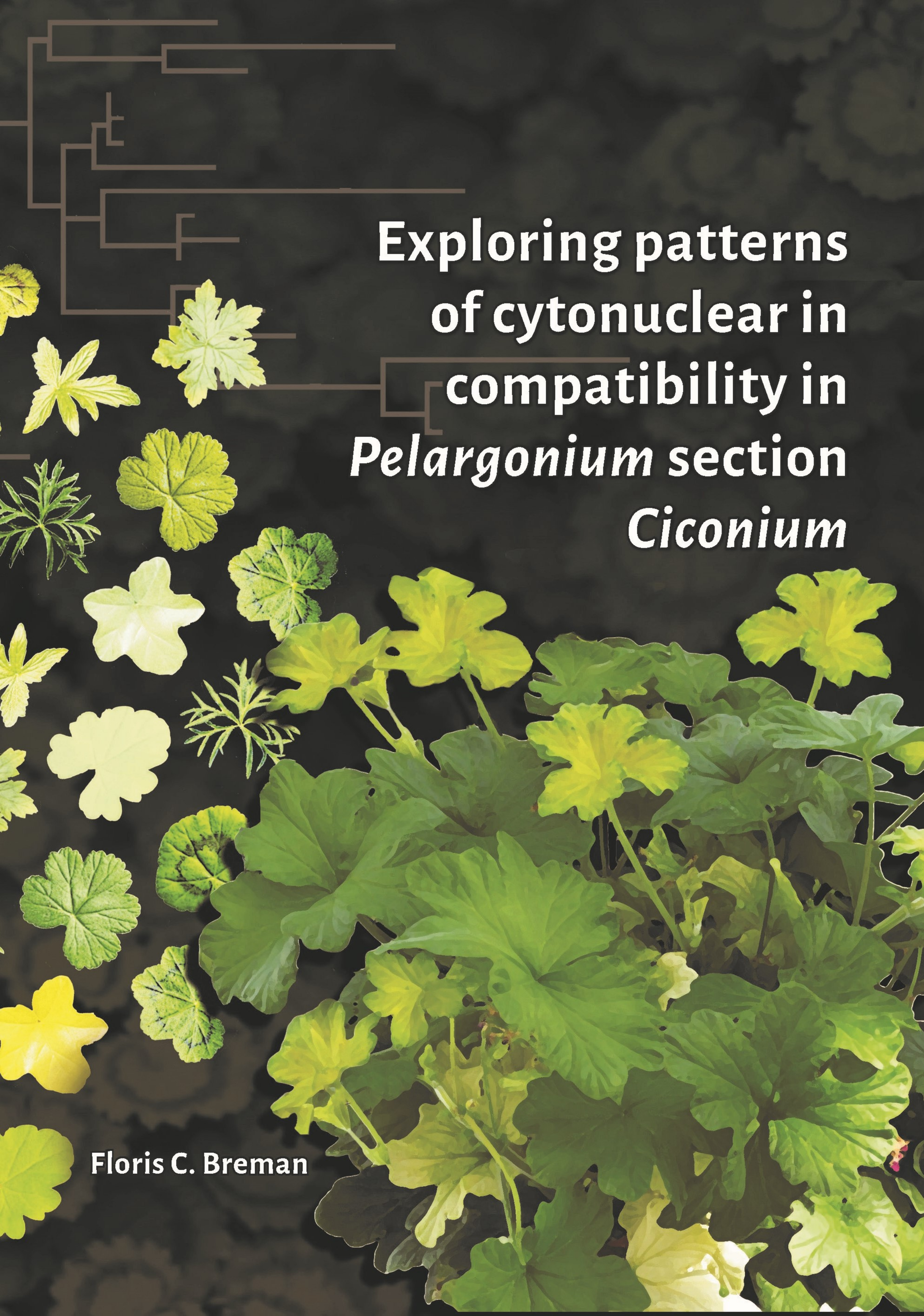


Exploring patterns of cytonuclear in compatibility in *Pelargonium* section *Ciconium*

Floris C. Breman

Propositions

1. Plastid encoded RNA polymerase plays a central role in *Pelargonium* section *Ciconium* plastome evolution through cytonuclear incompatibility
(this thesis)

2. Despite morphological differences and the occurrence of cytonuclear incompatibilities, *P.* sect. *Ciconium* species are highly compatible.
(this thesis)

3. Evolutionary biology requires one to simultaneously consider short term mechanisms, while maintaining an eon-long perspective.

4. Industrially-sized experiments allow us to address previously unanswerable questions in evolutionary biology.

5. Kartini's observations on emancipation are as relevant today as they were more than a century ago.

6. Current discussions on the use of nuclear energy indicate desperation as well as a lack of imagination.

Propositions belonging to the thesis, entitled

Exploring patterns of cytonuclear incompatibility in *Pelargonium* section *Ciconium*
Floris C. Breman
Wageningen, 13 September 2021

Exploring patterns of cytonuclear incompatibility in *Pelargonium* section *Ciconium*

Floris C. Breman

Thesis committee

Promotor
Prof. Dr M.E. Schranz
Professor of Biosystematics
Wageningen University & Research

Co-promotors
Dr F.T. Bakker
Assistant professor, Biosystematics group
Wageningen University & Research

Dr R.C. Snijder
Senior breeder Syngenta Seeds, Enkhuizen

Other members
Prof. Dr M.G.M. Aarts, Wageningen University & Research
Prof. Dr M. Gibby, Royal Botanic Garden Edinburgh, UK
Dr S. Dodsworth, University of Portsmouth, UK
Dr S. Heimovaara, Wageningen University & Research

This research was conducted under the auspices of the Graduate School Experimental Plant Sciences.

Exploring patterns of cytonuclear incompatibility in *Pelargonium* section *Ciconium*

Floris C. Breman

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
Monday 13 September 2021
at 11 a.m. in the Aula.

Floris C. Breman
Exploring patterns of cytonuclear incompatibility in *Pelargonium* section *Ciconium*, 196 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2021)
With references, with summaries in English and Dutch

DOI: https://doi.org/10.18174/551565
ISBN: 978-94-6395-936-0

Table of contents

Chapter 1	General introduction	**7**
Chapter 2	Interspecific hybrids between *Pelargonium* x *hortorum* and species from *P*. section *Ciconium* reveal biparental plastid inheritance and multi-locus cyto-nuclear incompatibility	**21**
Chapter 3	Repeatome evolution in *Pelargonium* section *Ciconium* (Sweet) Harvey	**47**
Chapter 4	Rapid plastome evolution in *Pelargonium* section *Ciconium* linked with altered plastidic ribosomes	**81**
Chapter 5	Possible causes of plastid induced CNI and organellar inheritance in *Pelargonium* section *Ciconium* (Sweet)	**117**
Chapter 6	General discussion	**145**
	References	**167**
	Summary	**184**
	Samenvatting	**186**
	Acknowledgements	**188**
	About the author	**191**
	List of publications	**192**
	Overview of training activities	**193**

Abbreviations

CWR:	Crop Wild Relatives
CNI:	Cyto-Nuclear Incompatibility
	pCNI: plastid-induced CNI
	mCNI: mitochondrially-induced CNI
plastome:	the entire genome contained in the chloroplast
mitome:	the entire genome contained in the mitochondrion
anterograde signaling:	nuclear control of organellar function
retrograde signaling:	signaling pathways emanating from the organelles and influencing the expression of the nuclear genome
CMS:	Cytoplasmic male sterility, caused by mitochondrially induced cyto nuclear incompatibilities (mCNI)
chlorosis:	'bleaching of leaves' caused by chloroplast induced by chloroplast induced cyto nuclear incompatibilities (pCNI)
repeat:	a DNA based sequence that has a repeating sequence pattern. Plays numerous, partially unknown, roles in the genome, dependent on the type
repeatome:	the collective repetitive part of the genome

Chapter 1

General introduction

Plant speciation and evolution

Humans have always had an intuitive notion about what a species is, and these can (especially in folklore) represent vast historical knowledge and experience (Berlin 1992). Species are the biological 'units' which we use to refer to a group of individuals. Moreover, good knowledge of what constitutes a species is critical in food safety and proper use of medicines. Simple mistakes in knowing which species is 'in front of you' can lead to the loss of income or worse, (Smith et al. 2011). Historical confusion among biologists between 'concept' and 'definition' has further obscured the issue (discussed in Hey 2006), with 'species definitions' frequently promoted to 'species concepts'. The theory of evolution provides us with the concept of the species that arises from variation (Darwin 1859) and most biologists would agree that 'the species' is a good concept (Wilkins 2003). Unfortunately, there is not one overarching definition with which to apply this concept and on which we can all agree (Wilkins 2003, Hey 2006), each definition having exceptions which defy it. As unsatisfactory as this is, species (be it definition or concept) do in any case represent a practical unit, which, enables scientific communication (Hey 2006).

As much as our opinions differ as to what defines a species, there is universal agreement on how species arise. They do so in the Darwinian framework of natural selection (Darwin 1859) which takes place when a group of organisms experiences an event or stress which is selected upon. Because of inherent natural variation in the genome, those individuals with the combinations that makes them best fit to changing environments have the highest change to survive and thus to procreate, resulting in higher fitness. Populations, derived from a single ancestor, that have become separated in space and time, acquire mutations independently from each other. Most mutations are repaired, but not all, and some of those may contribute to reproductive isolation.

Upon secondary contact, these mutations may prevent successful crossing to take place. This means that the populations are effectively kept separated, but now by genetic barriers rather than spatio-temporal ones. In other words, they have become incompatible. This is referred to as the Bateson-Dobzhansky-Muller (BDM) model of speciation (Bateson 1909, Dobzhansky 1936, Müller 1942) (see Fig. 1). As we will see in the coming paragraphs, BDM incompatibilities are not always absolute. Especially when populations rejoin after a relatively brief time, the acquired changes may not be sufficient for full separation and occasional crossing may take place. These may yield fit offspring, though often less fit, resulting in strong selection against such genomic combinations. To study the extent of BDM incompatibilities we can bring species into secondary contact in an artificial environment where even less-fit offspring can survive and, as will be discussed below, this aspect is the central approach of the work presented in this thesis.

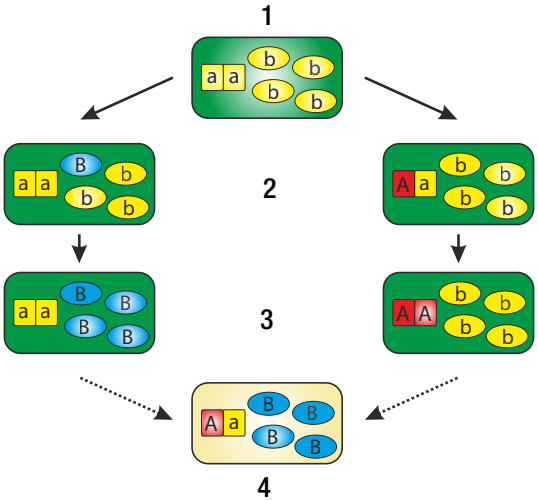


FIGURE 1 | Summary of the Bateson-Dobzhansky–Müller model of speciation adapted from Greiner et al. 2011 with permission. 1) ancestral population separates, 2) nuclear and organellar genomes co-evolve independently in each population 3) mutations become fixed in nucleus and organelles, 4) upon secondary contact populations have become incompatible.

Properties of the plant genome

Plants contain two types of organelles, those that carry their own DNA and those that do not. The Golgi apparatus, endoplasmatic reticulum (rough and smooth) and lysosome are non-DNA carrying organelles. Mitochondria and, in the case of plants and some bacteria, the chloroplasts carry their own genomes. The mitochondrion and chloroplast are endosymbionts that were incorporated as α-proteo-and cyanobacteria resp., by eukaryote ancestors (Gray and Archibald 2012). During evolution of eukaryotes numerous genes contained in the organellar ancestor were transferred to the nucleus (Martin and Herrmann 1998, Lloyd et al. 2012) reducing the number of genes from an estimated 5000, to 3-67 in mitochondria and to 23-200 in chloroplasts (Allen 2015). Transfer to the nucleus is thought the facilitate an escape from Muller's ratchet: By re-locating these genes to the nucleus they become subject to recombination thereby decreasing the chance of retaining detrimental mutations (Martin and Herrmann 1998, Blanchard and Lynch 2000). The genes retained in the organelles are essential for the highly reactive processes of photosynthesis and oxidative phosphorylation (the latter generates ATP). Transferring these to the nucleus would likely result in more detrimental mutations than can be corrected for with recombination. This is thought to be the reason for the small number of genes retained in the organelles. The genes that are retained in the organelles consist of the core set needed for photosynthesis, RNA polymerase (chloroplasts only), ribosomal RNA and ribosomal proteins, ATP synthase, tRNA's, and for mitochondria NADH dehydrogenase and cytochrome oxidase genes (Allen 2015). Among the transferred genes are regulatory genes such as those involved in gene transcription, DNA replication and

DNA repair (Martin and Herrmann 1998). The proper function of the organelles therefore depends on the signals from these regulatory genes reaching the organelle and vice-versa. This is called anterograde (from the nucleus to the organelles) and retrograde signaling (from organelle to nucleus) (Leister 2005, Pfannschmidt et al. 2020 and Fig. 2). These signals are composed of a variety of molecules, most prominent among which are nuclear encoded polymerases (Bychkow et al. 2021), sigma factors (Lerbs-Mache 2011) and PPR genes (Wang et al. 2021). Retrograde signaling, likewise consists of many molecules such as metabolites, H_2O_2, sigma factors and transcription factors (see the full list in a review by Singh et al. 2015). This implies that the nuclear and organellar genomes do not evolve independently from each other but in concert. (See Forsythe et al. (2021) for a review of plastid-nuclear coevolution 'signatures' across angiosperms). They form an integral and tightly-connected tri-partite genome that shapes the plant (Greiner et al. 2013 and Fig.1).

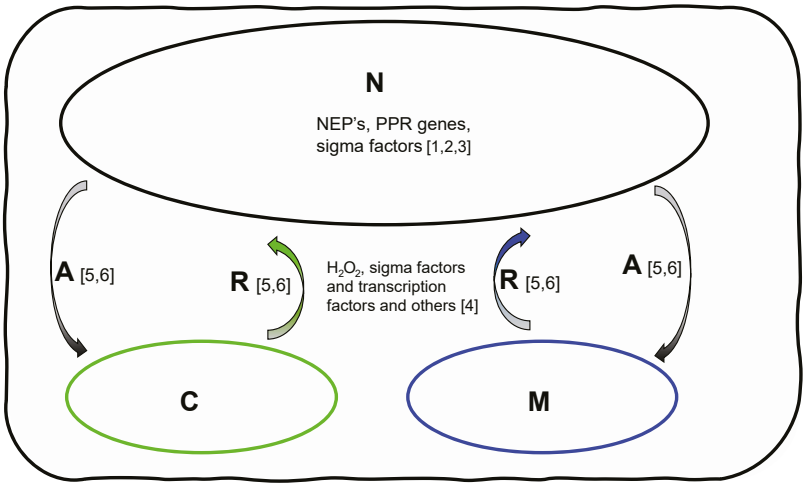


FIGURE 2 | Schematic overview of intracellular signalling. 'N' denotes nucleus, 'C' the chloroplast, 'M' the mitochondrion, 'A' denotes anterograde signals and 'R' the retrograde signals. References: 1-3 Bychkow et al. 2021, Wang et al. 2021, Lerbs-Mache 2011; 4 (Singh et al. 2015); 5 & 6 (Leister 2005, Pfannschmidt et al. 2020).

BDM incompatibilities

When the tight balance between different genomic compartments is upset, for instance because of secondary contact between separated populations, the effects of the independently acquired mutations become manifest as BDM incompatibilities (Fig. 1). These mutations can be acquired by either the nuclear genome or the organellar genomes but usually by all three as they evolve. Therefore, potential genomic barriers to successful crosses of previously separated populations can come in three kinds: The nuclear genomes of two species have become too different because critical alleles diverged; or chromosome numbers differ or chromosomes will not pair and further development or growth is impossible (Livingstone and

Rieseberg 2003, Chen et al. 2016). Such nuclear genomic-based BDM incompatibilities usually manifest themselves by hybrid necrosis caused by insurmountable differences in the plant immune systems (Wan et al. 2021). Differences in base chromosome numbers may also impair hybridization even though the nuclear genomes 'an sich' are compatible (Sharma and Sharma 2014). Polyploids represent a special case of chromosome number differences. Genomes of, closely related, (auto)polyploids are often still compatible and a cross may result in a plant with even a tetraploid or uneven a triploid ploidy levels. As long as the immune systems of the parents do not conflict the plant will be viable. Polyploids with an even-numbered level of chromosomes can even provide an additional escape from CNI (see below) and indeed polyploidy is known to be a powerful driving force in plant evolution (Soltis et al. 2009, Alix et al. 2017). Therefore, polyploidy is not an insurmountable barrier to hybridization and does not necessarily result in genomic incompatibilities (see for examples of different types of auto-polyploids Chapters 2 and 5).

The second and third kinds of potential BDM incompatibilities are caused by the organelles. When organelles and nucleus from different species are mixed this may result in (partial) lethality or reduced fitness because of impaired signaling as a result of BDM incompatibilities. Signaling between the nucleus and organelles is regulated by nuclear encoded proteins, and other molecules (examples mentioned above) resulting from metabolic processes. Mutations acquired in these genes may lead to changes in the physico-chemical properties of the molecules which in turn may lead to impaired or completely-blocked recognition of signals. When signaling is impaired between organelle and the nucleus this is referred to as Cyto-Nuclear Incompatibility (CNI). We distinguish mitochondrially induced CNI (mCNI) and plastid induced CNI (pCNI). While both types of CNI have effects on overall fitness of the plant, they differ in the way they act. Mitochondrially induced CNI manifests itself by occurrence of (partial) cytoplasmic male sterility (CMS), dwarf growth or both. Plastid induced CNI manifests itself by the occurrence of chlorosis (itself a symptom of impaired chloroplast development) (Schnable and Wise 1998). It is on CNI, or more precise, its expression in crosses, its effects on evolution, the underlying genetics ánd possible underlying causes that this thesis will focus.

Evolutionary consequences of CNI

The effects of CNI mentioned above have an effect on fitness and therefore on the evolution of plants as well. Cytoplasmic male sterility (CMS) results in the inability to pass on genes. Chlorosis, if too severe, will hinder or even prevent a plant from growing fast enough and from securing enough resources to develop to the reproductive stage. Upon secondary contact between two populations both mCNI and pCNI are likely to occur and will affect first generation offspring immediately. Selection for compatible nuclear alleles involved in organelle management will immediately take place for each organelle type present in the offspring (Barnard Kubow et al. 2017). The genetic basis of 'recovery' from CNI, especially

mCNI, has been well-studied (reviewed by Chase 2007). Recovery from CMS is usually brought about by changes in nuclear encoded, mitochondrially targeted genes, especially PPR genes (Kubo et al. 2020) encoding for pentratricopetide signaling proteins.

Genes or alleles identified as the basis of pCNI remain largely unidentified but nuclear encoded genes such as PPR, sigma factors and Whirly genes have been implicated (Wang et al. 2021, Wei et al. 2015, Isemer et al. 2012). This makes sense, as these genes are all involved in expression and maintenance of the chloroplast. The PPR genes and sigma factors are part of the anterograde signaling machinery whereas Whirly factors are involved in RNA editing and correcting micro repeat errors during replication (Cappadocia et al. 2010). The role of the chloroplast, which is at the receiving end of the signaling genes is still unclear, and possible plastome gene(s) explaining pCNI have not been identified so far. This is in contrast with plant mitochondria for which numerous mitochondrial genes have been identified as inducing mCNI (CMS) in specific offspring (Chase 2007, Horn et al. 2014). What is striking is that it is highly dependent on the nuclear genomic background of each cross whether mCNI occurs or not.

Regardless of the exact causes, CNI plays an important role in plant speciation (Burton et al. 2013, Greiner et al. 2011, 2013). While it may appear that nuclear genomes are the main source of circumventing BDM incompatibilities, there is another way in which plants can circumvent, or recover, from CNI. This is by maintaining more than one type of organelles, which in essence, would take away much of the effects of Mullers ratchet. Such heteroplasmy could allow a plant to 'switch' when encountering environments or circumstances that have different metabolic -photosynthetic or oxidative- requirements. Thus, having the ability to maintain more than one type of organelle would also increase adaptability and confer a selective advantage from an evolutionary point of view. No evidence has been found so far for these plastome switches to exist. While heteroplasmy was previously considered uncommon in plants (see Ramsey and Mandel 2019 for an overview of the occurrence of heteroplasmy across angiosperms) it was recently proposed to be 'common' and to be quantitative in nature, with one organelle type dominating (Gonçalves et al. 2020). Plant groups previously thought be heteroplasmic are merely displaying roughly equal and detectable quantities of each type of plastid present.

Pelargonium section *Ciconium* as a model system for studying CNI

Pelargonium L'Héritier ex Aiton is a mainly South African genus, known for its ornamental value, but also popular in science for it highly variable morphology, organellar genomes and the occurrence of heteroplasmy (see below). Heteroplasmy was first proposed to occur in *Pelargonium* by Baur (1909). Although the crossing experiments performed over the years by Kirk, Tilney-Basset and others (see below) gave already convincing results, it was Metzlaff et al. (1981, 1982) who used restriction enzyme analysis to show molecular evidence that variegated plants were indeed heteroplasmic. As argued in preceding paragraphs, heteroplasmy may

have a profound effect on evolution. Previous evolutionary studies on *Pelargonium*, though, have often focused on pre-zygotic selection as explanation for *Pelargonium* species diversity and evolution (Bakker et al. 2005, van der Niet et al. 2006, Jones et al. 2009, Martínez-Cabrera et al. 2012, Ringelberg 2012, Moore et al. 2018, van de Kerke 2020, Burger et al. in prep). These studies show correlations between nectar tube length, habitat variability and shift in flowering phenology with speciation.

In this thesis I use species from *Pelargonium* section *Ciconium* (Sweet) Harvey (1860: 298) (Geraniaceae) as a model system to study different aspects of CNI. In contrast with the pre-zygotic barriers outlined above, CNI is a post-zygotic process. Most *Pelargonium* species originate from the 'Cape Floristic Region' (CFR) in southern Africa where large variations in climate and niches occur at short geographical distances (Linder 1991, Cowling and Lombard 2002, Linder 2003), providing plenty of opportunities for speciation (Verboom et al. 2009). Sections in *Pelargonium* display a range of different lifeforms and base chromosome numbers (summarized by Bakker et al. 1999b) and have been considered to represent 'nested radiations' in the Cape Winter Rainfall region (Bakker et al. 2005).

Pelargonium sect. *Ciconium* likely evolved in the CFR ~5 Mya, based on plastome exon dating analysis using fossils, secondary calibrations (with split-timings derived from other studies) and historical ecological change as calibration points (Kerke et al. 2019, van de Kerke, 2019). That said, several of its species occur outside the CFR, in the Eastern Cape (*P. aridum* and *P. peltatum*), Eastern Africa (*P. multibracteatum*, *P. quinquelobatum*, *P. alchemilloides*, *P. somalense* and *P. insularis*) and the Arabian Peninsula (*P. yemenense* sp. nov., Gibby et al. in prep, *P. omanense* sp. nov. Knees et al. in prep). Plant forms in *Ciconium* range from shrubby herbs (e.g., *P. zonale*, *P. inquinans*, *P. frutetorum*) to herbs (*P. multibracteatum*, *P. aridum*) to semi-geophytic forms (e.g., *P. barklyi*, *P. articulatum*) to creeping plants (e.g., *P. peltatum*, *P. alchemilloides*) (van der Walt and Vorster 1988 and Röschenbleck et al. 2014). Species in section *Ciconium* show contrasting autecology's with some being drought resistant, decidedly fond of heat and thriving under direct sunlight (e.g., *P. acetosum*, *P. aridum* and *P. barklyi*) whereas other prefer more shaded and humid (e.g., *P. zonale*, *P. alchemilloides*, *P. quinquelobatum*) conditions, or occur preferably at high-altitude (*P. acraeum*) (van der Walt and Vorster 1988, http://pza.sanbi.org/pelargonium-acraeum). Photosynthesis in some *Pelargonium* species has evolved from C3 photosynthetic metabolism to so called 'CAM' idling' to cope with drought and heat. While no species displayed full CAM metabolism, there appeared to be signs of a transition induced by water or heat stress (Jones et al. 2003). Other possible adaptations to heat and drought may occur via leaf shape adaptations. For most leaf shapes occurring in *Pelargonium* no correlations with environment and phylogeny were found except for leaf dissection, which is likely an adaptation to reduce water loss (Nicotra et al. 2008, Jones et al. 2009, Nicotra et al. 2011). In *P.* sect. *Ciconium*, *P. aridum* has doubly-dissected leaves but *P. acetosum* and *P. peltatum*, and to a lesser extent *P. tongaense*, have succulent leaves with a thick wax layer. *Pelargonium barklyi* has evolved an almost geophytic lifestyle with few leaves appearing above ground. Floral variation in *Ciconium* is characterized by larger than average flowers, when compared to other *Pelargonium* species, and by the absence of petal markings in the majority of species (petals are 'paucisignate' sensu Röschenbleck et al. 2014).

Section *Ciconium* is the source of two popular ornamentals. The first are the 'garden geraniums' which are hybrids between *P. zonale*, *P. inquinans* and possibly *P. frutetorum*. The second are the 'hanging basket geraniums' which are hybrids of *P. peltatum* with contributions from *P. inquinans* and *P. zonale* (James et al. 2004, Ronald Snijder pers. comm). *Pelargonium* species are relatively easy to hybridize (Sweet 1820-1822, 1824) and, given their popularity with the general public, there is lots of experience available in keen amateur societies working with this group of plants (for a list see: https://www.geraniaceae-group.org/society/links/). *Pelargonium* systematics is well established (see below) allowing us to view our results in an evolutionary context. *Pelargonium* hybrids display CNI at varying levels and this depends on the species involved. More importantly, Geraniaceae organelle genomes, and in particular those from *Pelargonium*, have been demonstrated to be quite different from those found in most other angiosperms in terms of size, repeat content, structural rearrangements and substitution rates in exons (Bakker et al. 2006, Palmer et al. 2000, Mower et al. 2007, Ruhlman and Jansen 2018, 2021). This means that for studying CNI, given the peculiarities and variability of the *Pelargonium* organellar genomes, we can also expect to observe a wide spectrum of CNI related phenotypes. Achieving complete species-level sampling enables us to attempt each interspecific cross and thus allows us to study CNI in more detail. Another, intriguing, aspect of *Pelargonium* section *Ciconium* is the apparent ability to inherit chloroplasts, and possibly mitochondria as well, from either parent. For the chloroplast, the first to observe this was Baur (1909) who could not otherwise explain the variation in chlorosis he detected in the offspring of his *Pelargonium* crossings involving three *P.* x *hortorum* cultivars (collectively named "*P. zonale* hort" by Baur): "*P. zonale* J. C. Mapping", "*P. zonale* Miss Lennor 22", and "*P. zonale* Snowflake". Evidence that biparental inheritance may also be the case for mitochondria comes from studies on only a few crosses performed using two *Ciconium* species, *P. inquinans* and *P. zonale* (Weihe et al. 2009, Apitz et al. 2013). Biparental inheritance of organelles should, if generally occurring throughout the section, allow us to study multiple CNI effects in a single series of offspring.

Pelargonium systematics

Species belonging to genus *Pelargonium* as we now know it, were formally described for the first time by Linnaeus as part of *Geranium* in his Species Plantarum (first published in1752). Linnaeus based this classification on a few accessions, brought to Europe (from 1672 onwards) from the then 'Cape Colony', in which the nectar tubes were probably not conspicuous enough to warrant a new genus. Indeed, the type specimen for *Pelargonium* (herbarium voucher nr: WAG0272117), collected by Clifford (1685-1760) and designated as type by Linnaeus probably originates from the pre-Linnaean era. This specimen remains formally unidentified as flowers are absent or lost, but according to me, it is likely *P. zonale,* based on the overall specimen plant habit and leaf morphology. It was not until 1789 that Aiton published the name *Pelargonium* using the work of L'Héritier, who died during the French revolution, hence *Pelargonium* L'Héritier ex Aiton. Throughout the past two centuries, within *Pelargonium* several changes in systematics have been made (De Candolle 1824, Ecklon and Zeyher 1835, Harvey 1860, Knuth

1912, van der Walt et al. 1977, 1979, 1988, Röschenbleck et al. 2014). Sweet (1820, 1822, 1824), considered several of the sections to be genera (e.g., *Ciconium*), but this was later reversed by Ecklon and Zeyher, (1835) and this remains so to the present day.

The phylogenetic relationships among sections and species are generally well established and held up in the genomic age (Bakker et al. 1989, 1999a,b, Bakker et al. 2000, 2004, 2005, Weng et al. 2012, Röschenbleck et al. 2014 van de Kerke et al. 2019), although >30% of species still remains excluded from phylogenetic studies. The clade (A, B) C) structure as proposed by Bakker et al. 2004 is now commonly accepted as reference when discussing *Pelargonium* evolution. Section *Ciconium* is located in clade C2, sensu Bakker et al. (2004), and has 17 currently recognized species (Röschenbleck et al. 2014). One of these (*P. elongatum* x=4) branches off first and does not group with any of the other species in the section. This leaves 16 species that form the 'Core *Ciconium*' clade with x=9 for all species), further distributed over four supported clades (van de Kerke et al. 2019). In Chapter three a phylogenetic tree is presented based on the repeatome, which provides a nuclear genomic perspective of the evolutionary relationships in *Ciconium* for the first time.

History of *Pelargonium* breeding and research into plastid inheritance patterns

Pelargonium breeding took off in the early 19th century in Victorian England (Sweet 1820, 1822, 1824). From very early on CNI was observed and, although it was unclear at the time what the cause could be, the early breeders appreciated especially the variegated plants for their exotic appearance. Both the breeding and scientific communities remained curious about the cause. Then, as stated above, in the early 20th century Baur (1909) hypothesized that biparental inheritance of plastids occurs in *Pelargonium*. Later in the 20th century a number of studies were performed by Tilney-Basset to study the phenomenon and elucidate the genetics underlying this (see the following sentences for a description of these studies). These studies confirmed that in a *Pelargonium zonale* cultivar ("Flower of spring") plastids were indeed inherited biparentally (Tilney-Basset 1975). In follow-up experiments it was determined that in the cultivars "Dolly Varden" and "Flower of spring" the influence on the expression of variegation is for a large part, but not exclusively, controlled by the maternal nuclear genome (Tilney-Basest 1975, 1976). The author drew the conclusion that "plastid output has little to do with plastid input" (Tilney-Basset 1976) suggesting that plastids were more compatible with one genotype than with the other. The compatibility of genomes and plastomes was then confirmed by crossings, where heterozygous "type II" maternal genomes were more compatible with plastomes than the "type I" homozygous maternal genomes (Tilney-Basset et al. 1981). They further concluded that there was a random effect with respect to which plastomes become fixed in heteroplasmic zygotes. The plant material used had a known genotypic 'preference' inferred from previous experiments- for one or both plastid types, but this was not absolute (Tilney-Basset et al. 1981). In 1982 Tilney Basset et al. proposed that partial lethality of gametes containing the type II genotype was the cause of observed deviations from expected Mendelian segregation patterns. However, this still could not

explain the observed patterns and the hypothesis was subsequently refuted (Tilney-Basset et al. 1992). In the same study it was hypothesized that more than one nuclear-encoded gene was responsible for the observed patterns of incompatibility. This turned out to be the case, based on segregation patterns, for the majority of their crossings (though not for all). He concluded there were likely two independently assorting nuclear genes with different alleles involved in controlling the inheritance of plastids in these plants (Tilney Basset et-al 1992). The number of genes involved in other crosses remains to be tested. It could very well be that the number of genes involved change with increasing or decreasing phylogenetic distance. This is probably related to the time each species had to evolve their unique changes independently from each other as predicted under the BDM model of speciation.

Horn (1994) established that it is quite possible to create interspecific hybrids in *Pelargonium*. She further concluded that it is best to focus on intrasectional rather than intersectional crosses as with increased (phylo)genetic distance comes decreased change of obtaining any viable hybrids. With the exception of the well-known *P. fulgidum* (sect. *Ligularia*) x *P. oblongatum* (sect. *Hoarea*) crosses that are intersectional in nature and well-represented in amateur collections (https://www.geraniaceae-group.org/society/read/hybridizing-with-pelargonium-species-2/). Horn (1994) gave me confidence and justification that crosses among section *Ciconium* species might be possible, useful, and probably necessary to study CNI in detail.

Thesis aims

More often than not, 'crop wild relatives' (CWR's), have traits that could be very useful to us (Migicovsky and Myles 2017). For instance, traits that could improve crop performance when exposed to a-biotic stress factors such as salt tolerance, heat-resistance, or the ability to survive periodic flooding. But biotic traits can be enhanced also, such as: disease resistance, improved rooting, growth form and many others. Finally, CWR's may simply be useful to enhance the crop in terms of human desired traits, for instance by improving flavor of crops we eat or color and shape of flowers for our ornamentals. The occurrence of CNI, however, prevents us from simply taking the species with the desired traits and introducing their genetic material (containing their useful properties) into our crops. This thesis aims to investigate the extent of CNI in inter-specific crosses and to determine if certain cyto-nuclear combinations are possible despite species being diverged. Then, I would like to find out if it is possible to introduce organelles from one species into the nuclear genomic background of another, thereby opening the way for the introduction of particular traits into crops.

Thesis outline

This thesis contains, apart from a General introduction and General discussion, four research Chapters. These are discussed below and a visual overview with highlights is presented in Fig. 3.

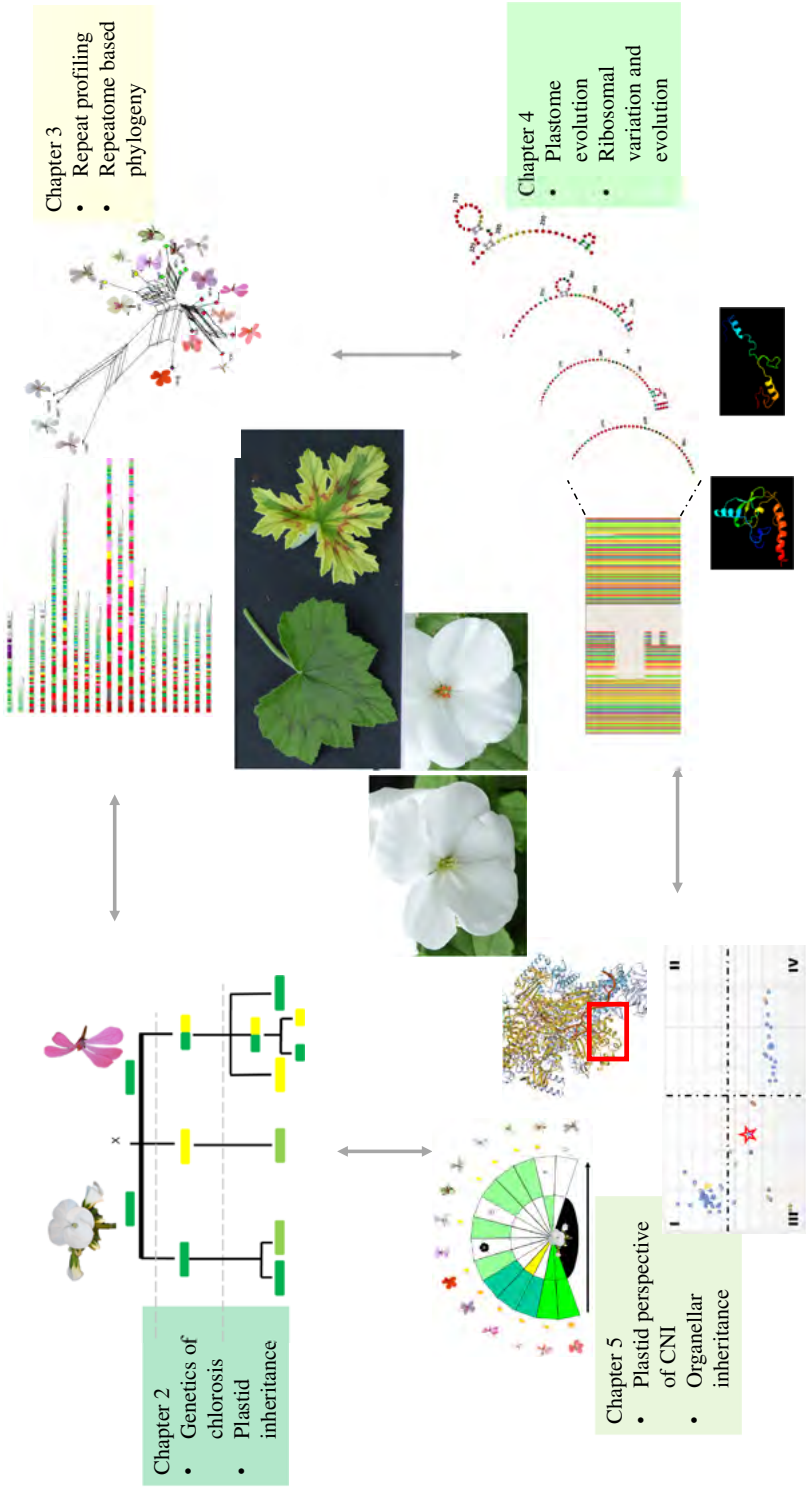


FIGURE 3 | Research Chapter overview. Keywords relating to the content of each Chapter are listed. The figures are drawn from each Chapter and detailed explanations can be found there. Central figures denote two flowers and leaves displaying mCNI and pCNI phenotypes as can be observed in *Ciconium*.

Chapter two presents a study on the genetic basis of chloroplast expression in a crossing series of all available *Ciconium* species with one homozygous cultivar, namely *P.* x *hortorum* 'Pinto White'. We used CNI-related phenotypes, backed up with genotyping, in crosses to determine which plastid was inherited to any offspring. We then established a second (F_2) generation, with known plastids and used CNI phenotype segregation ratios to determine the number of nuclear-encoded genes involved. We assessed segregation ratios of F_1 and F_2 crosses during the vegetative phase as well as during the fruit and seed development phase in order to capture as many CNI effects as possible.

In Chapter three we present our approach to genotype the nuclear genome in the *P.* sect. *Ciconium* species used. For this we used the most abundant parts of the repetitive part of the genome (repeatome) to establish repeat profiles. The repeatome allows, after correction for genomic size differences, comparison of cluster abundancies for each species. Such a profile can perhaps be considered as a kind of 'barcode' characterizing each accession of our study. We also tested if the repeatome could be used to reconstruct a nuclear genomic-based phylogenetic tree and compared this with the current knowledge on *Pelargonium* section *Ciconium* phylogeny. Finally, we explored to what extent certain repeat classes have evolved within the section. These results will, in the absence of annotated nuclear genomes, serve as a proxy for the distinction of genomic types present in the section.

In Chapter four we present novel insights on plastome evolution in *P.* sect. *Ciconium*. We tested if evolution of plastid-encoded genes correlate with each other. In other words, if a plastid-encoded gene in a certain species shows accelerated evolution does it show a similar pattern in other genes for the same species? We then proceeded to explore length variation in a gene (*rrn23*) encoding the RNA backbone of the ribosomal large subunit and investigate possible effects. We modelled the resulting 23S rRNA molecule and discussed the structural differences observed occurring between the species of section *Ciconium*. Finally, we provided evidence from two, de-novo, modelled protein structures that the ribo-proteome is likely undergoing adaptive structural evolution to retain a functioning ribosome in *Pelargonium* section *Ciconium*, given the observed 23S rRNA structural types.

The fifth Chapter addresses mitochondrial and plastid inheritance in three other crossing series in addition to the *P.* x *hortorum* series presented in Chapter two. The three series established, were based on: *P. acetosum*, *P. barklyi* and *P. multibracteatum*. This selection was phylogenetically informed to increase the changes of encountering CNI in our F_1 offspring. We also tested if mitochondria inherit biparentally as well and, if this was the case, we attempted to determine its extent across section *Ciconium*. We then proceeded to discuss a possible explanation for the observed patterns of plastid-related chlorosis (pCNI) in *Pelargonium* section *Ciconium* F_1 interspecific hybrids. To that end, we focused on the plastid-encoded RNA polymerase (PEP) and especially the nuclear controlled *rpo*B subunit of this enzyme. We compared overall rpoB peptide properties and amino acid sequence diversity, as well as model the rpoB peptide structure, to determine the possible effects the variation in this gene may have on pCNI and therefore on plastid expression.

In the final Chapter (six), I discuss the findings of this thesis in the light of *Pelargonium* evolution and speciation. I will integrate the findings of this thesis into pre-existing ideas on organellar evolution in *Pelargonium*, specifically with regards to CNI. *Pelargonium* speciation is often studied in the light of pollinator interactions and environmental variation. These act on the pre-zygotic life phase of the plant, but CNI occurs post-zygotic, and I will discuss the consequences of my findings on the ideas of *Pelargonium* speciation mainly from this perspective.

Chapter 2

Interspecific hybrids between *Pelargonium* x *hortorum* and species from *P.* section *Ciconium* reveal biparental plastid inheritance and multi-locus cyto-nuclear incompatibility

Breman FC[1], Snijder RC[2], Korver JW[1], Pelzer S[1], Sancho-Such M[2], Schranz ME[1], Bakker FT[1*]

[1]Biosystematics Group | Wageningen University and Research, Wageningen, the Netherlands
[2]Syngenta Seeds BV, Andijk, the Netherlands
Published in Frontiers in Plant Science, 2020, 11:614871. Doi: 10.3389/fpls.2020.614871. Slightly modified here.

FCB: ORCID https://orcid.org/0000-0002-9722-7564
FTB: ORCID https://orcid.org/0000-0003-0227-6687

2

abstract

The genetics underlying Cyto-Nuclear Incompatibility (CNI) was studied in *Pelargonium* interspecific hybrids. We created hybrids of twelve closely related crop wild relatives (CWR) with the ornamental *P.* x *hortorum*. Ten of the resulting twelve (F_1) interspecific hybrids segregate for chlorosis suggesting biparental plastid inheritance. The segregation ratios of the interspecific F_2 populations show nuclear interactions of one, two or three nuclear genes regulating plastid function dependent on the parents. We further validated that biparental inheritance of plastids is common in section *Ciconium,* using diagnostic PCR primers. Our results pave the way for using the diverse species from section *Ciconium*, each with its own set of characteristics, as novel sources of desired breeding traits for *P.* x *hortorum* cultivars.

Keywords: Interspecific hybridization, *Pelargonium*, biparental organelle inheritance, cyto-nuclear Incompatibility

Introduction

Several closely-related species from *Pelargonium* sect. *Ciconium* have been used for producing hybrids that are sold world-wide, commonly known as 'garden geraniums', that are some of the most popular and iconic ornamentals. However, there are genetic barriers to establishing crosses and making new hybrids, including Cyto-Nuclear Incompatibility (CNI). that can cause cytoplasmic male sterility (CMS), dwarf growth (DG) and chlorosis in hybrid offspring (Greiner et al. 2015, Zoe et al. 2020). Nearly all angiosperms have uniparental maternal organelle inheritance. Unusually, *Pelargonium* x *hortorum* as well as the species *P. zonale* display biparental inheritance of their organelles (Baur et al. 1909, Tilney-Basset et al. 1992, Weihe et al. 2009). Inheritance of organelles in plants with biparental transmission was found to be non-Mendelian in these studies, even though the expression of organelles is managed by the nuclear genome (Barkan et al. 2014, Börner et al. 2015, Zhang et al. 2019). Phenotypic effects of plastid types in otherwise equal nuclear genomic backgrounds were recently demonstrated in *Arabidopsis* (Flood et al. 2020), but other such studies are so far rare.

Pelargonium species are an attractive model system to study CNI as different organellar effects can be evaluated in offspring with equal nuclear-genomic backgrounds, using established crossing techniques. There is a long history of observations of CNI in *Pelargonium* starting in the 19th century when the foundations were laid for some of the cultivars we have today (e.g., Sweet 1820, 1822). Subsequently, more detailed studies of CNI, especially plastid-induced, were carried out in *Pelargonium* (Baur 1909; Tilney-Basset et al. 1973; 1974; 1975), which, based on segregation ratios, ultimately found support for a two gene model of complementary nuclear- genomic alleles that control the inheritance of organelles in *Pelargonium* (Tilney-Basset 1976, Tilney-Basset and Birky 1981, Tilney-Basset et al. 1982; 1984; 1988; 1989b).

To further advance our knowledge of CNI in *Pelargonium*, we have performed a section-wide survey of most of the crop wild relatives (CWR) of *P.* x *hortorum* and its supposed ancestors *P. inquinans* and *P. zonale* (James et al. 2004) to investigate the inheritance of organelles in general and plastids in particular. There are currently 17 species recognized in *Pelargonium* section *Ciconium* (Röschenbleck et al. 2014) which are all considered the CWR of *P* x *hortorum*. Phylogenetic relationships among these species have recently been reconstructed based on 76 plastome exon sequences (van de Kerke et al. 2019 and references therein). We further investigated if chlorosis in the hybrid offspring can be correlated with a particular plastid type (e.g., the combined plastid proteome, metabolome, and transcriptome inherited from one parent). Given the ubiquitous occurrence of chlorosis in crosses between species of *P.* sect. *Ciconium* and in other sections in the genus (Sweet 1820-1822, Horn et al. 1994, Breman pers. obs.), we expect that biparental inheritance of organelles is more common than is currently reported in the published literature.

