods. Front. Plant Sci. **6**: 802. doi:10.3389/fpls.2015.00802. PMID:26483817.

Yu, X.Q., Wang, X.X., Hyldgaard, B., Zhu, Z.B., Zhou, R., Kjaer, K.H., et al. 2018. Allopolyploidization in *Cucumis* contributes to delayed leaf maturation with repression of redundant homoeologous genes. Plant J. **94**(2): 393–404. doi:10.1111/tpj.13865. PMID:29421854.

Zhai, Y., Yu, X., Zhu, Z., Wang, P., Meng, Y., Zhao, Q., et al. 2019. Nuclear-cytoplasmic coevolution analysis of RuBisCO in synthesized *Cucumis* allopolyploid. Genes, **10**(11): 869. doi:10.3390/genes10110869. PMID:31671713.