Journal of Molecular Evolution (2019) 87:327–342
https://doi.org/10.1007/s00239-019-09916-1

ORIGINAL ARTICLE

Evolutionary Dynamics of Transferred Sequences Between Organellar Genomes in *Cucurbita*

Xitlali Aguirre-Dugua[1] · Gabriela Castellanos-Morales[2] · Leslie M. Paredes-Torres[3] · Helena S. Hernández-Rosales[3] · Josué Barrera-Redondo[3] · Guillermo Sánchez-de la Vega[3] · Fernando Tapia-Aguirre[3] · Karen Y. Ruiz-Mondragón[3] · Enrique Scheinvar[3] · Paulina Hernández[3] · Erika Aguirre-Planter[3] · Salvador Montes-Hernández[4] · Rafael Lira-Saade[1] · Luis E. Eguiarte[3]

Received: 29 April 2019 / Accepted: 28 October 2019 / Published online: 7 November 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Twenty-nine DNA regions of plastid origin have been previously identified in the mitochondrial genome of *Cucurbita pepo* (pumpkin; Cucurbitaceae). Four of these regions harbor homolog sequences of *rbcL*, *matK*, *rpl20–rps12* and *trnL–trnF*, which are widely used as molecular markers for phylogenetic and phylogeographic studies. We extracted the mitochondrial copies of these regions based on the mitochondrial genome of *C. pepo* and, along with published sequences for these plastome markers from 13 *Cucurbita* taxa, we performed phylogenetic molecular analyses to identify inter-organellar transfer events in the *Cucurbita* phylogeny and changes in their nucleotide substitution rates. Phylogenetic reconstruction and tree selection tests suggest that *rpl20* and *rbcL* mitochondrial paralogs arose before *Cucurbita* diversification whereas the mitochondrial *matK* and *trnL–trnF* paralogs emerged most probably later, in the mesophytic *Cucurbita* clade. Nucleotide substitution rates increased one order of magnitude in all the mitochondrial paralogs compared to their original plastid sequences. Additionally, mitochondrial *trnL–trnF* sequences obtained by PCR from nine *Cucurbita* taxa revealed higher nucleotide diversity in the mitochondrial than in the plastid copies, likely related to the higher nucleotide substitution rates in the mitochondrial region and loss of functional constraints in its tRNA genes.

Keywords Gene duplication · Inter-organellar DNA transfer · Molecular evolution · Paralogy · Nucleotide substitution rate · tRNA

Handling Editor: **Rafael Zardoya**.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00239-019-09916-1) contains supplementary material, which is available to authorized users.

✉ Xitlali Aguirre-Dugua
xaguirre@cieco.unam.mx

✉ Rafael Lira-Saade
rlira@unam.mx

✉ Luis E. Eguiarte
fruns@unam.mx

[1] Unidad de Biotecnología Y Prototipos, Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México, Av. De Los Barrios 1, Col. Los Reyes Iztacala, 54090 Tlalnepantla, Estado de México, Mexico

[2] Departamento de Conservación de La Biodiversidad, El Colegio de La Frontera Sur, Unidad Villahermosa, Carretera Villahermosa-Reforma km. 15.5, Ranchería El Guineo 2a Sección, 86280 Villahermosa, Tabasco, Mexico

[3] Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Circuito Exterior S/N Anexo Al Jardín Botánico, 04510 Ciudad de México, Mexico

[4] Campo Experimental Bajío, Instituto Nacional de Investigaciones Forestales, Agrícolas Y Pecuarias (INIFAP), Km 6.5 Carretera Celaya-San Miguel de Allende, 38110 Celaya, Gto., Mexico

Introduction

Comparative studies across plant lineages have shown that plastid and mitochondrial genomes (i.e., plastomes and chondromes, respectively) can exhibit different degrees of structural complexity (Smith and Keeling 2015). Plastomes display a quadripartite structure (with two inverted repeats [IRs], and two single-copy regions) whose main gene content and synteny have been conserved during land plant evolution. Most of the observed variation in the plastome architecture is due to rearrangements and gene losses associated to the transfer of gene clusters to the nuclear genome, to the expansion, contraction or loss of IRs, and to the reorganization of small dispersed repeats (Wicke et al. 2011). Gene gains in the plastome have occurred exceptionally, as with *matK* and *ycf1/2* during the early evolution of the plant kingdom (Wicke et al. 2011). The most dramatic departures in plastome architecture have been observed in lineages with a heterotrophic lifestyle, although many land plant groups have also undergone distinct plastome rearrangements (e.g., Sanderson et al. 2015; Chaw et al. 2018).

Plastome conservatism has been attributed to several factors, including the compact organization of genes in operons, copy-correction activity of IRs, constraints in trans-membrane protein-importing mechanisms and the action of natural selection over photosynthetically relevant genes (Wicke et al. 2011; Zhu et al. 2016).

When compared to plastomes, plant chondromes display a larger variation in size, gene number and organization, as well as greater intronic content and peculiar RNA editing mechanisms, many of which have been described as evolutionary 'eccentricities' (Sloan et al. 2012a; Smith and Keeling 2015). Chondromes exhibit a high propensity for fission and fusion (which is almost non-existent in plastomes), and are the subject of homologous recombination that can modify gene order among closely allied species and even among individuals from the same species (Alverson et al. 2011; Sloan et al. 2012b). It is noteworthy that plant chondromes have an active DNA import system that allows the uptake of foreign DNA (Scott and Logan 2011; Smith and Keeling 2015). The acquisition of large tracts by duplication and transfer from other genetic compartments within the cell, and even entire mitochondrial genomes from different species (Rice et al. 2013; Sanchez-Puerta et al. 2019), is one of the most distinguishing features of the plant mitochondrial genome, contributing to much of its structural complexity and size expansion (Adams et al. 2002; Goremykin et al. 2012; Knoop 2012).

Seed plant chondromes are also characterized by exceptionally low rates of nucleotide substitution in protein-coding genes when compared to other genetic compartments (about one third and one sixth of the plastome and nuclear silent substitution rates, respectively; Wolfe et al. 1987). To explain this pattern, Lynch et al. (2006) proposed the mutation pressure hypothesis, which postulates that the low mutational environment of the chondrome allows the acquisition and accumulation of non-coding sequences leading to the growth of mitochondria genome sizes. However, Alverson et al. (2010) reported an increased synonymous substitution rate in the coding sequences of the large *Cucurbita* chondrome when compared to the smaller *Citrullus* mitochondrial genome, challenging the expectations of the mutation pressure hypothesis of Lynch et al. (2006) in the Cucurbitaceae family.

The species of the *Cucurbita* genus show exceptionally high rates of molecular evolution on both their nuclear (Barrera-Redondo et al. 2019) and mitochondrial genomes. The mitochondrial genome of *Cucurbita pepo* (Cucurbitaceae; ca. 982 kb) has accumulated large amounts of plastome-like sequences (>113 kb or 11.5%), divided among 29 different regions (Alverson et al. 2010). This amount of plastid-like sequences represents about 1.7 to 29 times more plastid DNA than other fully sequenced plant mitochondrial genomes, including *Arabidopsis thaliana*, *Zea mays*, *Oryza sativa*, *Beta vulgaris*, *Nicotiana tabacum*, *Vitis vinifera* (Goremykin et al. 2009; Alverson et al. 2010), and *Silene*, which harbors the largest known plant mitochondrial genome (11.3 Mb; Sloan et al. 2012a, b). The *C. pepo* mitochondrial genome encodes 24 tRNAs in the plastid-derived sequences (15 of them intact), many of which are located in syntenic regions of the plastid sequence (Alverson et al. 2010).

Among the plastid-like mitochondrial regions, four have been observed to be highly similar to plastid regions harboring *rbcL*, *matK*, *rpl20*, *rps12, trnL*, and *trnF* genes. The plastid *rbcL* and *matK* genes, next to the *rpl20–rps12* spacer, the *trnL* intron, and the *trnL–trnF* spacer, have been widely used as markers for studying phylogenetic relationships both at deep (major lineages of angiosperms) and shallow (phylogeographic studies among closely related species) phylogenetic levels, not only within Cucurbitaceae but also in a wide array of land plant groups (Shaw et al. 2005; Borsch and Quandt 2009; Zheng et al. 2013). The usefulness of non-coding DNA (i.e., introns and intergenic spacers) from the plastome stems from the accumulation of a relatively higher number of microstructural rearrangements than other plastid regions, including simple sequence repeats, inversions, indels and nucleotide substitutions (Kelchner 2000; Borsch and Quandt 2009).

The aim of this study is to analyze the evolutionary dynamics of the mitochondrial and plastid copies of the *rbcL*, *matK*, *rpl20–rps12*, and *trnL–trnF* regions in *Cucurbita*, including cultivated (five taxa) and wild representatives (eight taxa). Recently, some patterns of the evolution

of the *Cucurbita* genus, including biogeographic, molecular clocks and diversification aspects, were analyzed by Castellanos-Morales et al. (2018) using different plastid regions. Other studies have explored the phylogeny and evolutionary patterns for the genus using chloroplast *rbcL*, *matK*, *rpl20–rps12*, and *trnL–trnF* sequences (Zheng et al. 2013), described nearly complete chloroplast genomes and archeological material (Kistler et al. 2015), and analyzed nuclear genes (Kates et al. 2017), or a mitochondrial intron (Sanjur et al. 2002). In particular, the presence of paralogs in the chondrome offers an opportunity to identify a novel marker useful for evaluating phylogenetic relationships previously assessed with plastid sequences (Bailey et al. 2003), to explore the recent molecular evolution of sequences in a new mitochondrial environment and the evolutionary forces acting on the intracellular transfer of DNA sequences.

Materials and Methods

Identification of Putative Homologous Regions

MAUVE whole genome alignment (Darling et al. 2004) was used to identify the locally collinear blocks between the *C. pepo* plastid (GenBank NC_038229; Zhang et al. 2018) and mitochondrial (NC_014050; Alverson et al. 2010) genomes. Locally collinear blocks (LCB) represent regions from two genomes that align to each other, presumably homologous and free of internal rearrangements. We used MAUVE's plugin in Geneious v9.1 software (Kearse et al. 2012) whose progressive algorithm was applied with default values of automatic seed weight detection and LCB score. We then identified the four LCBs containing the *rbcL*, *matK*, *rpl20–rps12,* and *trnL–trnF* regions. Additional plastome regions previously used in *Cucurbita* evolution studies such as *psbJ–petA*, *psbD–trnT*, and *atpI–atpH* (Castellanos-Morales et al. 2018) did not show putative homologous regions in the mitochondrial genome. For the four LCBs harboring each of *rbcL*, *matK*, *rpl20–rps12,* and *trnL–trnF* regions, we first extracted the plastid and chondrome sequences that form the entire LCB and aligned them using MAFFT v7.308 plugin (Katoh et al. 2002) from Geneious v9.1, with automatic algorithm selection, scoring matrix 1PAM/k = 2, gap open penalty set at 1.53 and Offset value at 0.123. Using the original gene annotations from the NC_038229 plastid genome sequence we identified the boundaries of the region of interest (v.g., *rbcL*), and then extracted the putatively homologous (based on the alignment) mitochondrial sequence.