Finally, based on segregation ratios over one of the crossing series, we deduced the underlying model of interacting genes which can explain the occurrence of chlorotic phenotypes in these crosses, and hence CNI. We did this by disentangling the effects of each possible plastome type on chlorosis in the F_1 species hybrids.

Material and methods

We established novel interspecific crossings between twelve related diploid species of *P.* section *Ciconium* and *P.* x *hortorum* (species and acronyms mentioned in Table 1). We verified the hybrid status of the offspring using phenotyping, especially by evaluating leaf morphology, as well as flower color and shape (for an example see Fig. 1 for all others see supplementary material. 4A-K). In addition, hybrid status and ploidy level of obtained F_1 hybrids were verified by flow cytometry using HORT as internal reference. Flow cytometry was performed by Iribov bv. (Heerhugowaard, The Netherlands) on freshly collected leaf material using a Partec CA-II flowcytometer according to de Laat et al., (1987). Nuclei were stained with a High-Resolution Kit (Partec).

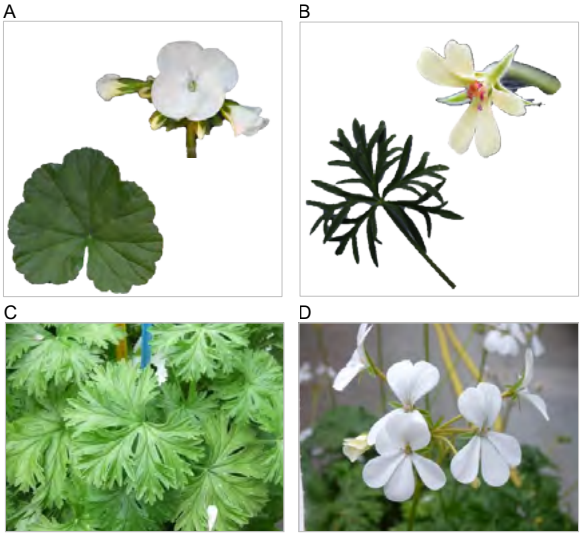


FIGURE 1 | Example of the phenotype of an F1 hybrid generated from the HORT x ARID cross. A) leaf and flower phenotypes of *P.* x *hortorum*.; B) leaf and flower phenotypes of *P. aridum*; C) leaf phenotype of F_1 HORT x ARID; D) the flower phenotype of F_1 HORT x ARID.

Interspecific F1 hybrids

The F_1 hybrids generated in this study were produced from the diploid HORT cultivar 'Pinto White' (PW) crossed with the species outlined above (listed in Table 1 and Fig. 2). In addition, we attempted one interspecific cross at the tetraploid level using HORT 'Tango White' (TW) and *P. articulatum* (ARTI). For all crosses, plants were moved to a pre-cleaned greenhouse and manually pollinated by using dedicated small paint brushes, made of animal hair, at one-day intervals from the moment of flowering, dependent on the species. When seed development did not take place or was impaired, embryo rescue (Table 2) was performed as follows: at two to three weeks after pollination, embryos were collected, dissected, and put on tissue culture in dedicated cabinets using an approach similar to Kamlah et al. (2019).

TABLE 1 | Plant materials used in this study.

Species	Herbarium Voucher#	Species acronym used in the text	Institute[1]
P. acetosum	1243	ACET	NHM
P. acraeum	1975	ACRA	STEU
P. alchemilloides	1885	ALCH	STEU
P. articulatum	1972055	ARTI	WAG
P. barklyi	1972061	BARK	WAG
P. frutetorum	1972062	FRUT	WAG
P. inquinans	0682	INQU	STEU
P. multibracteatum	2902	MULT	STEU
P. peltatum	1890	PELT	STEU
P. quinquelobatum	1972049	QUIN	WAG
P. ranunculophyllum	A3651	RANU	MSUN(*)
P. tongaense	3074	TONG	STEU
P. zonale	1896	ZONA	STEU
P. elongatum	0854	ELON	STEU
P. aridum	1972053	ARID	WAG
***P. x hortorum* 'Pinto white' (PW)**	1972048	HORT	WAG
***P. x hortorum* 'Tango White' (TW)**	NA	HORT	NA

Herbarium voucher information. 1STEU = Stellenbosch University, RSA; AL = Albers/MSUN = Münster, Germany, WAG = National Herbarium of the Netherlands, NHM = Natural History Museum London UK.

FIGURE 2 | Schematic overview of the experimental setup for the F_1 crosses put in a phylogenetic context. The tree displayed tree is based on van de Kerke et al., 2019. The position of *P. ranunculophyllum* is inferred from our plastome assembly data combined with those of van de Kerke et al. 2019 (Chapter six). On the top row are all accessions used, represented by their respective floral and leaf phenotypes and structured by their presumed phylogenetic distance to 'HORT' (van de Kerke et al. 2019). 'Generations established' indicates results obtained with 'F_1' indicating only F_1 plants were obtained, no F_2 could be generated; 'F_2' meaning the plants were fertile and could produce F_2 offspring; red asterisks denote crosses that failed for reasons explained in the text. The bottom row indicates the number of plants used across the seasons.

TABLE 2 | F_1 offspring overview.

Hybrid	Origins of plastid	Phenotype	#offspring obtained
HORT x ZONA	Maternal	Chlorotic	
HORT x ZONA	Paternal	Mostly Green	144
HORT x ZONA	Biparental	Variegated	
HORT x ACET	Paternal	Green Vir	7
HORT x ACET	Maternal	Lethal	
HORT x FRUT	NP	Green	72
HORT x INQU	NP	Green	2
HORT x ACRA	Paternal	Green	24
HORT x QUIN	Maternal	Lethal	
HORT x QUIN	Paternal	Chlorotic Vir	12
HORT x QUIN	Biparental	Variegated	
HORT x MULT	Paternal	Chlorotic Vir	21
HORT x ALCH	Paternal	Chlorotic Vir	8
HORT x TONG	Paternal	Chlorotic	36
HORT x ARID	Paternal	Chlorotic	10
HORT x ARID	Maternal	Lethal	
HORT x PELT	*	-	-
HORT x RANU	*	-	-
HORT x BARK	Paternal	Lethal	2
$HORT^4$ x ARTI	Either	Green	
$HORT^4$ x ARTI	Eihter	Chlorotic	-
$HORT^4$ x ARTI	Either	Lethal	

Vir indicates plants were virescent. * crosses failed either because (paternal) plants would not flower or no fruit was ever observed. **$HORT^4$** refers to a tetraploid cultivar from HORT called 'Tango White' All other HORT refer to the diploid 'Pinto White' cultivar.

Interspecific F_2 populations

In order to evaluate the nuclear background of CNI, we created F_2 progeny of particular F_1 individuals (Table 3). We selected F_1 plants which we assume to contain either one, or both parental cytotypes based on overall leaf coloration. We hypothesized that green and chlorotic plants contained one parental type (at that point unknown which one) and that variegated plants contained both (biparental). We selected from these a number of individuals for subsequent self-pollination to generate the F_2 populations: six plants in total representing three phenotypes encountered in the HORT x ZONA F_1 which includes two green (denoted as: HORT x $ZONA^G$'), 2 variegated (denoted as: 'HORT x $ZONA^V$') and two chlorotic (denoted as: 'HORT x $ZONA^C$') plants (see Figs. 3, 5). In addition, we included one cross (three green plants, the only surviving phenotype) involving *P. acetosum* (*ACET*). We then selected plants from a crossing involving *P. frutetorum* (*FRUT*) and *P. inquinans* (*INQU*) as positive controls for the evaluation. This is because *Pinto White* contains a plastid that is considered to have originated from the *P. inquinans* ancestor (James et al., 2004) and the plastid of *P. frutetorum* is indistinguishable from that of *PW* and *P. inquinans* (Chapter two, but see Chapters four and five noting small sequence variations in *rpoB*). Therefore, we expected these crosses not to display chlorosis in the F_2.

We also evaluated a subset of plants for evaluation of segregation for CNI phenotype patterns that are expressed during the pre-seedling phase. We selected three F_1 parents of HORT x ACET^G, four HORT x ZONA^G and one parent each of HORT x FRUT^G and HORT x ACRA^G. Because fruit-set was low this season for HORT x ZONA we pooled these to enable Chi^2 testing. We feel pooling was justified, because these plants share parentage, and have the same phenotype and plastid.

TABLE 3 | Genotypes detected in F1 and F2 offspring, using diagnostic PCR, for the HORT x ZONA cross.

pedigree	plant/cross	phenotype	Origin of plastid
F0	Hortorum	G	wildtype M
	Zonale	G	wildtype P
F1	**8542**	G	P
	8542	C	M
	8552	G	P
	8552	V	M
	8570	C	M
	8570	G	P
	8570	C	P
	8570	V	M
PEZ-BD8542	8618	C	G-P
	8618	G	G-P
	8618	C	NA
	8619	G	G-P
	8619	C	G-P
	8620	C	G-M
	8620	C	G-M
	8620	C	G-M
	8620	C	G-M
	8627	C	G-P
	8627	G	G-P
	8628	C	G-M
	8628	V	biparental
	8628	G	G-P
	8628	C	G-M
	8629	C	G-M
PEZ-BD85552	8621	C	NA
	8621	G	G-P
	8623	C	G-M
	8623	G	G-M
	8630	C	G-P
	8630	G	G-P
	8631	C	G-M
	8631	G	G-M
	8632	Lethal	G-M
	8632	C	G-M

pedigree	plant/cross	phenotype	Origin of plastid
PEZ-BD8570	8624	C	G-M
	8625	C	G-M
	8625	C	G-M
	8625	G	G-M
	8626	G	G-P
	8626	C	G-P
	8634	Lethal	G-M
	8634	G	G-P
	8634	C	G-M

Structured by cross (G)-P & (G)-M denotes (grand) paternally-and maternally inherited plastids.

Plant Rearing

Plants were grown in a greenhouse from seeds and leaf material was collected from the first primary leaves for DNA extraction. Herbarium vouchers from parental plants, as well as from representatives of F_1 offspring, were deposited at the Naturalis Biodiversity Centre which houses the National Herbarium of the Netherlands (WAG). See Table 1 for the full list of plant material used with Herbarium accession numbers and see Supplementary material 4A-K for representative phenotypes of each F_1 plant.

DNA extraction, primer design and genotyping

Genomic DNA was extracted from leaf material using a modified CTAB protocol (Bakker et al. 1998) followed by RNAse treatment. We designed specific primers for plastome-typing parents and F_1 offspring. We used the Long Single Copy region (LSC) of assembled plastomes (Breman et al. in prep) for *Pelargonium* section *Ciconium* species. The LSC has been shown to contain numerous indels (Chumley et al. 2006, Guisinger et al. 2008, 2011, Weng et al. 2017, Chapter four) which can be used to create genotype-specific primer sites. Visual inspection of sequence alignments, combined with parsimony analysis and using the 'Apomorphy list' command in PAUP*4b10 for windows (Swofford 2002), was performed to find suitable primer sites and to check for unique autapomorphies therein. We specifically scanned for regions with a unique indel or multiple unique substitutions, allowing for genotype-specific primers. Amplicon sizes were designed to be <500 bp, allowing for shorter PCR thermo-profiles. Candidate primer pairs were evaluated using Oligocalc (http://biotools.nubic.northwestern.edu/OligoCalc.html, Kibbe 2007) checking for differences between melting temperatures (ΔTm), self-priming and hairpin formation. Primers were accepted when ΔTm between forward and reverse primers was < 3 °C and with only one hairpin and/or one self-priming predicted. Further, we required a primer site to have a minimum Illumina read coverage of 20. A GC content of 40-50% was preferred, but this was not always possible. A GC content of 40-50% is considered best for ensuring stable binding during annealing and increase the primer pairs efficiency. Finally, we submitted the primers to a BLAST search (set for analyzing short sequences) to compare to all available *Pelargonium* sequences to verify

target-specificity. Occasionally a single primer would have a significant hit to *Pelargonium* species outside section *Ciconium*, but this never occurred for both primers of a pair.

Primers were tested in vitro, using a panel of 16 section *Ciconium* species representing the range of parental plastid variation we would encounter in our offspring. Primer candidates were evaluated using the target accession and an annealing temperature gradient ranging from 49- 60°C. Primers that amplified were subsequently tested against the panel of accessions at the highest possible temperature for which it showed amplification of the target. For PCR profiles and reaction conditions see supplementary material 3.

Phenotyping of F_1 and F_2 plants

Leaf color phenotyping was performed at the seedling stage (Figs. 4A, B). In order to consistently compare phenotypes across populations per cross, we took photos of seedlings at two-week intervals during the seedling stage until the development of the first two primary leaves (Fig. 3). We used the following four-leaf phenotyping categories based on a visual assessment of the phenotypes: 1) 'Green': leaf phenotype comparable to parents; 2) 'Chlorotic', plants are lighter green than either parent or even yellow; 3) 'White', plants germinate, but die within two weeks. Seeds that failed to germinate are added to this category; 4) 'Variegated', plants display more than one chlorotic phenotype in the same individual, presumably due to heteroplasmy (see Figs. 4B, C, and 5). Ratios of the four phenotypes for each tested F_2 population were compared and fitted to a one-, two- and three-gene model of inheritance of nuclear genomic alleles (calculated using the spreadsheet from Montoliu 2012). We assumed four phenotypes and combined these according to five different scenarios, each representing assumptions on expected phenotypic ratios and their expression. The first scenario tested considers four phenotypes (i.e., 'not affected', 'mildly', 'severely', and 'lethal'). The second and third scenarios consider there to be three phenotypes ('not affected', 'affected' and 'lethally affected'). Finally, the fourth and fifth scenarios consider only two phenotypes, ('affected' vs. 'not affected'). We then evaluated these five different scenarios by binning individuals differently. e.g., under scenarios two and three only green plants are considered to be unaffected but the lethal category consisted either of only the white or the white and severely affected plants (Table 4A) Thereby we further assumed different parental genotypes and their expected phenotypic ratios leading to eight testable phenotypic ratios representing models of one, two or three loci involved (Table 4B).

For evaluating seed phenotypes, we used a similar approach, distinguishing four phenotypes: 1) 'normal', not affected by CNI, 2); 'bleached', seed contains endosperm that is still filled, but the seed is bleached; 3) 'watery', in this case the endosperm is bleached and not properly filled; 4) 'lethal', seeds with this phenotype displayed early aborted or undeveloped embryos. For examples see Fig. 6. As for leaf phenotypes, for seed phenotyping we evaluated 5 scenarios as well. We assumed four ('normal', 'bleached', 'empty', 'lethal', scenario 1), three ('not affected', 'affected' and 'lethally affected', scenarios 2 and 3) and two ('affected' vs 'not affected', scenarios 4 and 5) phenotypes.

TABLE 4A | Crosses matching genetic models of inheritance.

scenarios	crosses meeting criteria	one gene model			two gene model			three gene model		observed ratios
		a	b	c	d	e	f	g	h	
Scenario 1	F2_hort_x_zona^{V}	M***	-	-	-	-	M***	-	M***	~4:5:1:3
Scenario 2	F2_hort_x_zona^{G}	-	-	M***	-	-	-	-	-	~7:7:1
	F2_hort_x_zona^{V}	M***	M***	-	-	-	M***	-	M***	~1:2:1
	F2_hort_x_frut	-	M***	-	-	-	M***	-	M***	~1:3:0
Scenario 3	None	-	-	-	-	-	-	-	-	-
Scenario 4	None	-	-	-	-	-	-	-	-	-
Scenario 5	F2_hort_x_zona^{G}	-	-	M***	-	-	-	-	-	~1:1 !$^{50:50}$
	F2_hort_x_zona^{V}	M***	M***	-	-	-	M***	-	M***	~1:2 !$^{75:25}$
	F2_hort_x_frut	-	M***	-	-	-	M***	-	M***	~1:3 !$^{75:25}$
					Fruit/seed phase					
Scenario 1	none	-	-	-	-	-	-	-	-	-
Scenario 2	none	-	-	-	-	-	-	-	-	-
Scenario 3	none	-	-	-	-	-	-	-	-	-
Scenario 4	F2_hort_x_frut	-	-	M***	-	-	-	-	-	1:1 !$^{50:50}$
	F2_hort_x_zona^{G}	-	M***	-	-	-	M***	-	M***	~1:3!$^{25:75}$
	F2_hort_x_acet	-	M***		-	-	M***	-	M***	~1:3 !$^{25:75}$
[2]Scenario 5	F2_hort_x_frut	-	M***		-	-	M***	-	-	~1:4
	F2_hort_x_acra	-	-	M***	-	-	-	-	-	~1:1.5
	F2_hort_x_zona^{G}	-	M***	-	-	-	M***	-	-	~1:1.3

M: mendelian model applies, ***P<0.001 under the χ2 test, ! indicates ratios matching particular model. # indicates not (yet) enough points for a reliable χ2 test. For all observed ratios and counts of phenotypes under each scenario please see SI 5.

TABLE 4B | Possible parental genotype and expected phenotypic ratios.

Lettercode	Genetic model and expected ratios
a	F1xF1=AaxAa 25:50:25
b	F1xF1=AaxAa 25:75
c	F1xF1=AaxAA 50:50
d	F1xF1=AaBbxAaBb 6.25:18.75:18.75:56.25
e	F1xF1=AABbxAaBB 25:25:25:25
f	F1xF1=AABbxAaBB 25:75
g	F1xF1=AaBbCcxAaBbCC 6.25:18.75:56.25:18.75
h	F1xF1=AaBbCcxAaBbCC 25:75

Genotyping F_1 plants and F_2 populations

We plastome-typed F_1 plants using our diagnostic primers described above (Table 3). In those cases where the F_1 population segregated for chlorosis, we tested accessions representing each phenotype. We then typed F_2 plants from each population, and plastome types were then associated with the measured leaf phenotypes to establish the correlation, and thus effect, of each plastid type in the segregating offspring.

Results

Crossing results

Interspecific F_1 hybrids

From thousands of pollination attempts we created a total of 314 F_1 hybrid plants from crossing our species panel to the ornamental cultivar *P.* x *hortorum* PW (see Table 2). Twelve interspecific crosses were successful in producing F_1 plants (Table 2). For three crosses embryo rescue (ER) was needed in order to produce scorable progeny, whereas three attempted crossings failed. Attempts to cross HORT with *P. elongatum* (ELON) failed, but this was expected given the difference in basic chromosome numbers between the two accessions (HORT x=9, *P. elongatum* x=4 (Gibby et al. 1986). The other two failed due to a lack of flowering HORT with *P. peltatum* (HORT x PELT) or poor greenhouse conditions (such as too high humidity or temperature) for the paternal source HORT with *P. ranuncluophyllum* (HORT x RANU). Except for HORT x ZONA (Baur 1909 and many others since), HORT x ACET and HORT x QUIN (Hondo et al. 2014, 2015), these crosses are novel and were never reported in literature before. Remarkably, in 10 cases the F_1 offspring displayed segregation for leaf color phenotype (e.g., chlorosis). When segregation did occur, it ranged from varying levels of chlorosis to nearly green for some crosses to spanning the full range of possible phenotypes from lethal white plantlets to nearly fully green plants (Fig. 3).

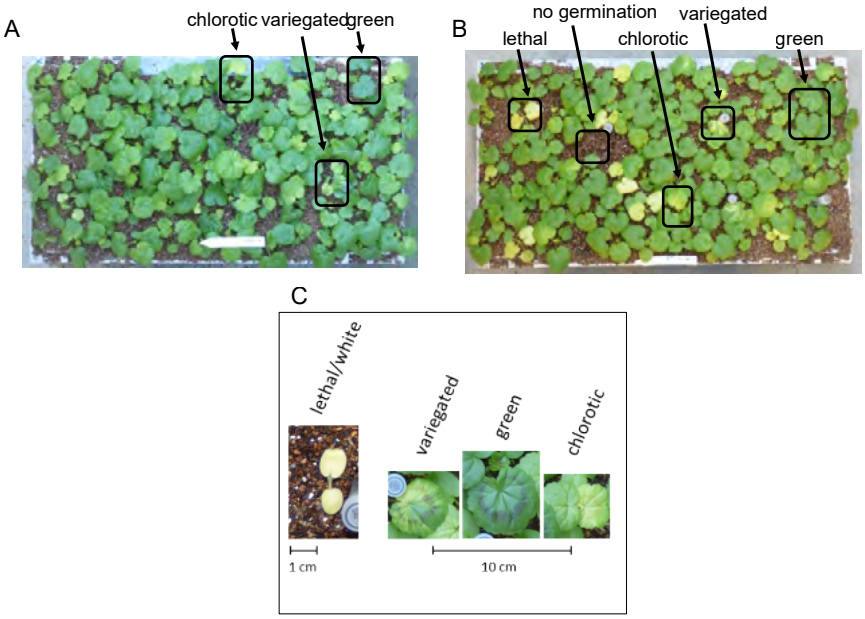


FIGURE 3 | Chlorosis phenotypes. A) Segregation for chlorosis among progeny in the F_2 HORT x FRUT. B) Segregation for chlorosis in F_1 HORT x ZONA. Examples of 'green', 'chlorotic' and 'variegated' plants are indicated, as well as 'lethal' phenotypes and seeds that failed to germinate. C) scaled closeup of examples of phenotypes.

Establishment of the F_2 populations

For all F_1 crosses we were able to obtain an F_2 generation with varying degrees of success e.g., the green F_1 'HORT x ZONA' cross (Fig. 4) used to produce the F_2 yielded significantly more offspring as well as a slightly higher germination success than the variegated or chlorotic parents did (Table 5). The seed phenotypes for F_2 crosses which were used in this study are shown in Fig. 6.

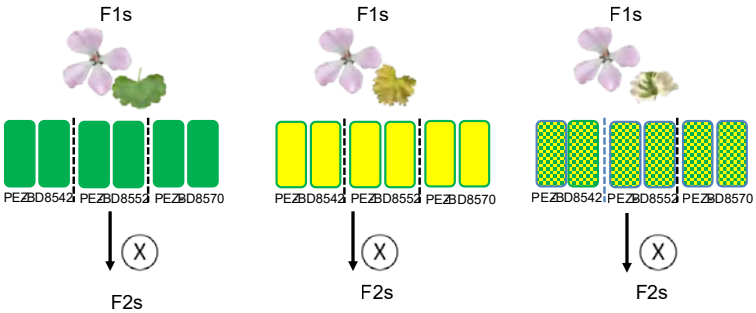


FIGURE 4 | Experimental setup for obtaining F_2 crosses of F_1 parents with different phenotypes.

TABLE 5 | F_2 material obtained using selection of plants from Table 1,

cross	G	V	C	lethal*	# seeds	sum	germination %	marker(s) found	ratio
F_2_hort_x_zona^C	8	1	51	34	125	94	0,78	5 FRIN, 1 ZONA	~1:6:4
F_2_hort_x_zona^G	139	3	138	20	332	300	0,84	1 FRIN, 5 ZONA	~7:7:1
F_2_hort_x_zona^V	52	4	79	48	230	183	0,76	2 FRIN, 1 ZONA, 2 FRIN & ZONA	~1:2:1
F_2_hort_x_acet^C	33	2	49	28	116	112	0,95	ACET	1
F_2_hort_x_frut^G	24	3	63	0	90	90	1,00	FRIN	1
F_2_hort_x_inqu^G	144	0	0	0	144	144	1,00	FRIN	1

Chlorosis phenotypes of F_1 parents and plant counts are given. '**G**' denotes a green plant, '**C**' a chlorotic one and '**V**' a variegated plant. For the description of the chlorosis categories see the text. * Lethal plants are counted without the seeds that failed to germinate.

2

Primer Design and PCR verification

We designed eleven primer-pairs targeting single accessions (e.g., genotype-specific primers) or a group of accessions (Table 6). All primer-pairs performed as expected, except BART, which amplified ARTI but not BARK. We therefore used this primer only for detecting ARTI. For gel photo documentation accompanying the primer pair evaluations we refer to supplementary materials. 2 and 3. All primers worked across a range of template DNA concentrations, (0.1 ng/µl up to >5ng/µl). A 1/10th dilution of the extracts generally increased PCR performance.

TABLE 6 | Primer Pair details.

Primer pair name	target(s)	sequence 5'-->'3	plastome region
FRIN	*P. inquinans, frutetorum & x hortorum*	AAAGGCCAGATTGGGCGGC GACGAATTCGGTCCGATTCAACAC	F: IGS & R; 5' of rna olymerase beta subunit 2 exon
ZONA	*P. zonale*	GAATTGTAATGCGGAGCTGC AAGAAAGAGGATATAGCCGGAC	F & R: IGS
ACET	*P. acetosum*	GAATCCCCACCTACACTACAC CCTTGACTAAAGCGCAATTTTG	F & R: MATK exon, 3' end
ACRA	*P. acraeum*	GACCCTATCTCTCTGTATTC TTTGGTCTCCGAAAAGAAAAGG	F & R: IGS
ALRA	*P. alchemilloides & ranunculophyllum*	GGATCTTATCTATTCTCTATTC CGATCTAGATCTAATTGTAC	F & R: IGS just downstream of trnK-UUU small exon
MUQU	*P. multibracteatum & quinquelobatum*	GGTTTCGCGTCAATTGC CTGAATTTAGCTATGATTTCG	F&R in flanking IGS's of atpH, atpH exon is entirley covered by fragment
ARID	*P. aridum*	CTGAACTGAACTCAAATGGA ATTGCGAGGATCCTACTTTG	F & R: in IGS, fragment contains trnH-IS and trnI-LE
BARK	*p. barklyi*	GAAAGATCTATTCGAGTCGAG GGGGCCTCATTACATTAATC	F: in IGS, R: in intron between tnrL-UAA exons

Primer pair name	target(s)	sequence 5'-->'3	plastome region
PELT	*P. peltatum*	CTCAAAAGAAGGGTAGAAGGG CCCTGTCTGCTCTTTCCAA	F & R: in IGS's surrounding trnS-GGA
TONG	*P. tongaense*	GATCTCAAAGCAAAGAGAGC CTTGGCTAGTGTATACCATTTG	F: IGS, R: in ndhJ exon
BART	*P. articulatum* & *P. barklyi*	GAATCCAAAAGAAATGAAATG AAAAGGAATAGGTTTTGTAG	F & R: IGS between atpB and rbcL

Plastome regions were identified using genbank ID: DQ897681.1 (*P.* x *hortorum*, Chumley et al., 2006). * was not in reference file, but was a ZONA unique region with no homologues in the other LSC assemblies. 'F' refers to forward, 'R' refers to reverse.

2

Phenotyping and genotyping the F_1 and F_2 population for HORT x ZONA and HORT x ACET

For a full overview of the tests for all scenarios under all eight genetic models (Table 4a and b) we refer to supplementary material 5. We discuss here those crosses that demonstrated Mendelian patterns of segregation as well as the models under which this applies. We found that the F_1 plants segregate for chlorosis, with no obvious Mendelian patterns of segregation (Table 4A and supplementary material 5), but that they are otherwise phenotypically homozygous, i.e., non-segregating. When genotyping the F_1 plants, we found that green individuals contained the *P. zonale* type plastid (ZONA), whereas chlorotic individuals contained that of *P. frutetorum* /*P. inquinans* (FRIN) (Tables 2 & 3 plastids of 'maternal origin'). A small minority (<5%) of the plants displayed (partial) variegation and this percentage reduced, for most, as the plant aged with most settling into a single phenotype. From these we detected either the FRIN or the ZONA plastids, but as we recovered both from the F_2 offspring (see below) they must have contained both. We have evaluated plastid types in all phenotypes of F_2 offspring (structured per F_1 cross, Fig. 5) for the HORT x ZONA cross series. We found FRIN and ZONA plastid types in the F_2 (Table 3) and, in general, F_2 offspring always contained the same plastid as was detected in the F_1 plant (for example, see Fig. 5), except for the variegated plants. In the F_1 HORT x ZONA variegated plants we found only one of the plastids, either FRIN or ZONA, but in the F_2 we detected both, even once in one variegated individual (Fig. 5A). We analyzed the bleached and green tissue from this plant and found that white tissue predominantly contained the FRIN type and green contained the ZONA type (Fig. 5A).

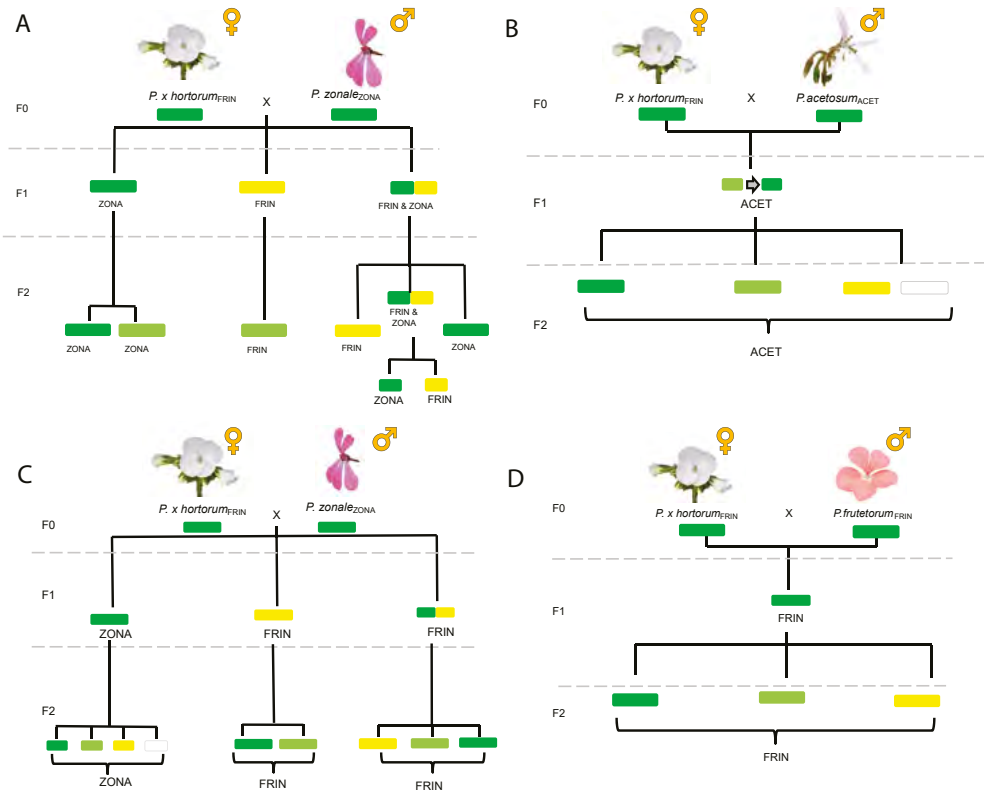


FIGURE 5 | Tracing plastids throughout the pedigrees. Correlation between phenotype and genotype, subsequent segregation in F2 offspring is also indicated. A) and B) two HORT x ZONA crosses; C) the HORT x ACET cross; D) the HORT x FRUT cross. Text in subscript denotes the plastid type found in the parental population.

When pooling the green and light green plants and treating these as one (scenarios 4 and 5) phenotype, subsequent testing for Mendelian patterns of segregation did not yield a clear pattern (supplementary material 5), as was the case for three phenotypic categories. When we categorized the phenotype ratios as 'affected' or 'not affected', we saw that they matched those expected under either a one- or two-gene model for all crosses assuming lethal interactions are also possible between alleles. The populations where the ratios conformed to the one gene model are F_2 HORT x ZONA^V and F_2 HORT x ZONA^C. The segregation ratios in this 'affected vs not-affected' analysis pointed to one lethal combination of alleles and two combinations that yield viable or affected plants. When pooling light green and yellow plants and subsequently testing for Mendelian patterns of segregation, a pattern emerges for the F_2 HORT x ZONA^V and the F_2 HORT x ZONA^G populations (scenario 2). In contrast, when analyzing the observations for the F_2 HORT x ZONA^C plants, there did not appear to be a pattern. The patterns for the F_2 HORT x ZONA^V and the F_2 HORT x ZONA^G populations did point to a genetic difference in the F_1 population (and therefore also in the F_o populations). With the green populations following the one gene model whereby the F_1 was Aa x AA.

The ratios for the plants phenotyped for F_2 seeds and their corresponding possible underlying genetic models are listed in Table 4A. We deduced that there were likely one (in HORT x FRUT and in HORT x ZONA) and two loci (in HORT x ACET)) interacting in this phase of plant development. Given that the phenotypic ratios under scenarios 1 to 4 did are similar to, but not exactly what would expect when of one, two or three genes interact. We suspect that more complex interactions, possibly involving more than two or even three genes, played are role or that the loci involved are linked in some cases with aggravating or moderating effects of linked loci. This appears especially to be the case for HORT x ACRA where ratios under scenarios 1 to 4 are: ~2:1:2:20 (3 loci); ~2:1:2 (2 loci); ~1:1:10 (3 loci) and ~1:12 (3 loci) respectively (see supplementary material 5 for more details).

Positive controls

Our positive controls HORT x INQU and HORT x FRUT yielded 100% green plants in the F_2. only for the F_2 HORT x INQU. Surprisingly, the F_2 of HORT x FRUT displayed segregation for chlorosis and seed phenotypes (Figs. 3B and 6,) indicative of the one gene model of segregation with a heterozygous parent with possible lethal combinations expressed in the pre-seedling phase as well. (Table 4a and Fig. 6).

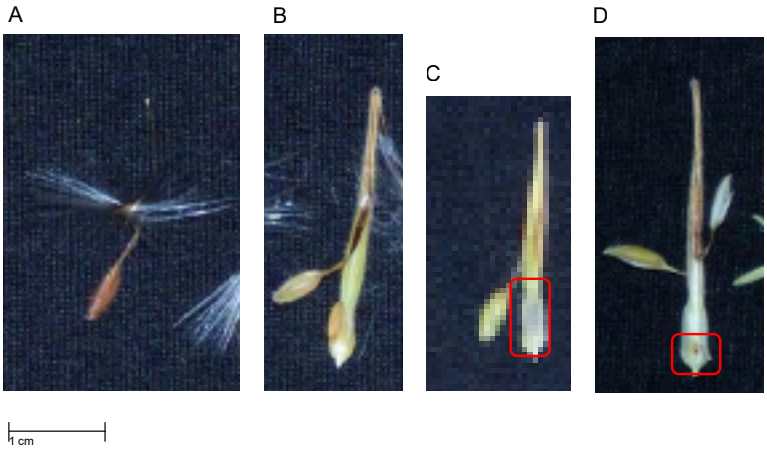


FIGURE 6 | Seed phenotypes displaying signs of Cyto-nuclear incompatibility (see text for further explanation). A) normal seed phenotype; B) bleached seed phenotype; C) seed containing no or watery endosperm; D) no embryo development.

Genotyping phylogenetically more distant F_1 hybrids

We recovered two plastid types in the offspring of F_1 of HORT x QUIN (Table 2). We found segregation for chlorosis and detected both the FRIN type as well as the MUQU type plastids in the offspring. None of these plants were fully green. In the F_1 HORT with *P. aridum* (HORT x ARID) we found segregation for chlorosis, with the majority of offspring lethal and one

plant surviving a full season. For F_1 HORT x ARID We detected FRIN and ARID plastids in the offspring. In the F_1 HORT x ALCH, F_1 HORT x TONG, F_1 HORT x ACRA, F_1 HORT x MULT and HORT x BARK, we detected only the paternal plastids (supplementary material 2). This is similar to the F_1 HORT x ACET cross in that we detected only one type in the offspring suggesting lethal interactions with the FRIN type plastid. In the F_1 HORT x ARTI cross we find segregation for chlorosis and no correlation between phenotype and genotype, we detected both the FRIN and ARTI type plastids. For an overview of all the results see Table 2.

Discussion

We show that biparental inheritance occurs throughout the section and that hybridization is relatively easy, both observations have important implications for interpreting current concepts of *Pelargonium* section *Ciconium* evolution. This study further demonstrates that using multiple interspecific crosses can be used to gain insight into the genetics underlying organelle management and expression, potentially uncovering drivers of speciation. Our studies expand on the two-interacting gene model found to regulate plastid inheritance in *Ciconium* which was inferred 50 years ago by Tilney-Basset et al. (1989b, 1992). While a limited number of crosses between *P.* x *hortorum* and section *Ciconium* have been previously reported (e.g., Hondo et al. 2014, 2015), we have greatly expanded on this by covering nearly all of the CWR in the section including those that are phylogenetically more distantly-related.

Biparental inheritance of plastids and evolutionary implications

We have found that maternal (*P. frutetorum /inquinans*; FRIN) and paternal (other *Ciconium* plastid types) inheritance occurs in nearly all our offspring indicating that the ability to inherit and express more than one plastid is the rule rather than the exception in *Pelargonium* section *Ciconium*. Even though it was demonstrated before on a limited scale, (Baur et al. 1909, Tilney-Basset et al. 1992, Weihe et al. 2009) it was never demonstrated to be so ubiquitous. This has important implications for the study of *Ciconium* speciation as biparental inheritance may provide an escape from the acquisition of deleterious plastid mutations (Mullers ratchet), because there is the possibility for a additional plastome types to occur in the individual plant. Also, it may allow to occupy new niches quicker and perhaps even allow populations that have become separate in space and time to reconnect (Barnard-Kubow et al. 2016, 2017, Apitz et al. 2013, Greiner et al. 2011, 2013, 2015, and Sobanski et al. 2019).

Plastid effects

We have found evidence that in our crosses the FRIN plastid caused bleaching in the HORT x ZONA crosses and that it was possibly lethal for the HORT x ACET cross given the absence of any offspring containing FRIN. The observation that ZONA plastids caused less chlorosis than FRIN in these types of crosses is not new in itself and this study confirms what was already

hinted at by Tilney Basset et al., 1989a and more recently confirmed by Weihe et al. 2009 who observed that the "inquinans plastid" caused bleaching. The F_1 HORT x ZONA plants were, in some cases viable when containing the FRIN plastid allowing us to evaluate the effects of both plastid types in subsequent generations. As to which part of the plastome is the root cause we can only speculate, but a number of genes have been demonstrated to be under selection in the Geraniaceae plastomes (Shikanai et al. 2001, Blazier et al. 2016a, 2016b, Weng et al. 2017, Ruhlman et al. 2017). More surprising was the find that the F_2 HORT x FRUT showed a segregation for chlorosis, even though the F_1 did not. This hints at a slight incompatibility between the FRIN type plastid and either the HORT or FRUT parent. This is surprising given that we cannot distinguish the plastids. Therefore, given the segregation ratios (Table 4A), one nuclear gene, either originating from HORT or FRUT, must be slightly divergent and must be responsible for this effect. Given that this segregation was not the case for The HORT x INQU F_2 population and no segregation occurs when selfing HORT, we deduce that one of the alleles originating from FRUT was responsible.

F_2 segregation pointing to two or three epistatically interacting genes

We demonstrate, in a second generation series of plants that, irrespective of plastid type, there was segregation for chlorosis. Chlorotic phenotypes of the F_2 did not appear to show Mendelian inheritance patterns under a one or two allele model in all cases. However nuclear alleles must be involved because the plastid backgrounds are the same for each plant (Stubbe 1959, 1989, Tilney-Basset 1994, Barr et al. 2010, Li et al. 2013). For the F_2 HORT x $ZONA^V$ population both the one gene model and the two gene model did seem to be equally good at explaining the results. The observed numbers conformed well to the F_1 HORT x $ZONA^V$ population being heterozygous. As outlined above, ratios for the three phenotypic categories do not shed much light on the underlying genetics, but when we categorize the phenotype ratios in a binary way, 'affected or not affected', we see the ratios for all crosses matching or approaching ratios for phenotypes that resemble the situation where one combination is lethal and two combinations of alleles are not. For the HORT x $ZONA^C$ population the ratio is more akin (10:1 under the two phenotypes scenario 5, supplementary material 5a) to the ratio's expected (9:1 under the two phenotypes scenario 4, supplementary material 5a) under a two gene interaction model whereby heterozygous combinations are lethal and the homozygous combinations of at least one allele are not. The ratios for the HORT x ACET cross hint at a possible trihybrid segregation, whereby two alleles interact in a lethal way, because of the following reasoning: If segregation was perfect, we would expect the following phenotypic ratio's under the three gene model; 27:9:9:9:3:3:3:1 but we observe 25::9:1:5 under the four phenotypes scenario 1 (supplementary material 5A). For this pattern to occur we would have to assume there are two alleles that interact in a lethal way, causing the deviation from the expected ratios, but also that there is a third allele which in turn moderates some of these effects or may cause extra lethality.

The ratios of CNI phenotypes observed in the seeds points to a similar type of interactions further explaining why we observe sometimes skewed segregation ratios. In the case of the HORT x ACET cross we observe mendelian segregation of under gene models b, f, h (25:75 phenotype ratios under the one, two and three gene models) with the majority of the individuals being lethal. When we view the ratios of all phenotypes for HORT x ACET and HORT x ACRA (10:3:2:1 and 20:2:1:2 respectively, supplementary material 5b) these, similarly as for the seedlings evaluated, reminiscent of ratios for the two gene model whereby heterozygous combinations are lethal and the homozygous combinations of at least one allele are not. Thus, combining the observations of both seed and seedling phase of plant development, would yield for the HORT x ACET cross a series of at least 5 loci involved in development and expression of organelles. For the HORT x FRUT at least two loci would be required to explain the observations, one acting in each stage of development we studied.

Model of >3 interacting nuclear genes

The observation that the HORT x ACET cross needs two and a three gene model to explain the observed patterns may indicate that those crosses which consist of combination that are phylogenetically further removed from HORT may be subject to the effects of more than three genes. As mentioned above CNI plays a role in embryo and fruit development as well. This in turn could point to a more complex model of genetic interactions involving more loci than we thus far proposed. The machinery for synthesis and management of organelles consists of numerous PPR genes that each act during a different step of these processes (Barkan et al. 2014, Börner et al. 2015, Zhang et al. 2019). These can perhaps be viewed as a genetic 'block chain' whereby no mismatch of combinations is allowed in order to result in a viable, green and self-sustaining plant. In our interspecific crosses there were ample opportunities for these mismatches to occur. While we have no hard evidence for this, we do see from the numbers of plants recovered from our crossing attempts decreases with increased phylogenetic distance. In other words, for the plants from the crosses of e.g., F_1 HORT x ARID we obtained one plant only using the same effort as was invested in the other crosses. This one plant may represent the rare, fortunate gene combination that allows the individual to survive under ideal conditions, while all other combinations are lethal. Given that phylogenetically close crosses (HORT x FRUT, HORT x ZONA, HORT x ACET) require the one-, two- or three- gene model with the assumption of lethality to explain the phenotypic ratios for both the seedling and seed phase we evaluated, we may just be viewing the tip of the iceberg for the phylogenetically more distant crosses. Generally, genes thought to be involved in chloroplast management and expression are Whirly genes, (Maréchal et al. 2009, Isemer et al. 2012, Krupinska et al. 2014, 2019) involved in importing proteins into chloroplasts (Krause et al. 2005, MacKenzie et al. 2020), and PPR genes, acting at the level of RNA editing (Takenaka et al. 2013, Wang et al. 2016a and b, Rojas et al. 2019, Small et al. 2020). These genes are good candidates to study in *Pelargonium* and a closer study of the proteins they encode for as well as the type of RNA editing taking place, may explain both biparental inheritance as well as early-stage processes of speciation.

Data quality

Our approach to phenotyping contains a number of potential sources of error possibly obscuring more nuanced phenotypic differences. We evaluated the seedlings at two points in time to correct for differences in development phase and possible environmental effects on the stability of the phenotypes. Differences in ambient temperature at each point can, potentially severely, affect the expression of chlorosis (pers. observations all authors). Furthermore, the interpretation of the photos, while allowing for reviewing the phenotyping afterwards is subject to interpretation. Defining a plant as 'affected' or not is sometimes context dependent. In the initial germination phase seedlings were germinated under controlled conditions and all at the same time to insure that we were comparing plants in equal phases of development. Great care was taken to make sure the photos of each set were taken at the same day to reduce chance of observing changed phenotypes when environmental conditions change. A further reduction of errors in interpretation can, in the future, be achieved by germinating seeds under even more controlled conditions and using automated imaging software, for interpretation of chlorotic phenotypes. (See for an example of this approach Flood et al. 2020).

Seed phenotypes in *Pelargonium* related to CNI have not been studied before. We have chosen very clear-cut categories and in doing so may have underestimated the actual level of CNI. Nevertheless, our phenotypes are reminiscent of what is regularly encountered in relation to mutated organelle expressing PPR genes in *Arabidopsis thaliana*. (Chi et al. 2008, Du et al. 2017, Zhang et al. 2017). Finally, in some cases we found a discrepancy in plastid types detected, between parents and the offspring of the variegated plants. Probably, variegated plants are able to manage and express both plastids and subsequently one type is outcompeted but not completely removed. This competition was demonstrated in *Oenothera* and occurs at a cellular level (Sobanski et al. 2019).

Crossings

In our study we have obtained at least one individual F_1 hybrid plants for the majority of interspecific crosses attempted (except for *P. ranunculophyllum*). Most were obtainable from seed showing high compatibility of the genomes and plastids. We attribute the two unsuccessful crosses to suboptimal greenhouse climate conditions as we observed that for a pollination to be successful abiotic factors such as climate and humidity are important. (reviewed by Lohani et al. 2020). The chance to obtain a (viable) F_1 plant further roughly correlates with previously published plastome based phylogenetic distances (Fig. 1).

Our approach in this study is reminiscent of the study recently published by Flood et al., 2020 who used cybrids to study the effects of different plastid types in equal nuclear genomic backgrounds. We have used F_1 generation crosses which, though different from the cybrids in the sense that the nuclear genome is hybrid, is still uniform and allows us to study the effects of different types of organelles. Our approach is different that this study focuses more on an evolutionary, rather than at the population level as was the case in Flood et al. (2020).

Conclusions and future applications

The insights from this study further open up possibilities for breeding of currently available *Pelargonium* cultivars with their crop wild relatives. Now we could, conceivably, start breeding in plastids that, for instance, perform better in warmer/colder/dryer climates allowing for the adjustment of cultivars to different climates (Deng et al. 2004, Westerbergh et al. 2018, Cortés et al. 2018) and other abiotic factors (Mezghani et al. 2019, Wang et al. 2019, Singh et al., 2020, Ribera et al. 2020). Especially, photosynthesis would be an interesting trait to focus on as differences between the species are, likely, more dramatic than those observed between the different populations of *A. thaliana* which has been the focus so far when studying the effects of plastid types and photosynthetic efficiency (Flood et al. 2011, Cruz et al. 2016, Flood et al. 2020). The fact that different types of plastids have a different effect in a similar nuclear background means that breeding efforts that wish to incorporate crop wild relatives to increase genetic diversity or introduce new traits should consider the organellar background of the material as well. Knowing the effects can aid in making more informed decisions as to which species to attempt a cross with and which not. This then can lead to more focused and mores successful breeding attempts.

Acknowledgements

We thank Frank Becker for his assistance with testing primers, Toni Lockers for her work on the countless pollinations, Dennis Reus Harald Gorter for the greenhouse management and Ewa Schoutsen and Anja Kolk for performing embryo rescue. Finally we thank the two anonymous reviewers who's comments improved this manuscript.

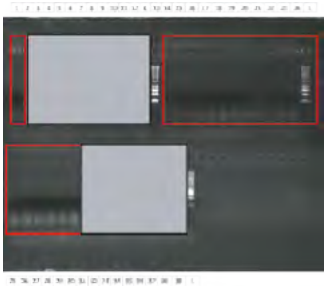
Author contributions

Conceived the study: FCB, FTB, RCS, MES. Carried out the analysis: FCB. Wrote the manuscript: FCB, FTB. Laboratory work PCR: FCB, SP, JWK. Laboratory work embryo rescue: MSM. All authors read the draft and gave feedback.

Funding

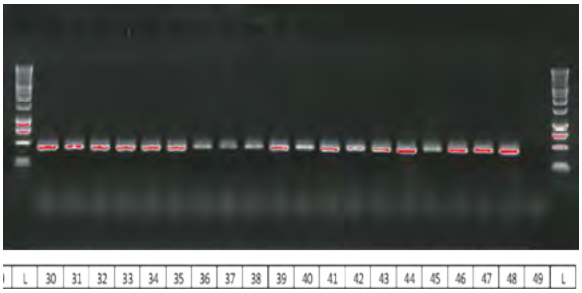
This research was funded by the Dutch Foundations for applied scientific research (TTW). Grant number: 14531 "*Pelargonium* genomics for overcoming cytonuclear incompatibility and bridging species barriers" of the Green Genetics program.

Supplementary material



CH2_SM1_top

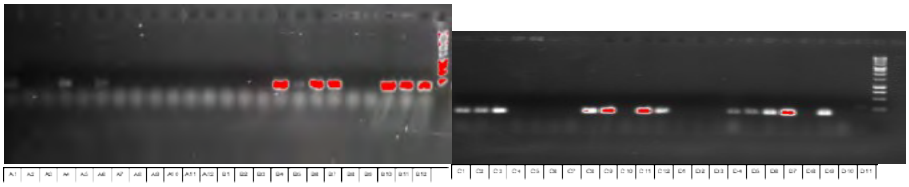
number	specimem
1	8534_08_lightgreen/green_F1_hort_x_acet
L	Lader 1kb
13	8711_green/yellow_F2_hort_x_acet
14	8711_green/yellow_F2_hort_x_acet
15	8711_white_F2_hort_x_acet
16	8711_light green_F2_hort_x_acet
17	8711_light green_F2_hort_x_acet
18	8711_white(+)_F2_hort_x_acet
19	8689_white(+)_F2_hort_x_acet
20	8689_green_F2_hort_x_acet
21	8689_yellow_F2_hort_x_acet
22	8689_green_F2_hort_x_acet
23	8689_light green_F2_hort_x_acet
24	8689_light green_F2_hort_x_acet
L	Lader 1kb
25	8690_green_F2_hort_x_acet
26	8690_green_F2_hort_x_acet
27	8690_yellow_F2_hort_x_acet
28	8690_yellow_F2_hort_x_acet
29	8690_white(+)_F2_hort_x_acet
30	8690_white(+)_F2_hort_x_acet



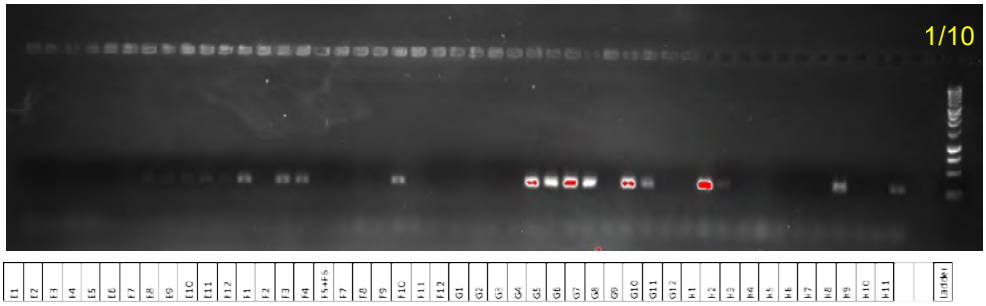
CH2_SM1_bottom

L	weight ladder
30	8534_08_lightgreen/green_F1_hort_x_acet
31	8711_green/yellow_F2_hort_x_acet
32	8711_green/yellow_F2_hort_x_acet
33	8711_white_F2_hort_x_acet
34	8711_light green_F2_hort_x_acet
35	8711_light green_F2_hort_x_acet
36	8711_white(+)_F2_hort_x_acet
37	8689_white(+)_F2_hort_x_acet
38	8689_green_F2_hort_x_acet
39	8689_yellow_F2_hort_x_acet
40	8689_green_F2_hort_x_acet
41	8689_light green_F2_hort_x_acet
42	8689_light green_F2_hort_x_acet
43	8690_green_F2_hort_x_acet
44	8690_green_F2_hort_x_acet
45	8690_yellow_F2_hort_x_acet
46	8690_yellow_F2_hort_x_acet
47	8690_white(+)_F2_hort_x_acet
48	8690_white(+)_F2_hort_x_acet
49	negative control
L	Lader 1kb

PCR20190626 evaluating the hort_x_zona F1 and their F2 descendants using the FRIN primer pair.



PCR20190626 evaluating the hort_x_zona F1 and their F2 descendants using the ZONA primer pair.



SUPPLEMENTARY MATERIAL 1 | PCR Evaluation of F_1 and F_2 HORT x ZONA material.

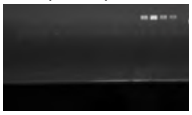
OSM table

PCR with ZI	PCR20190626				
PEZ-BD8517	ELWA0004	green	hort	E1	0
S1056	ELWA0006	green	zona	E2	0
8570_01	241	yellow	F1_hort_x_zona	E3	0
8542_34	242	yellow	F1_hort_x_zona	E4	0
8570_11	243	variegated	F1_hort_x_zona	E5	0
8552_23	244	variegated	F1_hort_x_zona	E6	0
8542_02	246	green	F1_hort_x_zona	E7	1
8570_08	247	yellow	F1_hort_x_zona	E8	1
8570_02	248	green	F1_hort_x_zona	E9	1
8552_13	249	green	F1_hort_x_zona	E10	1
8618	400	yellow	F2_hort_x_zona	E11	1
8618	401	green	F2_hort_x_zona	E12	1
8618	402	light green	F2_hort_x_zona	F1	0
8619	15	dark green	F2_hort_x_zona	F2	1
8619	17	light green	F2_hort_x_zona	F3	1
8620	406	yellow	F2_hort_x_zona	F4	0
8620	9	light green	F2_hort_x_zona	F5	0
8620	11	yellow	F2_hort_x_zona	F6	0
8620	13	light yellow	F2_hort_x_zona	F7	0
8621	120	light green	F2_hort_x_zona	F8	0
8621	121	green	F2_hort_x_zona	F9	1
8623	407	yellow	F2_hort_x_zona	F10	0
8623	118	green	F2_hort_x_zona	F11	0
8624	123	light green	F2_hort_x_zona	F12	0
8625	125	light green	F2_hort_x_zona	G1	0
8625	126	yellow	F2_hort_x_zona	G2	0
8625	127	green	F2_hort_x_zona	G3	0
8626	131	green	F2_hort_x_zona	G4	1
8626	132	light green	F2_hort_x_zona	G5	1
8627	134	light green	F2_hort_x_zona	G6	1
8627	137	green	F2_hort_x_zona	G7	1
8628	122	light yellow	F2_hort_x_zona	G8	0
8628	138	variegated	F2_hort_x_zona	G9	1
8628	141	green	F2_hort_x_zona	G10	1
8628	142	yellow	F2_hort_x_zona	G11	0
8629	143	light green	F2_hort_x_zona	G12	0
8630	145	light green	F2_hort_x_zona	H1	1
8630	148	green	F2_hort_x_zona	H2	1
8631	161	light green	F2_hort_x_zona	H3	0
8631	162	green	F2_hort_x_zona	H4	0
8632	158	white	F2_hort_x_zona	H5	0
8632	159	light green	F2_hort_x_zona	H6	0
8634	149	white	F2_hort_x_zona	H7	0
8634	153	green	F2_hort_x_zona	H8	1
8634	154	light green	F2_hort_x_zona	H9	0
negative	negative	negative	negative	H10	0
8618	4	light greer	F2_hort_x_zona	H11	1
negative	negative	negative	negative	H12	0

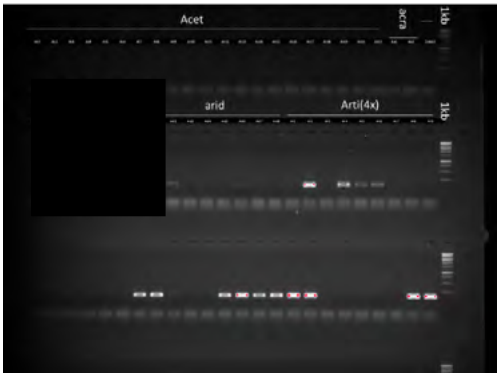
PCR with H	PCR20190626				
PEZ-BD8517	ELWA0004	green	hort	A1	1
S1056	ELWA0006	green	zona	A2	0
8570_01	241	yellow	F1_hort_x_zona	A3	1
8542_34	242	yellow	F1_hort_x_zona	A4	1
8570_11	243	variegated	F1_hort_x_zona	A5	0
8552_23	244	variegated	F1_hort_x_zona	A6	1
8542_02	246	green	F1_hort_x_zona	A7	0
8570_08	247	yellow	F1_hort_x_zona	A8	0
8570_02	248	green	F1_hort_x_zona	A9	0
8552_13	249	green	F1_hort_x_zona	A10	0
8618	400	yellow	F2_hort_x_zona	A11	0
8618	401	green	F2_hort_x_zona	A12	0
8618	402	light green	F2_hort_x_zona	B1	0
8619	15	dark green	F2_hort_x_zona	B2	0
8619	17	light green	F2_hort_x_zona	B3	0
8620	406	yellow	F2_hort_x_zona	B4	1
8620	9	light green	F2_hort_x_zona	B5	1
8620	11	yellow	F2_hort_x_zona	B6	1
8620	13	light yellow	F2_hort_x_zona	B7	1
8621	120	light green	F2_hort_x_zona	B8	0
8621	121	green	F2_hort_x_zona	B9	0
8623	407	yellow	F2_hort_x_zona	B10	1
8623	118	green	F2_hort_x_zona	B11	1
8624	123	light green	F2_hort_x_zona	B12	1
8625	125	light green	F2_hort_x_zona	C1	1
8625	126	yellow	F2_hort_x_zona	C2	1
8625	127	green	F2_hort_x_zona	C3	1
8626	131	green	F2_hort_x_zona	C4	0
8626	132	light green	F2_hort_x_zona	C5	0
8627	134	light green	F2_hort_x_zona	C6	0
8627	137	green	F2_hort_x_zona	C7	0
8628	122	light yellow	F2_hort_x_zona	C8	1
8628	138	variegated	F2_hort_x_zona	C9	1
8628	141	green	F2_hort_x_zona	C10	0
8628	142	yellow	F2_hort_x_zona	C11	1
8629	143	light green	F2_hort_x_zona	C12	1
8630	145	light green	F2_hort_x_zona	D1	0
8630	148	green	F2_hort_x_zona	D2	0
8631	161	light green	F2_hort_x_zona	D3	0
8631	162	green	F2_hort_x_zona	D4	1
8632	158	white	F2_hort_x_zona	D5	1
8632	159	light green	F2_hort_x_zona	D6	1
8634	149	white	F2_hort_x_zona	D7	1
8634	153	green	F2_hort_x_zona	D8	0
8634	154	light green	F2_hort_x_zona	D9	1
negative	negative	negative	negative	D10	0
8618	4	light greer	F2_hort_x_zona	D11	1
negative	negative	negative	negative	D12	0

SUPPLEMENTARY MATERIAL 1 (*continued*) | PCR Evaluation of F_1 and F_2 HORT x ZONA material.

ARTI primer pair on F1 HORT x ARTI



FRIN primer pair on F1 HORT x Ciconium



Paternal primer pairs on F1 HORT x Ciconium

FRIN (maternal) primer pair on F1 HORT x Ciconium

Paternal primer pairs on F1 HORT x Ciconium

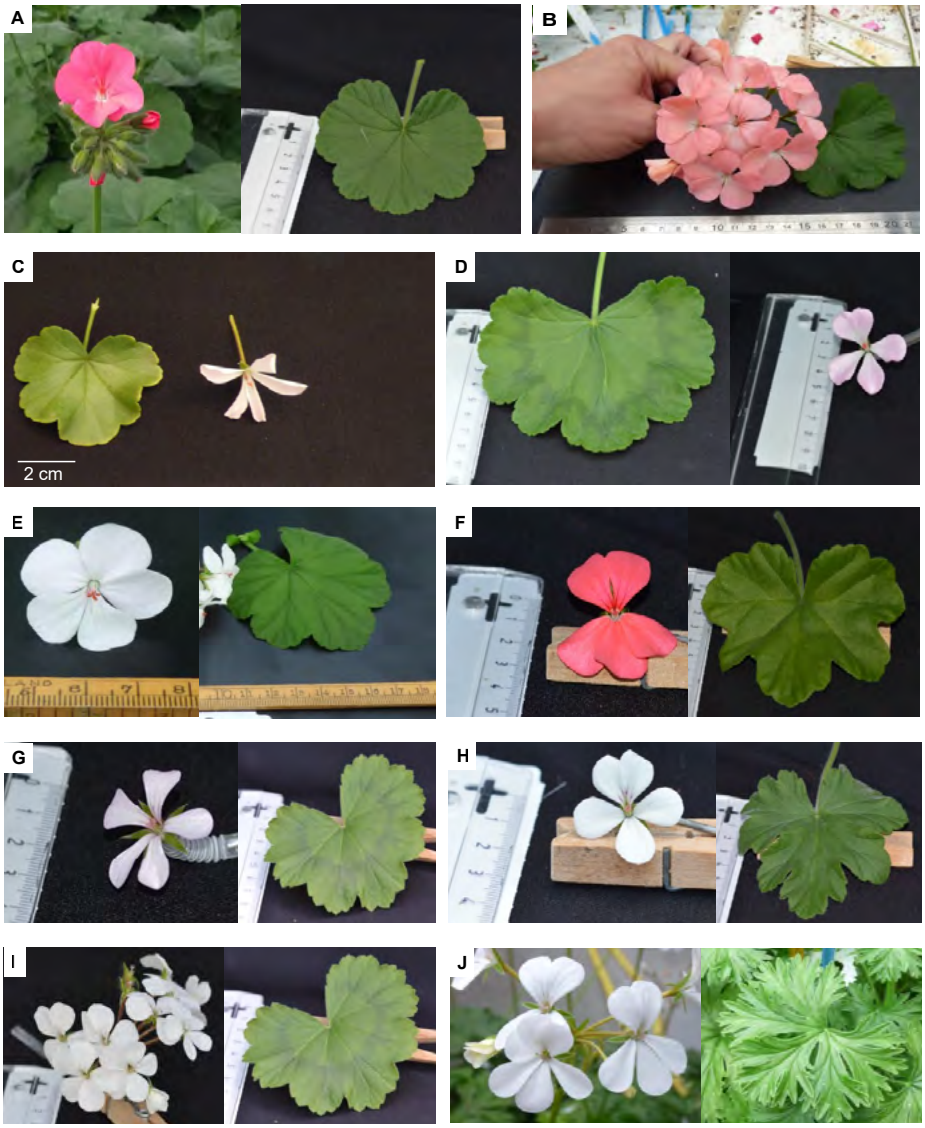
SUPPLEMENTARY MATERIAL 2 | F1 plastid marker evaluation using PCR.

Final concentrations		
PCR Buffer	1x	
MgCl2	0	mM
dNTP	0,2	mM
primerF	1	µM
primerR	1	µM
Taq	0,02	units/µl
µl DNA/reaction	1	µl

		Master Mix preparation		
16,00		reactions of	50	µl
160,00	µl	PCR Buffer	(5x)	
0,00	µl	MgCl	1,5	mM
32,00	µl	dNTP	5	mM
40,00	µl	Primer1	20	µM
40,00	µl	Primer2	20	µM
3,20	µl	Taq	5	Units/µl
512,00	µl	MQ H2O		
787,20				

PCR profile				Cycles
95	°C	2	min	
95	°C	30	sec	
49/60	°C	30	sec	25
72	°C	60	sec	
72	°C	7	min	

SUPPLEMENTARY MATERIAL 3 | PCR recipe for choroplast genotyping.



SUPPLEMENTARY MATERIAL 4A-K | Representative flower and leaf phenotypes for each F1 interspecific hybrid from the Hortorum crossing series.

A) F1 *P. x hortorum* x *P. inquinans*
B) F1 *P. x hortorum* x *P. frutetorum*
C) F1 *P. x hortorum* x *P. acetosum*
D) F1 *P. x hortorum* x *P. acetosum*
E) F1 *P. x hortorum* x *P. acraeum*
F) F1 *P. x hortorum* x *P. tongaense*
G) F1 *P. x hortorum* x *P. alchemilloides*
H) F1 *P. x hortorum* x *P. quinquelobatum*
I) F1 *P. x hortorum* x *P. multibracteatum*
J) F1 *P. x hortorum* x *P. aridum*
K) F1 *P. x hortorum* x *P. barklyi*

SUPPLEMENTARY MATERIAL 5A & B | Available upon request Spreadsheet with all scnearios tested and phenotypic ratios. Available upon request from FCB.

Chapter 3

Repeatome evolution in *Pelargonium* section *Ciconium* (Sweet) Harvey

Breman FC[1], Snijder RC[2], Chen G[1], Schranz ME[1], Bakker FT[1]

[1]Biosystematics Group | Wageningen University and Research, Wageningen, the Netherlands
[2]Syngenta Seeds BV, Andijk, the Netherlands.

FCB: ORCID https://orcid.org/0000-0002-9722-7564
FTB: ORCID https://orcid.org/0000-0003-0227-6687
MES: ORCID https://orcid.org/0000-0001-6777-6565

3

abstract

The repetitive part of the genome (the "repeatome") contains a wealth of often overlooked information that can be used to resolve phylogenetic relationships and test evolutionary hypotheses for clades of related plant species such as *Pelargonium*. We have generated genome skimming data for 18 accessions of *Pelargonium* section *Ciconium* and one outgroup. We analyzed repeat abundancy and repeat similarity in order to construct repeat profiles and then used these for phylogenetic analyses. We found that phylogenetic trees based on read similarity were largely congruent with previous work based on morphological and chloroplast sequence data. For example, results agreed in identifying a Core 'Ciconium´ group which evolved after the split with *P. elongatum*. We found that the Core *Ciconium* group was characterized by a unique set of repeats, which confirmed currently accepted phylogenetic hypotheses. We also found four species groups within *Ciconium* that reinforce previous plastome based reconstructions. A second repeat expansion was identified in a subclade which contained species that are considered to have dispersed from Southern Africa into Eastern Africa and the Arabian Peninsula. We speculate that the core *Ciconium* repeat set was correlated with a possible WGD event leading to this branch.

Keywords: repeatome, evolution, *Pelargonium*, speciation, phylogeny

Introduction

A large part of the eukaryote nuclear genome consists of repetitive DNA sequences (discovered by Britten et al. 1974, Flavell et al. 1974) and the collective repetitive DNA fraction of the genome is referred to as the 'repeatome' (Maumus and Quesneville 2014). In plants, the repeatome can make up >90% of the nuclear genome (Elliot et al. 2015, Novák et al. 2020, C value database by Leitch et al. at https://cvalues.science.kew.org/ accessed October 2020).

The repeatome has been shown to be a useful resource for phylogenetic markers, especially when studying closely related species (Weiss-Schneeweiss et al. 2015, Dodsworth et al. 2015). By using both repeat abundance and repeat similarity comparisons, repeatome dynamics and evolution can be studied in greater detail (Vitales et al. 2019). The repeatome is not homogenous but, consists of different classes of repeats such as various classes of transposons, ribosomal genes and tandem repeats, each with their own properties and dynamics (Wicker et al. 2007; Craig et al. 2015; Louzada et al. 2019, Paço et al. 2019, Hartley et al. 2020, Salim and Gerton 2019, Enriquez-Gasca et al. 2020). Furthermore, repeats occur at positions throughout the chromosomes, can be variable in abundance and can either be conserved, or rapidly evolving. Conserved repeats may have higher sequence similarity as measured across species, perhaps because they may be under functional constraints. Faster evolving repeat clusters on the other hand, are often Transposable Elements (TE's) that have 'escaped' from purifying selection and these can be highly mutagenic (Morgante et al. 2007, Deniz et al. 2019), caused by gene disruption as a result of their insertion in the genome. Their potential to acquire new substitutions is stochastic and may result, when neutral, in an escape from purifying selection because it confers no (dis)advantage (Oliver et al. 2013). We would expect to see selective pressures to be reflected in both abundancies and sequence similarities of repeats. An example of TE's under purifying selection with subsequent duplications comes from so called 'pack MULE' TE's (Hanada et al. 2019). Using repeatome dynamics in a phylogenetic context can shed light on the evolution of the different constituting repeat-classes, i.e., reconstruct 'repeat demography' or the expansion and contraction of repeat clusters through time.

In this paper we study repeatome evolution in the well-known clade *Pelargonium* (Geraniaceae) section *Ciconium* (Sweet) Harvey, which is the source of the iconic 'garden geraniums' (*P.* x *hortorum*) and 'hanging basket geraniums' (*P.* x *peltatum*), both originating from species from this section (James et al. 2004, and see below). We present the first analysis in *Pelargonium* of how nuclear genomic repeat families emerge and expand during cladogenesis. In doing so, we gain insight in *Pelargonium* section *Ciconium* phylogenetic relationships from the repeatome perspective.

Existing *Pelargonium* phylogenetic trees are based on highly conserved plastid and nuclear genes (Bakker et al. 1999a,b, Bakker et al. 2004, Jones et al. 2009, Weng et al. 2012, Röschenbleck et al. 2014). Two studies have incorporated ITS as a proxy for the nuclear genomic perspective alongside cp markers (van de Kerke et al. 2019; Bakker et al. 2004). Two studies also incorporated mitochondrial markers (Bakker et al. 2000, 2004), making the latter

the only *Pelargonium* phylogenetic study to date based on the perspective of all three genomic compartments.

From these studies a consensus emerged that five clades can be distinguished in *Pelargonium*: A1, A2, B, C1 and C2 (Bakker et al. 2004, van de Kerke et al. 2019). While there are differences as to the exact placement of individual species (e.g. *P. nanum*) within these clades, the general phylogenetic hypothesis seems to be stable and has been the basis of current taxonomic classifications which recognizes 15 sections in the genus (Röschenbleck et al. 2014).

Section *Ciconium*, which includes the previously recognized section *Eumorpha* (Röschenbleck et al. 2014), is grouped in clade C2 (Bakker et al. 2004, van de Kerke et al. 2019) and has a base chromosome number of x=9. It contains the ancestral parental species of two of the most commonly sold *Pelargonium* cultivars: the 'garden geranium' *P.* x *hortorum*, with *P. inquinans* and *P. zonale* as supposed ancestors (James et al. 2004), and the 'hanging basket' or 'ivy leaved geranium' derived from *P. peltatum* (James et al. 2004), with various contributions from *P.* x *hortorum*, (RS pers. comm and FCB pers. obs.). *Pelargonium*. sect. *Ciconium* represents a clade that proliferated ~5 mya, according to plastome exon dating analysis by van de Kerke et al. (2019). The *Ciconium* clade likely evolved in the Cape Floristic Region (CFR) of South Africa (van de Kerke 2019), with several of its species occurring outside the CFR in the Eastern Cape (*P. aridum* and *P. peltatum*), Eastern Africa (*P. multibracteatum*, *P. quinquelobatum*, *P. alchemilloides*, *P. somalense* and *P. insularis*) and the Arabian Peninsula (*P. yemenense* sp. nov, Gibby et al. in prep). These SW versus NE African *Pelargonium* (and *Ciconium*) disjunctions have been interpreted to reflect previous jump dispersal events, along the high-altitude rift mountain corridors (van de Kerke 2019). Despite the widespread occurrence of the section, many *Ciconium* species are restricted to specific niches (van der Walt et al. 1988 Vols I-III, Verboom et al. 2009) and may have gone through historical population bottlenecks.

We compared both abundancies and sequence similarities of nuclear genomic repeats in most species of *Pelargonium* section *Ciconium* in order to explore their evolution and utility as phylogenetic markers. We then combined these in a principal component analysis (PCA) to test if different repeats have different patterns of evolutionary change, which could be related to evolutionary age, chromosomal location or specific class.

Material and methods

Plants were grown from seed, in a climate-controlled greenhouse for 10 months after which leaf material was collected for DNA extraction (see Table 1). Current taxonomic opinion on *Pelargonium* sect. *Ciconium* (Röschenbleck et al. 2014) recognizes 17 species. Of these, we included 15 in our study plus *P. karooicum* (section *Subsucculentia*) as an outgroup. For one species (*P. alchemilloides*), two accessions were included, as both ploidy levels and morphology were different (Table 1 and Gibby et al. 1990), as well as their separate phylogenetic (polyphyletic) placement based on plastome exon sequence comparisons (van de Kerke et al. 2019). Furthermore, we included a new species from the Arabian Peninsula (*P. yemenense*,

Gibby et al. in prep). Finally, we included *P. salmoneum*, who's species status is uncertain as it may be a hybrid (Breman et al. pers. obs.).

Genomic DNA was extracted from leaf material using the modified CTAB protocol described by Bakker et al. (1998), now including RNAse treatment (RQ1 Promega), followed by cleaning on a silica column (Nucleo Spin Machery Nagel). DNA extracts were sent to Novogene Inc. (Cambridge and Hong Kong) for Illumina HiSeq sequencing (0.5-1 x coverage). Read libraries were generated from 1.0µg genomic DNA using NEBNext® DNA Library Prep Kit following the manufacturer's protocols, with genomic DNA randomly fragmented by shearing to ~350bp. Fragments, were subsequently subjected to end polishing, A-tailing and ligation of fragments to the NEBNext adapter for Illumina HiSeq sequencing.

Flow cytometry

Average total genomic content per cell (2C value expressed in pg) was determined using flow cytometry (Iribov SBW, the Netherlands) for all 19 accessions. As a reference for the size estimates, we used *P.* x *hortorum* PEZ-BD8517 with known ploidy (2x) and total genome size (2C = 2.33pg, see Table 1).

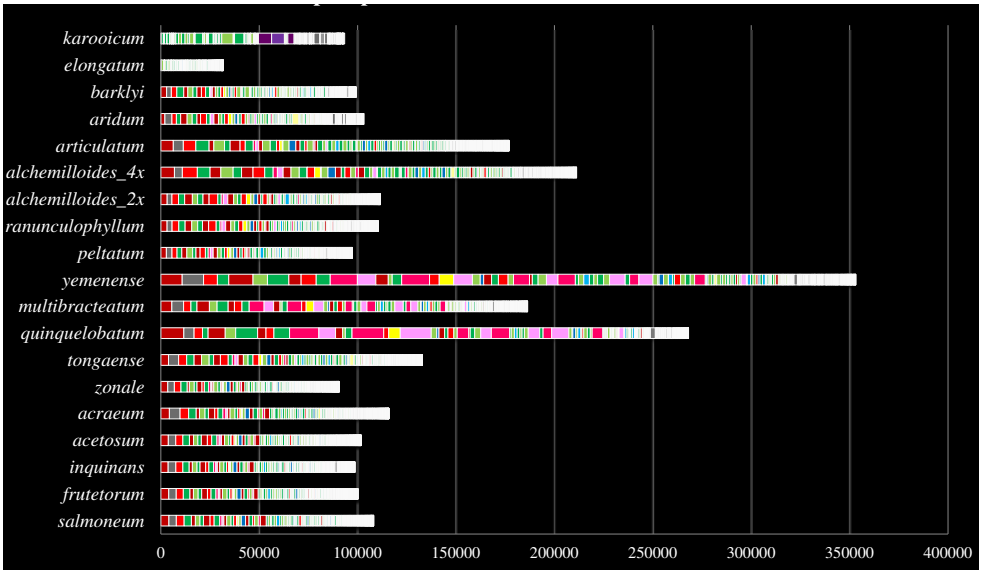


FIGURE 1 | Stacked histogram of homologous and abundance-ranked repeat clusters. The x-axis denotes cumulative read counts, the y axis denotes the accessions. The colors indicate the different cluster categories in terms of their occurrence on the *Ciconium* phylogenetic tree. Green denotes symplesiomorphic clusters, red: Core Ciconium clusters (*P. elongatum* and *P. karooicum* contribute less than 0.1% reads to the cluster), blue: *Ciconium* synapomorphic cluster (outgroup contributes <0.1% reads), yellow: autapomorphic clusters (one accession contributes >20% reads), pink: other synapomorphic clusters (two or more accession contribute >20% reads each), purple: outgroup specific autapomorphic clusters (*P. karooicum* contributes >50% reads), grey: other, small, clusters.

Multispecies comparative analysis

To perform a comparative analysis of genomic repeats from all 19 accessions, we combined a random subsample of reads from each of the Illumina read libraries into a combined multispecies read library (MSL), see below. Clustering of Illumina read pairs from the MSL was performed using the Repeat Explorer 2 (RE2, Novák et al. 2010, 2013) pipeline (version 2.3.7), implemented in the Galaxy server environment, using default settings (i.e., a minimum of 90% similarity over 55% of the read length will build a cluster). A cluster is a visual representation, using de Bruijn graphs, of relationships and overlap between reads. In here, single reads are 'nodes' and sequence overlaps (or relations) are 'edges'. In this way a cluster allows for the visualization of read differences as well as relations between reads. The addition of reads to a cluster is stopped when no more reads match the abovementioned criterium of similarity. For more details on the clustering process, see Novák et al. (2010, 2013).

As indicated above, we used read-subsampling for our MSL as implemented in RE2's default settings for filtering out poor-quality reads. We then set the number of read pairs to be sampled for each accession to correspond to the 1C value (half the 2C value) for each accession, in order to obtain a set of read pairs corrected for genome size (see Table 1). As a practical value, we used 100,000 read pairs per 1 pg of genomic DNA (Table 1) which amounts to ~1.5% genomic coverage. This is slightly more than the 1% shown by Dodsworth et al. 2015 to be sufficient to confidently recover read clusters and their abundancies.

In *Pelargonium* section *Ciconium* some species are polyploid, i.e. *P. alchemilloides* (2n=2x, 4x, 6x, 8x; Gibby et al. 1986), *P. articulatum* (2n=4x, this study) and *P. yemenense sp. Nov.* (2n=4x, this study) (Gibby et al. 1986, 1990). We did not reduce the number of read pairs selected from these polyploids in the manner outlined above, because we wanted to capture their genome dynamics post polyploidization, which can be profound (reviewed in Wendel 2015). It was shown for *Nicotiana* that post-polyploidization genomic variation can be captured by sampling the full size of the polyploid genome instead of reducing to the diploid level. Thos has an impact on the reconstruction of phylogenetic relationships in that it gives insight into evolution post polyploidization that would otherwise be missed (Dodsworth et al. 2015).

Comparative analysis

In order to visualize the repeat content unique to section *Ciconium* and/or specific accessions we used a cumulative, stacked histogram of read abundancies per cluster per accession (Fig. 1). We optimized occurrence of these clusters on a abundancy-based tree for *Ciconium* (see below and Fig. 3).

Phylogenetic tree reconstructions

Abundancies for selected clusters were recorded as counts per accession per cluster. Clusters containing ≥542 reads (or 0.01% of the genome) were retained for downstream analysis. For phylogeny reconstruction, we followed the approach of Dodsworth et al. (2017) and Vitalis

et al. (2019) for cluster abundancy- and cluster similarity-based analysis, respectively, and outlined below.

TABLE 1 | Plant material used in this study. Herbarium vouchers are deposited at L or STEU. Flow-cytometry values 1C as measured in this study, total reads used per accession in the RE analysis and percentage of clustered and non-clustered reads overall. * based on an average across three different measurements. † sp. nov.

Voucher/specimen code/ herbarium ref	*Pelargonium* species	1C(pg)	read pairs used	Reads in top clusters	Reads in singlets/ minor clusters
S1002/STEU1243	*acetosum*	2.43	121410	102187	140633
S1003/STEU1975	*acreaum*	2.44	121934	116451	127417
S1010/STEU1885	*alchemilloides*	2.26	112850	111973	113727
S1009/STEU1882	*alchemilloides (4x)*	4.14	206872	212131	201613
S1088/WAG1972053	*aridum*	2.23	111740	103642	119838
S1026/WAG1972055	*articulatum (4x)*	4.23	211706	177820	245592
S1027/WAG1972061	*barklyi*	2.34	117014	99603	134425
S1072/STEU1022	*elongatum*	1.30	64930	31862	97998
S1087/WAG1972062	*frutetorum*	2.23	111324	100655	121993
S1029/STEU0682	*inquinans*	2.32	115768	99017	132519
S1319/NA	*karooicum*	3.35	167610	94827	240393
S1032/STEU2902	*multibracteatum*	3.14	157040	186939	127141
S1034/STEU1263	*peltatum*	2.20	110230	97818	122642
S1044/WAG1972049	*quinquelobatum*	4.13	206494	268654	144334
S1045/WAG1972041	*ranunculophyllum*	2.10	104946	110818	99074
S1089/NA	*salmoneum*	2.45	122294	108640	135948
S1046/STEU3074	*tongaense*	2.59	129464	133304	125624
S1033/WAG1972037	*yemenense*†	6.07	303702	354143	253261
S1056/STEU1896	*zonale*	2.27	113440	91146	135734
PEZ-BD8517/WAG1972048	*P.* x *hortorum*	2.33*	_		

To perform a character-based analysis of the cluster data, treating each cluster as one continuously distributed character, we used Tree analysis using New Technology (TNT version, Goloboff and Catalano 2016). Cluster data was arranged into an 'accession X cluster' matrix with cluster abundancy as a continuous character state. The matrix was then cube-root transformed to reduce the effect of large abundancy differences and converted to TNT format using Mesquite (Maddison and Madisson, 2019). The continuous range of character states was then 'binned' by TNT into equally sized slices, the optimal amount of which was determined based on the distribution of reads across the cluster. In our case, TNT assigned 64 classes to the matrix, the maximal value for continuous character states (Goloboff et al. 2006, 2008). Tree inference and bootstrap resampling were performed using default 'traditional search' settings in TNT. Resampling analysis involved 'standard' bootstrapping with 1000 replicates (Fig. 2).

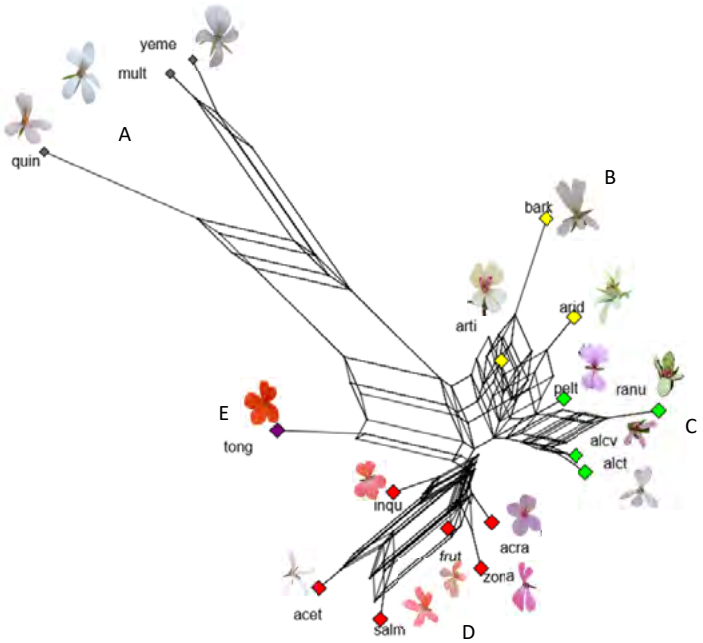


FIGURE 2 | Repeat cluster similarities. Consensus Network (with mean edge weights and conflict threshold 0.05) of 254 NJ trees based on all read clusters (including incomplete ones) that contain ≥ 542 reads; organelle-based clusters are excluded (see text for further details). The length of the edges corresponds to the number of splits supporting it (Holland et al. 2000). Informal /putative groups are indicated, i.e., the 'dispersing' group (A), the 'yellow-flowered' group (B), the 'creeping and climbing' group (C), the 'hort' group (D) and *P. tongaense* (E) unplaced.

For the similarity-based approach to tree building, of the same clusters, we followed Vitales et al. (2019) with the addition that we also included clusters to which not all accessions contribute reads, referred to here as 'incomplete clusters'. In the case of such incomplete clusters a tree based on a NJ analysis yields a polytomy or zero-distance branch for those accessions that do not contribute reads to a comparison. However, for summarizing the NJ trees in a CN this is not a problem as here only splits (not branch lengths) are used.

We inferred pairwise sequence similarities of observed/expected frequencies of reads between the clusters and produced a distance matrix by inverting the values in the similarity matrix. This provided a measure of relatedness based on an all-to-all read comparison per cluster. For every cluster, the distance matrix is then converted into Neighbour Joining (NJ) trees using the R-package APE (Paradis and Schliep 2018). NJ tree topologies were then summarized in a consensus network ('similarity-based CN') using SplitsTree v 4.14.6 (Huson et al. 2006) deploying split conflict thresholds of 10% and 5% (Fig. 4 and Supporting Materials 2).

Ciconium specific patterns

To assess the stability of the inferred relationships based on our repeatome data, we assessed congruence between the similarity-based CN and the abundancy-based cladogram. We further compared the obtained cladogram and CN to test whether there are unique or synapomorphic groups of repeats or clusters. We also compared our repeatome-based patterns with those in previously published, mainly plastome-based, phylogenetic trees to assess possible incongruencies (see below).

In order to study the possible expansion or shrinking of repeat clusters over evolutionary time, i.e. *cluster-demographics*, we use a supercluster (SC) approach (Supporting Material 5). A supercluster is a 'cluster of clusters' connected by edges based on reads from read pairs that occur in different clusters. Because these read pairs were the actual paired-end reads generated in the Illumina sequencing, they support the connection of these clusters (Novák et al. 2010, 2013). We identified SC1 which comprised of eight out of the ten largest clusters from the entire repeatome analysis. A supercluster is therefore, a useful object to evaluate the changes in contributions of all accessions included. Moreover, when you know the age of accessions, or can place accessions in a phylogenetic framework, a temporal context may be added as well. Changes of contributions per accession may then indicate expansion of reduction of a given repeat in one or more accessions.

Comparison with existing phylogenies for *Pelargonium* sect. *Ciconium*

To be able to efficiently discuss the current and past phylogenetic hypotheses, we have summarized all available published phylogenetic hypotheses for *P.* sect. *Ciconium* (James et al. 2004 (their Figs. 1 and 5), Jones et al. 2009 (which is a Bayesian version of Bakker et al. 2004), Röschenbleck et al. 2014, and van de Kerke et al. 2019). These studies are mostly plastome-based, each differing in their taxonomic sampling, therefore making supertree analysis a better option over consensus tree analysis. 'Best trees' from each of these studies, i.e. parsimony consensus trees, best ML trees or Bayesian consensus trees, were collected and decomposed into a MRP (matrix representation using parsimony) matrix in PAUP* with subsequent concatenation and parsimony reconstruction of the resulting super-MRP. The resulting set of equally most parsimonious resolutions were then summarized (Supporting material 1) in a Consensus Network (plastome-CN) using SplitsTree v 4.14.6.

Annotation of repeat clusters in *Pelargonium* sect. *Ciconium*

Annotation of nuclear genomic repeat clusters in RE2 is based on existing hierarchical classifications of repeat classes (Jurka et al. 2011, 2012, Wicker et al. 2007, Llorens et al. 2011). But, given the fast evolution and the limited knowledge of repeat classes across the plant kingdom many of our clusters may be *Pelargonium*-specific and could not be annotated. Therefore, we assigned names using the cluster numbering as they were assigned in the RE2 analysis, based on abundancy.

Contrasting abundance- and similarity-based patterns

For the purpose of discussing and describing the repeat profiles in a phylogenetic context we defined 'generic' or 'symplesiomorphic' clusters shared by all accessions. In contrast, 'Core *Ciconium*' synapomorphic clusters (occurring in *Ciconium* excluding *P. elongatum*) were defined as containing two or more species with each contributing at least 20% reads, and autapomorphic clusters (clusters containing ≥ 20% of the reads from one accession (for the full list see Table 2). We subsequently plotted the non-symplesiomorphic clusters over the abundancy-based tree (Fig. 3). Clades found in the abundancy-based tree were compared with the similarity-based CN's (using the CN with the split conflict threshold of 0.05 (Fig. 2) to determine conflicts between both approaches (Fig. 3).

We performed a PCA to test if different trends of abundancies and sequence similarities are present in different repeat clusters. We took abundancy differences and sequence similarities as variables and performed a PCA on a selection of clusters. We used the three largest cp-based clusters, representing clusters with a comparable genomic location, and therefore presumably subject to comparable trends, (#23, #44 & #45) and six selected clusters (#1, #4, #11, #16, #21, #73) that represent contrasting phylogenetic signals (Fig. 5A). We selected clusters: symplesiomorphic (#4), synapomorphic for section *Ciconium* (#21), synapomorphic for core *Ciconium* (#1), synapomorphic for a clade within section *Ciconium* (#16) and two that were autapomorphic (#16 & #73). We refer to Fig. 1 for the legend and to Table 2 for the cluster characterization. These clusters can also be found together with the clusters plotted over the abundancy-based phylogeny (Fig. 3). All PCA analyses were carried out in R studio Version 1.3.1073 using the libraries: 'FactoMineR' (Le et al. 2008) and 'factoextra V 1.0.7' (Kassambare and Mundt 2020). plots were visualized using the ggplot2 package (Wickham 2016).

Pattern Reproducibility

Repeatome-based phylogenetic reconstruction is claimed to be congruent with other methods of phylogeny reconstruction, efficient and reproducible (e.g. Dodsworth et al. 2015; Vitales et al. 2019). To explore the reproducibility and stability of the phylogenetic trees in relation to the percentage of genome representation that they are based on, we repeated the RE2 analyses using multispecies-sub libraries with (sub)sample sizes of 10%, 25%, and 50% of the reads from the read libraries respectively (see supporting material 4). We reconstructed CN's with edge weights mean: with threshold 0.1, of the NJ trees based on the top 50 clusters for each MS-sL (Fig. 4). We also compared trees obtained from overall repeat abundancies for each of the multispecies-sub libraries using the top 50 clusters (Supporting material 3).