In the case of *rbcL* and *matK* genes we evaluated if mitochondrial paralogs still harbored an open reading frame. Their dN/dS index was measured with DnaSP v6.12.01 (Librado and Rozas 2009) by comparing the plastome NC_038229 and the chondrome NC_014050 gene copies.

In the case of the *trnL–trnF* region, we included its mitochondrial homolog (designated as mt-*trnL–trnF*) for nine *Cucurbita* taxa obtained by PCR in this study (see next section) and an additional mt-*trnL–trnF* sequence from *Citrullus lanatus* (Cucurbitaceae) due to the report of a similar plastid-like region containing *trnL* and *trnF* genes in *Citrullus lanatus* (Cucurbitaceae) mitochondrial genome NC_014043 (Alverson et al. 2010). We performed an alignment using MAUVE between the mitochondrial genomes of *C. pepo* (NC_014050) and *Citrullus lanatus* (NC_014043) to identify the LCB harboring the mt-*trnL–trnF* region in *Citrullus lanatus*. Sequences from the LCB were extracted and aligned with MAFFT as previously described.

Amplification of Mitochondrial *trnL–trnF* Copies

Taberlet et al. (1991) *a-h* universal primers were aligned in silico using Primer3 v2.3.4 (Untergasser et al. 2012) in Geneious v9.0 against the mitochondrial genome of *C. pepo* NC_014050. Primers *c* and *f* align to positions 12,972 and 11,887, respectively, of the mitochondrial genome. Primer *f* matches perfectly, whereas primer *c* matches with a single mismatch (T/G). The remaining primers (*a, b, d, e, g, h*) did not match to any position. The mt-*trnL–trnF* region was thus expected to be 1,085 bp in length (Fig. 1). Originally designed to align with the plastid genome, primers *c* and *f* were also expected to amplify the plastid *trnL–trnF* region (here designed cp-*trnL–trnF*), including both the *trnL* intron and the *trnL–trnF* spacer in a single amplification reaction. In silico alignment of Taberlet's *c* and *f* primers to *C. pepo* subsp. *pepo* NC_038229 (Zhang et al. 2018), *C. moschata* NC_036506, and *C. maxima* NC_036505 (Zhu et al. 2017) complete plastomes indicated for the cp-*trnL–trnF* region an expected product of 1687, 1877, and 1376 bp, respectively. It is worth mentioning that published data on the cp-*trnL–trnF* region in *Cucurbita* have been generated using separate reactions for amplifying the cp-*trnL* intron (using Taberlet's primers *c* and *d*) and the cp-*trnL–trnF* spacer (with primers *e* and *f*) (Kocyan et al. 2007; Zheng et al. 2013). Furthermore, Zheng et al. (2013) used a modified version of primer *f* (here designed *f'*) to amplify the region in "certain taxa", as stated by the authors. In silico essays in *C. pepo* NC_038229, *C. moschata* NC_036506 and *C. maxima* NC_036505 show that *f'* aligns with the first of several *trnF* gene copies present in that region of the plastome, whereas Taberlet's original *f* primer aligns ca. 700, 400, and 880 bp downstream of the modified *f'* primer in *C. pepo, C. maxima,* and *C. moschata* plastomes, respectively. That being said, Kocyan et al. (2007) and Zheng et al. (2013) data

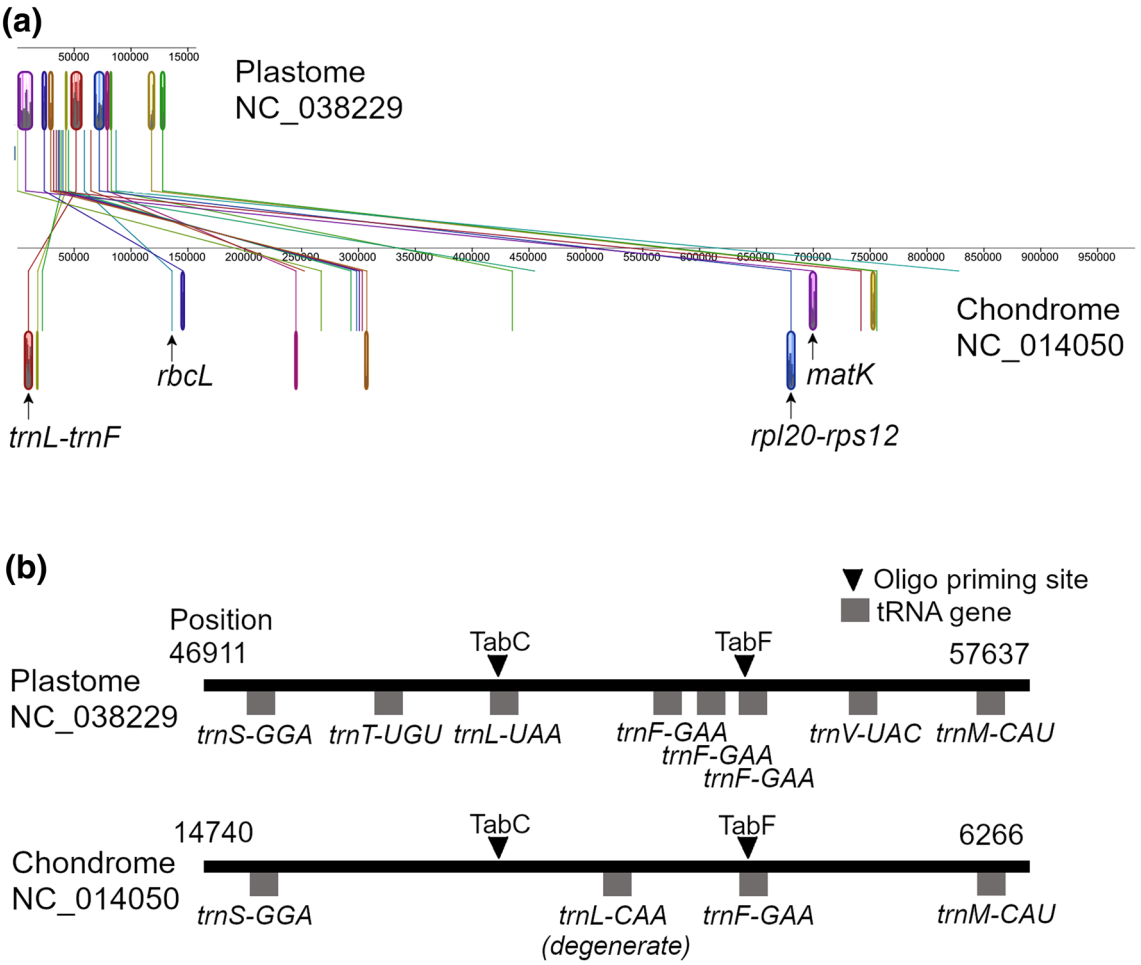

Fig. 1 **a** Schematic drawing depicting the 20 putatively homologous locally collinear blocks (LCB) identified between *Cucurbita pepo* plastid NC_038229 (top) and mitochondrial NC_014050 (bottom) genomes (shown as colored rectangles). The four LCBs containing the *rbcL*, *matK*, *rpl20–rps12* and *trnL–trnF* regions are shown with arrows. **b** Structural comparison, not to scale, between the LCB containing the *trnL–trnF* region in both organellar genomes (see also Table 2). Only tRNA genes are shown as from the original annotations of each reference genome, next to aligning sites of Taberlet et al. (1991) *c* (TabC) and *f* (TabF) primers

on the cp-*trnL–trnF* region of *Cucurbita* species (*trnL* intron plus *trnL–trnF* spacer) range in size from 913 to 942 bp.

We performed PCR essays in order to amplify the mt-*trnL–trnF* mitochondrial copy using total DNA extracted from leaf tissue using primers *c* and *f* from Taberlet et al. (1991). A temperature gradient was evaluated from 54 to 60 °C to identify the best conditions for the amplification of mt-*trnL–trnF* in 13 *Cucurbita* taxa, including eight wild and five domesticated (Table 1). PCR reactions were performed with dNTPs 10 mM, $MgCl_2$ 2 mM, primers 3 uM each, and Taq polymerase 2 U, in a final volume of 25 ul. Amplification reaction was as follows: 94 °C 5 min, 35 cycles of 94° 40 s, 54–60 °C 40 s, 72° 1:20 min, and a final elongation cycle of 72° for 7 min.

PCR products were visualized on 1% agarose gels stained with SafeGreen and were Sanger-sequenced in both directions at Macrogen (MA, USA). After verifying chromatograms quality, low quality calls on the 3′ and 5′ ends were suppressed. We performed standard nucleotide BLAST queries on the nr/nt nucleotide collection database from the National Center for Biotechnology Information (NCBI) using Megablast (optimized for highly similar sequences) without further search restrictions. Hit results were used to identify the genomic origin of the amplified product (i.e., *Cucurbita* plastome or *Cucurbita* chondrome), considering the smallest E-value and the greatest MaxScore, MaxIdentity and QueryCover. Based on the BLASTn results, two matrices were built, one for sequences of plastome origin (cp-*trnL–trnF*) and another for sequences of chondrome origin (mt-*trnL–trnF*) (Suppl. Table S1). Sequences obtained in this study are available at GenBank under accession numbers MN381966, MN381967, MN411619-MN411628, MN481396-MN481399.