Results

Paired-end read libraries (insert size 350bp) contained 5-7M reads of 150bp length each. The Multi Species library (MSL) was created by random sampling from each accession (see Table 1). The final MSL, comprising reads from all 19 accessions consisted of 5.42 * 10^6 reads. Of these, 4.31 * 10^6 reads clustered whereas 1.11 * 10^6 reads were not assigned to any cluster and are therefore considered 'singlets' (see Table 1 and Fig. 1). Analyzing the MSL on the Galaxy server using RE2 required ~ 3.5 days using default parameters for a 'long run', except for the RAM used by TAREAN, which was changed to 96000000. This yielded a total of 316,059 superclusters and 316,161 clusters of which 311 contained ≥542 reads, or 0.001% of the genome. After filtering out the 56 organelle-based clusters, the final number of clusters was reduced from 311 to 255. A comparative abundancy matrix (255 AMx) was compiled consisting of 255 clusters for 19 accessions, representing 18 *P.* sect. *Ciconium* accessions and an outgroup species: *P. karooicum*.

Flow cytometry

Flow-cytometry values are listed in Table 1. Values (1C) range from 1.3 pg for *P. elongatum* to 6.07 pg for *P. yemenense*. The largest diploid genome is *P. quinquelobatum* (4.13 pg) which is approximately the same size as the polyploid specimens of *P. articulatum* (4.23 pg) and *P. alchemilloides* (4.14 pg).

General overview and distribution of reconstructed read clusters

About 20% of total clustered reads were located in the top 20 clusters, except for *P. elongatum* and the outgroup. Within the core *Ciconium* group, some clear differences in terms of the read abundance per cluster per accession could be observed. These will be discussed in more detail below. The clustering results in Fig. 1 were ranked by size (in terms of total reads per clusters). A cluster was considered 'large' when it represented >1% of the total repeatome of an accession. It was considered 'very large' when it represented > 3% of an accession's repeatome. In the top 20 clusters #10 (*P. zonale*), #19 (*P. quinquelobatum*, *P. multibracteatum* and *P. yemenense*) clusters fell into the 'large or very large' category. Only *P. quinquelobatum*, *P. multibracteatum* and *P. yemenense* contained autapomorphic clusters that made up >3% of their respective repeatomes (see heatmap in Supporting material 7).

In terms of absolute read contributions, *P. yemenense* sp. nov. was the largest -and *P. elongatum* the smallest contributor of overall accession reads in the MSL (Fig. 1), which was in line with the flow cytometry measurements (see Table 1). The diploid accessions *P. multibracteatum* and *P. quinquelobatum* were the largest contributors with ~175K and ~250K (Fig. 1) reads resp., per accession, which was comparable to that of the tetraploid accessions. This was not unexpected given their large flow cytometry-based 2C values we found for these accessions (see Table 1). *Pelargonium yemenense*, *P. multibracteatum* and *P. quinquelobatum* further shared a number of potentially synapomorphic clusters in terms of abundance and in terms of similarity which is further discussed below.

Pelargonium karooicum and *P. elongatum* each had four clusters from the 'large' category (#6, #14, #15, & #20 for *P. elongatum* and #47, #63, # #88 & #95 for *P. karooicum*) in the top 100. For *P. karooicum* three of these clusters were autapomorphic with virtually no contributions from the other accessions reflecting its more ancient common ancestor with rest of the accessions. Even though *P. elongatum* does share these four clusters, it is conspicuous for the fact that it, just as for *P. karooicum*, contributed virtually no reads to the other large clusters that occurred in the core *Ciconium* accessions. The plotting of non -symplesiomorphic clusters over the abundancy-based tree (Fig. 3) yielded the following results: 174 *Ciconium* synapomorphic clusters, 17 of which are 'core' *Ciconium* synapomorphic clusters. We counted 22 synapomorphic clusters, and 17 autapomorphic clusters (see Table 2).

Superclusters

From the top 100 most abundant clusters, 26 belonged to supercluster 1 (SC1). Supercluster 1 contained virtually no reads from *P. elongatum* nor *P. karooicum* with respect to read contributions per cluster (<0.1% for both *P. elongatum* and *P. karooicum*, Fig. 1). We therefore did not consider these further and focused instead on 'core *Ciconium*' repeatome evolution.

1. Abundancy based approach

From the 255 AMx matrix we generated a cladogram based on the 19 accessions with 255 characters (Supporting material 8), 29 of the characters represented 'incomplete clusters'. The cladogram contained 16 clades with Bootstrap support values ≥50. Nine of these had BS values >85.

2. Similarity-based approach

From the similarity matrix, we used the same clusters as for the 255 AMx, but we removed one incomplete cluster which contained only reads for *P. karooicum*. From the resulting 254 NJ trees, we generated a consensus network (similarity-based CN) with edge threshold set to 0.1 (Supporting material 2). The similarity-based CN's for the overall dataset (Fig. 2) showed little conflict amongst the accessions, but was also poorly resolved and not informative with regards to the relations between the *Ciconium* accessions, nevertheless we could still distinguish the following four informal core groups in Fig. 2: a 'Dispersing' group (A) ((quin(mult,yeme)); a 'Yellow Flowered' group (B) ((arid(bark,arti)); a 'Creeping & Climbing' group (C) (ranu(pelt(alch2x,alch4x)))); and the 'Hortorum' or 'Red Flowered' group (D) ((acra,zona),(frut,inqu),(acet,salm)). *Pelargonium tongaense* (E) remained as a single branch, connected to the others via several splits. Thus, our similarity-based CN could be summarized in a tree as follows: (*P. karooicum* (*P. elongatum* (yellow flowered group(creeping and climbing group(dispersing group (*P. tongaense*, hortorum group))))). The hortorum and dispersing groups seem to be clearly separated from the rest of the section with relatively few conflicts occurring under any of the two evaluated thresholds. Between the yellow flowered and creeping & climbing group there remained a number of unresolved conflicts.

3. Abundancy vs. similarity

The overall phylogenetic patterns were similar when comparing the abundance- and similarity-based approaches, with only two clades differing in position between the two. In the similarity-based analysis the dispersing group was clearly separated from the hortorum group. However, in the abundancy cladogram, core *Ciconium* dispersing and hortorum groups (Fig. 3) were nested in a larger clade (BS = 85) with BS = 97 for the dispersing clade. This clade further contained *P. articulatum* and the tetraploid *P. alchemilloides*. Contrastingly, *P. articulatum* was located much closer to *P. barklyi* and *P. aridum* in the similarity analysis (Fig. 2). In the abundancy cladogram, *P. peltatum* formed a single branch and *P. yemenense* was now sister to *P. quinquelobatum*, while in the similarity analysis it was sister to *P. multibracteatum*. Even though the yellow flowered group and the creeping and climbing group in the similarity-based CN (Fig. 2) were not entirely resolved, the tetraploid *P. alchemilloides* grouped together with the diploid *P. alchemilloides* and *P. ranunculophyllum*. The creeping & climbing group also contained *P. peltatum* and while the relationships with the other accessions were not entirely clear, this does not contradict either the abundancy or similarity analyses. Therefore, it seemed likely that polyploids 'attract' each other by virtue of their large genome sizes (Large Genome Attraction, LGA) in the abundancy analysis and this was not the case when analyzing the similarities.

Ciconium specific patterns

We found that most of the diploid core *Ciconium* accessions contributed comparable amounts of reads to the overall analysis, except for *P. multibracteatum* and *P. quinquelobatum*, who contributed ~70-150% more reads. In *Ciconium*, Supercluster one (SC1) is unique (Supporting material 5) to the core *Ciconium* accessions (i.e., *Ciconium* without *P. elongatum*) whereas SC 2 consisted solely of plastome based clusters (not shown) and SC 3 is the largest symplesiomorphic supercluster consisting of nine clusters (OSM 5). SC1 consisted of 31 clusters in total with 26 of them in the top 100 and eight in the top ten. SC1 is expanded in and diagnostic for section *Ciconium* relative to *P. elongatum*. This supercluster is further expanded within *P. quinquelobatum*, *P. multibracteatum* and *P. yemenense* relative to the other core *Ciconium* accessions. These three accessions contain homologous clusters with the rest of the section, but six clusters (#16, #19, #32, #40, #43 & #148) from SC1 were unique to these three accessions (the other accessions contributing <0.1% reads). Taken together these made up ~10% of their respective genomes indicating expansion and possibly a relaxation of constraints on proliferation of these repeats. Therefore, it seemed that species with a high 2c value, i.e. *P. quinquelobatum*, *P. multibracteatum* and *P. yemenense*, had a different repeatome development as compared to diploid species, but similar to polyploid species.

Abundancies and sequence similarity trends

Comparison with existing phylogenies for *P.* sect *Ciconium*

The same four groups and four single lineages that were recoverable from the similarity-based repeat analysis were also in our Consensus Network summary of published phylogenetic trees ('plastome-CN') (Supporting material 1). The large number of unresolved splits between the 'creeping & climbing' group C and the 'yellow flowered' group B could indicate that these should be grouped. Experimental evidence from interspecific crosses also demonstrates that representatives of these groups yield fit, green plants albeit infertile (Breman et al. in prep.) indicating at least some incompatibility between the groups. Based on further evidence from plastome based phylogenetic trees (van de Kerke et al. 2019) and morphology (Röschenbleck et al. 2014) we treat these here as separate groups.

The abundancy-based tree (Fig. 3) was in conflict with the plastome-CN, caused chiefly by the 'polyploid branch attraction' (LGA) mentioned above. The similarity-based CN on the other hand approached the plastome-CN much closer, with *P. aridum* being included by us in the yellow flowered group rather than being considered a single lineage. However, it did not contradict the plastome-CN directly as *P. aridum* could be considered a single lineage as well as it has a few splits setting it apart from *P. barklyi* and *P. articulatum*. There was agreement between the plastome-CN, the similarity-based CN and the abundancy base tree with 100% bootstrap support for the placement of *P. elongatum* in *Ciconium*, separate from the Core *Ciconium*, which themselves formed a clade with 100% bootstrap support. Within the Core *Ciconium*, bootstrap supports >85% are found for the dispersing group nested in the hortorum group, effectively making the hortorum group paraphyletic. The two accessions for *P. alchemilloides* were split in our abundancy-based tree. This was, again, in conflict with the similarity-based CN (Figs. 2 & 3).

Pelargonium articulatum was sister to the tetraploid *alchemilloides* and this was also in conflict with the similarity-based CN and plastome-CN's. In the abundancy-based tree *P. aridum* and *P. barklyi* are also single branches. This was partly in conflict with both the similarity-based CN and the plastome-CN where *P. barklyi* was grouped with *P. articulatum*. *Pelargonium aridum* formed a single branch in the plastome-CN and the abundancy-based tree, but not in the similarity-based CN. *Pelargonium peltatum* formed a single branch (BS79) in the abundancy-based tree, but not in the similarity-based CN or the phylogenetic summary. The grouping of *P. inquinans, frutetorum, zonale, salmoneum* in one clade and of *P. multibracteatum* and *P. quinquelobatum* in another was supported by all three analyses.

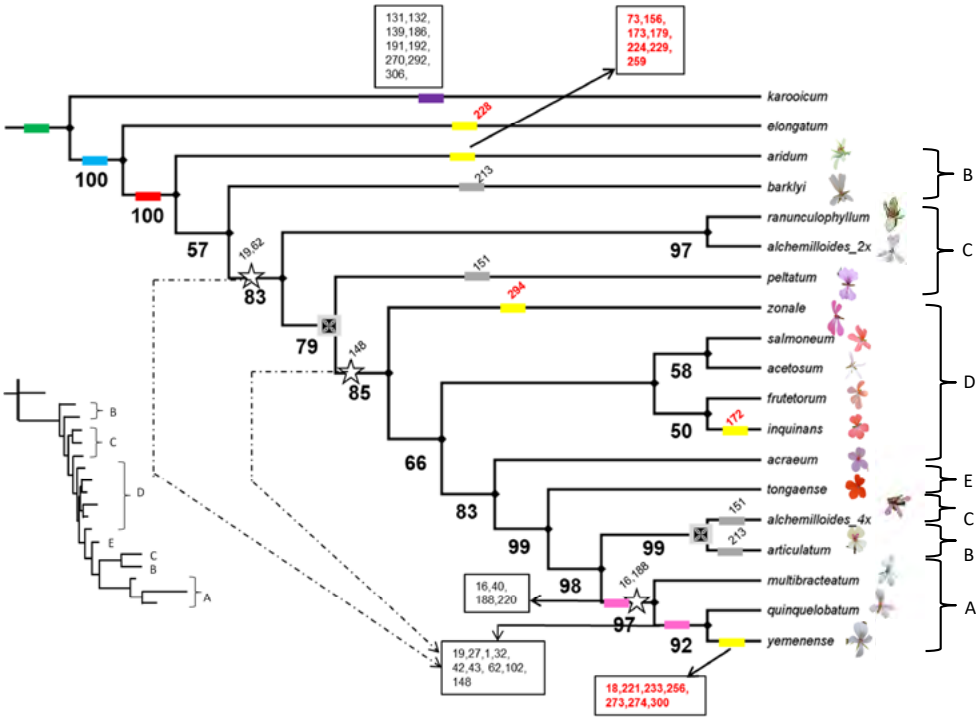


FIGURE 3 | Repeat cluster abundancies. Abundancy-based cladogram with clusters indicated in boxes. Symplesiomorphic clusters are indicated in green; autapomorphic clusters in yellow; synapomorphic clusters in blue (for *Ciconium*), in red for Core *Ciconium*, in pink for other clades, and polyphyletic clusters are in grey. For the corresponding clusters see Table 2B. A—E labels refer to informal groups as indicated in Fig. 2. Nodes not present in the similarity-based consensus networks are indicated by "✠". Bootstrap support values are labelled in bold type on the branches. Other numbers indicate cluster synapomorphies for nodes with 0.1% threshold for read contribution (with clusters indicated). The dashed arrows indicate clusters that shifted when the threshold for synapomorphy was changed from 0.1% read contribution to 20%. Inset: the same tree as additive tree.

Pelargonium yemenense is a new species and was not included in the summary of previously published phylogenetic trees. *Pelargonium insularis* was not included in our study (but was in Bakker et al. 2004), therefore their positions cannot be directly compared. However, they are likely closely related because in all analyses they are sister to *P. multibracteatum* and *P. quinquelobatum*. The positions of *P. acraeum* and *P. tongaense* vary, but all analyses agreed that *P. acraeum* is either a sister to, or part of the hort group in the plastome based CN (Supporting material 1) and the similarity-based CN respectively (Fig. 2B). *P. tongaense* is a single branch in all analyses and its position remains unresolved in the consensus networks (Fig. 3).

TABLE 2 | List of clusters per abundance category for the abundance based cladogram. Clusters are listed per clade/accession as indicated in Figure 3.

cluster types	cluster numbers	
Symplesiomorphic clusters	1, 4, 6, 7, 10, 14, 15, 20, 22, 24, 26, 29, 30, 33, 34, 36 - 39, 46, 47, 49, 51, 54, 56 - 58, 66, 69, 76, 80, 82, 87, 88, 90, 91, 95, 98, 100, 103, 107, 108, 111, 113, 114, 116, 117, 120 - 122, 126 - 129, 133, 134, 136, 137, 138, 140, 142 - 144, 146, 147, 149, 150, 152, 153, 154, 157 - 171, 175 – 178, 180 - 185, 187, 190, 193 - 211, 214 - 219, 222, 223, 225, 230, 231, 234, 237 - 239, 241, 244, 246 - 255, 257, 258, 260, 262 - 267, 269, 272, 275 - 280, 282 - 288, 291, 296, 298, 299, 301, 302, 304, 307, 310	
Synapomorphic clusters supporting section *Ciconium*	21, 35, 48, 50, 64, 106, 110, 112, 174, 235, 261, 271, 281, 289, 295, 305, 311	
Synapomorphic clusters supporting core *Ciconium* species	3, 5, 6, 8, 9, 11, 12, 13, 17, 25, 27, 28, 31, 32, 40, 55, 96, 109, 123, 172, 212, 243, 303	
Autapomorphic clusters containing >20% reads for ONE accession. These are labelled separatly in red and plotted over the tree	*aridum*	**73, 156, 173, 179, 224, 229, 259**
	elongatum	**228**
	zonale	**294**
	inquinans	**172**
	yemenense	**18, 221, 233, 256, 273, 274, 300**
Autapomorphic clusters outgroup, the outgroup contains > 50% reads	131, 132, 139, 186, 191, 192, 270, 292, 306	
Other synapomorphic clusters, indicating clades with accessions contributing >20% reads per accession	16, 19, 27, 31, 32, 40, 42, 43, 62, 102, 148, 188, 220	
Paraphyletic clusters. Two or more species contribute >20% reads per accession, but are not in a clade	151, 213	

Contrasting repeat abundance and similarity patterns

The primary axis of the PCA (Fig. 5A & B) explains both abundancy and similarity differences among the *Ciconium* accessions. Repeat abundancy showed a positive correlation with the primary axis whereas the repeat similarities were negatively correlated (see the biplot in supporting material 9). The abundancy differences were more pronounced, and they explained the largest part of the first axis (Fig. 5B). The 2nd axis is also explained by both the abundancy and similarity variables, but mainly by accessions from the dispersing clade indicated by the dots in the ellipse in Fig. 5A. Interestingly the 3rd axis is mainly explained by the abundancy differences and similarities from the polyploid and large diploid genomic accessions, perhaps suggesting a different trend (Fig. 5A & B). From our PCA we see that clusters in the plastome, symplesiomorphic and the *Ciconium* category are driven both by differences in similarity

and abundancy but there is no indication of expansion or contraction see the cutout in Fig. 5A, leaving sampling differences as main cause explaining the abundancy patterns. The core *Ciconium* cluster (red dots) in contrast, showed increased correlation with the first axis and an increase in abundancy seemed to explain these clusters better. These are possibly younger clusters that have not yet gone through curbing/restraining of their expansions. The synapomorphic clusters (red, orange and pink dots in Fig. 5A) were variable with both similarity and abundancy being responsible for the observed variation. This may indicate a lack of constraints on expansion and substitutions, making these possibly even younger than the core *Ciconium* clusters. These clusters corresponded to the clusters from SC1 that are synapomorphic for the clade of *P. yemenense*, *P. multibracteatum* and *P. quinquelobatum* whose estimated age of about ~1.5 my (van de Kerke et al. 2019) is indeed considerably younger than the rest of *Ciconium*.

The autapomorphic clusters also showed contrasting patterns, especially in the case of cluster 16, which belongs to the clusters that have expanded in the abovementioned three accessions. Cluster 73 in contrast, which is autapomorphic for *P. aridum*, does not appear to be inflated or diverged, it represents perhaps an older cluster that evolved uniquely in this species. Fig. 5A further shows that the plastome-based clusters are highly correlated, with all showing virtually equal trends, in contrast to the five repeatome based clusters added to the same analysis. This further suggests that different (or fewer) genomic constraints act on the six selected repeatome based clusters. Testing for correlation between similarities and abundancy differences, for the selected six clusters and five accessions, revealed no correlation between the two types of data (not shown, but see supporting material 10).

Reproducibility of cluster-based phylogeny reconstruction

We reconstructed similarity-based CN's and cladograms (abundancy-based trees) using the 10%, 25%, 50% and 100% MultiSpecies sub libraries (Supporting material 4). We have used the top 50 largest clusters of the MS sublibraries and these yielded 37/38/45/47 repeatome based NJ trees from which we reconstructed the similarity-based CN's (Fig. 4). The abundancy-based trees were based on 37/38/45/47 characters with repeatome based abundancy scores, respectively (Supporting material 3). When comparing the similarity-based CN's we observed that the patterns obtained are largely consistent across the 10, 25, 50 and 100% samplings. Overall, node resolution appears to increase with increasing sampling.

Discussion

Our study confirms earlier findings (Dodsworth et al. 2015, 2017) that the repeatome contains a treasure trove of information useful for studying phylogenetics and evolution. Combining sequence similarities and abundancies for a number of clusters appears to provide important information about changes in abundancy and sequence similarity for particular repeat classes. Vitales et al. (2019) found that in *Nicotiana* clusters containing <0.01% of input reads are more likely to show species-specific patterns, making them less suitable for detecting 'generic-level' clusters. We decided to consider these autapomorphic clusters as they could, potentially, shed more light on the evolutionary history of that particular accession. These types of clusters, by their very nature, reinforce the differences between accessions, some contribute reads to them and some do not. Nevertheless, these clusters probably represent actual biological phenomena of expanded repeats and we have opted to include these in our study as they are particularly informative with respect to repeatome evolution within the section.

Pelargonium sect. *Ciconium* phylogeny

Phylogenetic patterns are well studied in *Pelargonium* using mainly plastome-based characters (Bakker et al. 1998, 1999b, c, Bakker et al. 2000, James et al. 2004, Bakker et al. 2004, 2005, Jones et al. 2009, Röschenbleck et al. 2014, van de Kerke et al. 2019). But comparisons of chemical compounds (Lis-Balchin 1996, 1997) and karyology (Gibby et al. 1990) have also been used to asses relationships in *Pelargonium* in general and for section *Ciconium* in particular. Some non-chloroplast sequences have been used for phylogenetic reconstructions including nuclear genomic rDNA ITS sequences, (Bakker et al. 2004; van de Kerke et al. 2019, Jones et al. 2009, James et al. 2004) and mitochondrial encoded *nad6* exons (Bakker et al. 2000).

Our phylogenetic trees were based on the nuclear genomic repeat landscape, which are considered to be independent from specific constraints acting on individual genes or genomic regions (Dodsworth et al. 2015). Our similarity-based CN (Fig. 2) was comparable to the phylogenetic trees reconstructed by van de Kerke et al. 2019 and the previous four published studies indicated. Both the abundancy-based and similarity-based analysis confirmed the close relationships between the accessions of the hortorum group (Fig. 2). The grouping of Core *Ciconium* (or the splitting of *P. elongatum*) was supported by all analyses. In contrast, the cluster abundancy-based analysis conflicted with the previously published literature and our sequence similarity-based analysis. In this analysis the dispersing group was nested in the hortorum group together with *P. articulatum* and this was not found in any other analysis.

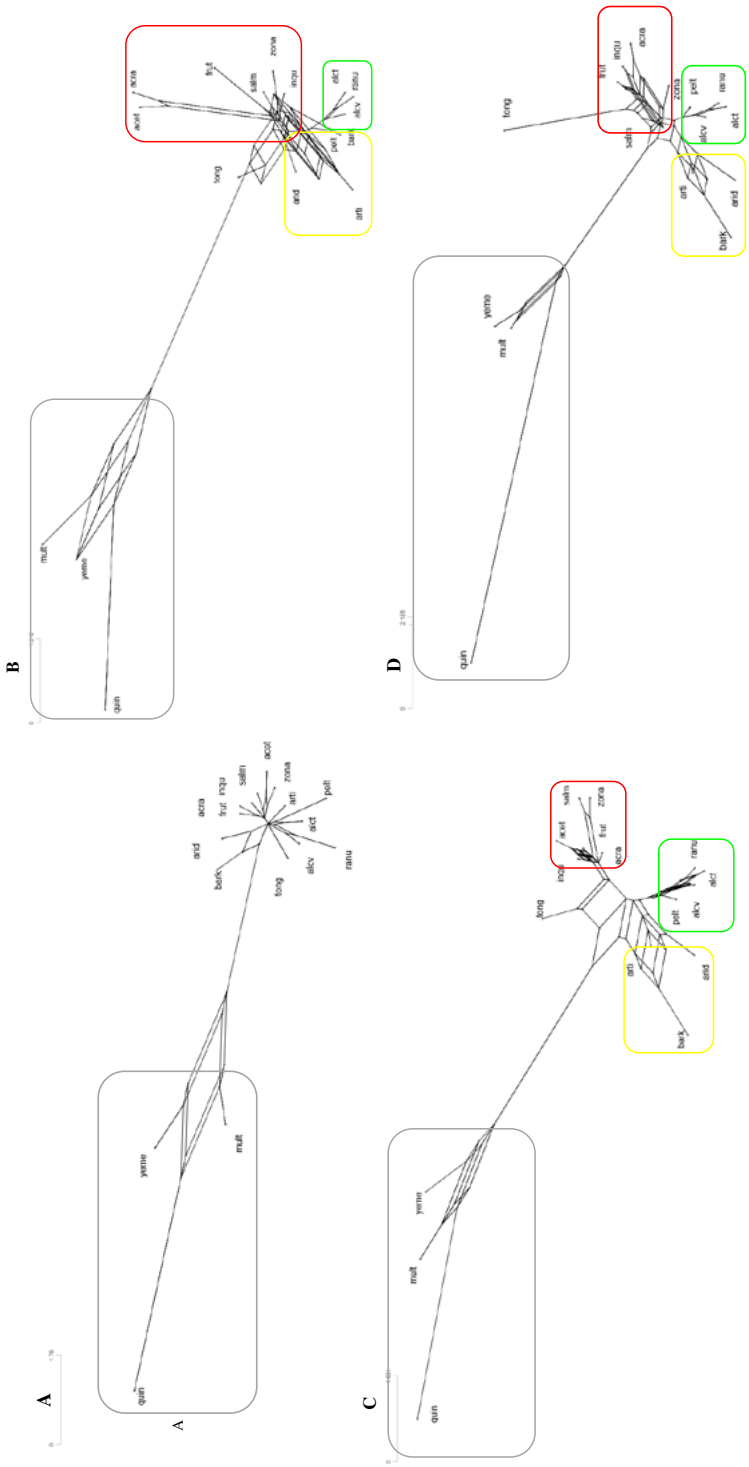


FIGURE 4 | Pattern reproducibility. Consensus networks of NJ trees that are based on top 50 clusters (excluding organelle-based clusters, including incomplete clusters) of the Repeat Explorer analysis using 10% (a), 25% (b), 50% (c) and 100% (d) of the multi-species read library (see text). We have excluded *P. karooicum* and *P. elongatum* as these species appeared on too long edges. Edge weights: mean: threshold 0.1. Colored boxes and group labelling (A—E) as in Fig. 3.

Abundancy vs similarity

Abundancies encountered in the repeatome have been reported to be phylogenetically informative (Dodsworth et al. 2015), meaning that related accessions will have similar abundancies. Our results call this into question and contrast with what was found in *Nicotiana* (Dodsworth et al. 2017) where accessions with comparable genome size differences were included. However, in the case of *Nicotiana* no major autapomorphic repeat clusters were found, or these were excluded from the analysis, and the repeat characteristics were, by and large, comparable over the range of accessions included. This, in itself, does not have to indicate that there is no phylogenetic signal in our abundancy data, but large genomes attract each other by virtue of their increased overall read contribution confusing the pattern. To mitigate these effects, we applied a square root transformation, but it did not completely remove Large Genome Attraction (LGA). In particular those repeats belonging to the Core *Ciconium* repeat SC1 may be responsible for LGA as these were found to be expanded across the Core *Ciconium* and even more so in the creeper and climber group (Fig. 2).

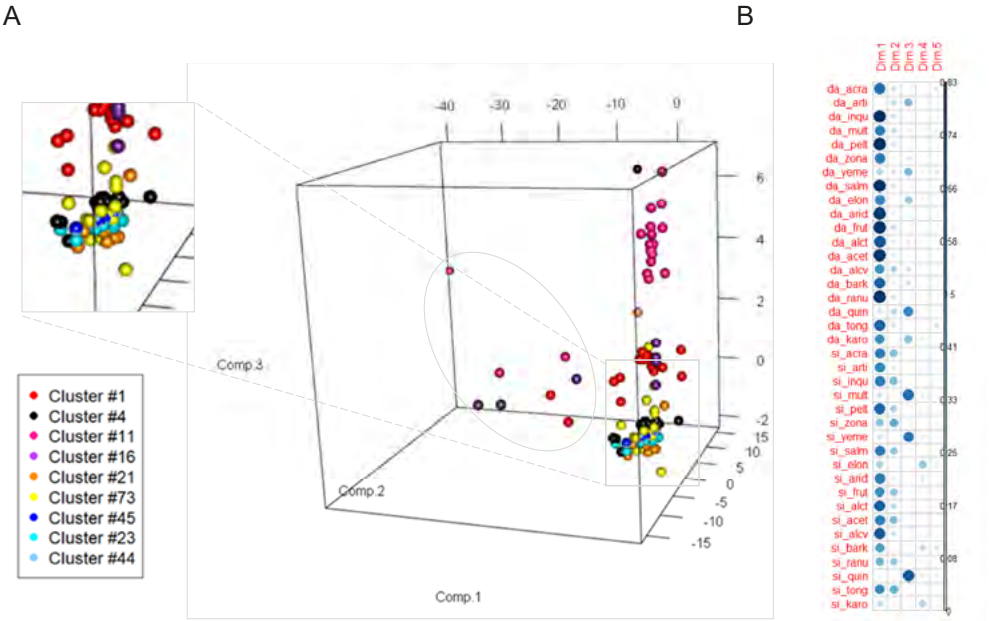


FIGURE 5 | A) Principal component analysis on contrasting repeat clusters similarity and abundance; a) six selected clusters (1, 4, 11, 16, 21, 73) along with three plastome-based clusters (23,44,45) for comparison (see the cutout). The ellipse and indicates clusters belonging to the dispersing group (A). The red dots denote a Core-*Ciconium* synapomorphic cluster, the orange dots represent a *Ciconium* synapomorphic cluster, the black dots represent a symplesiomorphic cluster, the pink dots represent a cluster synapomorphic for the dispersing group (A), the teal, light blue and blue dots represent the three chloroplast-based clusters, and the yellow dots represent an autapomorphic cluster. B) Contribution of each variable to the first 5 axis of the PCA. 'Dim' refers to axis 1-5 respectively. "da_####" denote the pairwise abundancy differences. "si_####" denote the pairwise similarities and a four-letter species acronym.

Another cause for these discrepancies if the abundancy data versus the similarity data is compared, is the inclusion of autapomorphic clusters. Which, if one or a few unrelated accessions carry some, could lead to long branch attraction purely based on these few, very abundant clusters. Therefore, we feel that the abundancy-based data, at least in our case, is more appropriately used to study repeatome evolution rather than phylogeny reconstruction. We found that the similarity-based approach provided a solid basis for studying the phylogenetic relationships between the accessions and avoids the kind of abundancy-related artifacts mentioned. We are cautious in interpreting the relationships found based on the abundancies alone for some diploid accessions (e.g., *P. quinquelobatum*) as well because the relationships of the tetraploid accessions (*P. alchemilloides*, *P. articulatum* and *P. yemenense*) are likely the result of their large read contribution. Our suspicion was further reinforced by the observation that the abundancy-based tree does not correlate well with the previously published phylogenetic trees nor with our similarity-based CN.

Ciconium phylogenetic patterns

The Greater Cape Floristic Region (GCFR) is a plant evolutionary model area, comprising plant species hyper-diversity. It contains a wide variety of micro-climates, soil types, fire regimes on a relatively small geographical region, which is thought to have spurred speciation in a number of angiosperm clades (Linder et al. 2003, Verboom et al. 2009, Warren et al. 2011) with *Pelargonium* being the 7th largest one of them. The GCFR includes an estimated 80% of known *Pelargonium* species (Van der Walt Vols. I-III 1988, Bakker et al. 2005, Verboom et al. 2009, Röschenbleck et al. 2014). As in other GCFR clades, *Pelargonium* species (Verboom et al. 2009), generally display highly local, and endemic distributions (van der Walt et al. Vols. I-III 1988, Marais 1994, 2005, 2014, Röschenbleck et al. 2014, gbif.org 2020). This may lead to very small effective population sizes and combined with the large differences in ecosystem parameters (climate, soil, pollinators), this may have led to the fixation of specific repeat clusters through genomic drift processes. As an example, especially in *P. aridum*, autapomorphic clusters occur that may be the results of fixation due to historical population bottlenecks (Fig. 3). The other example from this study came from the 'dispersing group B' containing *P. multibracteatum*, *P. quinquelobatum* and *P. yemenense* which was supported by six synapomorphies. Further, *P. yemenense* has an additional seven autapomorphic clusters. These stand out even more because they appear to be the result of a dispersal event either out of the CFR via the 'African track' (Linder 1992) and then via Socotra to the Arabian Peninsula or the other way around. They also possibly went through quite a population size bottleneck, possibly causing the fixation of the large, unique repeat clusters inferred here. We may in fact have underestimated the occurrence of these autapomorphic clusters because we used the threshold of 20% of reads in a particular cluster that must originate from a single accession for it to be considered an autapomorphic cluster. This is ~10x more than can be expected if the reads are distributed evenly per accession.

Pelargonium Sect. *Ciconium* repeatome evolution shaped by a possible WGD event?

WGD's are widespread in angiosperms and these have had a profound effect on the evolution of flowering plants (Soltis et al. 2009, 2015; Schranz et al. 2012; Wendel et al. 2015). One of the resulting effects is an increased speciation rate (Landis et al. 2018). Section *Ciconium* comprises a recently evolved group of species according to (van de Kerke et al. 2019) that emerged in the CFR ~5 mya. Whether we see an increased speciation rate in Core-*Ciconium* is tempting but still difficult to say. Perhaps multiple x=4 species went extinct already, and speciation analysis using, for instance, LTT (lineage-through-time) plots may not be applicable here.

Support for a WGD comes from the following evidence. We found > 25 unique Core *Ciconium* clusters and, the increased basic chromosome numbers of the Core *Ciconium* (from x=4 to x=9 see Gibby et al. 1996; Bakker et al. 2005) correlating with larger Core *Ciconium* Cx values (2.23 pg -4.13 pg) compared to *P. elongatum* (1.3pg, see Table 1). Finally, the occurrence of polyploids in some of our accessions (*P. articulatum* and *P. alchemilloides*, *P. yemenense sp. nov*) all may also support a WGD event. This propensity for polyploidization may further increase genome plasticity explaining the varied niches (Leitch et al. 2008) that *P.* section *Ciconium* species occupy. The occurrence of Core-*Ciconium* specific repeat clusters, especially those in SC1, could indicate incomplete genome reduction, which is common after a WGD event (Sankoff et al. 2010, Buggs et al. 2012). This is known to especially affect the non-coding, repetitive DNA (Eckardt 2001, Freeling et al. 2012). We do not see this effect in our repeat profiles when comparing them to *P. elongatum*. We would have expected more similarity between *P. elongatum* and the other species had this reduction progressed further. Previously Ren et al. (2018) deduced that there may have been a WGD (+/- 10 Mya, (HPD: 9.0-10.5 Mya) leading to the lineage *Pelargonium*. Our data hints that this may have taken place, but after the split of the Core *Ciconium* species from *P. elongatum*, placed at +/- 5 Mya (HPD 4.0-6.0 Mya) by van de Kerke et al. 2019. Perhaps more than one WGD event occurred in genus *Pelargonium* with one leading to the much younger lineage of section *Ciconium*. Finally, the occurrence of polyploids, derived from diploid inter-specific crossings (Breman et al. unpublished data and pers. obs. RS & FCB), shows that parental species generate polyploid pollen and have matching genomes.

The pattern of the *P.* sect. *Ciconium* proliferation driven by possible WGD (Leitch et al. 2008) is not unlike patterns observed in other angiosperm clades (at family-level) associated with WGD events such as Poaceae (Leitch et al. 2010), Brassicaceae (Johnston et al. 2005, Koo et al. 2011) and Asteraceae (Vitales et al. 2019b) but see (Puttick et al. 2015) for a critical review.

Analyzing patterns of genome/repeatome evolution

PCA's are commonly used to analyze large genomic datasets especially when answering questions of trends in populations or groups of closely related species such as: heritability of selected SNP markers in *Citrus* (Ollitrault et al. 2012); to detect convergent evolution of a gene in humans (Galinsky et al. 2016); to detect direction of evolution after hybridization in *Vitis* (Miller et al. 2013); or to detect genetic divergence between closely related species of *Camellia*

(Yang et al. 2016). All these analyses have in common that they aim to discover a trend in a large genomic dataset (often SNP based) finding which combination of characters provides the most explanation for patterns observed. We aimed to do the same using the abundancy differences and similarities as characters. We proposed that those clusters that did not show obvious trends (positive or negative) for abundancy and similarities could be considered conserved or 'evolutionar old', assuming some sort of streamlining through time. The symplesiomorphic clusters should especially display such a pattern. We expected that clusters that display other patterns (e.g., increased effects of abundancy or similarity) to be clusters that could be evolutionarily younger, having so far eluded constraints on their proliferation. Our results indicate that the PCA is indeed an appropriate method to explore the repeatome for clusters that have differing trends. Symplesiomorphic clusters do show highly similar trends, whereas those clusters (e.g., #11 and #12) that have expanded significantly, such as the syn-or autapomorphic ones in the dispersing group, show a trend for being especially affected by the expansion, but also by changes in similarities (therefore by substitutions) (see Fig. 5A). The Core *Ciconium* clusters also display such a pattern, but the effect of similarity and abundancy is smaller compared to those of the clusters expanding in the dispersing group. Given that the grouping of points in the PCA for these clusters do not show obviously different trends from the symplesiomorphic clusters suggests that some constraints are already acting on these both with regards to abundancy and similarity.

To be able to also detect possible positional effects (centromeric or telomeric) we would, ideally have an annotated *Pelargonium* genome to be able to map our repeat clusters to a specific region. Since this is not available, we have used the plastome, which has tight and, presumably, equal selective pressures acting on it. Plastids are highly constrained in their function (Wicke et al. 2011) and we expected clusters based on the plastome to show a roughly equal similarity and abundancy patterns. In other words, these cp-cluster behave as an 'old' repeat cluster would. Our evidence shows they do (Fig. 5A). The use of the plastome as an approximation of a single genomic region does require some caution because different selective pressures, such as increased substitution rates for specific regions (e.g., in *Oenothera* see Greiner et al. 2008, or in *Caragana* see Jiang et al. 2018) or structural re-arrangements (e.g., for the plastome in *Silene*, sese Sloan et al. 2014) have been reported. *Pelargonium* plastomes are also subject to increased re-arrangements and increased non-silent substitution rates (Weng et al. 2013, 2017, Ruhlman and Jansen 2018), thus we need to be cautious when interpreting the different plastome based clusters patterns. Nevertheless they, presumably, all belong to the same chromosome and stem from a much more functionally constrained region than the repeatome. Therefore, we choose plastome based as a baseline to compare other clusters against.

Acknowledgements

We thank Petr Novák and Jirí Macas for help with the RE server. We thank Daniel Vitales for advice on the data analysis.

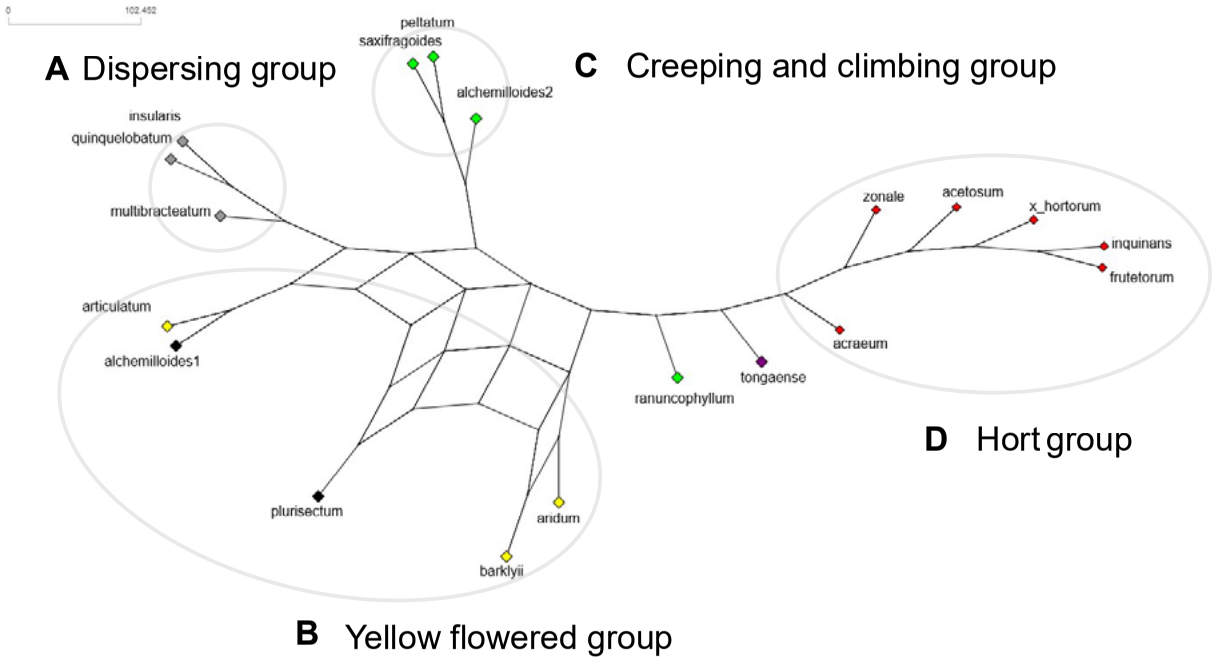
Author contributions

Conceived the study: FCB, FTB. carried out the analysis: FCB. Wrote the manuscript: FCB, FTB, MES. Informatics analyses: FCB, GC. All authors read the draft and gave feedback

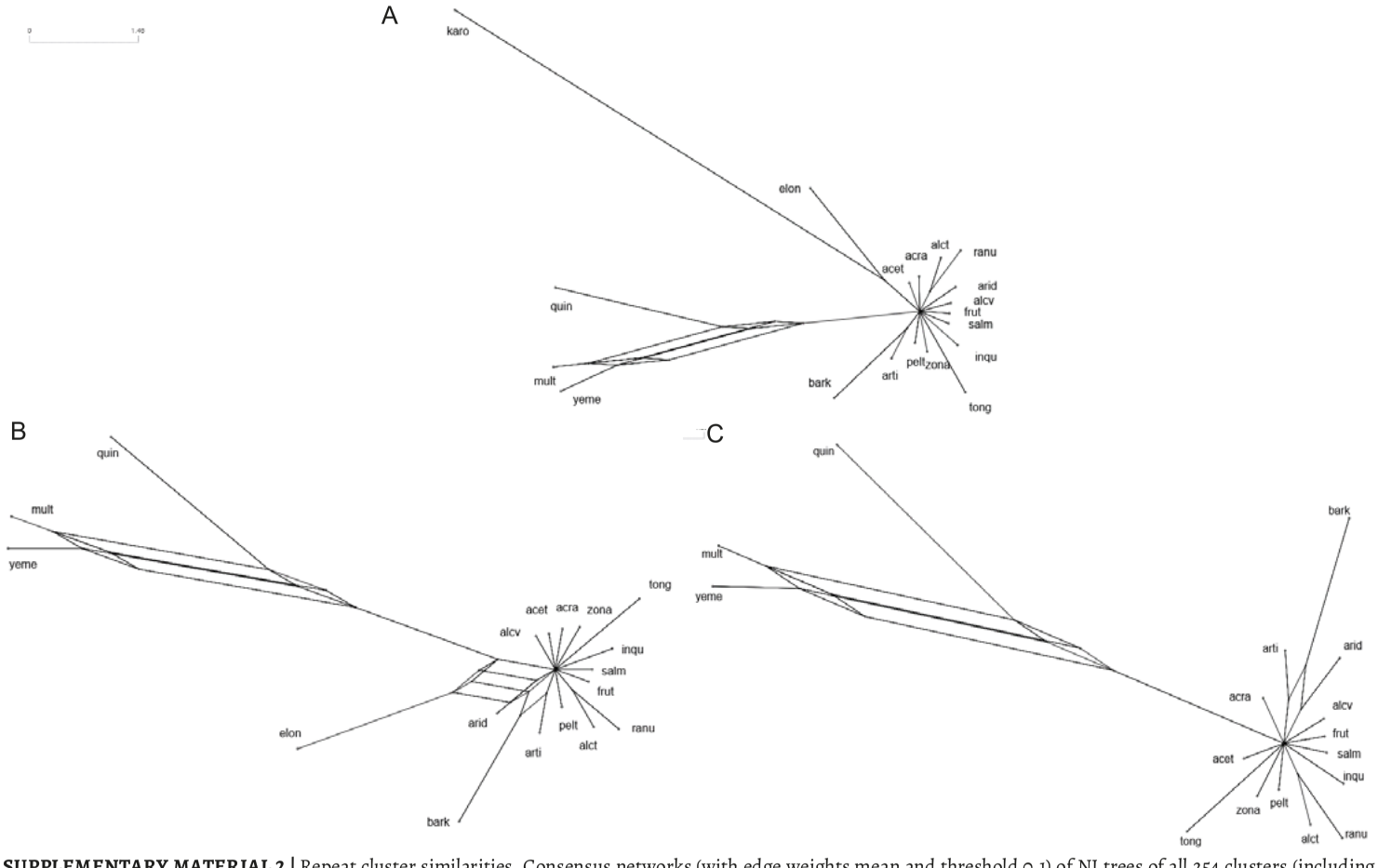
Funding

This research was funded by the Dutch Foundations for applied scientific research (TTW). Grant number: 14531 "*Pelargonium* genomics for overcoming cytonuclear incompatibility and bridging species barriers" of the Green Genetics program.