Table 1 Published *rbcL, matK*, *rpl20–rps12* and *trnL–trnF* sequences from *Cucurbita* plastomes (Kocyan et al. 2007; Zheng et al. 2013; Castellanos-Morales et al. 2018) included in this study

Habitat	Taxon	Status	GenBank accession number			
			rbcL	*matK*	*rpl20–rps12*	*trnL–trnF*
Mesophytic	*Cucurbita lundelliana*	W	HQ438619	HQ438601	MH470049	HQ438665
Mesophytic	*C. okeechobeensis* subsp. *martinezii*	W	HQ438621	HQ438604	MH470048	DQ536766
Mesophytic	*C. okeechobeensis* subsp. *okeechobeensis*	W	HQ438620	HQ438605	HQ438644	HQ438666
Mesophytic	*C. pepo* subsp. *fraterna*	W	HQ438622	HQ438608	MH470046	KT898816
Mesophytic	*C. pepo* var. *ovifera*	D	HQ438626	HQ438609	MH470047	HQ438672
Mesophytic	*C. pepo* subsp. *pepo*	D	HQ438623	HQ438611	MH470045	HQ438669
Mesophytic	*C. argyrosperma* subsp. *argyrosperma*	D	HQ438617	HQ438592	MH470043	HQ438663
Mesophytic	*C. argyrosperma* subsp. *sororia*	W	HQ438616	HQ438595	MH470044	HQ438662
Mesophytic	*C. maxima*	D	HQ438627	HQ438602	MH470050	HQ438673
Mesophytic	*C. andreana*	W	HQ438628	HQ438590	HQ438652	HQ438674
Mesophytic	*C. ecuadorensis*	W	HQ438630	HQ438598	MH470051	HQ438676
Mesophytic	*C. moschata*	D	HQ438614	HQ438603	MH470042	HQ438660
Xerophytic	*C. palmata*	W	HQ438634	HQ438606	MH470058	HQ438680
Xerophytic	*C. digitata*	W	HQ438636	DQ536664	MH470057	HQ438682
Xerophytic	*C. cordata*	W	HQ438635	HQ438597	MH470056	HQ438681
Xerophytic	*C. pedatifolia*	W	HQ438633	HQ438607	MH470053	HQ438679
Xerophytic	*C. foetidissima*	W	HQ438632	HQ438600	MH470054	HQ438678
–	*Peponopsis adhaerens*	W	DQ535766	DQ536716	DQ536580	DQ536848
–	*Polyclathra cucumerina*	W	–	DQ536717	DQ536628	DQ536849
–	*Anacaona sphaerica*	W	EU036995	DQ536756	DQ536528	DQ536785
–	*Penelopeia suburceolata*	W	DQ535834	DQ536713	DQ536578	DQ536847
–	*Citrullus lanatus*	D	DQ535791	DQ536650	DQ648162	DQ536798
–	*Cucumis melo* subsp. *melo*	D	DQ535800	DQ536659	NC015983	NC015983
–	*Bambekea racemosa*	W	DQ535783	–	DQ536541	DQ536788
–	*Ibervillea lindheimeri*	W	DQ535821	DQ536690	DQ648178	DQ536832
–	*Dendrosicyos socotrana*	W	–	AY973018	–	–

GenBank sequences for *rpl20–rps12* (Castellanos-Morales et al. 2018) correspond with the same individuals from which we amplified the mt-*trnL–trnF* copies. Sequences obtained from organellar genomes of Kistler et al. (2015) and Zhang et al. (2006) were extracted using their original annotations. Status *W* wild, *D* domesticated. Habitat as from Lira-Saade (1995)

Phylogenetic Analyses

For each marker, we built a matrix based on available sequences of *Cucurbita* found on GenBank (Table 1) corresponding to the plastid copies, and we added the corresponding *C. pepo* mitochondrial sequence representing the putative paralog. The sequences include representatives of all *Cucurbita* main groups, including xerophytic and mesophytic species (Kocyan et al. 2007; Zheng et al. 2013; Kistler et al. 2015; Castellanos-Morales et al. 2018). The former are perennials characterized by the presence of storage tuberous roots, whereas the latter harbor fibrous root systems and are annual or short-lived perennials (Lira-Saade 1995). We also included plastid marker sequences of *Peponopsis, Polyclathra, Anacaona,* and *Penelopeia* from tribe Cucurbiteae, *Citrullus* and *Cucumis* from Benincaseae, and *Ibervillea*, *Bambekea,* and *Dendrosicyos* from Coniandreae (Table 1; Kocyan et al. 2007; Zhang et al. 2006).

For each marker, the entire matrix (plastid sequences of *Cucurbita* and related genera, plus the mitochondrial *C. pepo* putative paralog from reference genome NC_014050) was submitted to a phylogenetic analysis aimed at identifying sister relationships between organellar copies. For each matrix the alignment was performed with MAFFT as previously described, and carefully verified manually (Kelchner 2000). We used jModelTest v2.1.9 (Guindon and Gascuel 2003; Darriba et al. 2012) to identify the best substitution model of nucleotides, with default parameters (11 substitution schemes, allowing for differences in base frequencies, rate variation, base tree ML optimized, and NNI base tree search). The corrected Akaike Information Criterion (AICc) was used to select the best model fitting the data.

Genealogical relationships between the plastome and chondrome copies were estimated through a Bayesian Markov chain MonteCarlo (MCMC) approach as implemented in MrBayes v3.2.6 (Ronquist and Huelsenbeck, 2003), using the complex gap-coding approach of Müller (2006) (with Seqstate v1.4.1; Müller, 2005) to add as a second partition of discrete characters to the data matrix. A majority rule consensus tree was built after running one to two million generations of the MCMC chain, a burn-in fraction of 0.25 and sampling every 500 generations. Convergence was considered to be attained when the minimum effective sample size (ESS) was > 200 and the average standard deviation of split frequencies was below 0.01. We also built a maximum likelihood tree as implemented in PhyML v3.1 (Guindon and Gascuel, 2003) with one thousand bootstrap replicates to confirm tree topology.

We evaluated a set of competing hypothesis on the position of the mitochondrial paralog along the phylogenetic tree through the Approximately Unbiased (AU) test of Shimodaira (2002). This test performs a multiscale bootstrap of site-wise log-likelihoods in order to build a null distribution of the log-likelihood difference of each candidate tree from the best tree in the set. As the tree set is specified a priori, this distribution represents the null hypothesis that candidate tree scores are the same, and is used to calculate the *p*-value of each tree. This *p*-value is considered a measure of the possibility that the tree is the true tree. The set of trees that are not rejected by the test (whose *p*-value is not smaller than the confidence level, $\alpha = 0.05$) is known as the confidence set. Here, the site-wise log-likelihoods were calculated with PAML v4.9 h (Yang 2007), and the bootstrap and *p*-value calculations were performed with CONSEL v. 0.1i (Shimodaira and Hasegawa 2001). For *rbcL*, *matK* and *rpl20–rps12* we compared 12 possible trees where the mitochondrial paralog arose in different interior branches, from a very recent transfer event within *C. pepo* subsp. *pepo* lineage, to a very ancient event in the common ancestor of tribes Cucurbiteae, Benincaseae and Coniandreae (Fig. 3a). For mt-*trnL–trnF* we considered 10 competing trees, considering from 9 independent transfer events within separate lineages, to a single transfer event in the common ancestor of tribes Cucurbiteae, Benincaseae and Coniandreae (Fig. 3b).

Substitution rates along tree branches of the phylogeny were estimated with BEAST v1.10.1 (Suchard et al. 2018) based on the GTR + G model, under a Yule-speciation process and a log normal relaxed clock model. The relaxed clock model allows a large amount of rate variation across the tree by assigning each branch with its own nucleotide substitution rate (Drummond et al. 2006). We time-calibrated the trees using as secondary calibration points the estimated dates of divergence between genus *Cucurbita* and *Peponopsis* (mean ± SD = 16 ± 4 mya) and between tribes Cucurbiteae and Benincaseae (30 ± 4 mya) of Schaefer et al. (2009). We also set a maximum age constraint of 20 mya to the node leading to *Penelopeia suburceolata* and *Anacaona sphaerica*, two endemics found in the Hispaniola island (Antilles). This is the maximum age estimated for Dominican amber, which represents a proxy for the establishment of tropical forest in the island (Iturralde-Vinent and MacPhee 1996). Nucleotide substitution rates were obtained by selecting output trees with branch lengths in substitution units (substitutions/site/million years). Two independent runs of one thousand million generations were performed to confirm convergence of the parameter values and attain effective sample sizes (ESS) > 200 during the sampling process. Results from both runs were merged with LogCombiner v1.10.1.

A likelihood-ratio test was applied to formally test whether chondrome copies had significantly higher nucleotide substitution rates than plastome copies. The null model of a global clock (every branch with the same rate) was compared to an alternative local clock model where mitochondrial copies had a different rate (with PAML v4.9). Number of segregating sites (*S*) and number of parsimony informative sites (*I*) were calculated as indicators of variability in each marker with DnaSP v6.12.01.

The nucleotide diversity π (Nei 1987) and average number of nucleotide differences *k* (Tajima 1983) within each group of *trnL–trnF* organellar copies were calculated with DNAsp v6.12.01.

tRNA Detection in the mt-*trnL–trnF* Region

The entire sequence of the LCB of *C. pepo* NC_014050 mitochondrial genome harboring the mt-*trnL–trnF* region (Fig. 1b) was analyzed using ARAGORN v1.2.38 (Laslett and Canback 2004) and tRNAscan-SE v2.0 (Lowe and Chan 2016) to confirm Alverson et al. (2010) annotations. Potential tRNA genes found on the mt-*trnL–trnF* copies obtained by PCR in this study were also annotated using the aforementioned programs.

Results

Putative Homologous Regions Between Organellar Genomes

Twenty locally collinear blocks (LCBs) were identified with MAUVE as putative homologous regions between the *C. pepo* plastome and chondrome (Fig. 1). All twenty mitochondrial sequences with a homologous plastid region harbored Alverson et al.'s (2010) annotations on their putative plastid origin. Nine regions from Alverson et al.'s (2010) NC_014050 chondrome originally annotated as plastid-like

were not identified with MAUVE as putatively homologous to NC_038229 plastome, very likely due to their small size < 700 bp. The positions and lengths of lineally collinear blocks harboring the four regions analyzed in this study are shown in Table 2.

Specifically, the mitochondrial *rbcL* sequence (mt-*rbcL*) represents only a fraction of the plastome *rbcL* gene (cp-*rbcL*). While cp-*rbcL* gene is 1,449 bp long, the mt-*rbcL* measures 711 bp, which are homologous to 671 bp of the cp-*rbcL* copy. In other words, mt-*rbcL* lacks the first 47 bp (including the start codon) and the last 731 bp of cp-*rbcL*. A codon-based alignment (without indels and considering that stop codons could code for a rare aminoacid) allowed us to calculate a dN/dS index between *rbcL* paralogs of 1.5823. The mitochondrial copy of *matK* (mt-*matK*) is also shorter (924 bp) than its plastome counterpart (cp-*matK*, 1,535 bp) and despite the presence of a start codon, a series of indels led to a premature stop codon after 270 bp. The dN/dS index between *matK* paralogs was 1.4693.

Phylogenetic Analyses

Differentiation of xerophytic and mesophytic species groups within the genus was resolved in *matK* and *trnL–trnF* trees, as reported by previous studies based on plastome sequences (Zheng et al. 2013; Castellanos-Morales et al. 2018), whereas sister relationship between tribes Cucurbiteae (including *Cucurbita, Peponopsis*, *Polyclathra*, *Anacaona*, and *Penelopeia*) and Benincaseae (including *Citrullus* and *Cucumis*; Kocyan et al. 2007) was resolved in *rbcL*, *matK* and *trnL–trnF* trees (Fig. 2).

The *rpl20–rps12* tree was the one with the lowest overall resolution, followed by the *rbcL* tree, where xerophytic and mesophytic groups are not separated. The lack of resolution in these two markers is in accordance with a lower number of segregating and parsimony informative sites (for *rbcL* $S = 55$, $I = 12$; *rpl20–rps12* $S = 45$, $I = 7$), whereas the *matK* and the *trnL–trnF* regions displayed the largest number of informative sites (*matK* $S = 55$, $I = 20$; *trnL–trnF* $S = 124$, $I = 103$). This result concurs with a lower nucleotide substitution rate in *rbcL* and *rpl20–rps12* regions when compared to *matK* and *trnL–trnF* (see below). In contrast, the *trnL–trnF* region displayed the highest support values along the tree (Fig. 2). The higher phylogenetic signal harbored by *trnL–trnF* when compared to the remaining three plastome regions was already pointed out by Zheng et al. (2013). Maximum Likelihood (ML) phylogenetic trees were characterized by nearly identical topologies to those obtained with the Bayesian approach, although with low bootstrap support values (Suppl. Figure S1).

The majority rule consensus trees for each of the four regions differ in the position of the mitochondrial paralog and in the support of their branches (Fig. 2). In the *rbcL* and *rpl20–rps12* trees the putative mitochondrial paralog was grouped with the remaining *Cucurbita* and *Peponopsis* plastome copies, and nodes were characterized by low support values (posterior probabilities < 0.90), indicating that mitochondrial copies could have arisen before or after the *Cucurbita-Peponopsis* divergence. The AU test indicates that for *rbcL* the best tree is the one where the mitochondrial paralog arose in the ancestor of the *Cucurbita* genus (tree #7), excluding the possibility that the transfer event occurred before the rise of Cucurbiteae tribe; in contrast, the origin of *rpl20–rps12* mitochondrial paralog in the common ancestor of *Peponopsis* and *Cucurbita* (tree #8) or in the ancestor of *Cucurbita* only (tree #7) displayed a very similar likelihood (Table 3).

In the case of *matK* and *trnL–trnF*, the mitochondrial paralogs were sister to plastid copies belonging to the mesophytic *Cucurbita* clade only (posterior probabilities of 1), which suggests that transfer from the plastome to the chondrome occurred in the common ancestor of the mesophytic group, after its divergence from xerophytic species. In the case of the mt-*trnL–trnF* sequences obtained from nine *Cucurbita* taxa (Suppl. Table S1), the phylogenetic tree revealed that all the mt-*trnL–trnF* copies form a strongly supported monophyletic group (Bayesian posterior probability of 1), sister to the remaining cp-*trnL–trnF* of mesophytic taxa. The origin of the mt-*trnL–trnF* copy in the common ancestor of mesophytic clade could explain the absence of mt-*trnL–trnF* sequences on our PCR essays on xerophytic

Table 2 Positions and sizes of the sequences forming locally collinear blocks (LCB) between organellar genomes containing putatively homologous regions *rbcL*, *matK*, *rpl20–rps12*, and *trnL–trnF*

	Cucurbita pepo				*Citrullus lanatus*	
	Plastome NC_038229		Chondrome NC_014050		Chondrome NC_014043	
	Start/end positions	Size (bp)	Start/end positions	Size (bp)	Start/end positions	Size (bp)
rbcL	58,743 / 59,552	810	136,873 / 136,163	711	-	-
matK	483 / 14,605	14,123	696,349 / 703,758	7,409	-	-
rpl20–rps12	68,011 / 77,142	9,131	684,854 / 676,285	8,569	-	-
trnL–trnF	46,911 / 57,637	10,726	14,740 / 6,266	8,474	139,271 / 142,812	3,541

When end position precedes start position, the mitochondrial sequence has an inverted orientation relative to the plastid putative homologous sequence

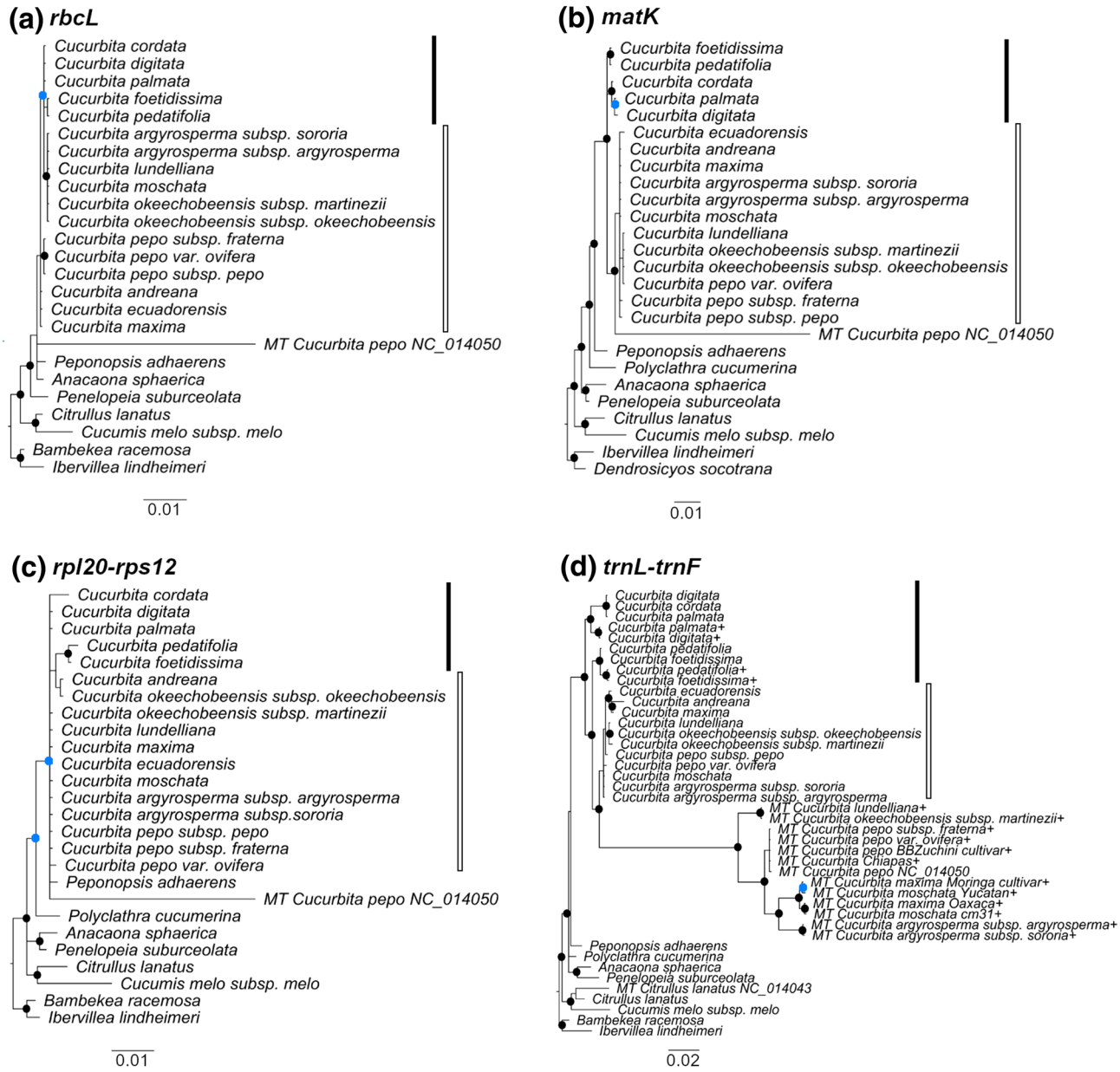

Fig. 2 Majority rule consensus trees obtained with MrBayes including putative homologous plastome and chondrome copies of **a** *rbcL* (matrix size: 1,491 characters), **b** *matK* (1,202 characters), **c** *rpl20–rps12* (798 characters) and **d** *trnL–trnF* (1,287 characters) regions. Chondrome copies are shown with the *MT* prefix, all others represent plastome copies. Sequences obtained in this study are shown with a + symbol, whereas the remaining sequences were obtained from GenBank (see Table 1). Posterior probabilities are shown with dots: black (>0.95), blue (0.90–0.95), green (0.85–0.90). Vertical bars indicate the habitat of the taxon: xerophytic (black) or mesophytic (white). The best substitution model identified for each of the analyzed regions is shown in Suppl. Table S2 (Color figure online)

C. palmata, *C. digitata*, *C. pedatifolia,* and *C. foetidissima* (Suppl. Table S1). On the other hand, the mt-*trnL–trnF* region of *Citrullus* appears as sister to the cp-*trnL–trnF* region of the same species, suggesting an independent event of transfer of this region between organellar genomes in tribe Benincaseae (Fig. 2d). These patterns were confirmed by the AU test, given that the best tree for *matK* was the one where the mitochondrial paralog arose in the ancestor of mesophytic *Cucurbita* (tree #5), and a transfer event preceding the origin of the *Cucurbita* genus was excluded from the confidence tree set. However, despite the general agreement between the AU test results and Bayesian phylogenetic trees of Fig. 2, the AU test did not exclude from the confidence set those trees where the mitochondrial paralogs arose in more recent events (trees #1–6 in Table 3). The possibility therefore remains (although with lower probability) that

Table 3 Approximately Unbiased (AU) test for tree selection comparing different topological positions of mitochondrial paralogs (see Fig. 3a)

Candidate tree	*rbcL*			*matK*			*rpl20–rps12*		
	T_i	*p*-value	rank	T_i	*p*-value	rank	T_i	*p*-value	rank
1	3.9784	0.056	10	3.7760	0.056	7	1.7322	0.157	6
2	3.5400	0.081	9	*2.3469*	*0.013*	*4*	1.7321	0.161	3
3	0.4044	0.533	4	1.0176	0.255	2	1.7322	0.160	5
4	0.4149	0.573	6	1.0184	0.250	3	1.7322	0.156	4
5	0.4068	0.482	5	**− 1.0176**	**0.920**	**1**	1.7324	0.156	7
6	0.8958	0.293	8	2.3583	0.218	6	1.7325	0.159	8
7	**− 0.0256**	**0.789**	**1**	2.3561	0.218	5	**− 0.0014**	**0.901**	**1**
8	0.1059	0.565	3	*8.9845*	*0.009*	*8*	0.0014	0.902	2
9	0.0256	0.602	2	*11.3130*	*0.002*	*9*	*5.9123*	*0.007*	*11*
10	0.4722	0.541	7	*15.2563*	*0.001*	*10*	*6.0657*	*0.007*	*12*
11	*6.9891*	*0.034*	*12*	*17.0657*	*0.001*	*11*	3.8659	0.293	10
12	*6.4879*	*0.022*	*11*	*17.0666*	*0.001*	*12*	3.8656	0.296	9