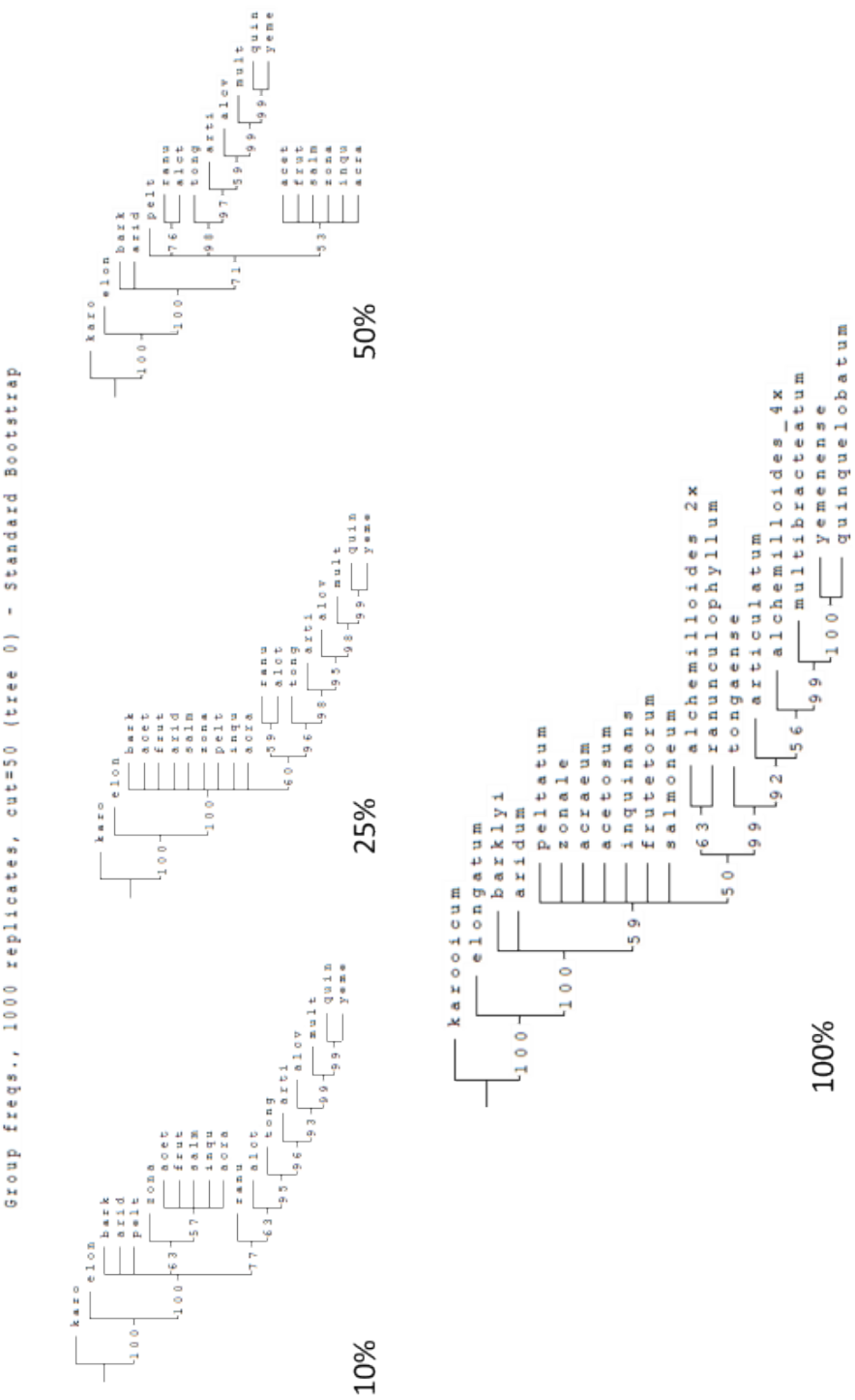
Supplementary material



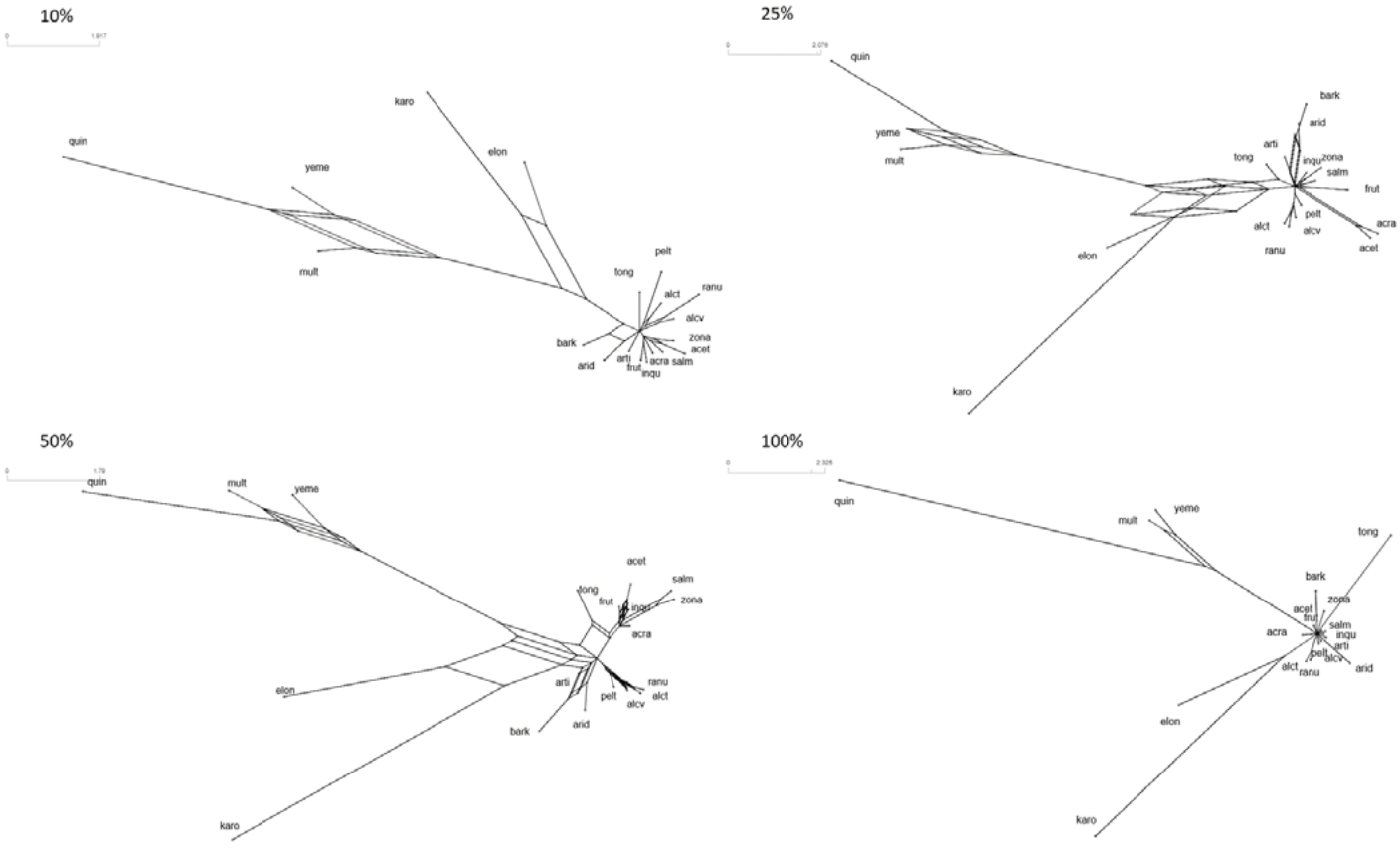
SUPPLEMENTARY MATERIAL 1 | MRP supertree analysis. Most parsimonious resolution of Matrix Representations of four plastome-based phylogenetic trees (MRP supermatrix analyzed using parsimony), summarized into a Consensus Network. Informal species groups proposed here are indicated and labeled. Where possible the labels overlap with Fig. 5 in the main paper.



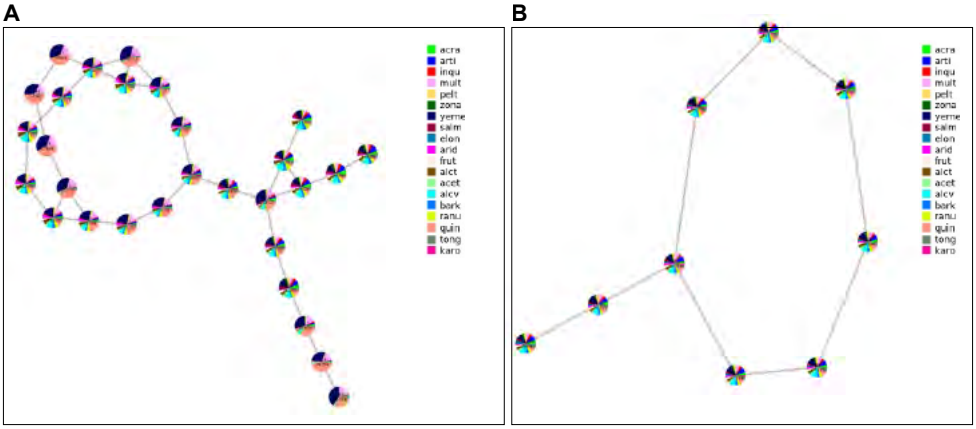
SUPPLEMENTARY MATERIAL 2 | Repeat cluster similarities. Consensus networks (with edge weights mean and threshold 0.1) of NJ trees of all 254 clusters (including incomplete clusters) containing ≥ 542 reads, but excluding organelle-based clusters of the RE analysis using 100% of the MuSp read library including: a) *P. karooicum* and *P. elongatum*; b) Excluding *P. karooicum*; c) excluding *P. karooicum* and *P. elongatum*. The length of the branches corresponds to the number of splits supporting this branch (Holland et al. 2000). This is indicated in the scale at the top left part of the figures.



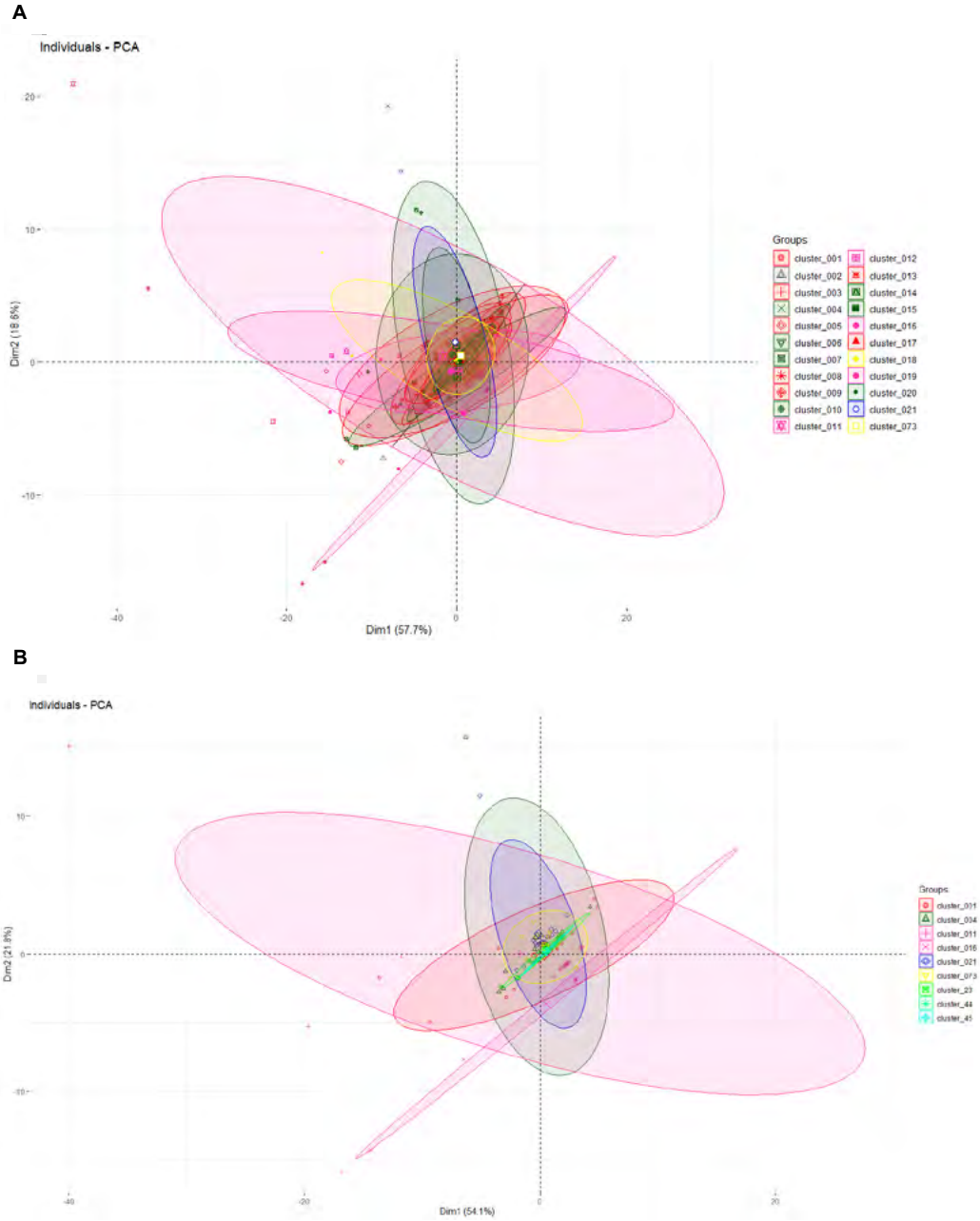
SUPPLEMENTARY MATERIAL 3 | Reproducibility analysis. Comparison of patterns from the abundancy dataset cladogram based on different MuSp sub libraries of the RE analysis using 10,25,50% and 100% of the multi-species read library. Shown are parsimony trees (TNT) based on continuously-distributed abundancy characters including *P. karooicum* and *P. elongatum*.



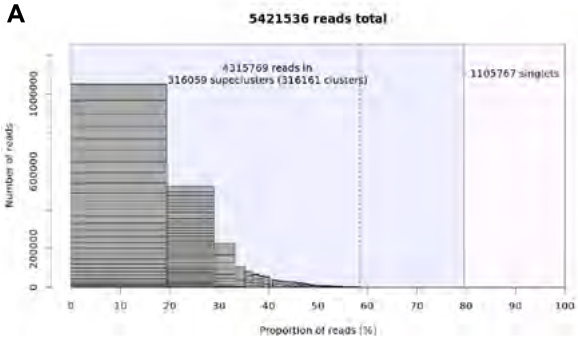
SUPPLEMENTARY MATERIAL 4 | Consensus networks of NJ trees of the read similarity based analysis on top 50 clusters. These are excluding organelle-based clusters, including incomplete clusters. RE analysis using 10,25,50% and 100% of the multi-species read library. Including *P. karooicum* and *P. elongatum*. Splits tree: Consensus Network: Edge weights mean: threshold 0.1. The length of the branches corresponds to the number of splits supporting this branch (Holland et al. 2000). This is indicated in the scale at the top left part of the figures.



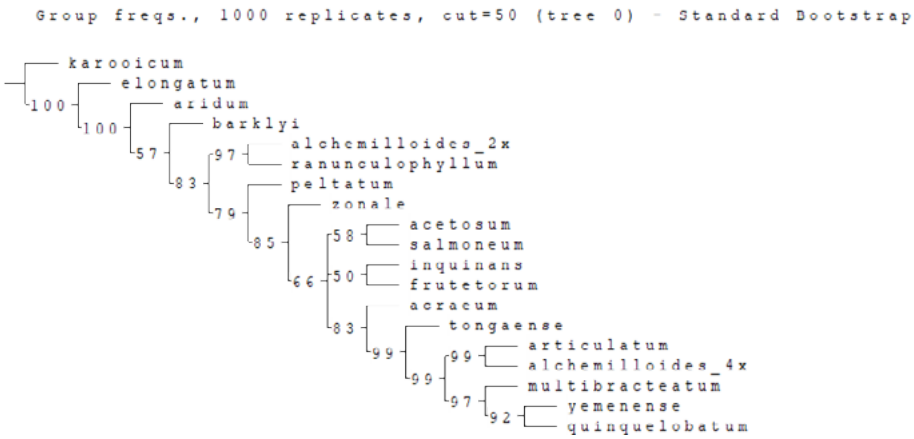
SUPPLEMENTARY MATERIAL 5 | Two superclusters. A) SC1. This supercluster consists of the Core *Ciconium* clusters; B) SC3. This supercluster consists of the symplesiomorphic clusters with contributions from all accession including the outgroups. In both figures the pie charts represent clusters connected through mates (read pairs). the colors in the pie charts indicate contributions from each accession to a particular cluster.



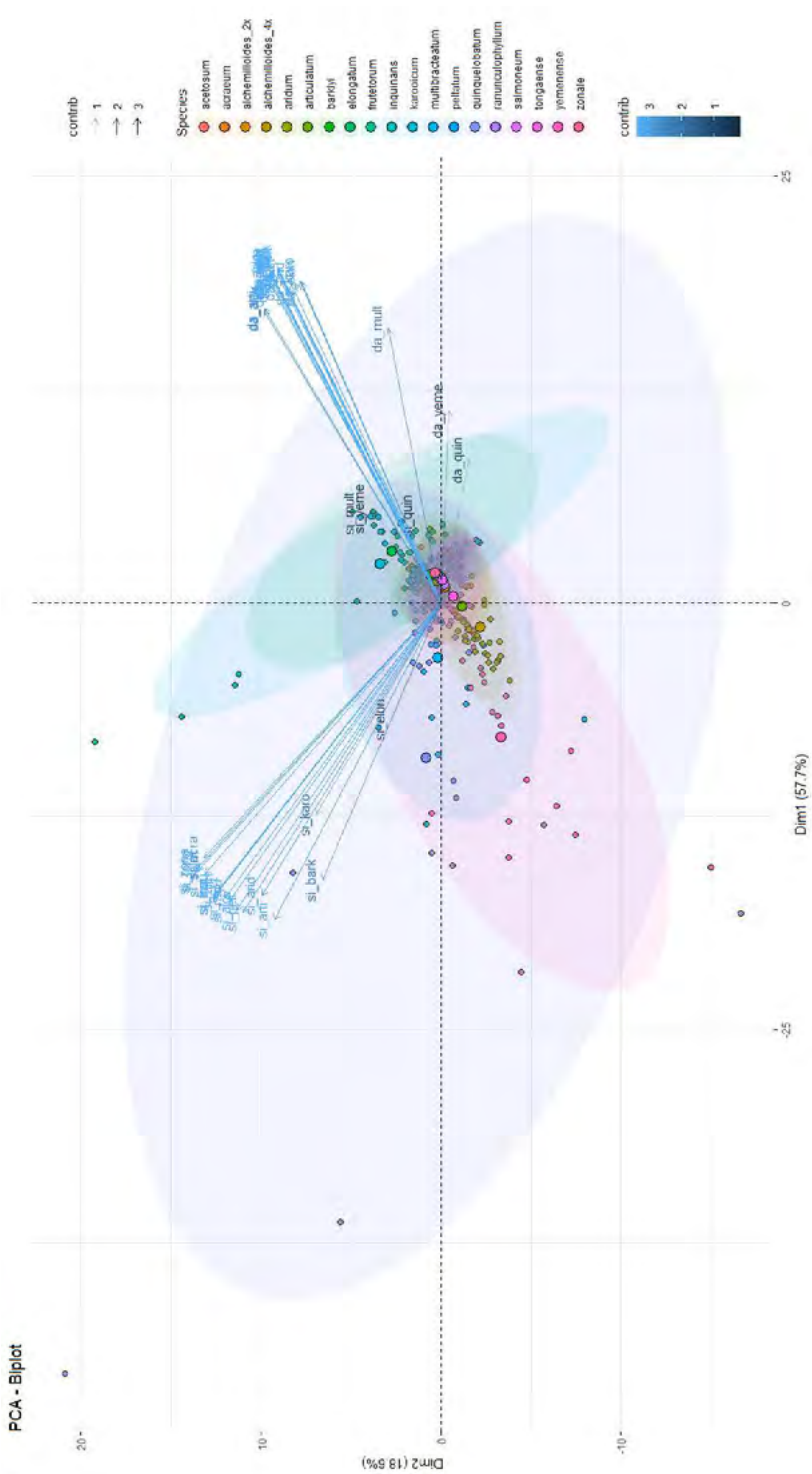
SUPPLEMENTARY MATERIAL 6 | Principal component analysis on 22 and 9 clusters. A) Principal component analysis on the abundancy differences and similarities for 22 repeat clusters. Clusters are colored by class as defined in Fig. 1. The legend has been partially reproduced to facilitate interpretation. B) as A, but now for previously selected Clusters: 1, 4, 11, 16, 21, 73 with selected cp clusters, 23,44,45.



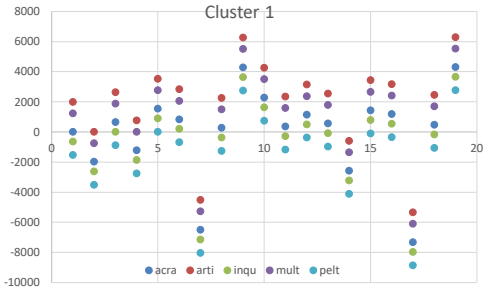
SUPPLEMENTARY MATERIAL 7 | Read use and heat map of cluster make up per accession. A) Distribution or reads used for clustering and singletons (not used for clustering). B) Heatmap displaying repeatome makeup normalized for clustered reads per accession.



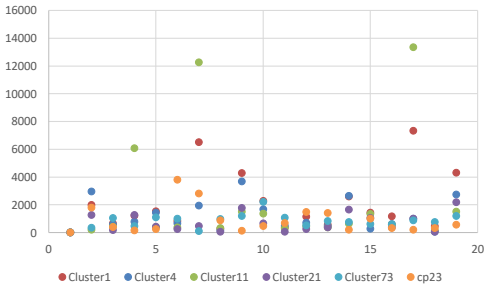
SUPPLEMENTARY MATERIAL 8 | Parsimony analysis. BS consensus tree for 255 clusters after cube root transformation; bootstrap consensus tree based on the abundancy dataset after cube root transformation for 255 clusters. Bootstrap values ≥ 50% are displayed, nodes with lower support values are collapsed.

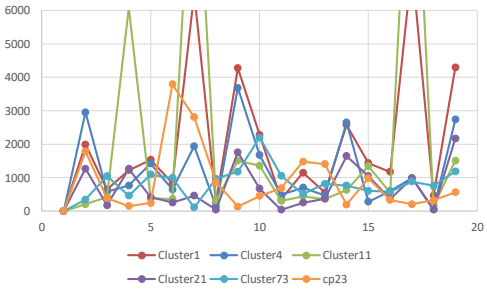


SUPPLEMENTARY MATERIAL 9 | Biplot with trendlines added for similarity and abundancy data. With trendlines added for both abundancies and similiarites per species.



SUPPLEMENTARY MATERIAL 10 | Correlation of similarities and abundancies for six selected clusters and five accessions. X-axis displays clusters ranked by read similarity. Y-Axis displays read count abundancies.





Supporting materials can be viewed at https://figshare.com/s/fbdc94037892998f6b49

Chapter 4

Rapid plastome evolution in *Pelargonium* section *Ciconium* linked with altered plastidic ribosomes

Breman FC[1], Schranz ME[1], Greiner S[3], Snijder RC[2], Bakker FT[1]

[1]Biosystematics Group | Wageningen University and Research, Wageningen, the Netherlands
[2]Syngenta Seeds BV, Andijk, the Netherlands
[3]Max Planck Institute of Molecular Plant Physiology, Wissenschaftspark, Golm Am Mühlenberg 1, 14476 Potsdam – Golm, Germany

FCB: ORCID https://orcid.org/0000-0002-9722-7564
FTB: ORCID https://orcid.org/0000-0003-0227-6687
MES: ORCID https://orcid.org/0000-0001-6777-6565

abstract

Geraniaceae plastomes are very large, highly rearranged, have accelerated nucleotide substitution rates and several genes show signs of positive selection. We have focused on the well-defined and horticulturally important *Pelargonium* section *Ciconium* (Geraniaceae) to study plastome evolution between closely related species. We tested plastome encoded genes for positive selection and co-evolution. We examined the effects of gene length variation on rRNA and ribosomal peptide structures. Strikingly, *rpoB*, *rpoC1*, *rpoC2*, *uS19c* and *ycf1* (a large ORF of uncertain function) are under positive selection. We found that the *rrn23* gene, encoding the structural 23S rRNA of the plastid, displays greater sequence length variation between *Ciconium* species than it does between angiosperm orders. We also found that changes in *rrn23* correlate with those in the *rpo* subunits encoding the plastid encoded RNA polymerase *(PEP)*. We further identified correlations between length variation of ribosomal proteins with *rrn23*. We find that the ribo-proteins uL32c and uS19c have undergone structural changes to maintain overall ribosomal structural integrity and function.

Keywords: *rrn23*, positive selection, plastid genes, coevolution, *Pelargonium*.

Introduction

The organellar genomes of Geraniaceae are remarkable as they include both some of the largest and smallest known angiosperm plastomes (Ruhlman and Jansen 2018). They show remarkable structural changes: for example, the genus *Monsonia* has a contracted inverted repeat (IR) (Blazier et al. 2011; Guisinger et al. 2011); *Erodium* completely lacks the IR (Downie and Palmer 1992, Blazier et al. 2016a); and *Pelargonium* has a significantly expanded IR (Chumley et al. 2006, Weng et al. 2017). The variability of Geraniaceae plastomes includes other elements such as the abundance of smaller repeats and accelerated nucleotide substitution rates (reviewed in Guisinger et al. 2008, and Ruhlman and Jansen 2018). These changes are thought to be the result of altered repair, replication and recombination (RRR) genes (Zhang et al. 2016). Although it has been reported that nucleotide substitution rates of genes are lower when located inside the IR (Curtis and Clegg 1984, Wolfe et al. 1987, Wolfe 1991, Gaut 1998, Shaw et al. 2007), this does not seem to be the case in *Pelargonium* expanded IR (Weng et al. 2017).

Several plastid-encoded genes have been shown to be under positive selection across the Geraniaceae, including ribosomal encoding protein genes, NAD(P)H-quinone oxidoreductase (*ndh*) and *rpo* subunits that make up the plastid encoded RNA polymerase (PEP). (Guisinger et al. 2008, Weng et al. 2016). Genes not only show positive selection, but also length differences between exons (both protein-coding and non-protein coding such as the *rrn* genes) introns and intergenic spacers (IGS) (Chumley et al. 2006, Guisinger 2011, Weng et al. 2016, Weng et al. 2017). All these changes to Geraniaceae plastomes did not evolve in isolation, but rather co-evolved; especially the co-evolution between both nuclear encoded and plastid encoded ribosomal proteins (collectively known as the ribo-proteome) was found to be significant (Weng et al. 2016).

To date, studies have examined patterns of plastome evolution across the Geraniaceae, but the extent of changes at the genus or sub-genus level has not been investigated. We studied plastome evolution in *Pelargonium* section *Ciconium* (Sweet) Harvey (1860: 298), using a near complete taxon coverage. The section as currently recognized contains 18 species and occurs both in and outside the Cape Floristic Region (CFR). We have added two previously undescribed species to our study (*P. yemenense* sp. nov. Gibby et al. in prep and *P. omanense* sp. nov. Knees et al. in prep). Species from *Ciconium* have been used as a model to study Cyto-Nuclear Incompatibility (CNI) and heteroplasmy for more than a century (Baur 1909, Tilney-Basset et al. 1992, Zhang et al. 2015, Breman et al. 2020). Having a comprehensive overview of plastome evolution in a series of closely related species can elucidate the role of plastid changes on the evolution and control of CNI (see Chapter five).

Comparing co-evolutionary patterns across species within a clade could shed light on evolution of genes involved, be it from the perspective of protein structure or selection on individual amino acid residues. Further, possible correlation of changes between genes, whether they have length differences, are under positive selection or display both, has

not yet been established in *Pelargonium* plastomes. Especially the *rpo* subunits, which show the highest d_N/d_S (ω) values (Guisinger et al. 2008), may be important as PEP plays a central role in plastome gene transcription (Börner et al. 2015). Another important aspect are the changes noted by (Weng et al. 2016, Chapter 5) in the *rrn23* gene as this encodes the 23S rRNA backbone of the ribosome and changes in this gene are probably relevant in explaining the observed positive selection of ribosomal protein coding genes.

In this study, we focused on the *rpo* genes and the 23S rRNA-encoding *rrn23* gene, as co-evolutionary analysis of these genes with other has so far been neglected in the studies listed above. We have examined the changes in, and correlations thereof, of plastid-encoded genes across *Ciconium* by performing branch length tests of gene trees (based on both substitutions and indel patterns). We also performed comprehensive tests of positive selection on sites/residues in several genes that displayed length variation, high substitution rates or a combination of both. We did so to compare evolutionary patterns previously identified at larger phylogenetic distances (across Geraniaceae), with patterns between closely related species (in section *Ciconium)*. We hypothesize that not all genes under positive selection in Geraniaceae are under positive selection in *Ciconium*, but that those that are, co-evolve.

To address the changes in *rrn23* we have used the *Spinacia oleracea* model of the chloroplast ribosome generated by Bieri et al. (2017) and Perez-Boerema et al. (2018) for comparison. We explored to what extent the secondary structure of the *Ciconium* 23S rRNA encoded by *rrn23* is affected by the length variation observed. We then hypothesize that if the 23S rRNA molecule has undergone structural changes then at least some of the ribosomal protein structures (such as from *uS19c*) are affected as well. We modelled the structures of two ribo-proteins accordingly to test this.

Material and methods

Plant material and growth conditions

Plants were grown from seeds in a climate-controlled greenhouse. Leaf material for DNA extraction was collected after 10 months. *Pelargonium* strains used were published previously (Breman et al. 2020) and summarized in the supplementary materials (supplementary material 1).

Taxonomy and nomenclature

For currently recognized species in section *Ciconium* and their correct names, we followed the taxonomy of Röschenbleck et al. 2014. For nomenclature of the ribo-proteome, we adopted a recently published nomenclature which better reflects function and genomic location of the subunits (Ban et al. 2014).

Sequencing and contig assembly

We assembled partial plastomes for 22 accessions using Hi-seq Illumina based sequence reads. One specimen was a herbarium collection (*P. somalense*) and two were from silica-dried, but freshly collected material (*P. omanense* and *P. insularis*). All other accessions came from freshly grown plant material. For the Illumina HiSeq sequencing, we followed the same protocol as outlined in Chapter three. Plastome assemblies were performed using the GetOrganelle pipeline (Jin et al. 2019) using default setting, except for the assumed insert size which was set to 350bp. Contigs were visualized and assessed for gene content and read coverage using BandAge (see supplementary material 3) (Wick et al. 2015). Final contigs were pruned of sequences not shared by all accessions, to retain only homologous pieces, and subsequently concatenated using MEGA7 (Kumar et al. 2016). One IR was removed from the overall collection of contigs for downstream analyses. Alignment of the concatenated sequences was performed using MAFFT (Katoh et al. 2019) and evaluated using MAUVE 2.3.1 (Darling et al., 2004) to scan for large rearrangements. Annotation was performed using a two-step process. First, GetOrganelle automatically performed annotations based on its internal database and we used this as a proxy. Second, to get the final annotation and to extract all exons, we submitted the concatenated file from GetOrganelle to GeSeq (https://chlorobox.mpimp-golm.mpg.de/geseq.html) (Tillich et al. 2017) as implemented on the CHLOROBOX server using the *Pelargonium* x *hortorum* annotated plastome (GenBank Accession Number DQ897681.1, Chumley et al. 2006) as reference. Occasionally GeSeq failed to extract the correct genes when large insertions were present or stop codons, indicating possible rearrangements which may have resulted in stop codons occurring outside the reading window. In those cases, we manually added sequences per gene for each accession.

Scan for variable hotspots in the partial plastome alignment

To identify regions with increased levels of substitutions in our assemblies, we determined Pi (π)-values (a measure of changes per site), as implemented in DnaSP v6 (Rozas et al. 2017), calculated pairwise as the number of changes in a 'step' (number of base pairs) across a specified 'sliding window' (see below). Calculating π-values was previously demonstrated to be a useful approach for identifying regions with increased substitution rates across an alignment (Greiner et al. 2008a & b). We compared all accessions pairwise and used a sliding window of length 1000 bp and a step size of 10 for sequence comparison. Indel regions were included in the analysis as these occur across the *Pelargonium* plastomes and they are therefore important in estimating the overall pattern of change. We have calculated the average π-value for the entire alignment and identified regions that showed values between 5x (π = 0.02151) and >10x (π= 0.0432) for which the averages are displayed in Fig. 1. [A breakdown of the π-plot into functional regions (LSC, IR and SSC) can be found in the supplementary material 3.]

Gene tree reconstructions

Gene sequences were aligned using MAFFT (Katoh et al. 2013, Katoh et al. 2019) and the resulting alignments were manually refined if necessary. Phylogenetic analyses of our datasets were performed under ML using IQ-TREE with standard settings, at the IQ-TREE web server (iqtree.cibiv.univie.ac.at) (Hoang et al. 2018, Kalyaanamoorthy et al. 2017, Nguyen et al. 2015, Trifinopoulos et al. 2016) and used model selection (ModelFinder Kalyaanamoorthy et al. 2017) as well as ultrafast bootstrapping (UFBoot) to generate 1000 trees. Indels were ignored by IQ-TREE but as these contain potentially phylogenetic information, trees were additionally generated based on their indel sites. We are aware that in phylogeny reconstruction gaps are usually treated as a single character (Simmons 2000) irrespective of length, however the aim here was not to infer phylogeny, but to comparatively quantify length variation and test correlations between genes or exons. Therefore, indels were scored in a binary fashion per sequence. We took the length of the indel into account whereby each nucleotide position in the indel was assigned a '1' (insertion) or a '0' (deletion). We built our indel matrices using only the variable parts of the alignment. The binary matrices were subsequently analyzed using IQ-TREE with the JC2 model selected for binary datasets. In all cases, except for *rpoB*, the model selected for analysis of the majority of the binary matrices was Jukes-Cantor (JC2) + equal codon frequencies (FQ) and ascertainment bias correction (ASC). For *rpoB* rates were allowed to vary (R2) in addition to the same model parameters mentioned above. (See http://www.iqtree.org/doc/Substitution-Models for details on each model parameter). Sequence-based trees were visualized using Figtree 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) whereas indel-based trees were exported to Newick format and subsequently visualized in Mesquite v3.61 (Madisson et al. 2019).

Tests for positive selection

Site tests

Reading frames in the gene sequence alignments were optimized using Mesquite v3.61, using the 'minimize stop codons' function. For each gene, d_N/d_S ratios (ω) were calculated using the gene trees and alignments described above, but especially for genes with increased π-values (*rpoB*, *rpoC1*, *rpoC2*, *ycf1*, and *uS19c*). In addition we calculated (ω) in genes for which positive selection was previously reported in Geraniaceae (Guisinger et al. 2008, Weng et al. 2016), as well as for genes with low π-values (*ccSA*, *atpF*, *matK*, *uL2c*, *uL22c*). We performed the calculation of (ω) for all selected genes using CodeML as implemented in PAML (Yang et al. 2007, Xu et al. 2013). For the 'site test' the following settings were employed: codon frequency model was F3x4; kappa (κ) and omega (ω) values estimated (non-fixed); alpha (α) fixed at 0; NSites = 0, 1, 2, 7, 8; we assumed no clock; Mgene was set to 'rates', freq estimation was set to 'observed freq' as found in the data; branch lengths were not fixed, the genetic code set to 'universal'; and 'clean data' was not used (i.e., we applied pairwise deletion of codons in case of length variation); the standard error of site changes was also estimated. Significance of models selected were tested by using the 'Hierarchical Likelihood Ratio Test' (hLRT) which

uses a Chi^2 test of significance with df being the difference in free parameters. The full panel of genes tested as well as comparing these using the Chi^2 test are listed in Table 2 and supplementary material 6. Bayesian Empirical Bayes and Naive Empirical Bayes criteria are calculated as default.

Branch tests

Using the method described by Yang (54), a null model (H0; branch model ω 0), where one d_N/d_S ratio was assumed across *Ciconium* branches, was compared with an alternative model (H_1; branch model ω 2), where the different clades were allowed to have a different d_N/d_S.

Branch-site tests

For *rpoB* and *rpoC1* we performed a branch-site test to ascertain which amino acid sites are under positive selection for each branch/terminal across our trees. All settings were as mentioned above, but NSites =2. Gene trees were used to formulate different scenarios of having two different categories of ω values for different parts of the tree (see Results). For *rpoB* four scenarios were evaluated (see Fig.2a). Scenario 1) ω different for the Core *Ciconium* versus the outgroup; 2) ω was different for terminal branches only; 3) five known major *Ciconium* clades (van de Kerke et al. 2019) had a different ω each; and 4) the three longest single branches had their own ω (*P. peltatum*, *P. acetosum*, *P. barklyi*), and all other branches were considered to share one ω. For *rpoC1* scenarios 3 and 4 were changed. Scenario three is now: 3) the clades (acet,frut(HOINSA)), and (bark,arti) had their own ω. Scenario 4) was adjusted. with *P. peltatum*, *P. acetosum* and *P. aridum* (instead of bark) each having their own ω. We then performed the hLRT to evaluate and compare the LnL values for each scenario per exon.

Branch length correlation tests

We tested if branch lengths across gene trees and across accessions were correlated i.e., if an accession had a longer/shorter branch lengths in one tree did it have the same pattern in another gene tree? Patterns of accessions' branch lengths for the substitution and indel based gene trees were compared in a pairwise fashion per gene tree. Wilcoxon singed rank tests (WSR) were used to test the correlation of branch lengths per gene tree and p-values were adjusted by using the Bonferroni correction for multiple tests. The WSR-test was selected here because of the limited number of accessions (21), since it does not assume an *a-priori* distribution type and because we are interested in the ranking of each accession in a gene tree. Branch lengths were considered to correlate strongly if the test statistic 'R' for comparing two gene trees with each other resulted in a *p*-value of <0.01 (see Table 1). The full calculations of branch lengths, ratios and the pairwise WSR-tests can be found in supplementary material 4.

Structural modelling of *rrn23* in *Ciconium*

To evaluate the effects of the length differences in the *rrn23* gene on ribosomal structure, sequences were aligned together with selected representatives from ten major angiosperm clades including *Spinacia oleracea* (Bieri et al. 2017, Perez-Boerema et al. 2018). Currently

spinach has the best-corroborated model for 23 S rRNA structure and the angiosperm plastid ribosome in general. We included *Amborella trichopoda* (GenBank RefSesNC_005086.1) as an outgroup (Fig. 3A-C) and used Mesquite v3.61 to visualize the alignment. We adopted the division of the *S. oleracea* 23S rRNA structure into six domains (I-VI) used by Bieri et al. (2017). We then modelled the secondary structure of length variable regions using the software 'RNA structure' (version 6.0.1.) which is based on free energy minimization for inferring RNA secondary structure. We modelled *S. oleracea* together with the other nine angiosperm 23S structures as well as for *Ciconium* separately, using the 'RNA turbofold' option with default settings. This approach uses the comparison of multiple sequences to assess the probabilities of the structures modelled. It generally yields higher probabilities for conserved structures and thus increases the reliability of the resulting structure. We visualized our results in Structure Editor v1.0 (Reuter et al. 2010; see Figs 4-6) and labeled stem-loops and helices according to homology with the *S. oleracea* model. *Pelargonium*-unique structures were labelled with a unique alpha numerical code (e.g., 'Pel-1-λ1' refers to *Pelargonium* unique structure 1 in domain 1, 'λ' refers to an extra loop within this structure). Finally, to assess the effect of *rrn23* variation on the overall structure we also modelled a s full 23S structure for *P. barklyi* and highlighted the (structural) changes compared to those in *S. oleracea* in Fig.7.

Ribosomal protein modelling

To view effects of the *rrn23* changes on the 23S structure we modelled the secondary and tertiary structure of two *Ciconium* ribosomal proteins, one of which is affected by substitutions and indels (*uS19c*) as well as one that mainly contains indels across multiple accessions (*bL32c*). Both were taken from *P. barklyi* as this species was also used for the modelling of the full 23S model indicated above. We then compared these to the published *S. oleracea* models from Sharma et al. 2014 in the worldwide Protein Data Bank (https://www.wwpdb.org, Berman et al. 2003, 2007, wwPDB consortium 2019). We used the Phyre2 server which uses advanced remote homology detection methods (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index; Kelley et al. 2015). We used the 'intense' option as this will attempt homology-based reconstruction first, but if that is not available it will turn to an *ab-initio* calculation for the aa sequence supplied, using a hidden Markov-model (HMM) (Kelley et al. 2015). The HMM employs known sequences and attempts to model structures based on comparable sequences from other taxa, with known protein structures in the database.

Results

Plastome assemblies

We obtained partial plastome assemblies for all 22 *Pelargonium* section *Ciconium* accessions that we sequenced. The total MAFTT alignment of all 22 plastomes was 107,992 bp positions long. The shortest assembly was for *P. multibracteatum* (104,587bp), the longest one was

for *P. barklyi* (105,300bp). Due to the highly rearranged nature of *Pelargonium* plastomes, probably caused by the presence of numerous (small) repeats, the use of short sequencing reads will have difficulties to assemble full plastomes. Our assemblies lacked in total ~14 kb (see supplementary material 3) when compared with the annotated *P.* x *hortorum* plastome (Genbank accession DQ897681.1, Chumley et al. 2006) and this region was confirmed to be problematic by Ruhlman et al. (2017). In *Monsonia emarginata* a comparable ~2000bp non-assembled gap region was found to contain repeats. When compared with the annotated *P.* x *hortorum* plastome (DQ897681.1) the non-assembled regions should include the following genes: *cemA*, *petG*, *petL*, *psaI*, *psaJ*, *psbE*, *psbF*, *psbJ*, *psbL*, *rpl33*, *rpoA*, *uS12c*, *uS14c*, *ycf2*, and *ycf4*. The absence of specific genes differed per accession, for example, *uS14c* was sometimes incorporated into the larger contigs but not always. We further analyzed only genes that were assembled in all accessions.

Scans for variable hotspots

The average π-value, used to estimate sequence divergence across the entire alignment, was 0.0043 changes per reading window. Regions that showed values for between 5x and >10x the average value of the entire alignment are displayed in Fig. 1. Our analysis showed numerous 'hotspots of variation' with at least five times the average value for π. Thirteen hotspot regions were located in intergenic spacers, ten of which were within <50bp of exons and two within introns. Hotspots were found in nine exons with increased π-values (see Table 4), most notably in the following genes: *rpoB*, *rpoC1*, *rpoC2*, *rrn23*, *uS19c* and *ycf1*.

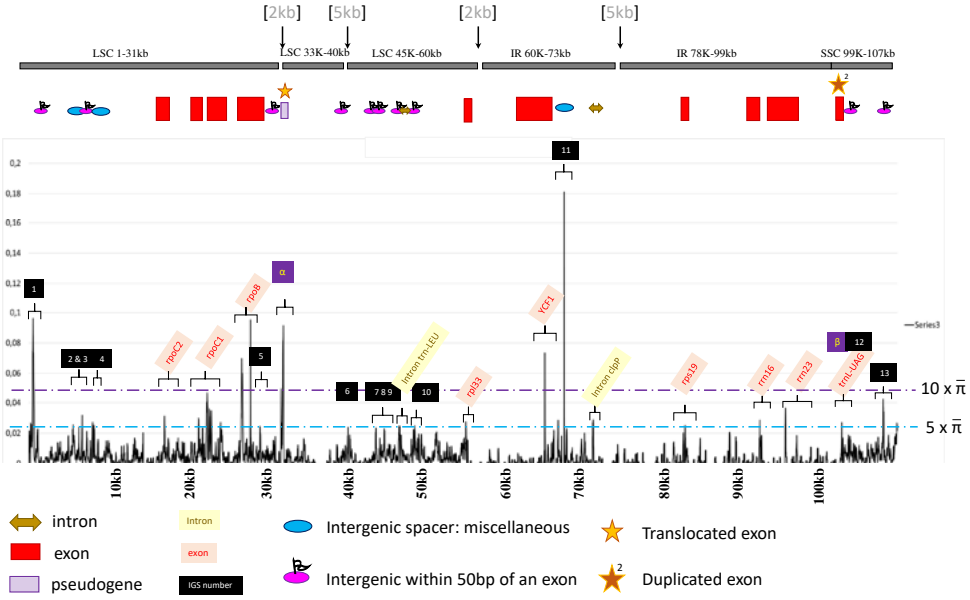


FIGURE 1 | Annotated π-plot of the partial plastome in an all to all comparison. Legend indicates the identity of the region corresponding to the peak (also listed in Table 4). Contig lengths are indicated with plastome functional regions (LSC, IR and SSC). Approximate gap lengths are indicated in grey between '[] brackets' at the top of the figure. 'π̄' indicates 5 and 10 times average pi values.

TABLE 1A | Correlation of substitution and indel patterns across gene trees. Results of the branch length correlation test on a) the substitution-based trees and b) the indel-based trees. Table 1a Wilcoxon signed rank tests of overall branch lengths for substitution patterns.

	rpoB<->rpoC1	rpoB<->rpoC2	rpoB<->rrn23	rpoB<->YCF1	rpoC1<->rpoC2	rpoC1<->rrn23	rpoC1<->YCF1	rpoC2<->rrn23	rpoC2<->YCF1	rpoB<->uS#c	rpoC1<->uS#c	rpoC2<->uS#c	rrn23<->uS#c	rrn23<->YCF1	rpoB<->uL#c	rpoC1<->uL#c	rpoC2<->uL#c	rrn23<->uL#c	uS#c<->uL#c
Σ + ranks	187	81	0	4	137	52	118	42	85	160	137	123	168	186	129	100	65	148	66
Σ - ranks	-42	-148	-229	-225	-89	-174	-108	-184	-141	-67	-90	-104	-55	-39	-99	-126	-161	-77	-159
test statistic	42	81	0	4	89	52	108	42	85	67	90	104	55	39	99	126	161	77	159
correlation?	yes	no	yes	yes	no	weak	no	yes	no	no	no	no	weak	yes	no	no	no	no	no
p	<0,01	NS	<0,01	<0,01	NS	0,02-0,05	NS	<0,01	NS	NS	NS	NS	0,02-0,05	0,01	NS	NS	NS	NS	NS
p^c	no	no	yes	yes	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no

TABLE 1B | Wilcoxon signed rank tests of overall branch lengths for indel patterns.

	rpoB<->rpoC1	rpoB<->rpoC2	rpoB<->rrn23	rpoB<->YCF1	rpoC1<->rpoC2	rpoC1<->rrn23	rpoC1<->YCF1	rpoC2<->rrn23	rpoC2<->YCF1	rpoB<->uS#c	rpoC1<->uS#c	rpoC2<->uS#c	rrn23<->uS#c	rrn23<->YCF1	rpoB<->uL#c	rpoC1<->uL#c	rpoC2<->uL#c	rrn23<->uL#c	uS#c<->uL#c
Σ + ranks	86	151	6	17	146	10	137	3	134	12	69	32	205	213	100	157	143	213	217
Σ - ranks	-141	-74	-220	-209	-53	-213	-75	-214	-72	-216	-159	-193	-20	-5	-125	-56	-58	-8	-13
test statistic (V)	86	74	6	17	53	10	75	3	72	12	69	32	20	5	100	56	58	8	13
correlation?	no	no	yes	yes	yes	yes	no	yes	no	yes	no	yes	yes	yes	no	weak	weak	yes	yes
p	NS	NS	<0,01	<0,01	0,02-0,05	<0,01	NS	<0,01	NS	<0,01	NS	<0,01	<0,01	0,01	NS	0,02-0,05	0,02-0,05	<0,01	<0,01
p^c	no	no	yes	yes	no	yes	no	yes	no	yes	no	yes	yes	yes	no	no	no	yes	yes

Gene tree reconstructions

In order to compare changes per gene, we reconstructed trees for each gene showing elevated π-values. The resulting gene trees are presented in Fig. 8, scaled for comparison, and indicating that *uS19c*, *rpoB* and *rpoC1* have the longest total branch lengths. The general topology of the gene trees is similar, but *rpoC1* differed in the positions of *P. aridum* and *P. multibreacteatum/yemenense*. For the other genes, some of the accessions shared a genotype. This is indicated in the Fig. 2.

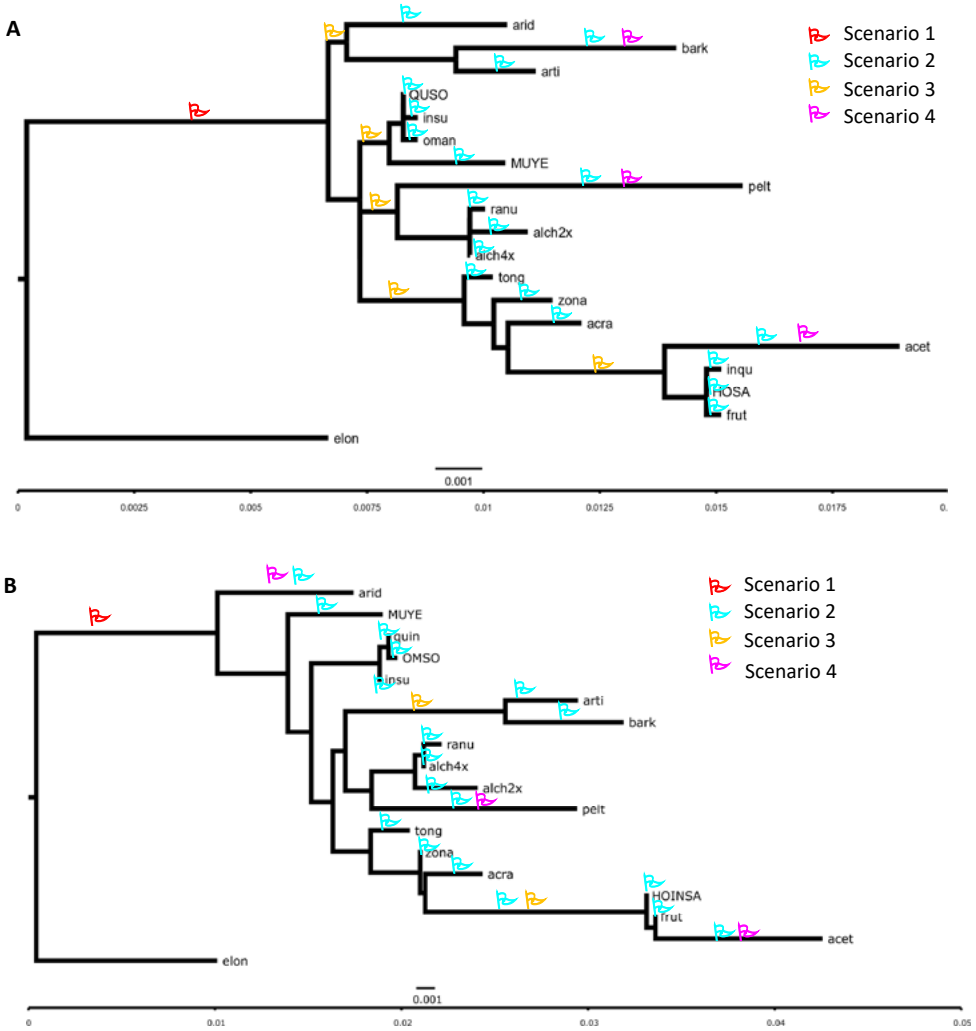


FIGURE 2 | Branch-site scenarios of positive selection plotted over *rpoB* (A) and *rpoC1* (B) trees.. Red flags indicate a separate ω for Core *Ciconium* vs. the outgroup. Teal flags indicated branches tested had separate and shared ω. Orange flags assume separate ω per clade based on van de Kerke et al. 2019. Purple flags denote separate ω for long branches only.

Test for positive selection

Site tests

All ten genes with low and high π-values (*rpoB*, *rpoC1*, *rpoC2*, *ycf1*, *uS19c*, *ccSA*, *atpF*, *matK*, *uL2c* and *uL22c*) yielded alignments with intact open reading frames. Of those, the genes *rpoB*, *rpoC1*, *rpoC2*, *ycf1*, and *uS19c* fitted models assuming positive selection significantly better than models assuming no or nearly neutral evolution (see Table 2 and the next paragraph for their respective ω-values) across branches. For the remaining genes (*ccSA*, *atpF*, *matK*, *uL2c*, *uL22c*) positive selection cannot be assumed using the site test and thus, these were not considered further. Only sites that were under positive selection under the Bayesian Empirical Bayes (BEB) criterium, which corrects for sampling errors, were selected. The Naïve Empirical Bayes (NEB) criterium potentially underestimates the number of sites under positive selection (Yang et al. 2007).

Branch tests

For *rpoB*, *rpoC1*, *rpoC2*, *ycf1*, and *uS19c* d_N and d_S trees (with non-silent and silent branch lengths respectively), are shown in Fig. 9. These show that *rpoB*, *rpoC1* and *uS19c* d_N trees have longer branch lengths, indicating higher non-silent rates. For *rpoC2* the d_N and d_S tree are about equal length. For *ycf1*, the d_S tree has longer branch lengths indicating more localized selection in the exon. The average ω-values for each of these genes, as calculated under the branch test, were: 2.1 (*rpoB*), 2.3 (*rpoC1*), 0.9 (*rpoC2*), 0.6 (*ycf1*) and 3.6 (*uS19c*), respectively.

Branch-site tests

For the branch-site test for *rpoB* we evaluated four scenarios (see Fig. 2A: 1) The best LnL scores were for scenario three (values not shown but available on request) and based on the HLRT, this scenario outperformed the others significantly (see supplementary material 6). Therefore, *rpoB* appears to be under varying levels of positive selection in the five clades recognized by van de Kerke et al. (2019) than over the entire section. For *rpoC1* results were slightly different, as the *rpoC1* gene tree topology is in minor conflict with other topologies (see above). Scenarios two to four (Fig. 2B) provided only a weakly, or no, significant improvement over scenario one. Even though the best LnL was for scenario two (all branches have separate ω's), it was not significantly better than the more conservative assumptions of scenario one. We therefore deduce that the ω for *rpoC1* is uniform throughout the clade and branch specific ω's cannot be assumed (not shown but available upon request). Individual sites listed as being under positive selection by the branch-site tests are largely the same as the ones detected under the site model for each gene (see Tables 2 & 3).

TABLE 2 | Summary of tests for positive selection (site test), including model 0: one-ratio ω, model 1: nearly neutral (2 categories), Model 2: Positive Selection (3 categories), Model 7: beta (10 categories) and Model 8: beta & ω >1 (11 categories). * Indicates that model 7 presented an improved explanation over model 2, but even though 8 had better LnL values (see supplementary material six) these were not significant over either 2 or 7. **bm**: indicates the model with the best lnL value regardless of significance. **#s+**: Indicates the number of sites under positive selection. **#s++**: indicates number of sites under positive selection with p>95_/_p>99. **s**: indicates the site positions under positive selection with p>95_/_p>99. **fm**: indicates final model assumed after the Chi² tests.

gene	bm	*p*	#s+	#s++	s		fm	ϰ/ω
atpF	2	-	NA	NA	-	-	model 0	0.69/0.56
ccsA	2	-	NA	NA	-	-	model 0	0.94/0.29
matK	8	-	NA	NA	-	-	model 0	0.98/0.47
rpl2	1	-	NA	NA	-	-	model 0	2.59/0.50
rps4	2	-	NA	NA	-	-	model 0	1,08/0,77
rpl22	2	-	NA	NA	-	-	model 0	0.57/0.27
uS19c	2	<<0,001	14	3_/_2	5,8,93	25,98	Model 2	0,24/3,26
rpoB	2	<<0,001	33	2_/_13	239,1075	42,47,215,216,218, 238,268,271,274, 275,378,815,1022	Model 2	1.27/2.08
rpoC1	2	<<0,001	59	11_/_20	50,145,178,180, 329,333,337,339, 481,549,712	24,44,47,59,140, 172,190,194,233, 254,342,343,347, 362,369,538,545, 552,653,708	Model 2	1.16/2.30
rpoC2	2	<<0,001	31	3_/_-	615, 1026, 1319	-	Model 2	0.81/0.87
YCF1	8	<<0,001	18	9_/_5	4,10,19,42,107,212, 329,1123,2303	60, 351, 828, 1717,2296	Model 7*	0.21/0.57

TABLE 3 | Sites under selection based on the branch-test. Scenarios are explained in the main text. 'best scenario' refers to sites in **bold** type were recovered with the site test as well. **bm**: indicates the model with the best lnL value regardless of significance. **#s+**: Indicates the number of sites under positive selection. **#s++**: indicates number of sites under positive selection with p>95_/_p>99. **s**: indicates the site positions under positive selection with p>95_/_p>99. **fm**: indicates final model assumed after the Chi² tests. 'Best scenario' refers to the scenario with the best LnL score after applying the hierarchical likelihood ratio tests.

gene	best scenario	#s+	#s++	s	
rpoB	scenario 3	23	1_/_5	276	**42,215,268,274,378**
rpoC1	scenario 1	25	2_/_3	33,**145**	**44**,198,**347**

4

TABLE 4 | Intergenic spacer list. List of intergenic spacers/special sites with elevated π-values as displayed in Fig. 1.

lable in Fig. 1	located close to gene	other properties
IGS 1	5' *trnI* CAU	-
IGS 2	miscellaneous	-
IGS 3	between trnQ-UUG and *psb*K	insertion unique to *P. zonale*
IGS 4	miscellaneous	-
IGS 5	5' *rpoB*	-
IGS 6	5' *psbD*	absent in *P. inquinans* & *frutetorum*
IGS 7	3' *psbZ*	-
IGS 8	*trnS* GGA	-
IGS 9	3' *trn*-LEU	-
IGS 10	5' *trn*-LEU	-
IGS 11	miscellaneous	-
IGS 12	*trnL*-UAG	-
IGS 13	5' *ndhI*	-
α	*trnG*-GCC translocated in pelt and *bl33c* is a pseudogene in *P. multibracteatum*, *yemenense*, *quinquelobatum*, *insularis*, *omanense* and *somalense*	
β	*trnl*-UAG pelt is duplicated in a *P. peltatum* unique insertion	

4

Branch length correlation test

The branch length correlation test results for the substitution-based trees are reported in Table 1A and for the indel-based trees in Table 1B. The latter display stronger correlation between genes than the substitution-based trees. For the substitution-based trees only the correlation between *rpoB/C1* remains intact after Bonferroni correction. When using the uncorrected *p*-value of 0.01 the strongest correlations are between *rpoB/C1/C2*, *rrn23*, and *ycf1*. The analysis for the substitution-based *rpo* trees showed no correlation with each other or with any of the other gene trees evaluated. For the trees based on the indels the correlation patterns remained intact even after correction. Nearly all tree lengths correlated with each other except for *rpoC1* which did not correlate with other genes. For a full overview see Table 1A. Surprisingly, branch lengths for *rpoB* did not correlate with those for *rpoC2*. Branch lengths for both types of changes in the *rpo* subunits correlate with those in *rrn23* signifying an unknown connection. Both types of changes in *rpoB* further correlate with those in *ycf1*, which is not the case for *rpoC1* and *rpoC2*. The branch lengths for the substitution datasets of the concatenated *uSc* and *uLc* genes do not correlate with those of either *rpoB*, *rpoC1*, *rpoC2*, *ycf1* or *rrn23*, nor do they correlate with each other (WSR-test p 0.01). *Rrn23* indel patterns, in contrast, do correlate with those in *uSc* and *uLc*, as do *uSc* and *uLc* with each other WSR-test (p 0.01).

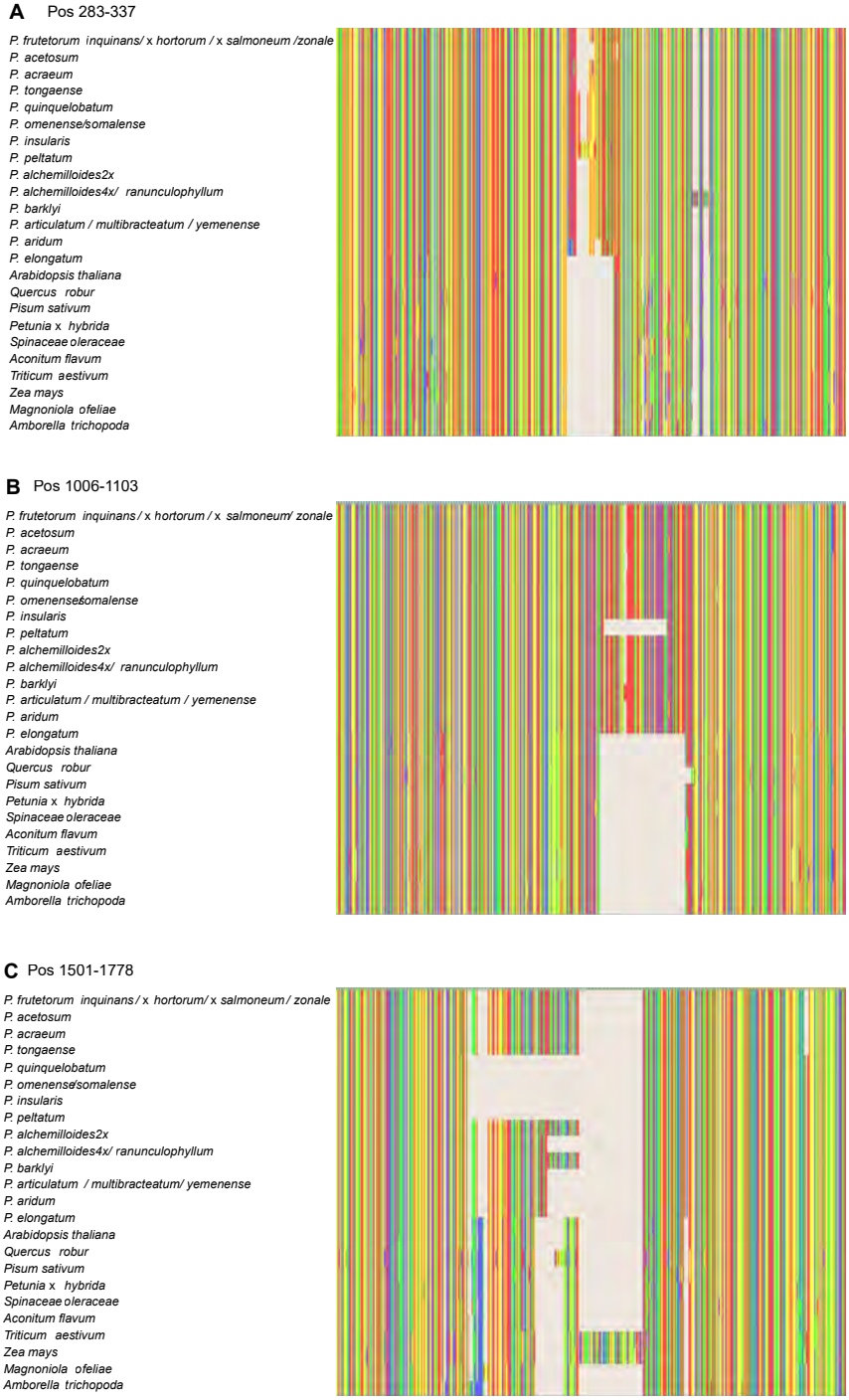
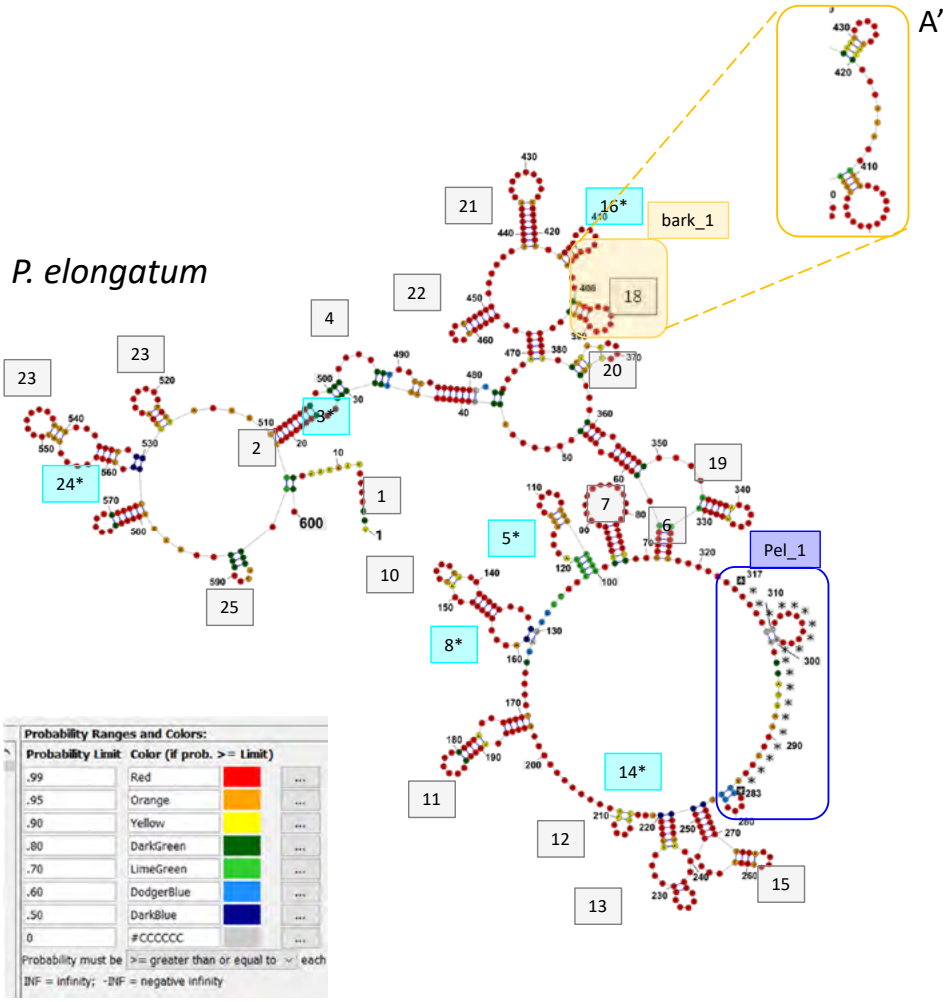


FIGURE 3 | Three *rrn23* regions from *Ciconium* compared with angiosperm outgroups: shown are large *Ciconium* unique indels. The outgroup *A. trichopoda* is listed at the bottom in each figure. A) contains domain I, B) domain II, C) domain III.

rrn23 indel variation in *Ciconium* compared to angiosperms

The *rrn23* gene ranges in length, for the included non-*Pelargonium* angiosperms, from 2810 bp in *S. oleracea* to 2888 bp in *Triticum aestivum*. Among *P.* sect.*Ciconium* species the length ranges from 2812 bp in *P. peltatum* to 2982 bp in *P. barklyi*, which is nearly twice as much variation as encountered across angiosperms. *Ciconium*-unique indels relative to the outgroups per domain (as defined by Bieri et al. 2017) are shown in Figs. 3A-C. We observe three different indel-regions where length variation is most prominent and apparently unique to *Ciconium*. The sizes of these are ~50, ~100 and ~100 to -10bp and are located in domains I, II, and III, respectively. For most accessions there are insertions relative to *S. oleracea*, but the indel region in domain III contains a 'gap' of 10 bp for *P. quinquelobatum*, *peltatum*, *omanense*,



somalense and *insularis* relative to *S. oleracea*. While sequence lengths and bp composition differ less in the *Ciconium* unique indels in domains I and II when compared to those in domain III, we find more variation in the estimated RNA secondary structures (Figs. 4-6). In domain III the length differences range from -10 to 30 bp for the 'short' indel region (e.g., *P. quinquelobatum*) relative to *S. oleracea*. For other accession the 'large' insertion ranges from 80 to 105 bp (e.g., *P. barklyi*). For domain III, the main structures correlate with these size difference classes mentioned above. I.e., a large insertion yielded one structure regardless of minor sequence length variation. The short indel region yielded one structure also, regardless of minor sequence length variation. The variation is contained in a stem-loop (for the large insertions) or just a large loop in 'Pel_3' for the shorter sequences (Fig. 6).

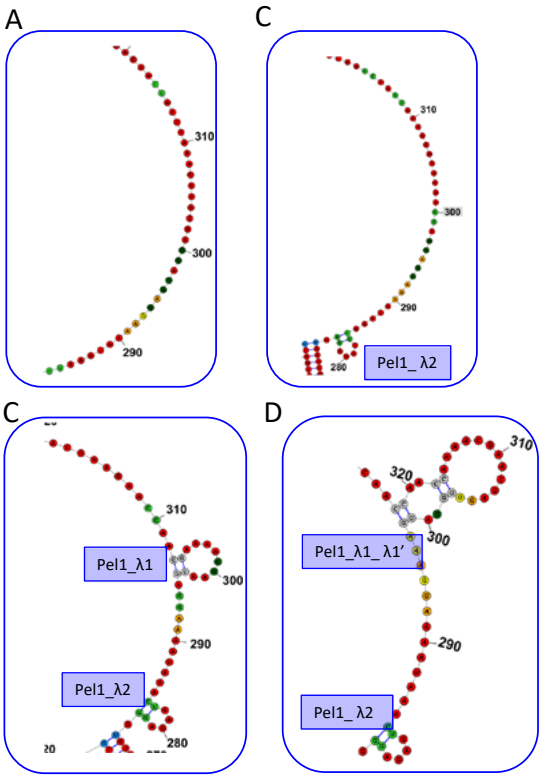
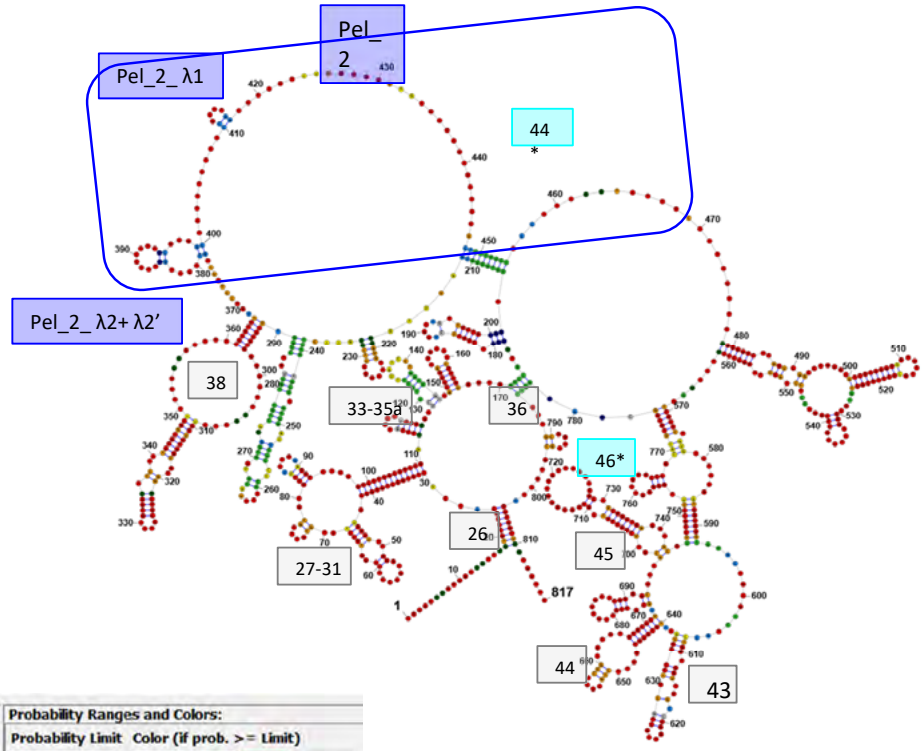


FIGURE 4 | Panel of 23S domain I modelled RNA structures for A) *P. barklyi* and B) *P. aridum*, *P. articulatum*, *P. multibracteatum*, *P. yemensene*, C), *P. frutetorum*, *P. inquinans*, *P. zonale*, *P. acraeum*, *P. tongaense*, *P. peltatum*, *P. alchemilloides*(2x), *P. ranunculophyllum*, *P. alchemilloides*(4x) and D) *P. omanense*, *P. somalense*, *P. quinquelobatum*. Colors of the base positions indicate probabilities. All were compared with *P. elongatum*. Numbers in the *P. elongatum* structure correspond to structures found in Bieri et al. 2017. Numbers in the grey boxes correspond to structures found in the *Pelargonium* structural model when compared with *P. oleraceae*. Numbers in the light blue boxes represent structures that were not recovered in the *Pelargonium* structural model. Finally structures with an * were not recovered in our analysis for either the *S. oleracea* comparison or the *Pelargonium* comparison. The legend indicates the confidence values for each color. The labels Pel_#λ indicate stem loop structures.

P. elongatum



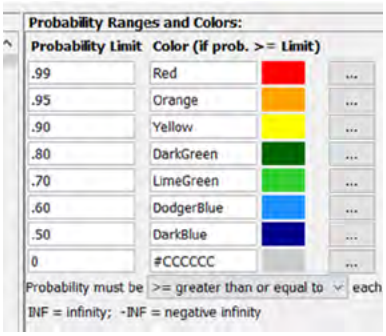
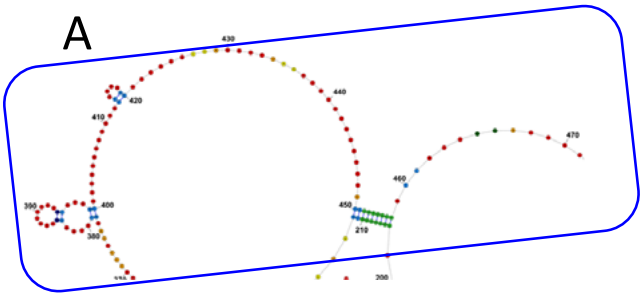
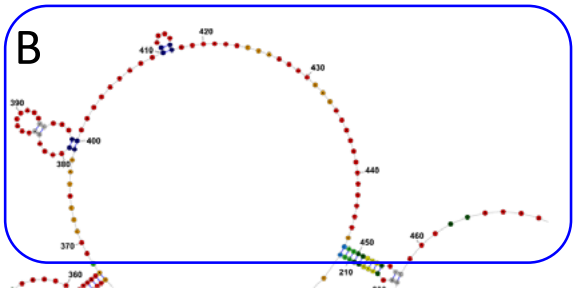
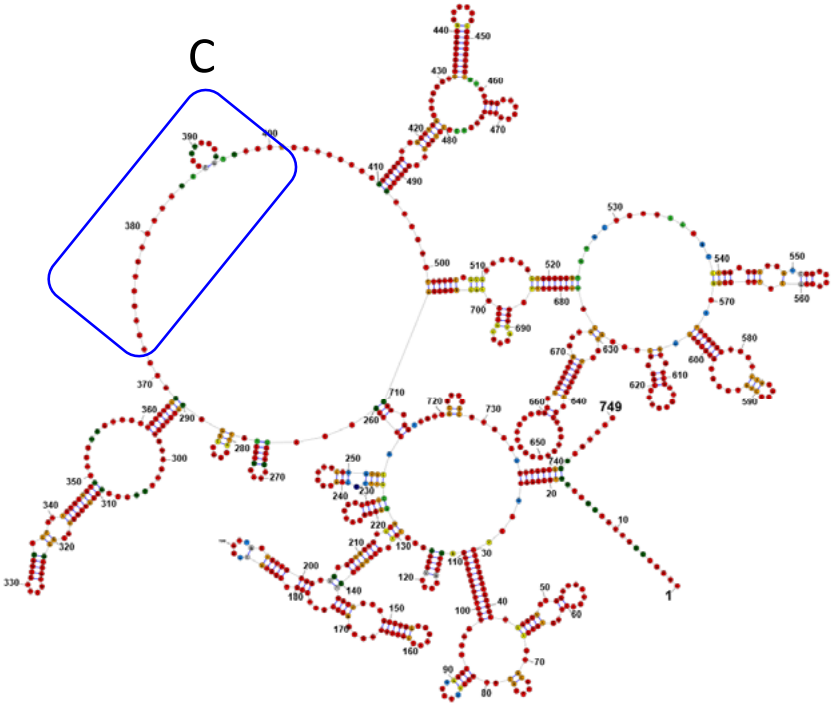


FIGURE 5 | 23S, domain II, modelled RNA structures for A) *P. articulatum*, *P. multibracteatum* and *yemenense* B) *P. acetosum*, *P. frutetorum*, *P. inquinans*, *P. zonale*, *P. acraeum*, *P. tongaense*, *P. alchemilloides*(2x), *P. ranunculophyllum*, *P. alchemilloides*(4x) *P. omanense*, *P. somalense*, *P. quinquelobatum*, *P. insularis* and c). *P. peltatum*. Coloring as in Fig. 4.







4

S. oleraceae

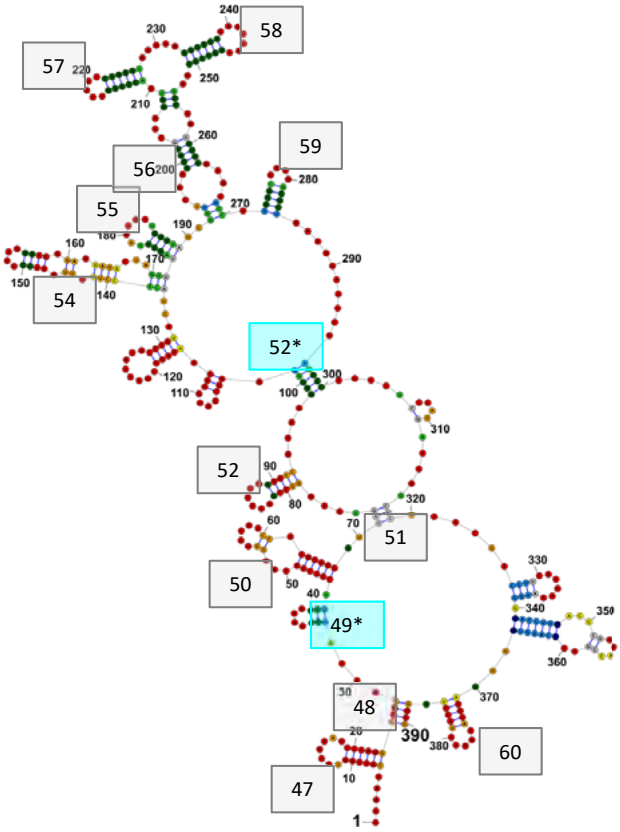
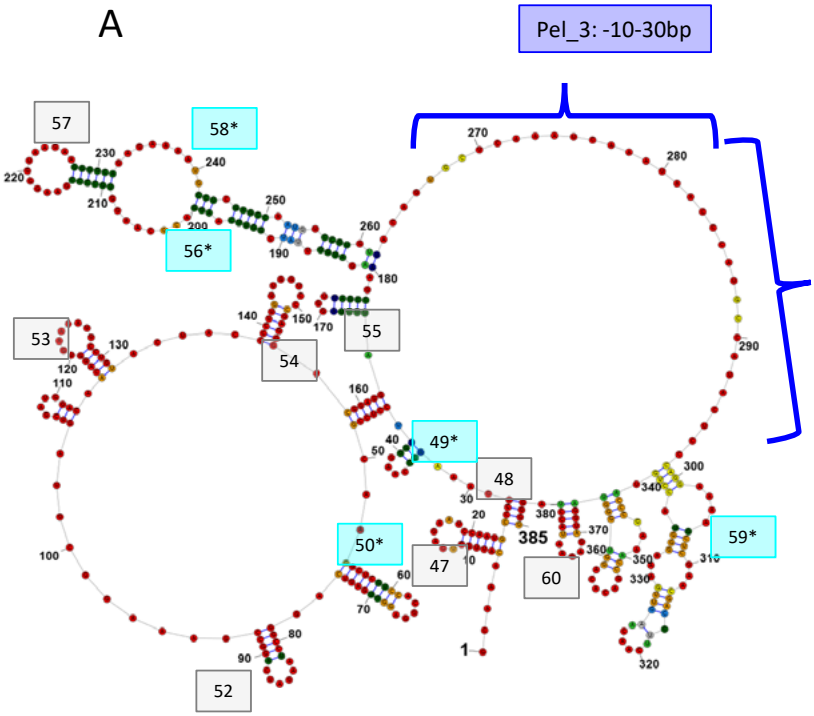
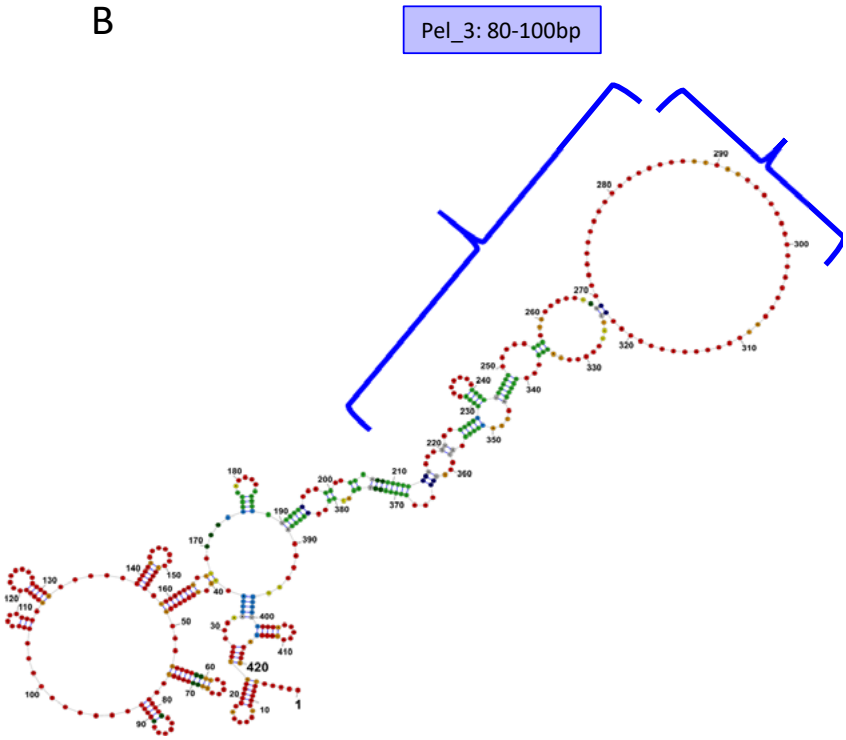


FIGURE 6 | *Pelargonium* 23S domain III compared with *S. oleraceae* A) *P. elongatum, P. aridum, P. articulatum, P. multibracteatum, P. yemensene, P. ranunculophyllum, P. alchemilloides*(4x), *P. omanense, P. somalense, P. quinquelobatum, P. peltatum, P. insularis*. B) *P. barklyi, P. frutetorum, P. inquinans, P. zonale, P. acetosum, P. acraeum, P. tongaense, P. alchemilloides*(2x). Coloring as in Fig. 4.

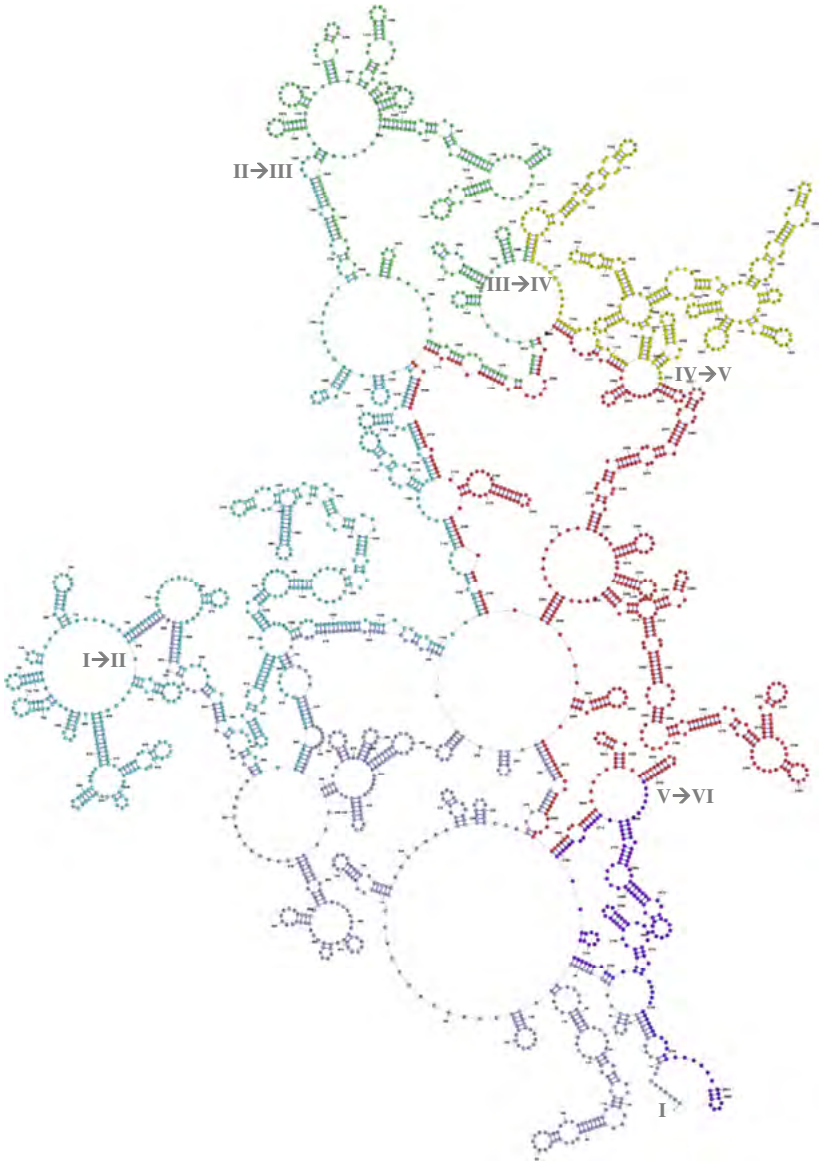
A


Pel_3: -10-30bp
57
58*
56*
53
54
55
49*
48
50*
47
60
59*
52
385
1

B


Pel_3: 80-100bp
420
1

A



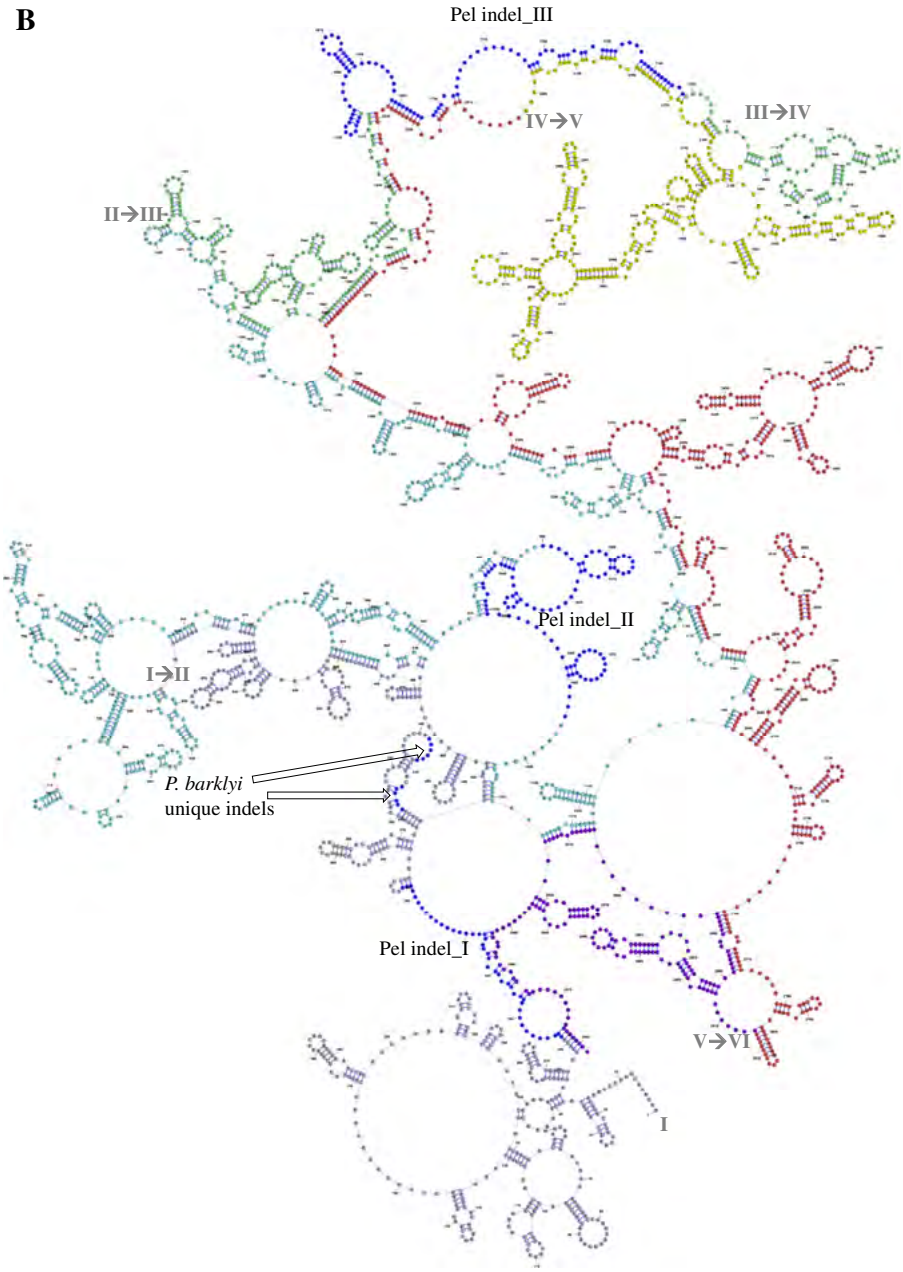
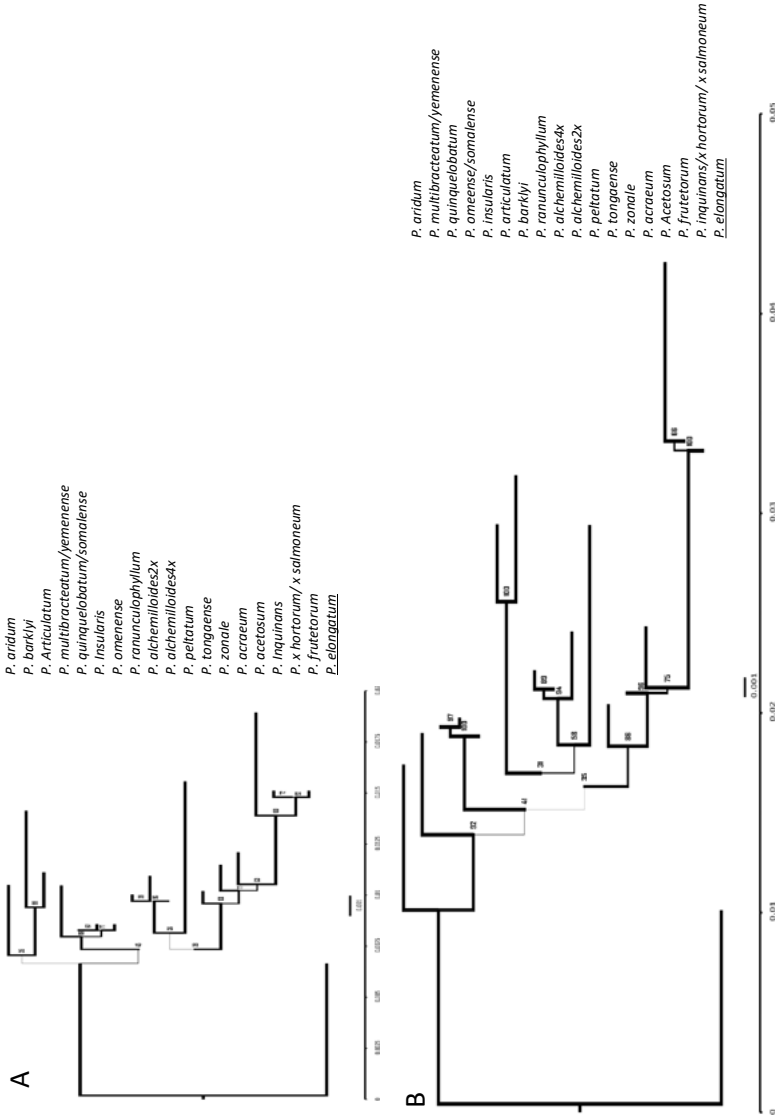


FIGURE 7 | Full 23S modelled RNA structure for A) *S. oleracea* and B) *P. barklyi*. Colors of each domain correspond to the domain colors assigned by Bieri et al. 2017. The dark blue regions are *Ciconium* unique insertions. Latin numbers in grey denote begin and end of each domain. Figure with confidence values is available upon request.