The tree with the largest *p*-value is shown in bold, whereas the trees excluded from the confidence set ($p < 0.05$) are shown in italics. Ranks correspond to increasing values of *Ti*. *Ti*: log-likelihood difference between each tree and the tree with the maximum log-likelihood

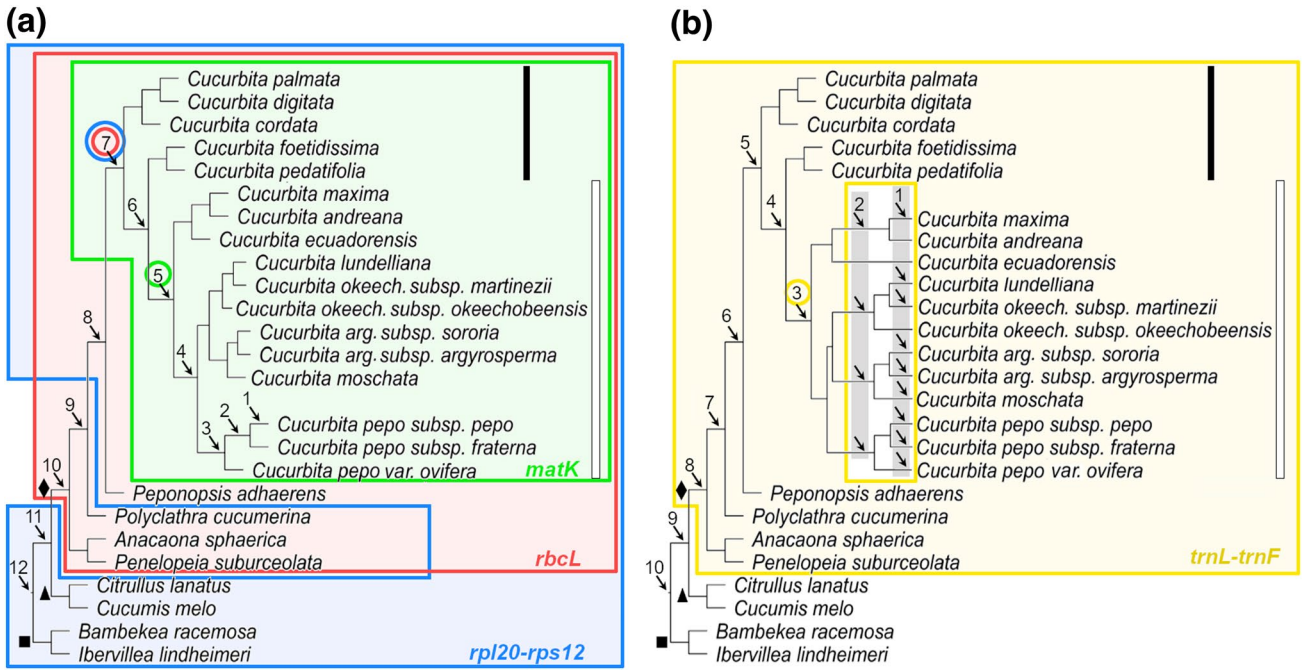

Fig. 3 Assessment of 12 different topological positions of **a** *rbcL*, *matK* and *rpl20–rps12* and b *trnL–trnF* mitochondrial paralogs with the Approximately Unbiased (AU) test of Shimodaira (2000). Each hypothesized position of the mitochondrial paralog among the remaining plastid sequences is shown with an arrow and corresponds to a different candidate tree. In **b** trees 1 and 2 represent 9 or 4 independent transfer events, respectively, whereas trees 3–10 represent a single ancient transfer event. Color boxes represent the confidence tree set for each marker whereas circles outline the best tree among the competing trees (*rbcL*: red, *matK*: green, *rpl20–rps12*: blue, *trnL–trnF*: yellow; see Tables 3 and 4). Black and white vertical lines depict xerophytic and mesophytic taxa, respectively. Black symbols depict branches subtending tribes Cucurbiteae (diamond), Benincaseae (triangle) and Coniandreae (square). The backbone plastid topology is based in Castellanos-Morales et al. (2018) and Kocyan et al. (2007) (Color figure online)

Table 4 Approximately Unbiased (AU) test for tree selection comparing different topological positions of mt-*trnL–trnF* paralog (see Fig. 3b)

Candidate tree	*trnL–trnF*		
	T_i	*p*-value	rank
1	*1803.8609*	*9e-008*	*10*
2	*695.7967*	*3e-008*	*9*
3	**− 2.3729**	**0.803**	**1**
4	2.3729	0.422	2
5	3.8725	0.392	3
6	7.3080	0.143	5
7	7.1564	0.322	4
8	11.4288	0.170	6
9	*16.0046*	*0.018*	*8*
10	*11.8976*	*0.012*	*7*

The tree with the largest *p*-value is shown in bold, whereas the trees excluded from the confidence set ($p < 0.05$) are shown in italics. Ranks correspond to increasing values of *Ti*. *Ti*: log-likelihood difference between each tree and the tree with the maximum log-likelihood

Table 5 Estimated nucleotide substitution rates (substitutions/site/million years) for each putatively homologous organellar copy obtained with BEAST

	Nucleotide substitution rate	
	Plastome copy	Chondrome copy
rbcL	0.0004 (0, 0.0014)	0.0039 (0.0017, 0.0070)
matK	0.0028 (0.0001, 0.0086)	0.0282 (0.0043, 0.0777)
rpl20–rps12	0.0032 (0, 0.0098)	0.0177 (0.0030, 0.0473)
trnL–trnF	0.0061 (0, 0.0186)	0.0577 (0.0075, 0.1502)

The substitution rate shown corresponds to the branch subtending the genus *Cucurbita.* The 95% highest posterior density (HPD) of each estimate is shown in parentheses. See also Suppl. Figure S2

rbcL, *matK* and *rpl20–rps12* transfer events occurred later within specific *Cucurbita* lineages.

In contrast, the AU test in *trnL–trnF* excluded from the confidence set those trees where mitochondrial sequences were the result of recent independent transfer events within different *Cucurbita* lineages (trees #1 and #2) and those trees where the transfer occurred before the rise of Cucurbiteae tribe. Between these extremes, the tree with the highest likelihood was the one with a single ancient inter-organellar transfer in the common ancestor of all mesophytic taxa (Fig. 3b, Table 4).

The divergence of the mitochondrial copies was associated to markedly long branches (Fig. 2). Nucleotide substitution rate estimates show that, once acquired by the chondrome, mitochondrial copies suddenly underwent an increase in substitution rate of one order of magnitude (Table 5). This increase was similar whether the transfer apparently occurred in the common ancestor of *Peponopsis* and *Cucurbita* (*rbcL* and *rpl20–rps12*) or later in the ancestor of *Cucurbita* mesophytic taxa (*matK* and *trnL–trnF*).

However, a decrease in nucleotide substitution rate near the tips was observed both in the clade formed by mt-*trnL–trnF* sequences and in the clade of cp-*trnL–trnF* sequences (Suppl. Figure S2). A Spearman's correlation test between branch length and branch depth revealed that this decrease was significant in both types of copies ($S = 140.55$, $p < 0.05$ in the mt-*trnL–trnF* clade; $S = 1709.2$, $p < 0.05$ in the cp-*trnL–trnF* clade), a pattern that holds as well in the plastome trees of *rbcL* ($S = 445.47$, $p < 0.05$), *matK* ($S = 511.37$, $p < 0.05$), and *rpl20–rps12* ($S = 599.47$, $p < 0.05$).

Diversity of the mt-*trnL–trnF* Region

The PCR amplification reaction was successful in all the surveyed species. A single band was observed during gel electrophoresis and, more importantly, Sanger chromatograms were characterized by clean calls (i.e., we observed no evidence of overlapping peaks pointing to the potential simultaneous sequencing of two different DNA molecules). The BLASTn search revealed that sequences from the six mesophytic taxa were of mitochondrial origin, whereas sequences obtained from the four xerophytic taxa were of plastome origin (Suppl. Table S1).

In mesophytic taxa, the preferential amplification of the mt-*trnL–trnF* copy over the plastome copy during PCR could be explained by differences in the sizes of the expected fragments (>1300 bp in the plastome copy vs. ca. 1000 bp in the chondrome copy; see Methods and Suppl. Table S1). The larger size of the plastome copy is associated to the occurrence of several *trnF* copies in the cp-*trnL–trnF* region (Fig. 1b). Therefore, the shorter chondrome fragment would be favored during the first PCR cycles.

Analysis of the matrix of cp-*trnL–trnF* sequences showed the presence of the duplication of a GAAAT motif in all the taxa of *Cucurbita*. Kocyan et al. (2007) signaled this duplication as a synapomorphy of the genus *Cucurbita*, based on the comparison of a small subset of *Cucurbita* taxa to genus *Peponopsis* and *Polyclathra*. This duplicated motif was also present in our samples of *C. digitata* and *C. palmata*, although it was absent from *C. digitata, C. cordata*, and *C. palmata* sequences from Zheng et al. (2013) (Table 1; Suppl. Fig. S3). We also detected the 43 bp inversion in our cp-*trnL–trnF* sequences of *C. digitata* and *C. palmata,* as previously reported by Kocyan et al. (2007). A 10 bp inversion was identified in the mt-*trnL–trnF* sequences of *C. moschata*, *C. pepo*, *C. argyrosperma*, and *C. maxima* (Suppl. Figure S3). Inversions were reverse-complemented before aligning the sequences in order to avoid bias in branch length

and nucleotide substitution rates, and were included as discrete characters in the second partition of the data matrix. Bayesian phylogenetic reconstruction based on the first partition only (i.e., DNA sequences without coded indels and inversions) produced identical topologies to the ones obtained with two partitions (data not shown).

Relationships between mesophytic taxa were less resolved in the phylogenetic tree based on cp-*trnL–trnF* sequences than in the tree based on mt-*trnL–trnF* copies (Fig. 4). Plastome sequences did not resolve relationships between *C. moschata* and *C. argyrosperma* samples, while *C. pepo* taxa were not grouped in a single group. In contrast, mitochondrial sequences showed longer branch lengths and greater clade resolution, including a single clade for *C. pepo* and *C. argyrosperma* taxa, and the divergence of a clade formed by *C. lundelliana* and *C. okeechobeensis* subsp. *martinezii* (Fig. 4).

Nucleotide diversity π and average number of nucleotide differences *k* were higher in mitochondrial sequences when compared to plastome sequences, not only within the mesophytic group (the only one with the mitochondrial duplication) but also when compared to the cp-*trnL–trnF* sequences at the genus level (Table 6).

tRNA Genes Annotation on mt-*trnL–trnF* Mitochondrial Copies

When we conducted the tRNA annotation of *C. pepo* NC_014050 mitochondrial LCB sequence containing mt-*trnL–trnF* region (Fig. 1b) with tRNAscan-SE v2.0, we detected the four genes consistent with the original annotations of Alverson et al. (2010), who also used this tool for annotating their chondrome. In contrast, ARAGORN detected only three of them (*trnS* [recognized as *trnP*], *trnF*, and *trnM*), and did not detect the putative degenerate *trnL-CAA*. Two tRNA genes, *trnT* and *trnV*, located on the putatively homologus plastome sequence of the LCB

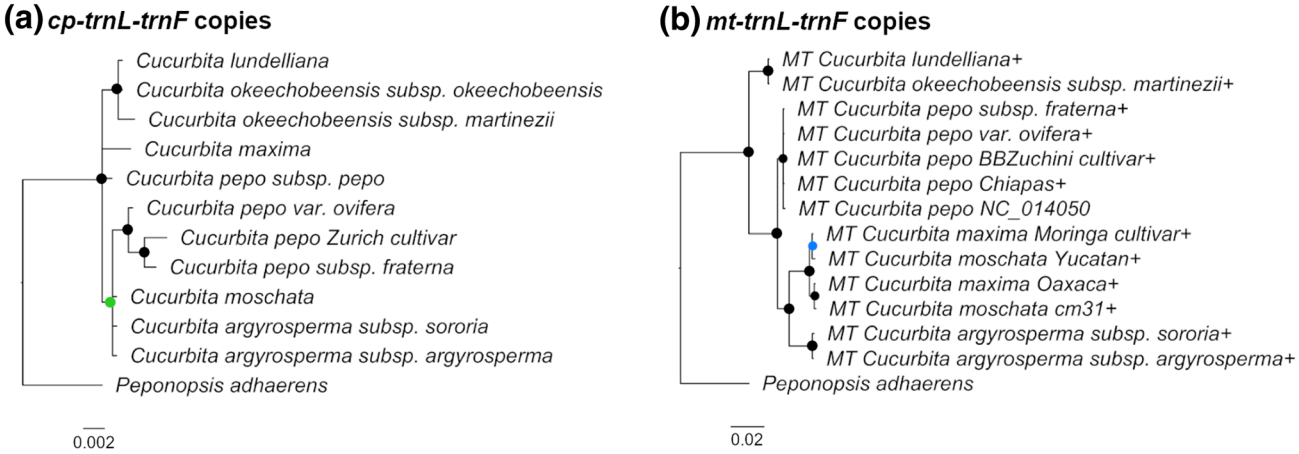

Fig. 4 Majority rule consensus trees of mesophytic *Cucurbita* species based on plastome (**a**) and mitochondrial (**b**) copies of the *trnL–trnF* region obtained with MrBayes. Notice that scale bars are different. Chondrome copies are shown with the *MT* prefix, whereas all others represent plastome copies, including *Peponopsis*. Sequences obtained in this study are shown with a + symbol, whereas the remaining sequences were obtained from GenBank (see Table 1). Posterior probabilities are shown with dots: black (>0.95), blue (0.90–0.95), green (0.85–0.90) (Color figure online)

Table 6 Nucleotide diversity per site π and average number of nucleotide differences *k* between groups of homologous *trnL–trnF* copies from different organellar genomes in *Cucurbita* taxa. *n*: sample size

Organellar genome	*Cucurbita* genus (xerophytic + mesophytic)					*Cucurbita* mesophytic group				
	n	Matrix length	π	*k*	Pairwise identity (%)	*n*	Matrix length	π	*k*	Pairwise identity (%)
Plastome	16	986	0.01206	10.3917	95.9	11	897	0.00362	2.545	99
Chondrome	–	–	–	–		13	1075	0.02209	20.717	93.1

The mitochondrial copy only amplified in mesophytic species

(Fig. 1b) were not found in the corresponding sequence of the mitochondrial genome by either program.

We did not detect any recognizable tRNA gene in the mt-*trnL–trnF* sequences obtained in this study by PCR with any of the two programs used. A more detailed inspection of the mt-*trnL–trnF* alignment reveals that the loss of the identifiable tRNA sequence is related to different mutations in each of the mitochondrial lineages of Fig. 4b. In the *C. lundelliana-C. okeechobeensis* lineage there is a GATT insertion and nucleotide substitutions at three positions, whereas in the lineage formed by *C. moschata-C.argyrosperma-C. maxima* nucleotide substitutions at six positions are observed (Suppl. Figure S4).

Discussion

Inter-organellar Transfer Events and Molecular Evolution of Mitochondrial Copies

The position in the phylogenetic trees of the mitochondrial paralogs supports that regions harboring *matK* and *trnL–trnF* copies were likely transferred from the plastid to the mitochondrion genome in the common ancestor of *Cucurbita* mesophytic taxa (ca. 3.6 MYA, see Castellanos-Morales et al. 2018). In contrast, *rbcL* and *rpl20–rps12* trees did not have enough resolution to determine if the mitochondrial paralog was transferred in the common ancestor of *Peponopsis* and *Cucurbita* (ca. 16 MYA; Schaefer et al. 2009) or later. Results on the topological position of the mitochondrial paralogs were observed to be consistent among the different approaches used, given the similarity between trees obtained with Bayesian (Fig. 2) and Maximum Likelihood (Suppl. Fig. S1) methods. The contrast between high posterior probabilities in Bayesian trees and low bootstrap support in ML trees points to the presence of very short estimated edge lengths that, in fact, represent true polytomies (Lewis et al. 2005). Yet, the rankings obtained with the AU test (which is based on a ML approach) confirm that the ML trees shown in Suppl. Fig. S1 are the ones with the highest likelihood when compared to alternative trees that consider a different position of the mitochondrial paralogs along the phylogeny (Tables 3 and 4).

The presence of a mitochondrial *rbcL* copy in *Cucurbita pepo* and *Cucumis sativus*, next to the likely absence of a mitochondrial copy in *Cucumis melo* and *Citrullus lanatus*, was first reported by Stern (1987) using DNA probe hybridization. Later, also based on hybridization analysis, Cummings et al. (2003) reported a mitochondrial *rbcL* copy in *Cucurbita maxima* and corroborated the absence of mitochondrial *rbcL* in *Cucumis melo*. Although sequences from these previous studies are not available to our knowledge and were not directly included in our phylogenetic analysis, the position of *Cucurbita pepo rbcL* mitochondrial paralog in Fig. 2a—after the Cucurbiteae and Benincaseae divergence—and the exclusion of its origin before the origin of Cucurbiteae (Fig. 3b) supports the hypothesis that the heterogeneous distribution of mitochondrial *rbcL* copies among members of Cucurbitaceae is due to two independent transfer events within the family, as opposed to a single ancient transfer event followed by the loss of the mitochondrial copy in *Citrullus lanatus* and *Cucumis melo*.

The repeated acquisition of *rbcL* sequences by mitochondrial genomes in young lineages has also been reported in *Zea* and *Oryza* (Poaceae; Cummings et al. 2003), and *Ipomoea* (Convolvulaceae; Cummings et al. 2003). Similarly, we found evidence of two independent inter-organellar transfer events of *trnL–trnF* within Cucurbitaceae, one in *Cucurbita* and the other in *Citrullus lanatus*. Occurrence of *trnL* and *trnF* paralogs has been reported in *Oryza* (Notsu et al. 2002) and Annonaceae (Pirie et al. 2007). In fact, an increased mutation rate in the second copy of the *trnL–trnF* region led Pirie et al. (2007) to infer that the paralog was likely located in the nuclear genome, based on the premise that mitochondrial mutation rates were even lower than those of the plastome (as based from synonymous substitution rates of codifying genes; Wolfe et al. 1987). Yet, our results on the increased nucleotide substitution rate of *trnL–trnF* sequence after its acquisition by *Cucurbita* chondrome challenge this assumption.

Once transferred, the four mitochondrial copies examined experienced a sharp increase in nucleotide substitution rate and partial sequence losses when compared to their plastome counterparts, supporting the notion that these plastid DNA insertions are non-functional. From the four analyzed regions, two of them are intergenic regions (*rpl20–rps12*, *trnL–trnF*) and two genes (*rbcL* and *matK*) that apparently lost their functional roles. In the case of *matK*, the presumed encoded protein is very small (89 aminoacids long) when compared to the original plastid protein (503 aminoacids). Moreover, we did not find evidence of the transcription of this putative gene from a BLASTn search on the TSA database of GenBank (data not shown). Similar ratios of non-synonymous to synonymous substitutions in non-coding mt-*rbcL* (dN/dS = 1.5823) and coding mt-*matK* (dN/dS = 1.4693) as well suggest that mutations in mt-*matK* are not constrained by natural selection, but due solely to high mutation rates.

The contrast between the high mutation rate of plastid-like sequences in *Cucurbita* chondrome and the low mutation rate of its native protein genes (Alverson et al. 2010) support the notion of a broad and variable mutational spectrum within this organellar genome. Christensen (2013) proposed that low mutation rates within mitochondrial genes could be explained by the occurrence of gene conversion as the main DNA repair mechanism, whereas high mutation

rates, rearrangements and drift in non-coding regions (such as the four plastid-like regions analyzed in this study) could be associated to a different, error-prone, repair mechanism like break-induced replication.

Overall, the mitochondrial copies examined in the present study are of recent origin, since transfer events occurred at most in the common ancestor of Cucurbiteae (*rbcL*, *trnL–trnF*) or *Cucurbita* (*matK*) (Fig. 3). But whether the analyzed mitochondrial paralogs were transferred independently or during the same intracellular transfer event cannot be ascertained from our data. The four regions analyzed lack their original relative positions from the plastome, are separated by large tracts of chondrome sequence interspersed with plastid-like sequences (on the order of hundreds of thousands bp), and some are inverted (Fig. 1). Given the compatibility of the confidence tree sets of the markers (Fig. 3), we cannot rule out the possibility that the regions analyzed were transferred in a single event, and later split and rearranged due to recombination (Alverson et al. 2011). Yet, on the other extreme, a single transfer event for each of the twenty LCBs between *Cucurbita* plastid and mitochondrion genomes would translate into twenty independent transfer events, which seem unlikely. Sequencing of complete chondromes from additional *Cucurbita* species and structural comparative analyses among them are necessary steps towards a better understanding on the number of inter-organellar transfer events and probable rearrangements in the chondrome of *Cucurbita*, especially because it was not possible to entirely exclude more recent transfer events with the AU test in the case of *rbcL*, *rpl20–rps12*, and *matK*.