A
P. aridum
P. barklyi
P. Articulatum
P. multibracteatum/yemenense
P. quinquelobatum/somalense
P. Insularis
P. omenense
P. ranunculophyllum
P. alchemilloides2x
P. alchemilloides4x
P. peltatum
P. tongaense
P. zonale
P. acraeum
P. acetosum
P. Inquinans
P. x hortorum/ x salmoneum
P. frutetorum
P. elongatum
B
P. aridum
P. multibracteatum/yemenense
P. quinquelobatum
P. omeense/somalense
P. insularis
P. articulatum
P. barklyi
P. ranunculophyllum
P. alchemilloides4x
P. alchemilloides2x
P. peltatum
P. tongaense
P. zonale
P. acraeum
P. Acetosum
P. frutetorum
P. inquinans/x hortorum/ x salmoneum
P. elongatum
0.001
0
0.01
0.02
0.03
0.04
0.05

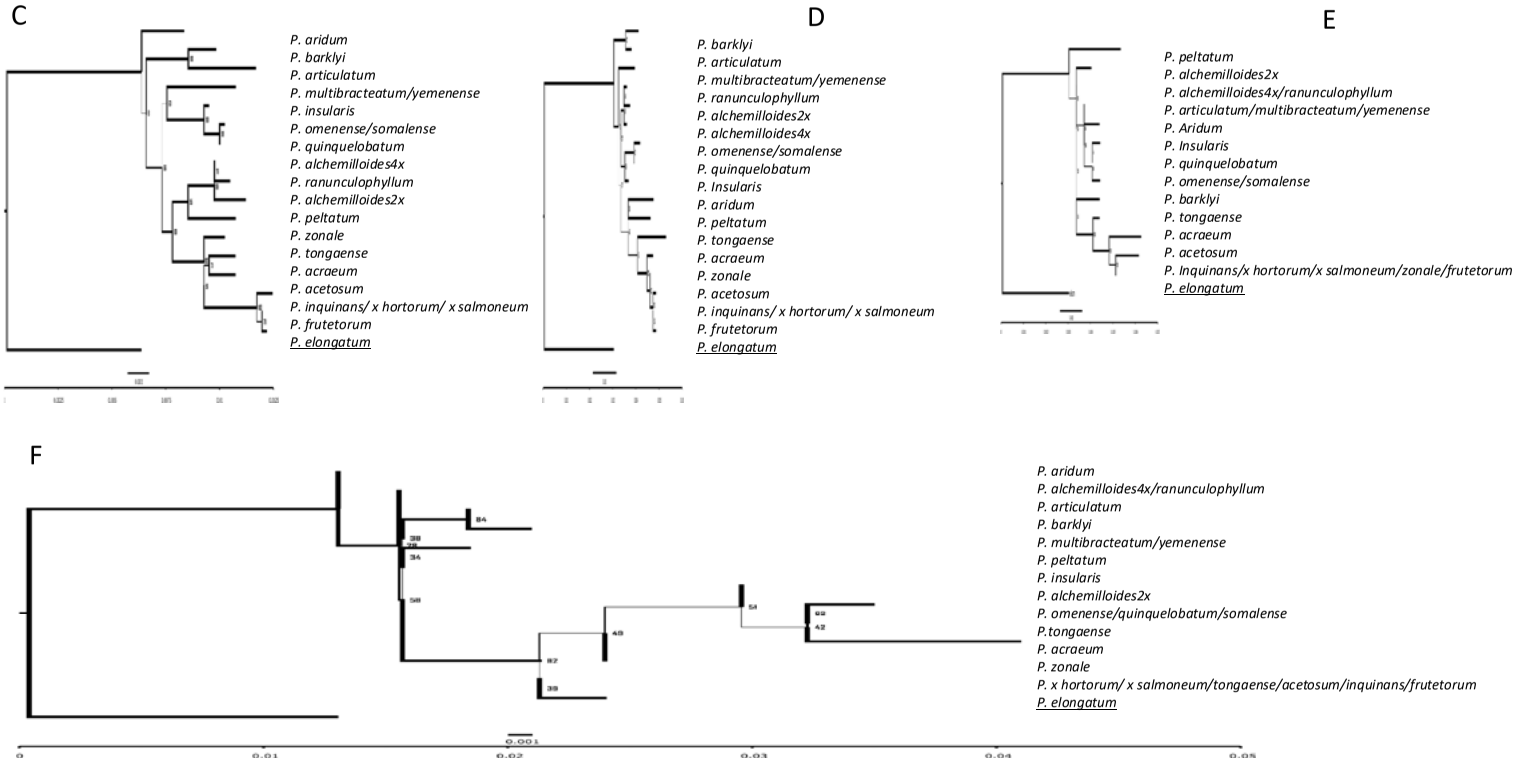


FIGURE 8 | dN/dS trees for, A: *rpoB*, B: *rpoC1*, C: *rpoC2*, D: *rrn23*, E: *YCF1* and F: *uS19c* scaled trees, based on substitution patterns. All scale bars indicate 0.001 substitutions per site.

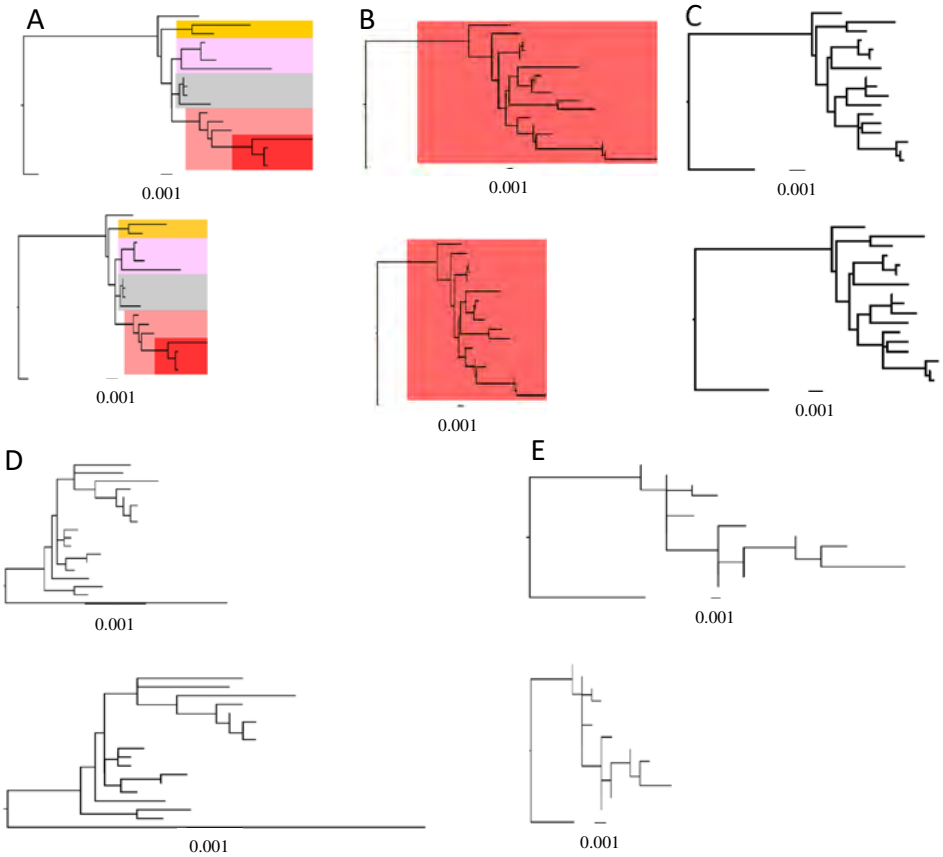


FIGURE 9 | Silhoutes of dN/dS trees for A: *rpoB*, B: *rpoC1*,C: *rpoC2*, D: *YCF1* and E: *uS19c* top dN trees, bottom dS trees. In A) Five clades are highlighted with elevated dN/dS ratios (ω>1) under the optimal scenario (three) for *rpoB*. In B) Core *Ciconium* is highlighted as being under + selection (ω>1) for *rpoC1* under the optimal scenario (one) in the B-S test.

Ribosomal protein structure modelling

The length differences of *P. barklyi* bL32c and uS19c peptides compared to *S. oleracea* are 29 and 40 aa respectively (or 38% and 31%) (Figs. 10 & 11a). These were modelled *ab initio*, as no reliably annotated homologue was available for either gene. For the other 62%/69% of the sequence the structure model was >90% accurate as it could be compared with annotated homologue *S. oleracea* that also represented the closest match for both genes. Together, this resulted in an overall accuracy was <70% for both genes, making the reconstruction 'unreliable' according to the authors of the Phyre2 software (Kelley et al. 2015). For those parts of the sequences for which homologues structures were found, the reconstruction was 'highly reliable'. For both examples the *Ciconium* proteins have acquired extra structures: bL32c and uS19c have mainly acquired hydrogen turns and bends (determined by electrostatic interactions) resulting in the addition (relative to *S. oleracea*) of alpha helices to the C- terminal (Figs. 10B & 11B). The full protein modelling reports are available upon request from the authors.

Discussion

Our study shows that the plastid gene encoding 23S rRNA is more variable between species of *P*. sect. *Ciconium* than it is compared to the *rrn23* gene of entire angiosperm plant orders (Fig. 3). This gene is usually highly conserved even among phyla. Even though hints of length variation in Geraniaceae rRNA genes were published previously (Weng et al. 2016), we found the high sequence length variation in this gene among *Ciconium* species remarkable. The variation is located in domains I-III (Figs. 3-6) as originally defined by Yusupov et al. 2001. As these domain definitions are based on ribosomal morphological traits, we speculate that the *Ciconium* ribosomes likely have different RNA folding structure. Given that these three domains cover half of the 23S molecule the effects on the overall ribosomal structure maybe quite profound. The first is that the ribosome may have been altered in the way it functions. The second consequence is that compensatory changes in the ribo-proteome have evolved to maintain a normal function. We think the first explanation is highly unlikely as this would imply changes in the entire genetic pathway. We believe the second explanation to be more plausible in the light of the evidence from this study. We compared our structural models with the best characterized 23S rRNA structure model to date from *S. oleracea* (Bieri et al. 2017 and Perez-Boerema et al. 2018). Our structural models of domains I-III, our two protein structural models and the full 23S model indeed show *Ciconium* specific structural variation in both the ribosomal 23S backbone and the bL32c and uS19c protein structures, which may suggest unique and/or altered species-specific ribosomal structures.

Correlation of branch lengths across gene trees

Our findings suggest that the patterns of insertion and deletion correlate more strongly with each other than substitution patterns do. That being said, we may have underestimated the correlation because we have used a very strict method for *p*-value correction. Even so, after the Bonferroni correction, the branch lengths correlations for the substitution-based trees stayed intact for both *rpoB* and *rpoC1*, with *ycf1* (WSR-test $p<0.01$). This suggests a tight co-evolution between the two, which is, however, not obvious from a functional perspective. *ycf1* maybe involved in protein import into the chloroplast (Kikuchi et al. 2013), but this is still controversial (e.g., Bölter and Soll, 2017). The fact that the indel patterns show stronger correlation than the substitutions suggests a co-evolution of structural changes brought on by rearrangements acts in *Ciconium* as it does across Geraniaceae (Blazier et al. 2016b).

Other indel correlations can be seen between *rpo* subunits and *rrn23*/ribosomal genes (Table 1B). This suggests co-evolution of PEP and *rrn23*. Because *rrn23* is solely transcribed by PEP (Börner et al. 2015) this is perhaps not surprising, although the actual mechanism remains unclear. Given the obvious functional relationships and demonstrated co-evolution between *rrn23* and the ribo-proteome (Weng et al. 2016), the correlation between these comes as no surprise.

Positive selection

We show that in *P.* sect. *Ciconium*, plastid-encoded *rpoB*, *rpoC1* and *uS19c* are under positive selection. For *rpoC2* and *ycf1* the average across sites (ω) values were below one, nevertheless three sites for *rpoC2* and nine for *ycf1* appear to be under positive selection, suggesting more localized or lineage specific adjustments of the gene. Interestingly the genes under positive selection occur both in the IR and the LSC. The SSC seems to be unaffected, although the intergenic spacers (IGS) are affected there as well (see Fig. 1 and supplementary material 2 for details). The fact that gene evolution is not limited by occurrence in a specific plastomic regions further substantiates the findings by Weng et al. (2017) that the IR does not mitigate mutational effects in *Pelargonium* as would be expected by the occurrence of two copies per plastome (Weng et al. 2016). The genes *bL32c*, and *uS11c* mainly contain indels, but are listed as being under positive selection in the Geraniaceae (Guisinger et al. 2008, Weng et al. 2014, 2016). As these studies were conducted at the family level, it remained to be seen which genes would remain when testing a in a smaller group of more closely related species. At the sectional level we show that correlation between branch lengths of different gene trees occurs as well. In other words, accessions with longer branches for one gene should also have longer branches in other genes. Probably as a response to the changed 23S rRNA molecule, ribosomal proteins are under positive selection, or they have obtained indels. This would suggest to us that the nuclear part of the ribo-proteome is, partially, under positive selection in *Ciconium* as it is in Geraniaceae as well (Weng et al. 2016).

The *Ciconium* ribosomes

There appear to be two ways, apart from translocation to the nucleus (e.g., *bL32c* in *Populus*; Ueda et al. 2007, *uS7c*, *uS16c*, *uL20c*, *uL22c* and *bL32c* in *Passiflora*; Shesthra et al. 2020), in which ribosomal proteins can respond to the changed rRNA backbone. The first is altered patterns of protein amino acid substitutions. The second is the occurrence of small indels within protein coding regions (small stretches of deletions or duplications of particular amino acids). We see both occurring, dependent on which protein we examined. The following proteins show mostly altered substitution patterns: *uS2c*, *uS3c* (red flowered clade only) *uS4c*, and *uS16c*. None of these are under positive selection in *Ciconium*, but the substitutions did lead to aa changes in most of these genes some of which are listed as being under positive selection in Geraniaceae (Guisinger et al. 2008, Weng et al. 2016). The second option occurs in *bL32c* and *uS11c*, and a combination of both occurs in *uS19c*, *uL22c* and *uL23c*. Interestingly, the *uL23c* exon is broken into two pieces in one accession (*P. elongatum*) with an intron-like sequence being present for this taxon only. We hypothesize that, as part of the larger plastomic upheaval in Geraniaceae, the ribosomal proteome underwent structural changes. This is supported by the occurrence of length variations and substitutions in our sequences as well as by our ribosomal protein structural modes. The ribosomal proteins probably co-evolved to cope with the effects of the changed rRNA structure. One of the altered ribosomal proteins (bL32c) in *Ciconium* is known to be important in maintaining structural integrity (by connecting 4.5S rRNA to 23S rRNA) of the ribosome (Bieri et al. 2017). Extra alpha helices in bl32c are associated with compensatory changes that accommodate ribosomal structural changes (Bieri et al. 2017) and our protein model of bl32c shows exactly that (Fig. 10).

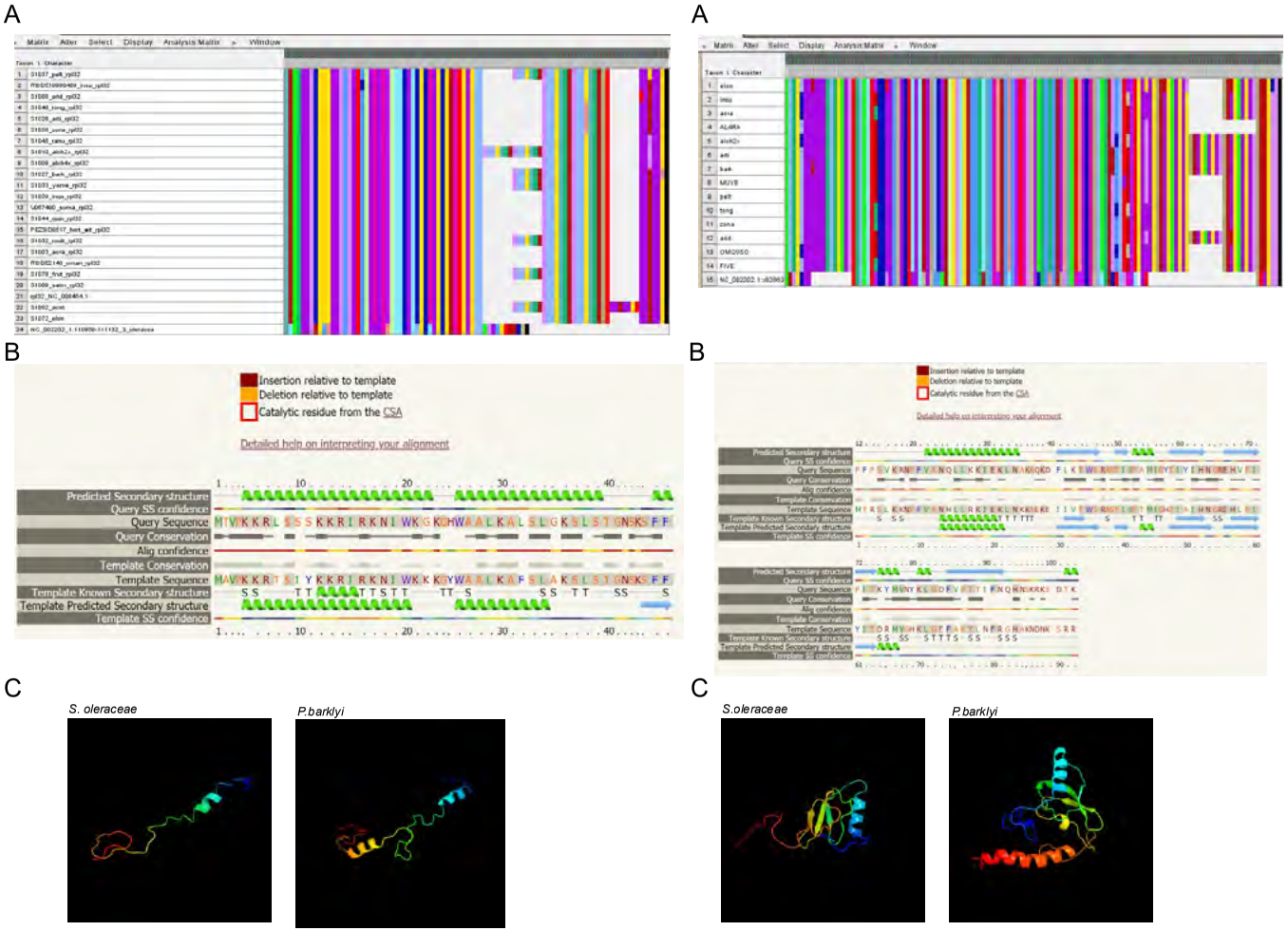


FIGURE 10 AND 11 | Protein modelling bl32c (left) and uS19c (right). A) Alignments of *bL32c* and *uS19c* of *S. oleracea* (bottom) with *Ciconium*, with coloration based in amino acid identity. B) Secondary structure summary and C) 3D structure comparisons for each protein structure *of S. oleracea* and *P. barklyi*.

Nuclear genomic genes interacting with the plastome

Especially in later developmental stages, PEP is the main polymerase responsible for transcribing rRNA in the chloroplast (Siniauskaya et al. 2016, Börner et al. 2016), and plays a central role in its development (Kremnev and Strand 2014). However, PEP is aided by the so called 'interactome' (Westrich et al. 2012) which is a range of, mainly, nuclear encoded proteins that fulfill numerous roles during assembly and maturation of PEP (Shikanai and Fujii 2013). *Rrn23* structural evolution seems to co-evolve with *rpo* subunits as well as with the ribo-proteome. Unfortunately, the exact physico-chemical interactions and processes involved are not completely known therefore, we cannot know exactly how the changed PEP structure affects the ribosomal assembly. We do not know whether changes in PEP caused altered rRNA structures or whether the proteins in the interactome might also contribute (Williams-Carrier et al. 2014). It is thought that the nucleoid provides the scaffold for subunit assembly, and that so called 'ribosomal assembly co-factors' are needed for proper ribosome assembly (Bohne 2014).

A recent study of the numerous interactome proteins required for proper plastome expression found that nuclear encoded *PPR* genes assist in ribosome assembly (Westrich et al. 2012). *PPR* genes play major roles in many steps of plastome and ribosome development, (reviewed by Wang et al. 2021) and are often targeted specifically at organelles (Barkan 2014). *PPR* genes also play a role in compensating for changes in plastid genes via RNA editing (Ichinose 2017, Small et al. 2020), but suppressing errors in 'micro-homology', which may occur in *Pelargonium* plastomes due to the rearrangements and repeats, is the domain of the nuclear encoded *Whirly* genes (Maréchal et al. 2009, Isemer et al. 2012). The role of sigma factors in ribosome assembly is unclear, but they do play an important role in gene expression, closely interact with PEP and were shown to co-evolve with plastomes in Geraniaceae (Zhang et al. 2015). Because of the important role of the three groups of genes discussed above they should not be overlooked and a detailed study of the nuclear genomic perspective should be undertaken in *Ciconium* as a follow up to this study.

Evolutionary rate covariation affects the ribosomal machinery in *Ciconium*

When correlated changes in rates of sequence evolution across a phylogeny in functionally related proteins occurs, this is referred to as 'evolutionary rate covariation' ('ERC', Forsythe et al. 2021), which was found to occur across angiosperms. Two results support that ERC is occurring in Geraniaceae and *Ciconium* as well: the co-evolution between functionally related genes and the occurrence of CNI explained by nuclear-encoded plastid targeting genes. Numerous studies point to nuclear genomic alleles being responsible for the occurrence of CNI in *Ciconium* (Tilney-Basset 1976, 1984, Metzlaff et al. 1982, Tilney-Basset and Almouslem 1989, Tilney-Basset et al. 1992, Apitz et al. 2013, Horn 1994, Zhang et al. 2015, Breman et al. 2020) and co-evolution between functionally related genes was shown to occur (Zhang et al. 2015, Weng et al. 2016, this study). In Geraniaceae, and certainly in *Ciconium,* ERC seems to occur in genes controlling the ribosomal machinery (referred to here as: rERC) rather than

the photosynthetic machinery (referred to here as: phERC). Changes in the photosynthetic machinery are not unknown though, and an effect of phERC is thought to be visible in a shift towards parasitism (Forsythe et al. 2021).

In summary, we have demonstrated that *Ciconium* plastomes are highly variable but the effects of the genomic changes and the plastids' role in CNI remain under studied. This study provides a detailed overview of the plastomic variation and thus, a good starting point to discuss the plastid role in the occurrence of CNI (Chapter five). Further, we found that the ribosomes and ribo-proteomes of closely related species from *Ciconium* are more variable than between other angiosperm orders. There are other cases whereby the organellar genomes and specifically the ribosomal proteins show relaxed constraints (e.g., in *Silene*: Sloan et al. 2013), but not on the scale as occurs in *Ciconium*.

Acknowledgements

We thank Sara van de Kerke, with initial assistance in plastome assembly and Nenad Ban for help with the interpretation of the Figs and structures from Bieri et al. 2017. We thank the Royal Botanic Garden Edinburg researchers Alan Forest, Mary Gibby and Sabina Knees for supplying herbarium specimens of *P. somelanse*, greenhouse material of *P. insularis*, as well as material of the new species *P. omanense* sp. nov (Knees et al. in prep).

Author contributions

Conceived the study: FCB, FTB, SG, MES. Carried out the analysis: FCB. Wrote the manuscript: FCB, FTB. Informatics analyses: FCB. All authors read the draft and gave feedback.

Funding

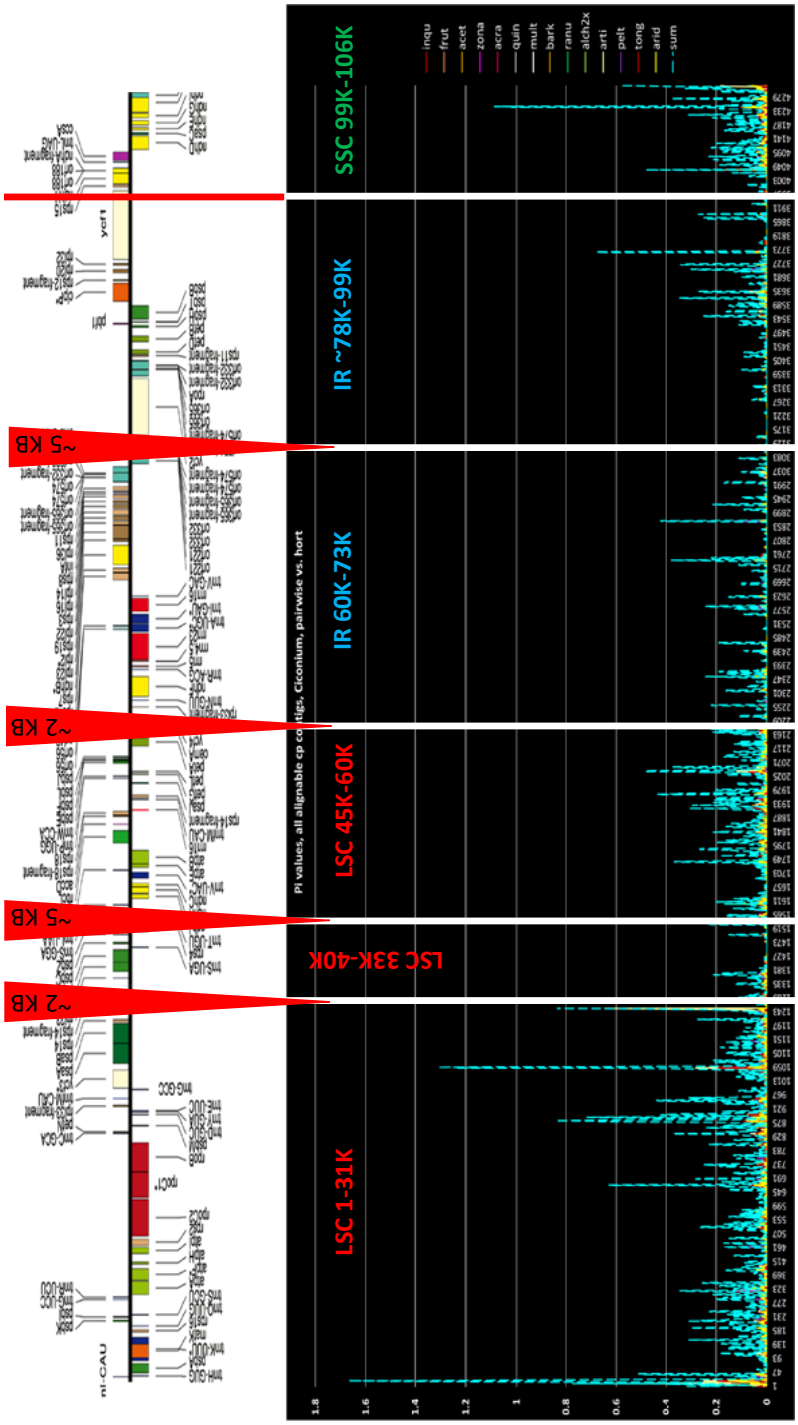
This research was funded by the Dutch Foundations for applied scientific research (TTW). Grant number: 14531 "*Pelargonium* genomics for overcoming cytonuclear incompatibility and bridging species barriers" of the Green Genetics program. SG is supported by the Max Planck Society.

Supplementary material

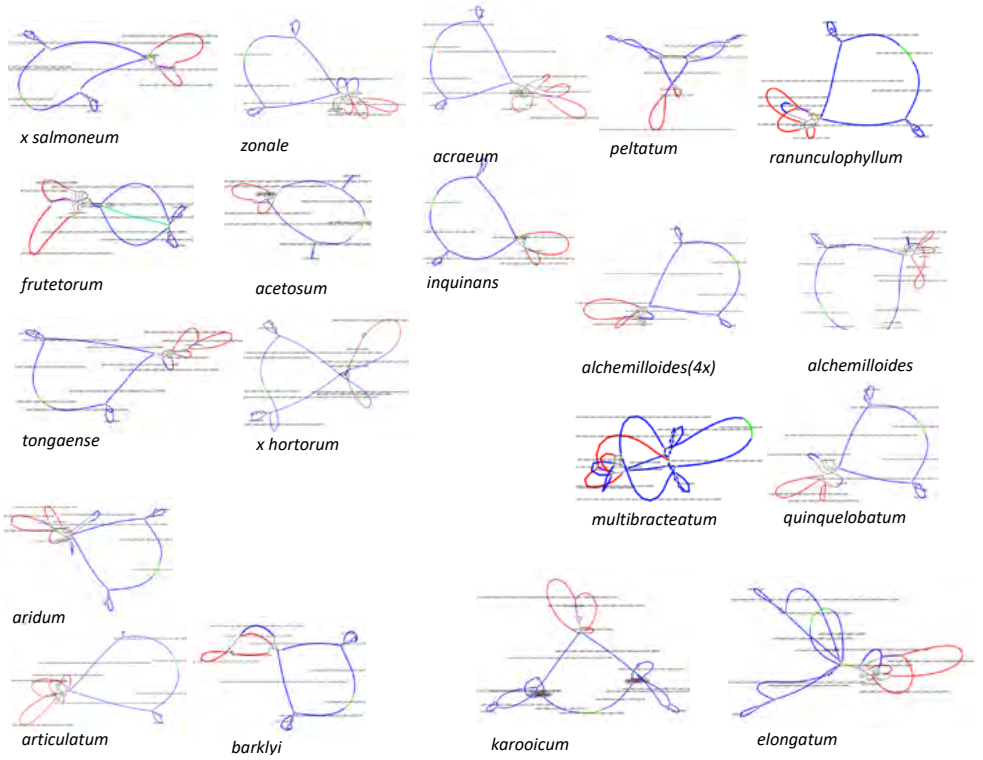
SUPPLEMENTARY MATERIAL 1 | Plant material used in this study.

Species	Herbarium Voucher accession	Institute[1]
P. acetosum	1243	NHM
P. acraeum	1975	STEU
P. alchemilloides	1885	STEU
P. alchemilloides	1882	STEU
P. articulatum	1972055	WAG
P. barklyi	1972061	WAG
P. frutetorum	1972062	WAG
P. inquinans	0682	STEU
P. multibracteatum	2902	STEU
P. peltatum	1890	STEU
P. quinquelobatum	1972049	WAG
P. ranuncolophyllum	A3651	MSUN*
P. tongaense	3074	STEU
P. zonale	1896	STEU
P. elongatum	0854	STEU
P. aridum	1972053	WAG
P. insularis	19990489	RBGE
P. yemenense sp. Nov	1972037	WAG
P. omanense sp. Nov	2184	RBGE
P. somalense	V-067490	V

[1]STEU = Stellenbosch University, RSA; AL = Albers/MSUN = Münster & *Bakker et al. 2004. WAG = National Herbarium of the Netherlands. V = Uppsala herbarium



SUPPLEMENTARY MATERIAL 2 | Summary π -plot comparing all sequences with each other inculding an overal sum indicated by the dashed line. Top part of the figure is a linear representation of the annotated *P. x hortorum* plastome (Chumley et al. 2006). 'LSC' denotes long single copy, 'IR' inverted repeat, 'SSC' the short single copy region. Numbers indicate (in K or Kbp) the length of homologous contigs from the alignment. LSC denotes the long single copy region of the plastome, IR the inverted repeat and SSC the short single copy region.



SUPPLEMENTARY MATERIAL 3 | Partial plastome reconstructions results from GetOrganelle with large contigs indicated and identified. The LSC fragments are colored red, the IR fragments in blue, the SSC in green, *uS14c* in yellow, *uS18c* in orange and *rbcl* in purple.

SUPPLEMENTARY MATERIAL 4 | Wilcoxon Singed rank tests. Spreadsheet with calculations of branch length ratios and the WSR-test.

SUPPLEMENTARY MATERIAL 5 | Table with genbank accessions of angiosperm rrn23 sequences.

Gb accession number/ref	species	order	clade*	rrn23 length
NC_005086.1	*Amborella trichopoda*	Amborelales	outgroup	2816
NC_020318.1	*Magnolia grandiflora*	Mangoliales	Magnoliids	2804
MT571464.1	*Aconitum flavum*	Ranunculales	Eudicots	2810
AB042240.3	*Triticum aestivum*	Poales	Monocots, Comelinids	2888
NC_001666.2	*Zea mais*	Poales	Monocots, Comelinids	2884
NC_000932.1	*Arabidopsis thaliana*	Brassicales	Eudicots, core Eudicots, RosidII	2810
NC_046388.1	*Quercus robur*	Fagales	eudicots, core Eudicots, RosidI	2809
MK_455838.1	*Pisum sativum*	Fabales	eudicots, core Eudicots, RosidI	2811
this study	*Pelargonium sp*	Geraniales	eudicots, core Eudicots, RosidI	2812-2982
AJ400848.1	*Spincea oleracea*	asterales	eudicots, core eudicots,asteridII	2810
NC_040178.1	*Petunia x hybrida*	solanales	eudicots, core eudicots,asteridI	2810
NC_028519.1	*Plantago maritima*	lamiales	eudicots, core eudicots,asteridI	2814

* Following the APG naming system

SUPPLEMENTARY MATERIAL 6 | Overview of hierarchical likelihood ratio tests. A spreadsheet with the full HLRT and results for positive selection tests.

SUPPLEMENTARY MATERIAL 7 | d_N/d_S trees in full detail. d_N/d_S trees are given in high resolution for *rpoB*, *rpoC1*, *rpoC2*, *ycf1* and *uS19c*.

Online data
All supporting materials can also be viewed at: https://figshare.com/s/4bfda6a08dbd9dd11583

Chapter 5

Possible causes of plastid induced CNI and organellar inheritance in *Pelargonium* section *Ciconium* (Sweet)

Breman FC[1], Korver JW[1], Snijder RC[2], Sancho-Such M[2], Schranz ME[1], Bakker FT[1]

[1]Biosystematics Group | Wageningen University and Research, Wageningen, the Netherlands
[2]Syngenta Seeds BV, Andijk, the Netherlands

FCB: ORCID https://orcid.org/0000-0002-9722-7564
FTB: ORCID https://orcid.org/0000-0003-0227-6687
MES: ORCID https://orcid.org/0000-0001-6777-6565

abstract

We studied the inheritance of mitochondria and chloroplasts in four series of interspecific crosses of species from *Pelargonium* section *Ciconium* (Geraniaceae). They were used to investigate the potential causes of cytoplasmic-nuclear incompatibilities (CNI), seen as chlorosis, in interspecific hybrids. We confirmed that chloroplasts were inherited maternally, paternally and bi-parentally and found that mitochondria were inherited primarily maternally, but that paternal and bi-parental inheritance does occur. We next examined the correlation of plastid-encoded *rpoB* sequence length variation and resulting physico-chemical properties of *rpoB* peptide variants with chlorosis phenotypes in F_1 interspecific hybrids. We generated structural protein models of the plastid-encoded (RNA) polymerase (PEP), subunit rpoB, to study the location and the possible effect of the variation encountered in *rpoB* in our accessions. We speculate that the variable *rpoB* genotypes and corresponding physio-chemical properties of the *rpoB* PEP subunit explain the occurrence of plastid-induced chlorosis observed in our crosses. We conjecture that PEP variable sites are in direct contact with the template DNA as well as nuclear encoded expression factors. This implicates *rpoB* variation as the causative agent of chlorosis, but also provides a possible, if partial, explanation for the observed elevated rate of structural evolution of the *Pelargonium* plastome.

Keywords: *Pelargonium*, evolution, plastome typing, mitome typing, KASP, *rpoB*, CNI, F_1 chlorosis prediction

Introduction

The commonly known 'garden geranium' is the product of interspecific hybridization between two species from *Pelargonium* sect. *Ciconium* (Sweet) Harvey (1860: 298), *P. inquinans* and *P. zonale* (James et al. 2004). Its origins dates back to the 19th century, as intense hybridization efforts were undertaken since the early 1800s (Sweet 1820-1822), especially in Victorian England. Early breeders noticed the frequent occurrence of chlorosis in offspring of species from section *Ciconium* (hereafter referred to as '*Ciconium*'), and later it was established that chloroplast inheritance was causal (Baur 1909).

Establishing interspecific hybrids between many species of *Pelargonium* is relatively easy (Horn 1994, Breman et al. 2020); however, the resulting offspring often display signs of Cyto-Nuclear Incompatibility (CNI). When caused by mitochondria ('mCNI') this manifests itself as the occurrence of dwarf growth or (partial) male sterility (Schnable and Wise 1998). Plastid induced CNI ('pCNI') is manifested by the occurrence of bleaching of the leaves (chlorosis) which occurs regularly in F_1 hybrids of interspecific crosses in *Pelargonium* (Baur 1909, Horn 1994, Breman et al. 2020). This is caused by the fact that chloroplasts can be inherited from either parent in *Pelargonium* (Baur 1909, Metzlaff et al. 1981, Guo and Hu 1995, Weihe et al. 2009, Apitz et al. 2013). While such biparental inheritance of organelles (mainly plastids) is uncommon across angiosperms (Greiner et al. 2015), it is widespread throughout *Ciconium* with all 17 species displaying the ability to transmit plastids from either parent (Breman et al. 2020). Given the ubiquity of cytoplasmic biparental inheritance in *Ciconium*, it could be widespread throughout the genus, though confirmation of this assertion is not yet available. The level of chlorosis in interspecific offspring is dependent on the plastome inherited from either of the parents as well as the nuclear genome of either (Tilney-Basset 1984, Tilney-Basset et al. 1992, Breman et al. 2020).

Even though bi-parental inheritance of plastids may be considered uncommon across angiosperms, bi-parental (or paternal) inheritance of mitochondria is considered to be much rarer (Greiner et al. 2015). Examples of paternally inherited mitochondria were found in cucumber (Havey 1997) and wheat (Kitagawa et al. 2002). An example of bi-parental inheritance of mitochondria was found in a reciprocal cross of *P. inquinans* x *P. zonale* (Weihe et al. 2009, Apitz et al. 2013). The plastome and mitome of *Pelargonium* species are both re-arranged and have elevated substitution rates (for the mitome see Bakker et al. 2000, 2006, Mower et al. 2005, Palmer et al. 2000 and see Ruhlman and Jansen 2018 for a review of Geraniaceae plastome properties). *Pelargonium* plastomes are among the largest, and smallest, known for angiosperms (~275kb Chumley et al. 2006; Weng et al. 2012, 2017). We do not know the size of *Pelargonium* mitomes, but typical plant mitomes are larger than even the largest plastomes (~700kb with outliers to 1-2 Mb) (Gualberto et al. 2014). In the work presented here, we have studied the inheritance of chloroplasts and mitochondria in interspecific crosses. We focused on the plastome to find a possible explanation for the observed pCNI in hybrid offspring.

The regulation and expression of organelles is coordinated by the nuclear genome (Zoschke and Bock 2018). For chloroplasts, anterograde (from nucleus to organelle) signals initially

target *rpoB* which then initiates expression of genes contained in the plastome (Börner et al. 2015). The *rpoB* sub-unit is part of the plastid-encoded (RNA) polymerase (PEP), which is not the only polymerase involved in plastid gene expression. There is also a nuclear-encoded polymerase (NEP) which is involved in the expression of several plastid genes. Generally, PEP is (almost) solely responsible for the expression of rRNA and photosystem I and II (*psba* and *psbb*) genes (Demarsy et al. 2006). All other plastid genes are, at least partially, expressed via NEP (Demarsy et al. 2006). In Geraniaceae, *rpoB* is highly variable in length (Chapter four and Fig. 1) and is among the genes that show signs of strong positive selection (Guisinger et al. 2008, Chapter four). In contrast to the plastids, mitochondria do not carry their own polymerase and mitochondrial genes are expressed via the same NEP as the plastids, as well as by a dedicated nuclear-encoded, mitochondrially-targeted NEP (Zoschke and Bock 2018).

Given the common occurrence of bi-parental inheritance of plastids in *Ciconium* and the above-mentioned occurrence of bi-parental inheritance of mitochondria (Weihe et al. 2009, Apitz et al. 2013), we hypothesize that mitochondria can be inherited biparentally in much the same way as plastids do. We therefore used previously published PCR markers for plastid detection (Breman et al. 2020) and developed mitochondrial KASP markers for four series of interspecific crosses. We further hypothesize that pCNI patterns, observed in >30 interspecific crosses, can be explained by variation in the plastid-encoded *rpoB* gene, as this is the gene through which organelle expression is initiated. Finally, we explored what the possible effects of the *rpoB* length variation may be on *rpo*B peptide physico-chemical properties and *rpo*B subunit peptide structure.

Material and methods

Plant material, DNA extraction and sequencing

The list of plant material, DNA extraction protocol, and standard Illumina HiSeq sequencing protocol are the same as those given in Chapter three. Additional plant material was collected from herbaria and living collections (Table 1) and these were subjected to the same treatment with respect to DNA extraction and sequencing. For the sake of convenience, throughout the text four letter acronyms for each accession will be used. The corresponding species names and their acronyms can be found in Table 1.

TABLE 1 | Plant materials used in this study, along with herbarium voucher information.

Species	Herbarium Voucher accession	Institute[1]	Acronym used in text
P. acetosum	1243	NHM	ACET
P. acraeum	1975	STEU	ACRA
P. alchemilloides	1885	STEU	ALCH2x
P. alchemilloides	1882	STEU	ALCH4x
P. articulatum	1972055	WAG	ARTI
P. barklyi	1972061	WAG	BARK
P. frutetorum	1972062	WAG	FRUT
P. inquinans	0682	STEU	INQU
P. multibracteatum	2902	STEU	MULT
P. peltatum	1890	STEU	PELT
P. quinquelobatum	1972049	WAG	QUIN
P. ranuncolophyllum	A3651	MSUN(*)	RANU
P. tongaense	3074	STEU	TONG
P. zonale	1896	STEU	ZONA
P. elongatum	0854	STEU	ELON
P. aridum	1972053	WAG	ARID
P. insularis	19990489	RBGE	INSU
P. yemenense sp. nov	1972037	WAG	YEME
P. omanense sp. nov	2184	RBGE	OMAN
P. somalense	V-067490	V	SOMA

STEU = Stellenbosch University, RSA; AL = Albers/MSUN = Münster & *Bakker et al. 2004. WAG = National Herbarium of the Netherlands. V = Uppsala herbarium.