In the same sense, our data do not allow us to determine if each of the mt-*trnL–trnF* sequences are located in the same physical position among *Cucurbita* lineages given the occurrence of structural rearrangements in the chondrome at short evolutionary timescales (Bartoszewski et al. 2004; Alverson et al. 2011). Yet, possible recombination does not weakens our conclusions regarding the transfer event itself and the monophyletic origin of the mt-*trnL–trnF* copies evaluated. Sanchez-Puerta et al. (2015) showed that mitochondrial recombination events are followed by the retention of one single allele per locus in most cases, allowing the evolutionary history described by the locus to be preserved. Locus integrity is manifest in our data through a 100% pairwise identity between mt-*trnL–trnF* sequences from independent *C. pepo* samples, including two wild subspecies geographically isolated (*C. pepo* subsp. *fraterna* and *C. pepo* subsp. *ovifera*) and two cultivated *C. pepo* subsp. *pepo* samples (including "Black Beauty Zucchini" cultivar and a native Mexican sample). Alverson's et al. (2010) *C. pepo* subsp. *pepo* "Dark Green Zucchini" cultivar NC_014050 copy differs by a single gap from the aforementioned sequences (99.96% pairwise identity).

Inferences regarding the timing of inter-organellar transfers are based on the extracted mitochondrial paralogs from a single taxon, *C. pepo*, which has the only sequenced chondrome available to date for this genus. However, some inferences can be drawn from the phylogenetic patterns reported here. For example, the position of *C. pepo* mitochondrial *rbcL* in our phylogeny (Fig. 2a) predicts the presence of a mitochondrial copy in the remaining *Cucurbita* species. Cummings et al. (2003) reported a mitochondrial *rbcL* copy in *C. maxima*, and future studies shall test its presence in other taxa of the genus. Similarly, a phylogenetic tree including only *C. pepo* mt-*trnL–trnF* copy next to homologous plastome copies places the *C. pepo* mitochondrial paralog at the root of mesophytic *Cucurbita*, in the exact same position as on the tree shown in Fig. 2d (Suppl. Figure S5). As would be expected from this topology, the amplification of mt-*trnL–trnF* copies obtained by PCR from nine *Cucurbita* taxa revealed a single monophyletic group supported by a posterior probability of 1 (Fig. 2d), and by the lack of product in our PCR essays on xerophytic species that diverged before the transfer event.

Tree Testing with Mitochondrial Paralogs

The higher nucleotide substitution rate of mt-*trnL–trnF* copies allowed for increased resolution of genealogical relationships between mesophytic species when compared to cp-*trnL–trnF* copies. For example, whereas *C. pepo* infraspecific taxa were not grouped in the plastid tree and the nodes supporting their positions were characterized by low posterior probabilities (Fig. 3a), they were clearly grouped as a single clade in the mt-*trnL–trnF* tree with strong support (posterior probability = 1, Fig. 3b), clearly separated from the closely allied *C. lundelliana* and *C. okeechobeensis* subsp. *martinezii*. The relationships between *Cucurbita* mesophytic taxa based on mt-*trnL–trnF* (Fig. 3b) are consistent with the phylogeny of Sanjur et al. (2002), who used the mitochondrial *nad1* intron; the only difference is the position of South American *C. maxima*, which resulted as sister to *C. moschata* in our analysis, whereas it was retrieved as sister to the remaining mesophytic *Cucurbita* of North America by Sanjur et al. (2002), Kates et al. (2017) based on nuclear data, and Kistler et al. (2015) and Castellanos-Morales et al. (2018) based on plastome data. The mode of inheritance of mitochondria in *Cucurbita* is unknown, but there are reports of paternal inheritance in the related *Cucumis* (Havey 1997; Shen et al. 2013). Considering the alternate grouping of *C. maxima* and *C. moschata* samples in the mitochondrial clade of Fig. 3b, we may be detecting the breeding history of *C. maxima* cultivars used in this study, since *C. moschata* has been used in breeding programs for improving *C. maxima* varieties (e.g., Rouphael et al. 2012; Ara et al. 2013).

Using paralogs in phylogenetic reconstruction has been proposed as a critical route to test species trees (Wolfe and Randle 2004). Our analysis supports sister relationships between domesticated *C. moschata* and the wild/domesticated pair *C. argyrosperma* subsp. *sororia* and *C. a.* subsp. *argyrosperma*, and between wild *C. okeechobeensis* and *C. lundelliana* previously identified with both plastome (Zheng et al. 2013; Castellanos-Morales et al. 2018) and nuclear data (Kates et al. 2017). Yet, while nuclear data point the (*C. okeechobeensis*, *C. lundelliana*) clade as the most closely related to *C. pepo*, plastome data place this clade as more closely related to *C. moschata* and *C. argyrosperma*, and mitochondrial data place it as sister to all mesophytic species of North America. These differences are likely associated to contrasting modes of inheritance of the markers analyzed.

Our data show that the mitochondrial copies represent deep paralogs (i.e., those resulting from duplication and divergence prior to speciation events; Bailey et al. 2003), which implies a high potential for the inadvertent inclusion of a mixture of copies in the same phylogenetic matrix, leading to erroneous species trees (Palmer 1992; Wolfe and Randle 2004). As suggested by our data, the risk of amplifying a mitochondrial paralog is directly related to the degree of conservation of universal primer aligning sites after the inter-organellar transfer, independently of the conservation (or not) of the open reading frames and of the observed high nucleotide substitution rates. Moreover, even in the two mesophytic mitochondrial lineages that suffered mutations in the expected tRNA sequence, priming sites were still conserved.

Given the high nucleotide substitution rate shown by the mt-*trnL–trnF* copy (Table 3), preliminary analyses show that this mitochondrial region can be highly useful for the identification of gene pools at the intraspecific level in *C. moschata* and *C. argyrosperma* (>4 haplotypes detected in 10 individuals from different localities in Mexico; Hernández-Rosales et al. submitted; Fernando Tapia, pers. com.) suggesting its potential as a genetic marker for population studies in these taxa.

Conservation Versus Loss of tRNA Genes in the *trnL–trnF* Paralog

The increased tRNA gene content of plastid-derived tracts in *Cucurbita* and *Cucumis* chondromes has raised the possibility of functional replacement of native mitochondrial tRNAs by tRNAs "imported" from the plastome (Alverson et al. 2010, 2011). Here, we confirmed the presence of three intact tRNAs in the mitochondrial LCB harboring mt-*trnL–trnF*. However, tRNA genes are also characterized by short sequences, which could accumulate less mutations when compared to longer sequences with the same mutation rate per nucleotide. Pseudogenization of nine from the 23 plastid-derived tRNAs in *Cucurbita pepo* chondrome (Alverson et al. 2010), and the common occurrence of independent mutations on the putative degenerate *trnL-CAA* gene in different *Cucurbita* lineages indicate that genetic drift can play an important role in the fate of these transferred genes. Detailed future studies on the correct transcription and editing of tRNA genes embedded within transferred tracts of plastid DNA are needed to clarify if acquired genes are functional and potentially adaptive (Leon et al. 1989; Dietrich et al. 1996).

Conclusions

Inter-organellar transfer of large and small plastome regions to the chondrome may occur in different moments along the evolutionary history of a plant lineage, leading to uneven distribution of mitochondrial paralogs in descendant species; in this study, between xerophytic and mesophytic *Cucurbita* taxa.

The extremely high nucleotide substitution rates detected on mitochondrial paralogs of plastid origin, next to low mutation rates previously reported on mitochondrial coding genes, support the notion of a broad and variable mutational spectrum within the chondrome.

The high nucleotide substitution rates of the mt-*trnL–trnF* region turn it into a useful new marker for intraspecific studies in *Cucurbita*, while supporting genealogical relationships among the main groups of *Cucurbita* mesophytic taxa. However, special attention must be given in not mixing copies of different organellar origin during phylogenetic reconstruction. The erroneous inclusion of plastome markers paralogs (which have been shown to occur in different plant lineages, especially for *rbcL* and *trnL–trnF*) may produce misleading conclusions on the evolutionary history of taxa.

An increasing number of full mitochondrial genomes will undoubtedly contribute to identify the frequency of such intracellular transfers and their role in organellar genome evolution.

Acknowledgements This manuscript includes in part the results of the Bachelor's degree thesis of FTA, LMPT, PH, and KYRM, and postdoctoral work of XAD at Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México. We are grateful to D. Piñero and V. Souza for supporting the project, and Laura Espinosa-Asuar for her help in laboratory work. Funds were provided by Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (Conabio) Project KE004 "Diversidad genética de las especies de *Cucurbita* en México e hibridación entre plantas genéticamente modificadas y especies silvestres de *Cucurbita*" and Project Conabio PE001 "Diversidad genética de las especies de *Cucurbita* en México. Fase II. Genómica evolutiva y de poblaciones, recursos genéticos y domesticación", as well as Consejo Nacional de Ciencia y Tecnología (CONACyT) Problemas Nacionales grant number 247730 to Daniel Piñero (Instituto de Ecología, UNAM). XAD had a fellowship from Programa de Becas Posdoctorales de la Dirección General de Asuntos del Personal Académico (DGAPA), Universidad Nacional Autónoma de México.

Compliance with Ethical Standards

Conflict of interest The authors state they have no competing interests to declare.

References

Adams KL, Qiu Y-L, Stoutemyer M, Palmer JD (2002) Punctuated evolution of mitochondrial gene content: High and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc Natl Acad Sci USA 99:9905–9912. https://doi.org/10.1073/pnas.042694899

Alverson AJ, Wei X, Rice DW et al (2010) Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). Mol Biol Evol 27:1436–1448. https://doi.org/10.1093/molbev/msq029

Alverson AJ, Rice DW, Dickinson S et al (2011) Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. Plant Cell 23:2499–2513. https://doi.org/10.1105/tpc.111.087189

Ara N, Nakkanong K, Lv W et al (2013) Antioxidant enzymatic activities and gene expression associated with heat tolerance in the stems and roots of two cucurbit species ("*Cucurbita maxima*" and "*Cucurbita moschata*") and their interspecific inbred line "Maxchata". Int J Mol Sci 14:24008–24028. https://doi.org/10.3390/ijms141224008

Bailey CD, Carr TG, Harris SA, Hughes CE (2003) Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. Mol Phylogenet Evol 29:435–455. https://doi.org/10.1016/j.ympev.2003.08.021

Barrera-Redondo J, Ibarra-Laclette E, Vázquez-Lobo A et al (2019) The genome of *Cucurbita argyrosperma* (silver-seed gourd) reveals faster rates of protein-coding gene and long noncoding RNA turnover and neofunctionalization within *Cucurbita*. Mol Plant 12:506–520. https://doi.org/10.1016/j.molp.2018.12.023

Bartoszewski G, Malepszy S, Havey MJ (2004) Mosaic (MSC) cucumbers regenerated from independent cell cultures possess different mitochondrial rearrangements. Curr Genet 45:45–53. https://doi.org/10.1007/s00294-003-0456-6

Borsch T, Quandt D (2009) Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA. Plant Syst Evol 282:169–199. https://doi.org/10.1007/s00606-009-0210-8

Castellanos-Morales G, Paredes-Torres LM, Gámez N et al (2018) Historical biogeography and phylogeny of *Cucurbita*: Insights from ancestral area reconstruction and niche evolution. Mol Phylogenet Evol 128:38–54. https://doi.org/10.1016/j.ympev.2018.07.016

Chaw S-M, Wu C-S, Sudianto E (2018) Evolution of gymnosperm plastid genomes. Adv Bot Res 85:195–222. https://doi.org/10.1016/bs.abr.2017.11.018

Christensen AC (2013) Plant mitochondrial genome evolution can be explained by DNA repair mechanisms. Genome Biol Evol 5:1079–1086. https://doi.org/10.1093/gbe/evt069

Cummings MP, Nugent JM, Olmstead RG, Palmer JD (2003) Phylogenetic analysis reveals five independent transfers of the chloroplast gene rbcL to the mitochondrial genome in angiosperms. Curr Genet 43:131–138. https://doi.org/10.1007/s00294-003-0378-3

Darling ACE, Mau B, Blattner FR, Perna NT (2004) Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9:772

Dietrich A, Small I, Cosset A et al (1996) Editing and import: Strategies for providing plant mitochondria with a complete set of functional transfer RNAs. Biochimie 78:518–529. https://doi.org/10.1016/0300-9084(96)84758-4

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. PLoS Biol 4:699–710. https://doi.org/10.1371/journal.pbio.0040088

Goremykin VV, Salamini F, Velasco R, Viola R (2009) Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. Mol Biol Evol 26:99–110. https://doi.org/10.1093/molbev/msn226

Goremykin VV, Lockhart PJ, Viola R, Velasco R (2012) The mitochondrial genome of Malus domestica and the import-driven hypothesis of mitochondrial genome expansion in seed plants. Plant J 71:615–626. https://doi.org/10.1111/j.1365-313X.2012.05014.x

Guindon S, Gascuel O (2003) A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. Syst Biol 52:696–704

Havey MJ (1997) Predominant paternal transmission of the mitochondrial genome in cucumber. J Hered 88:232–235. https://doi.org/10.1055/s-0031-1299652

Kates HR, Soltis PS, Soltis DE (2017) Evolutionary and domestication history of *Cucurbita* (pumpkin and squash) species inferred from 44 nuclear loci. Mol Phylogenet Evol 111:98–109. https://doi.org/10.1016/j.ympev.2017.03.002

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059

Kearse M, Moir R, Wilson A et al (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647

Kelchner SA (2000) The evolution of non-coding chloroplast DNA and its application in plant systematics. Ann Missouri Bot Gard 87:482–498

Kistler L, Newsom LA, Ryan TM et al (2015) Gourds and squashes (*Cucurbita* spp.) adapted to megafaunal extinction and ecological anachronism through domestication. Proc Natl Acad Sci USA 112:15107–15112. https://doi.org/10.1073/pnas.1516109112

Knoop V (2012) Seed plant mitochondrial genomes: complexity evolving. In: Bock R, Knoop V (eds) Genomics of chloroplasts and mitochondria. Springer, New York, pp 175–200

Kocyan A, Zhang LB, Schaefer H, Renner SS (2007) A multi-locus chloroplast phylogeny for the Cucurbitaceae and its implications for character evolution and classification. Mol Phylogenet Evol 44:553–577. https://doi.org/10.1016/j.ympev.2006.12.022

Laslett D, Canback B (2004) ARAGORN, a program for the detection of transfer RNA and transfer-messenger RNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16

Leon P, Walbot V, Bedinger P (1989) Molecular analysis of the linear 2.3 kb plasmid of maize mitochondria: apparent capture of tRNA genes. Nucleic Acids Res 17:4089–4099

Lewis PO, Holder MT, Holsinger KE (2005) Polytomies and bayesian phylogenetic inference. Syst Biol 54:241–253. https://doi.org/10.1080/10635150590924208

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452. https://doi.org/10.1093/bioinformatics/btp187

Lira-Saade R (1995) Estudios taxonómicos y ecogeográficos de las Cucurbitaceae latinoamericanas de importancia económica. International Plant Genetic Resources Institute, Rome

Lowe TM, Chan PP (2016) tRNAscan-SE On-line: integrating search and context analysis of transfer RNA Genes. Nucleic Acids Res 44:W54–57. https://doi.org/10.1093/nar/gkw413

Lynch M, Koskella B, Schaack S (2006) Mutation pressure and the evolution of organelle genomic architecture. Science 80-(311):1727–1731

Müller K (2005) SeqState—primer design and sequence statistics for phylogenetic DNA data sets. Appl Bioinform 4:65–69

Müller K (2006) Incorporating information from length-mutational events into phylogenetic analysis. Mol Phylogenet Evol 38:667–676. https://doi.org/10.1016/j.ympev.2005.07.011

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Notsu Y, Masood S, Nishikawa T et al (2002) The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: Frequent DNA sequence acquisition and loss during the evolution of flowering plants. Mol Genet Genomics 268:434–445. https://doi.org/10.1007/s00438-002-0767-1

Palmer JD (1992) Mitochondrial DNA in plant systematics: applications and limitations. In: Soltis PS, Soltis DE, Doyle JJ (eds) Molecular systematics of plants. Chapman and Hall, London, pp 36–39

Pirie MD, Vargas MPB, Botermans M et al (2007) Ancient paralogy in the cpDNA trnL-F region in Annonaceae: Implications for plant molecular systematics. Am J Bot 94:1003–1016. https://doi.org/10.3732/ajb.94.6.1003

Rice DW, Alverson AJ, Richardson AO et al (2013) Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm. Science 80-(342):1468–1473

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Rouphael Y, Cardarelli M, Rea E, Colla G (2012) Improving melon and cucumber photosynthetic activity, mineral composition, and growth performance under salinity stress by grafting onto *Cucurbita* hybrid rootstocks. Photosynthetica 50:180–188. https://doi.org/10.1007/s11099-012-0002-1

Sanchez-Puerta MV, Zubko M, Palmer JD (2015) Homologous recombination and retention of a single form of most genes shape the highly chimeric mitochondrial genome of a cybrid plant. Genome Res 206:381–396. https://doi.org/10.1111/nph.13188

Sanchez-Puerta MV, Edera A, Gandini CL et al (2019) Genome-scale transfer of mitochondrial DNA from legume hosts to the holoparasite *Lophophytum mirabile* (Balanophoraceae). Mol Phylogenet Evol 132:243–250. https://doi.org/10.1016/j.ympev.2018.12.006

Sanderson MJ, Copetti D, Búrquez A et al (2015) Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): loss of the ndh gene suite and inverted repeat 1. Am J Bot 102:1115–1127. https://doi.org/10.3732/ajb.1500184

Sanjur OI, Piperno DR, Andres TC, Wessel-Beaver L (2002) Phylogenetic relationships among domesticated and wild species of *Cucurbita* (Cucurbitaceae) inferred from a mitochondrial gene: Implications for crop plant evolution and areas of origin. Proc Natl Acad Sci USA 99:535–540. https://doi.org/10.1073/pnas.012577299

Schaefer H, Heibl C, Renner SS (2009) Gourds afloat: a dated phylogeny reveals an Asian origin of the gourd family (Cucurbitaceae) and numerous oversea dispersal events. Proc R Soc B 276:843–851. https://doi.org/10.1098/rspb.2008.1447

Scott I, Logan DC (2011) Mitochondrial dynamics. In: Kempken F (ed) Plant mitochondria. Springer, New York, pp 31–63

Shaw J, Lickey EB, Beck JT et al (2005) The tortoise and the hare II; relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am J Bot 92:142–166

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

Shimodaira H (2002) An Approximately Unbiased test of phylogenetic tree selection. Syst Biol 51:492–508. https://doi.org/10.1080/10635150290069913

Shimodaira H, Hasegawa M (2001) CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17:1246–1247

Sloan DB, Alverson AJ, Chuckalovcak JP et al (2012a) Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol 10:e1001241. https://doi.org/10.1371/journal.pbio.1001241

Sloan DB, Müller K, McCauley DE et al (2012b) Intraspecific variation in mitochondrial genome sequence, structure, and gene content in *Silene vulgaris*, an angiosperm with pervasive cytoplasmic male sterility. New Phytol 196:1228–1239. https://doi.org/10.1111/j.1469-8137.2012.04340.x

Smith DR, Keeling PJ (2015) Mitochondrial and plastid genome architecture: Reoccurring themes, but significant differences at the extremes. Proc Natl Acad Sci USA 112:10177–10184. https://doi.org/10.1073/pnas.1422049112

Stern DB (1987) DNA transposition between plant organellar genomes. J Cell Sci. https://doi.org/10.1242/jcs.1987.Supplement_7.11

Suchard M, Lemey P, Baele G et al (2018) Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol 4:vey016. https://doi.org/10.1093/ve/vey016

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 17:1105–1109

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437–460

Untergasser A, Cutcutache I, Koressaar T et al (2012) Primer3-New capabilities and interfaces. Nucleic Acids Res 40:e115. https://doi.org/10.1093/nar/gks596

Wicke S, Schneeweiss GM, dePamphilis CW et al (2011) The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. Plant Mol Biol 76:273–297. https://doi.org/10.1007/s11103-011-9762-4

Wolfe AD, Randle CP (2004) Recombination, heteroplasmy, haplotype polymorphism, and paralogy in plastid genes: implications for plant molecular systematics. Syst Bot 29:1011–1020. https://doi.org/10.1600/0363644042451008

Wolfe KH, Li W-H, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84:9054–9058. https://doi.org/10.1073/pnas.84.24.9054

Yang Z (2007) PAML 4: a program package for phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591

Zhang LB, Simmons MP, Kocyan A, Renner SS (2006) Phylogeny of the Cucurbitales based on DNA sequences of nine loci from three genomes: implications for morphological and sexual system evolution. Mol Phylogenet Evol 39:305–322. https://doi.org/10.1016/j.ympev.2005.10.002

Zhang C, Zhu Q, Liu S et al (2018) The complete chloroplast genome sequence of the *Cucurbita pepo* L. (Cucurbitaceae). Mitochondrial DNA Part B 3:717–718. https://doi.org/10.1080/23802359.2018.1483766

Zheng YH, Alverson AJ, Wang QF, Palmer JD (2013) Chloroplast phylogeny of *Cucurbita*: Evolution of the domesticated and wild species. J Syst Evol 51:326–334. https://doi.org/10.1111/jse.12006

Zhu A, Guo W, Gupta S et al (2016) Evolutionary dynamics of the plastid inverted repeat: The effects of expansion, contraction, and loss on substitution rates. New Phytol 209:1747–1756. https://doi.org/10.1111/nph.13743

Zhu,Q., Gao,P., Liu,S., Wang, X., Qu,S. and Luan,F. (2017). The complete chloroplast genome sequence of the *Cucurbita moschata* Duch. Direct Submission (18-Dec-2017) National Center for Biotechnology Information. Reference Sequences NC_036506 and NC_036505.