Establishment of four F_1 crossing series

The method for generating and establishing F_1 hybrids were described in Breman et al. (2020) for the domesticated *P.* x *hortorum* crossing series. We have generated three additional crossing series. We selected three wild species representing phylogenetic diversity across the *Ciconium* clade (van de Kerke et al. (2019) and Breman et al. (submitted)) with which we crossed allavailable section members (Table 1). The three species used were: *P. barklyi* ('BARK'), *P. multibracteatum* ('MULT') and *P. acetosum*, ('ACET') and a visual overview of the total of four crossing series is given in Fig. 1. Using this phylogenetically diverse crossing approach we aimed to optimize our analysis of segregation patterns in offspring affected by CNI. Several crosses were also analyzed in order to examine additional chlorosis patterns (listed in tables in supplementary materials 3 and 4). We used embryo rescue (ER) of all F_1s to maximize the number of offspring for evaluation of chlorosis phenotypes (thus eliminating hybrid incompatibilities caused by failures of endosperm development).

FIGURE 1 | Experimental setup for the four comprehensive *Ciconium* crossing experiments. The four species indicated at the left were selected as mother plants. The *P.* x *hortorum* series is from Breman et al. 2020. The paternal accessions are listed to the right and their respective floral and leaf phenotypes are shown below the species names. The paternal accessions are listed based on phylogenetic distances relative to *P.* x *hortorum*.

5

Categories of CNI

We distinguished between pCNI and mCNI. The former manifests itself by the appearance of chlorosis (with varying levels of virescence) in hybrid offspring, whereas mCNI causes (partial) male sterility and/or dwarf growth. Care must be taken as sterility can also be the result of nuclear genomic incompatibility. Nevertheless, we found this distinction useful as it provides a way of estimating the level of mCNI, especially in crosses between closely related taxa. In the case of uniparental inheritance of organelles, the correlation of phenotype with genotype is straightforward. Given the known F_1 nuclear genomic background, a direct connection can be established between the observed pCNI phenotype and the chloroplast genotype responsible. In the case of variegated offspring, which we suspected contained chloroplasts inherited from both parent s, we tested, when possible, green or white parts of leaves separately. Phenotypes reflecting severity of mCNI and pCNI are listed in Table 2. For pCNI, we distinguished the following syndromes: I) 'green' -no virescence or chlorosis observed; II) 'near green' -virescence occurs under extreme physiological conditions; III) 'mildly chlorotic' plants were always chlorotic, but never lethal; IV) 'severely chlorotic' – plants were always chlorotic and easily turned yellow or lethal; and V) 'lethal' – plants were always

yellow or white. Mitochondrially induced CNI phenotypes were categorized as follows: I) fertility is equal to or exceeds wildtype fertility, no dwarf growth; II) plants are partially male fertile, spontaneous fertilization takes place occasionally, no dwarf growth; III) plants are always male sterile and/or dwarf growth may occur; IV) plants are always male sterile, stamens poorly developed, flowering may be impaired and dwarf growth may occur; V) no flowering observable, presumably completely sterile.

Flow cytometry

Flow cytometry of parental and hybrid material was performed as described in Breman et al. 2020 with the same reference material (*P.* x *hortorum* PEZ-BD8517) as used in that study. We refer to this paper for details of ploidy and genome size determination in our plants.

Mitome and plastome assembly

Mitome and plastome assemblies were made using the GetOrganelle pipeline (Jin et al. 2019) using default settings, except for the assumed insert size which we set to 350. Contigs were visualized and assessed using BandAge (see supplementary material 2) (Wick et al. 2015) and final contigs were concatenated using MEGA7 (Kumar et al. 2016) and aligned using MAFFT (Katoh et al. 2019). For the results of the plastome assemblies, we refer to Chapter four. For the mitome, as a first step an overall assembly was performed to recover fragments with sufficient base coverage (>10). When insufficient coverage (<10) for a given fragment was obtained, custom references were used, such as target exons obtained from our own previous runs on other accessions. This was necessary because of the high substitution rates (Bakker et al. 2006), differences in presence and absence of introns (Bakker et al. 2006, Choi et al. 2021) and low-frequency presences of mitome sequences in the read libraries. The first two reasons probably resulted in a poor match between target and reference and the low presence of mitome reads decreased the chance of finding reads that fell directly into the conserved parts of the targets. Therefore, these custom references were added to the GetOrganelle pipeline, and the assembly was re-attempted. For the resulting fragments, reading frames and overall identity, were cross-checked with published mitochondrial genes available. We used for *Ciconium* species (mainly *P.* x *hortorum*) if possible, but otherwise with *P. citronellum* fragments (all from Grewe et al. 2016) were used.

TABLE 2 | Categories of mCNI and pCNI.

	pCNI in F1 offspring		mCNI in F1 offspring
I	Green	I	male fertile
II	Green Vir	II	partial male sterile/reduced fertility
III	mildly affected; never fully green, but otherwise fine	III	male sterile and/or dwarf growth
IV	severely affected, always light green to yellow	IV	male sterile and/or impaired flowering and/or dwarf growth
V	Lethal, white	V	no flowering observable, presumably completely sterile

If the fragment still had a coverage between 5x and 10x it was included, but labelled as 'moderately reliable' (supplementary material 1). A targeted assembly of individual fragments often resulted in 'inflated' assemblies, whereby other, non-target regions were also assembled. The 'inflated' assemblies often contained large stretches of plastome sequences as well. This may be caused by so-called *mtpt*'s (Loyd et al. 2012), i.e., plastomes sequences having been transferred to the mitome, and in case of rDNA: because rDNA is partly similar between both organelles, which may have led to these parts being hybrid assemblies. Also, the search for overlapping reads then led to the assembly 'spilling over' into plastome assembly. We empirically determined the optimal number of assembly rounds needed for reliable recovery of a single target without the resulting inflation. We then used the output from the series with optimized number of assembly rounds as basis for final targeted assembly. The full results for each fragment and for each accession are displayed in the supplementary material 1. We extracted the following seven mitochondrial genes: *atp1*, *COX2*, *COX3*, *cytb*, *nad1-exon 1*, *nad5* and *nad7*.

Organellar genotyping using PCR and KASP markers

We used diagnostic PCR to genotype the inherited chloroplasts following the methods and primers developed by Breman et al. 2020. For mito-type detection we used Competitive Allele Specific PCR (currently named 'Kompetitive Allele Specific PCR', or KASP™) from Kbioscience or LGC Genomics (http://www.lgcgenomics.com). KASP was proven to be a reliable and fast technique for genotyping material and is now considered a benchmark technique for SNP calling (Hiremath et al. 2012, Semagn et al. 2014, Islam et al. 2015, Ka et al. 2018). KASP is a qPCR-based assay designed to detect SNP variants by using two forward primers (containing the SNP) and one reverse primer. Our F_1 and F_2 progeny had two parents that differed by SNPs at various positions. The mother of each crossing series was considered the 'target' and the paternal parent can be any other species from section *Ciconium* and is therefore referred to as 'Ciconium' in our assays. KASP primers were designed on (partial) assemblies of mitome exons as outlined above. Using introns for marker development is standard, as these are generally more variable yet close to conserved exon sequences. But, in the case for *Pelargonium*, mitome introns have been found to be absent (Bakker et al. 2000, Grewe et al. 2016, Choi et al. 2021). We therefore focused on exons, which have been shown to have elevated rates of substitution in *Pelargonium* (Bakker et al. 2000; 2006; Mower et al. 2007, Choi et al. 2021) and therefore may be good candidates for SNP marker development.

The following exons were used as a source: *cox2*, *cox3*, *cytb*, *NAD1*-exon 1, *nad7*, *NAD5* and *atp1*. We selected SNPs unique to a target, if possible, to be able to distinguish it from the other Ciconium species. Our KASP thermo profile was as follows: 5 minutes denaturation at 94°C, 10 cycles of 94°C 20s, 61-55°C annealing and extension for 60s (dropping 0.6°C for the annealing temperature each cycle), followed by 30 cycles of 94°C and 55°C extension for final amplification. Both F-primers (containing the SNP) and R-primers were added at 10 µM concentration. These were mixed with the KASP 2x master mix (LGC Genomics. (2013) and mQ water. Template DNA concentrations must be in the range of 0.5-10 ng/µL and these

need to be added to the mix per each reaction. For KASP we used two positive control samples which contained only the maternal and paternal genotype, and non-template controls (NTC') to be able to discriminate fluorescence signal caused by primer interactions from those with the templates. As stated above, KASP is a qPCR based technique that can determine the relative concentration of each genotype as well. In order to test if there are differences in concentrations between alleles, we determined the DNA concentrations using SYBR green (Zipper et al. 2004) and mixed these with extracts of target samples to obtain series of known concentration ratios(10:90, 25:75, 50:50, 75:25 & 90:10). This allowed us to determine if there were different ratios of mitochondria of either parent present in the hybrid accessions and thus enable quantification of inherited mitochondria. The full results are visualized in Fig. 5 as allelic discrimination plots.

Ciconium rpoB variation

rpoB was shown to be under positive selection in *Ciconium* (Chapter 4) and given its regulation by the nucleus, it may be the most promising gene to explain chlorosis. Thus, we assessed amino acid sequence variation in *rpoB*, recognizing a series of *rpoB* types, based on both length and aa substitution variation. We describe various physico-chemical peptide properties (Table 6) for the entire *rpoB* aa sequence as well as for 100 aa sequence (aa position 200-300, corresponding to base positions 600-900 in the *rpoB* alignment spanning a variable *rpoB* indel region (Fig. 2) referred to here as the '*rpoB*-vind' region. We also characterized the *rpoB*-vind variation by its aa content (Table 3). We used the R-package 'Peptides' (Osorio et al. 2015) to calculate the weight (Da), the grand average of hydrophobicity (GRAVY) index, the aliphatic index ('aI'), the iso-electric point of zero charge ('IEPoZC') (pH) and the net charge (C at pH 7) for each aa-sequence in over the full-length exon as well as for *rpoB*-vind. For the polarity we used the R-package 'Alakazam' (Gupta et al. 2015). To characterize the *rpoB*-vind types, we created heatmaps of the polarity and charge distribution over the *rpoB*-vind region (Figs. 6A & B).

We also modelled the structure of one subunit of PEP from *P. acetosum* using homology assessment (Waterhouse et al. 2018) as implemented in the Swiss-Model (Bienert et al. 2017) (Fig. 7). The *P. acetosum* sequence was selected because it contains the largest deletion and the highest number of aa changes in *rpoB*-vind region when compared to the other accessions (Breman et al. in prep and see, Fig. 1). Structure and distances were assessed using the models (ProMod3) (Bertoni et al. 2017, Studer et al. 2020, 2021). To assess the potential interactions of the *rpoB*-vind with nuclear genomic factors, we also assessed the structure and placement of *rpoB*-vind in the PEP subunit protein relative to the other operons which make up PEP (alpha, beta' and beta") using the 'compare' options of the Swiss-Model and the bacterial homologue for PEP as published and annotated by Yang et al. 2012. (Figs. 7 A & B). When discussing the sequence or operon, we refer to the sequence with Latin letters and the operon using Greek letters: *rpoA* is the gene encoding operon rpoα; *rpoB* is the gene encoding operon rpoβ; *rpoC1* is the gene encoding operon rpoβ' and *rpoC2* is the gene encoding operon rpoβ").

Results

Confirmed F_1 hybrids

We obtained, over four crossing seasons, a total of 30 verified F_1 interspecific hybrids (see Table 5) crossing *P. acetosum*, *P. barklyi* and *P. multibracteatum* with all other *Ciconium* species. In addition, there are crossings available from our previous study (Breman et al. 2020) involving *P.* x *hortorum* as well as the genotypes for each offspring from that study. Using embryo rescue, we obtained offspring for nearly all interspecific crosses for which fruit and seed set was observed. All displayed various modes of plastid inheritance (paternal, maternal and biparental) (Table 5). In most cases at least some individuals of an offspring died quickly after germination or transplantation to the greenhouse. This was especially the case for MULT x ZONA, MULT x ACET, ACET x ZONA and BARK x QUIN from which a maximum of one, but usually no plants, survived transplantation to the greenhouse.

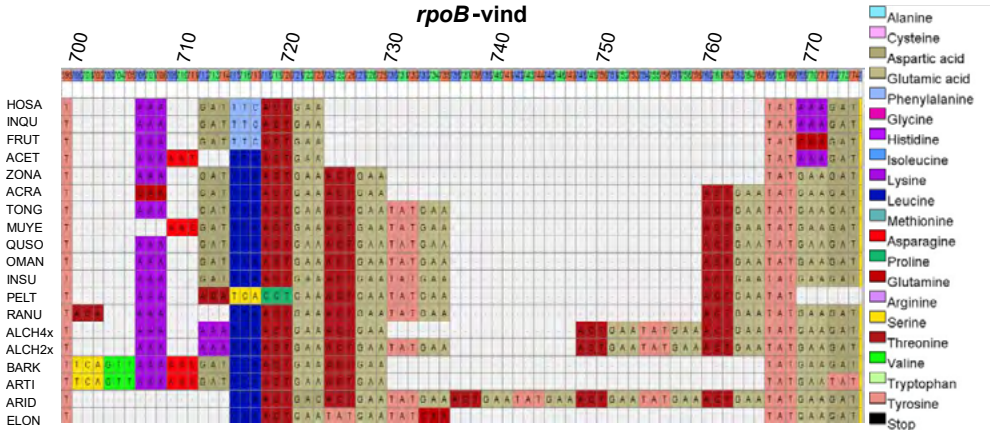


FIGURE 2 | Nucleotide sequence alignment of *Ciconium rpoB*-vind region. Color codes for amino acids listed at the right of the figure. The *rpoB*-vind region is flanked by non-length variable regions (not shown). The top bar lists aa codon positions (blue for first, green for second and red for third). The numbers show the base positions in the *rpoB* alignment.

For the *P. acetosum* series, we obtained six F_1 interspecific hybrids (see Table 5). Two were from reciprocal crosses (MULT x ACET and PELT x ACET) and these were chlorotic and sterile. Progeny of one cross was always lethal, as the plants never flowered (ACET x ZONA) and three yielded plants that were green or near green and were partially male-fertile (ACET x TONG, ACET, x FRUT and ACET x INQU).

For the *P. barklyi* series, we obtained four verified F_1 plants (F_1 BARK x FRUT, F_1 BARK x MULT, F_1 and F_1 BARK x QUIN). For one, we were unable to verify the status visually based on phenotype (BARK x INQU). Four more were obtained from reciprocal crosses, (F_1 FRUT x BARK, F_1 MULT x BARK, F_1 HORT x BARK, F_1 ALCH4x x BARK). Of these, three were chlorotic

and infertile (BARK x MULT, BARK x QUIN and ALCH4x x BARK), one was severely chlorotic, they flowered only once, and the plants were dwarfs (BARK x FRUT). Two of the plants were lethal and never flowered (BARK x INQU, BARK x HORT).

TABLE 3 | AA inserts for the rpoBI-vind region.

Accession	Type designation	Overall aa code
HORT	A-42dα	--K-DF(TE)(YK)(DS)
INQU	A-42dα	--K-DF(TE)(YK)(DS)
SALM	A-42dα	--K-DF(TE)(YK)(DS)
FRUT	A-36dα	--K-DF(TE)(YQ)(DS)
ACET	B-42dα	--KN-L(TE)(YK)(DS)
ZONA	C-36dα	--K-DL(TE)^2(YE)(DS)
ACRA	D-30d	--Q-DL(TE)^3(YE)(DS)
TONG	C-24dα	--K-DL(TE)^2(YE)(TE)(YE)(DS)
MULT	E-24dα	---N-L(TE)^2(YE)(TE)(YE)(DS)
YEME	E-24dα	---N-L(TE)^2(YE)(TE)(YE)(DS)
RANU	F-24dβ	T-K--L(PE)(TE)(YE)(TE)(Y-)
ALCH2x	G-12dβ	--K-KL(TE)^2(YE)(TEYE)2(DS)
ALCH4x	G-12dα	--K-KL(TE)^2(TEYE)^2(DS)
PELT	H-24dβ	--K-TS(PE)(TE)(YE)(TE)(Y-)
QUIN	C-24dα	--K-DL(TE)^2(YE)(TE)(YE)(DS)
INSU	C-24dα	--K-DL(TE)^2(YE)(TE)(YE)(DS)
SOMA	C-24dα	--K-DL(TE)^2(YE)(TE)(YE)(DS)
OMAN	C-24dα	--K-DL(TE)^2(YE)(TE)(YE)(DS)
ARTI	I-36dβ	SVKNDL(TE)^2(YE)(YS)
BARK	I-36dα	SVKNDL(TE)^2(YE)(DS)
ARID	J-0d	------(TE)^2(YE)(TEYE)^3(DS)
ELON	J-24dγ	------(TE)(YE)(YQ)(YE)(DS)

A -J describe types based in first 18bp of *rpoB*-vind. Number+'d' refers to size of deletion, α or β then refer to aa motif contained in second part of the *rpoB*-vind region.

For the *P. multibracteatum* series, we obtained eight F_1 hybrids. Two of these were green or near green and partially fertile (MULT x ALCH, MULT x QUIN). Two of these were chlorotic (MULT x ARID, MULT x BARK) and infertile. Two were severely chlorotic (MULT x ACET and MULT x ZONA) and flowering was never observed. *Pelargonium multibracteatum* crosses with INQU, FRUT, and TONG did not yield a plant despite two full seasons of crossing attempts. *Pelargonium multibracteatum* crosses with ACET or ZONA rarely yielded progeny and when they did the plants nearly always carried the MULT plastid. We found one case where the F_1 MULT x ACET also carried the ACET plastid, alongside the MULT plastid.

From the other crosses we obtained a further 15 F_1s, using additional *Ciconium* species as parents. From these series, the crosses with *P. yemenense* sp. nov. or *P. alchemilloides* (4x) with other accessions (*P. barklyi* and *P. frutetorum*) stand out as thy were performed using

parental accessions with different ploidy levels, resulting in polyploid offspring (see Fig. 3/ supplementary material 2).

All chlorosis- and fertility-related phenotypes for the four crossing series are displayed in Figs. 4A-D and in Table 5. All individual plants, together with their chlorosis phenotype and plastid and mitochondrial genotypes, are listed in supplementary materials 3 and 4 (see below).

Flow cytometry

The results of flow cytometry are listed in supplementary material 2. We were unable to obtain flow cytometry for all verified hybrids because some of these never made it past the embryo rescue (ER) stage or died quickly after germination or transplantation to the greenhouse, which left us with insufficient material to perform the analysis on. Flow cytometry results show that F_1 hybrids have a value that is intermediate between the two parents. The most striking result is the frequency of polyploids, especially for the MULT-series (Fig. 3) and the other crosses.

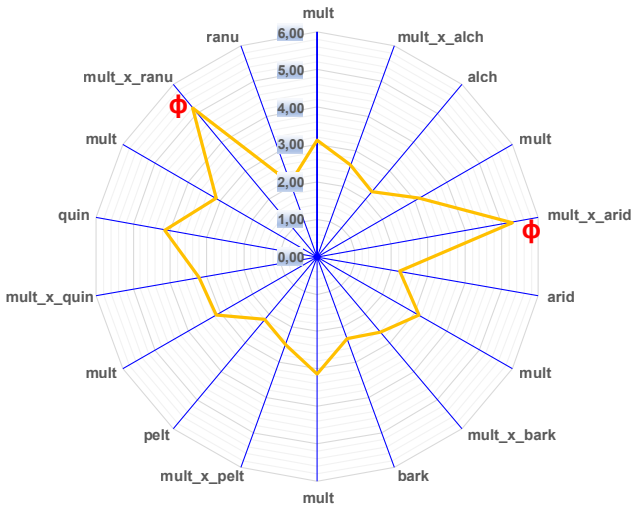


FIGURE 3 | Spiderweb diagram displaying Cx-values for the F_1 *Pelargonium multibracteatum* crossing series showing parental and F_1 hybrid genome sizes. The vertical axis displays the obtained Cx-values. The polyploids are indicated with the ϕ symbol. Species and accessions are more or less phylogenetically arranged. Two parental species flank each F_1 hybrid (i.e., a *P. multibracteatum* parent crossed with another species from *Ciconium*).

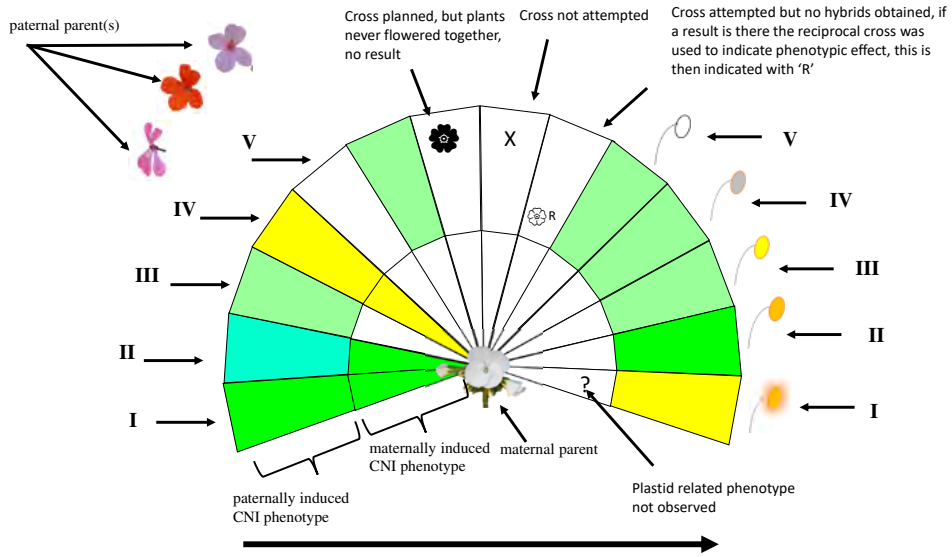
Mitome assembly

We were able to assemble mitome fragments with a base coverage >10 for most accessions (see OMS1). For some accessions, markers were assembled with a base coverage of 5-10 (OMS1). The moderately reliable sequences were incorporated for practical and comparative purposes, but not used for primer design. Our final dataset, only counting exons, was 4,353 bp long and the assembly lengths for each fragment were: *COX2*: 687 bp; *COX3*: 1028 bp; *cytb*: 487 bp; *nad1 exon 1*: 311 bp; *nad5*: 502 bp; *nad7*: 582 bp and *atp1*: 756 bp. Length differences were observed in *nad1 exon 1* and *nad5*. For *P. yemenense*, the *cytb* sequence remained incomplete despite repeated attempts at assembly.

Plastotyping and Mitotyping

Our genotyping of plastids (plasto-typing) demonstrated that plastomes can be inherited from either parent. We found maternal, paternal and biparental plants across all F1 offspring (Table 5) and clear differences in some cases with respect to plastid type inherited and resulting chlorosis in the offspring. The mito-typing experiments showed a different picture as can be observed from the allelic discrimination plots (Fig. 5). Mito-typing offspring identified plants with predominantly maternal mitotypes, as these plants, with a non-target *Ciconium* mother and a target *Ciconium* father (e.g. FRUT x BARK) nearly always ended up in quadrant I (Fig. 5 A-C) and those with the target as mother and non-target *Ciconium* as father in quadrant IV (Fig. 5 A-C). One population (F_1 HORT x QUIN) displayed biparental inheritance with plants displaying either the QUIN (paternal) or HORT (maternal) genotype. One accession, (F_1 of TONG x ACET), has a KASP reading close to the non-target *Ciconium*: target *acetosum* ratio of 25:75 (indicated by a star in Fig. 5B). In this single case there is more of the paternal type, but the maternal type is also present. Three F_1 samples show a strictly paternal genotype: These are F_1 BARK x QUIN, F_1 ACET x ZONA and F_1 BARK x INQU. Two F_2 plants of HORT x ACET also displayed a paternal (ACET) genotype (see supplementary material 4).

The *P. multibracteatum* crossing series showed ambiguous results. All F_1 accessions were plasto-typed as containing a *Ciconium* plastid, even the ones with the MULT mother. The positive MULT control did yield a 'MULT' call (this was the actual sample on which the primers were designed). The reciprocal comparison of the results for the other markers with the MULT marker were contradictory (supplementary material 4). The subsequent reciprocal comparison between the other markers (ACET, BARK and HORT) clearly showed that these three confirmed each other, leaving us with the MULT markers' ambiguous results. If for instance the BARK KASP assay would show that BARK was not present in the hybrid, the MULT KASP assay would indicate it was. To assess performance issues of the MULT KASP marker, we re-analyzed the *P. multibracteatum* series with a second KASP marker and we included another *P. multibracteatum* accession This again resulted in a near universal '*Ciconium*' call and this time the positive control also resulted in a '*Ciconium*' call. We therefore conclude that we have no reliable results for the mito-typing of the *P. multibracteatum* series, possibly caused by innate heteroplasmy of *P. multibracteatum* (see the discussion). The full overview of all reactions and raw data fluorophore measurements can be found in the supplementary materials 3 and 4.



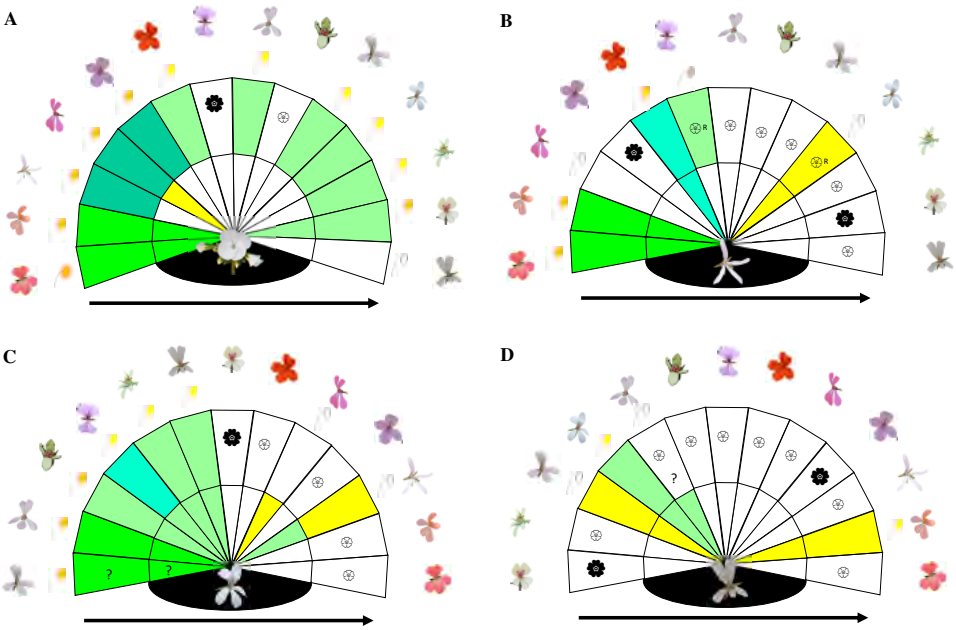


FIGURE 4 | Fan diagrams with plastid related phenotypes obtained for F_1 plants of four crossing series. A) the HORT series, B) the ACET series, C) the BARK series and D) the MULT series. The legend is listed at the top in an explanatory figure (with a fictional example). {R indicates cross attempted, but no hybrids obtained, if a result is there, the reciprocal cross was used (when this did yield a plant), to indicate phenotypic effect. This is then indicated with 'R'. | Cross planned, but plants never flowered together, no result. '?' indicates plastid related phenotype not observed. Flower shapes relate to each species as indicated in Figure 1. Roman numerals indicate the fertility/chlorosis class as indicated in Table 2. The arrow below the figure indicates phylogenetic distance between paternal and maternal plants (van de Kerke et al. 2019).

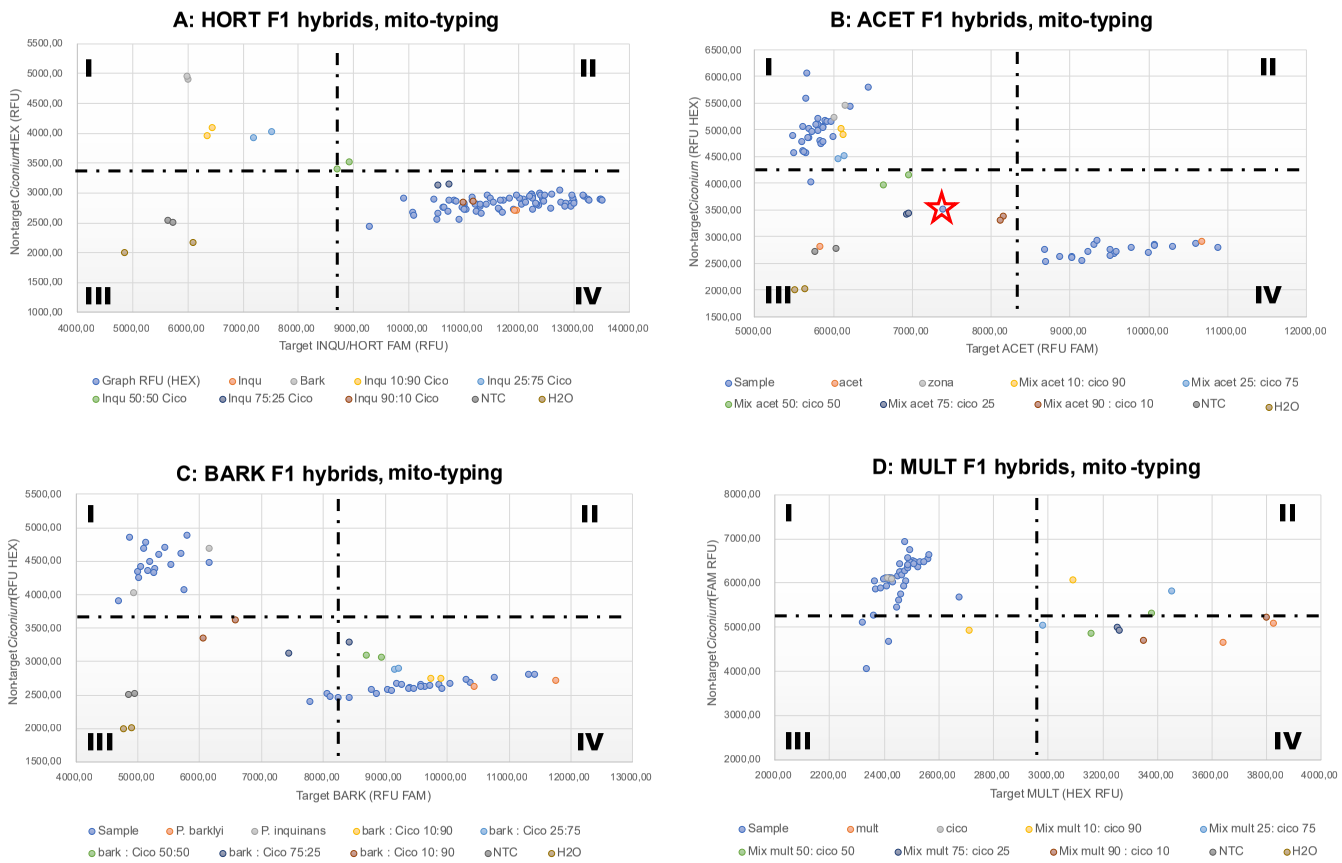


FIGURE. 5 | KASP allelic discrimination plots of *Ciconium* F_1 mitochondrial genotypes. A) HORT, B) ACET, C) BARK and D) MULT F_1 hybrid samples. In this figure we divided each graph into four quadrants (I-IV). In quadrant I the non-target *Ciconium* genotypes are located because the y-axis displays the values for the fluorescent labels assigned to the non-target *Ciconium* genotypes and quadrant IV holds the target genotypes and the target fluorescence values are displayed on the x-axis. Quadrant III either contains the control samples or samples that displayed neither genotype. The star in 5B indicates the biparental, F_1 TONG x ACET progeny.

TABLE 4 | In bold green are the sequence tails. The discriminating SNP is at the end of each primer in bold and blue.

Set name	Name	Sequence	Target(s)	Fluorescent dye	gene
MUYE	MultYeme_KU_I_F1	**GAAGGTGACCAAGTTCATGCTC**ATGAACAACCCTTTGTGGTAGACT**A**	*Ciconium*	FAM	*COX2*
	MultYeme_KU_I_F2	**GAAGGTCGGAGTCAACGGATT**ATGAACAACCCTTTGTGGTAGACT**C**	MULT/YEME	HEX	
	MultYeme_KU_I_R1	GTCTTAGCTCGTGCTTTGTGGCATT			
ACET	Acet_KU_I_F1	**GAAGGTGACCAAGTTCATGCTC**ATAAGTCCAATACCACTGATG**C**	ACET	FAM	*COX2*
	Acet_KU_I_F2	**GAAGGTCGGAGTCAACGGATTACTCAT**AAGTCCAATACCACTGATG**A**	*Ciconium*	HEX	
	Acet_KU_I_R1	GAGATAGTGGTCAATCCTGCAATTACTAT			
BARK	Bark_KU_I_F1	**GAAGGTGACCAAGTTCATGCT**GGAGATAAGAATGTTAGAACAAGTAACA**G**	BARK	FAM	*NAD7*
	Bark_KU_I_F2	**GAAGGTCGGAGTCAACGGATTA**GGAGATAAGAATGTTAGAACAAGTAACA**T**	*Ciconium*	HEX	
	Bark_KU_I_R1	GGGAAGTCTCTCGGTTCTTTTCACAT			
INQU	HoInSa_dnp7654_A1	**GAAGGTGACCAAGTTCATGCT**AATACAAATAAGAGTTGTTCCATGAATTTGAA**T**	INQU, HORT, SALM , *	FAM	*NAD5*
	HoInSa_dnp7654_A2	**GAAGGTCGGAGTCAACGGATTACA**AATAAGAGTTGTTCCATGAATTTGAA**G**	*Ciconium*	HEX	
	HoInSa_dnp7654_C1	GCATCTCTACCAAGCGGCTCCTT			

TABLE 5 | Summary of genotypes and phenotypes from F_1 interspecific crosses from *Ciconium*. 'small' or 'large' indicates F_1 smaller/larger throughout life than either parent. Phenotype comments: C) chlorotic, G) green, L) lethal seedlings, V) variegation occurred, +) plant never made it past ER. * Flowers only at temperature interval of ~20-25°C, ° indicates that fertility is temperature dependent ~20-25°C, ^ indicates that not all individuals of these crosses displayed spontaneous seed set, p: indicates fertility only for polyploid, estimated level indicated. For a visual overview of the cytonuclear interactions please see Fig. 5.

hybrid	count	n/o plants tested	mCNI class	pCNI class maternal	pCNI class paternal	count of F_1 plastid types per plant maternal / paternal / biparental	chlorotic phenotypes	other phenotypic remarks
ACET x FRUT	3	3	II	I	I	0/1/2	C,G,V	
ACET x INQU	5	1	II	-	I	0/1/0	G	
ACET x ZONA	>20	9	V	-	V	0/7/2	C,V,+	
BARK x FRUT	4	4	IV	IV	-	4/0/0	C,G	small, weak plant
BARK x INQU	2	2	V	?	?	0/2/0	G,+	died after seedling phase
BARK x MULT	10	3	III	III	III	0/2/1	V,C	small plants
BARK x QUIN	2	2	V	IV	IV	1/0/1	C,V,L,+	plantlet only
MULT x ACET	2	2	V	IV	IV	1/0/2	C,V	small, weak plant
MULT x ALCH	>20	13	II°^	I	I	8/3/3	C,G,V	often large
MULT x ARID	>20	3	III	I	-	3/0/0	C,G,V	one dwarf plant
MULT x BARK	10	6	III	III	III	6/0/1	C,V	small plants
MULT x PELT	2	2	III	III	II	0/0/2	C,V	
MULT x QUIN	10	6	II°^	?	?	-/-/-	G	large
MULT x RANU	10	9	II°^	I	I	1/7/1	C,G,V	large
MULT x ZONA	1	1	V	IV	-	1/0/0	C	small, weak plant
ALCH4X x BARK	2	2	IV	II	-	2/0/0	G,V	
ALCH4X x FRUT	2	2	IV	II	III	1/0/0	C	small
ALCH4X x YEME	3	3	III-II^{p6}	I	I	2/1/0	C,G,V	
ARID x QUIN	>20	3	IV*	II	-	3/0/0	C,G	often large plants
FRUT x ACET	5	1	I-II	I	I	0/0/1	G	fertile pollen
FRUT x BARK	6	6	IV*	III	III	5/0/1	C	small, weak plant
INQU x TONG	3	1	V	III	-	1/0/0	C	
PELT x ACET	4	4	IV	II	II	2/0/2	C,V	small
PELT x ALCH	2	2	V	I	-	2/0/0	G	very small
PELT x QUIN	1	1	V	-	III	0/1/0	C	small, weak plant
QUIN x ARID	10	3	IV*	II	-	3/0/0	C,G	reciprocals equal
TONG x ACET	7	7	III	II	II	6/0/1	G	
YEME x ALCH4X	>20	1	I-II	I	-	1/0/0	G	
ZONA x MULT	1	1	V	III	III	0/0/1	C	small, weak plant
ZONA x QUIN	15	6	IV*	II	?	6/0/0	C,G	often large

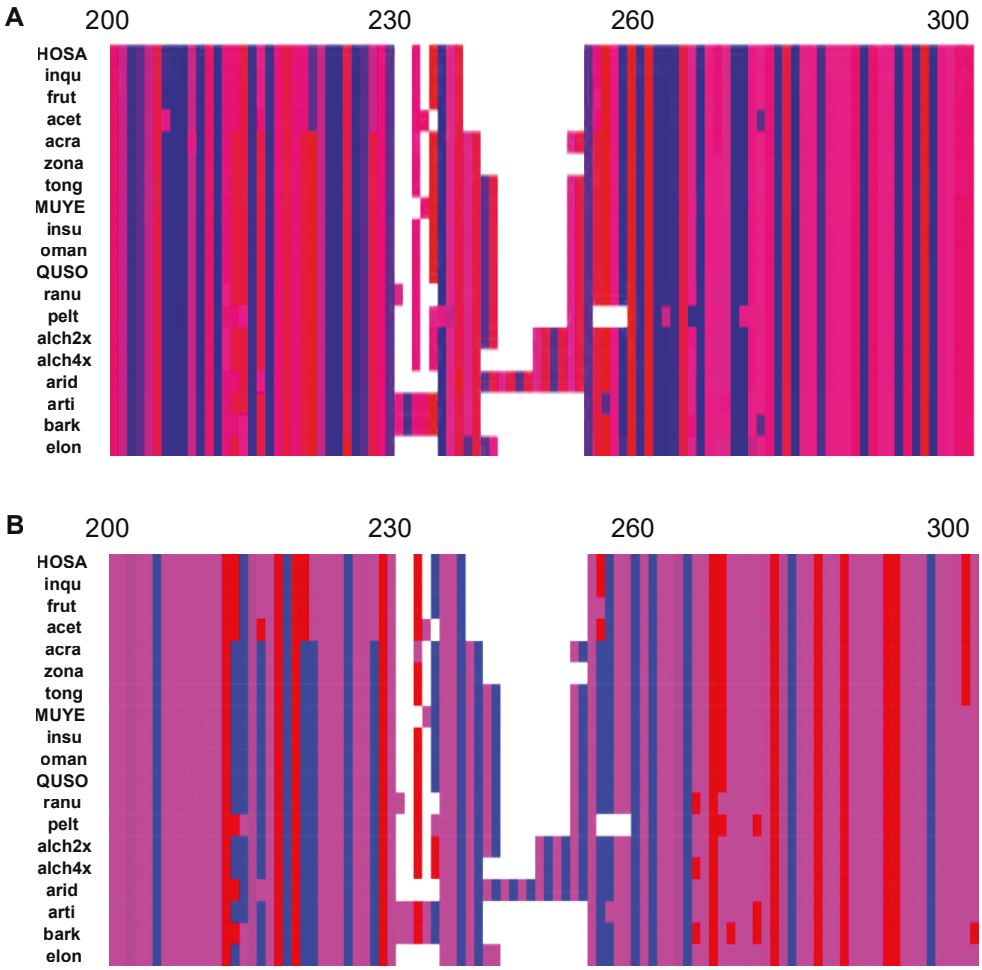


FIGURE 6 | Polarity and charge heatmaps of the *rpoB*-vind region. A) heatmap showing the polarity of all accessions over the *rpoB*-vind region. Red indicates aa residues with high polarity, blue with low polarity and pink neutral. The white area represents the gaps in the *Ciconium rpoB*-vind region in both figures. The numbers correspond to the aa position in the overall rpoB aa alignment. B) heatmap showing the charge of all accessions over the *rpoB*-vind region. Red denotes positive charge, blue negative charge, pink are aa with neutral charge.

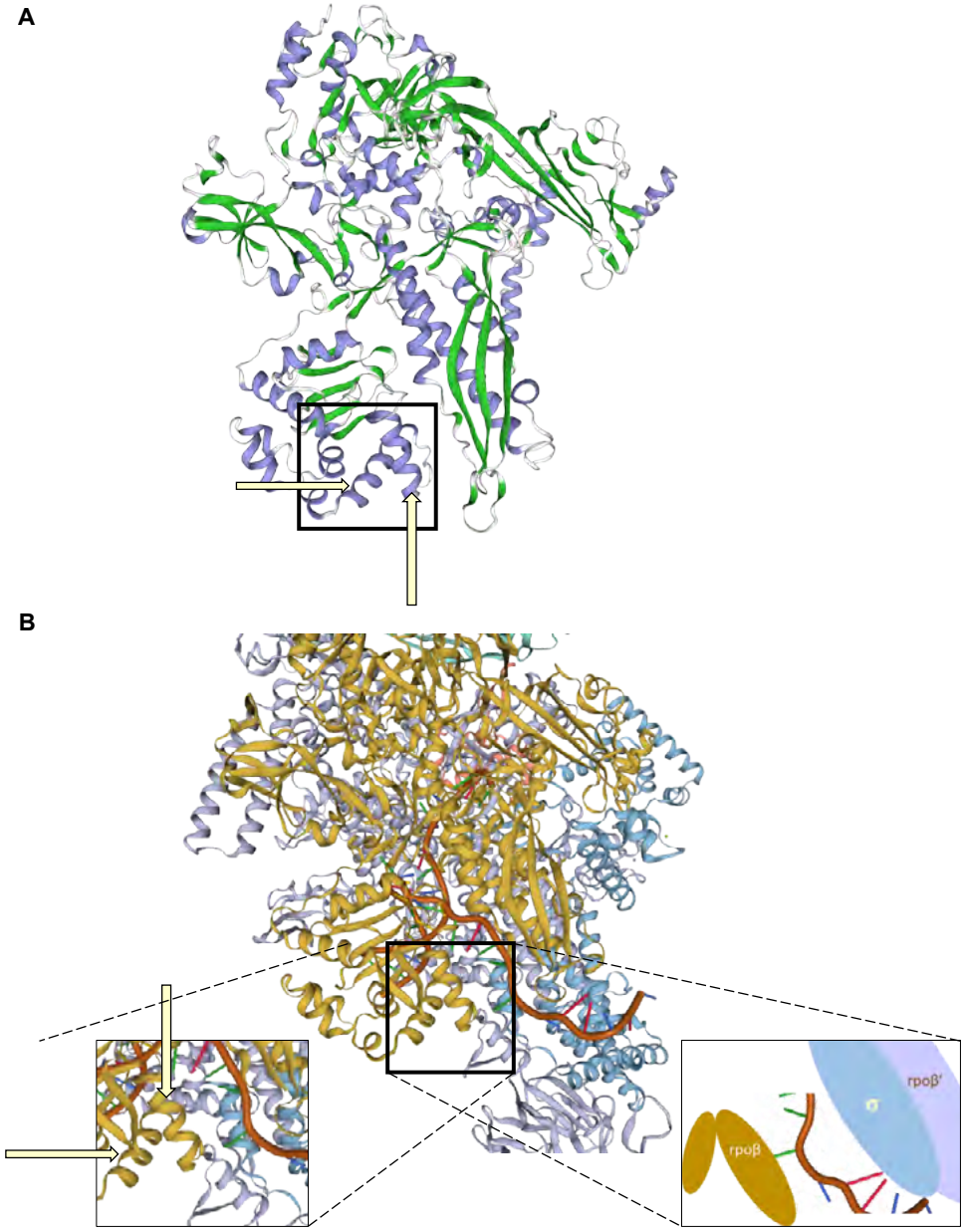


FIGURE 7 | structural models for *P. acetosum* rpoB and the entire PEP operon for *Thermus thermophilus*. A) structure of rpoB for *P. acetosum*. Green denotes beta sheets, blue the alpha helices, white areas are unordered helical structures. The arrows indicate the *rpoB*-vind region. B) model of the Crystal structure of *Thermus thermophilus* transcription initiation complex. Color legend: teal: rpo α; gold: rpo β; Violet: rpo β'; red: rpo β''; light blue: RNA polymerase σ- factor; brown: template ds DNA. Arrows denote the homologous region in *T. thermophilus* to the rpoB-vind region in *Ciconium*. The Left inset zooms in on the *rpoB*-vind region and the interactions with a sigma factor and template DNA. The right inset shows a schematic illustration of the area highlighted in the left inset.

Ciconium rpoB variation

The aa content of *rpoB*-vind is displayed in Table 3. The most striking feature is the occurrence of insertions of threonine ('T'), aspartic acid ('E') and tyrosine ('Y) in various combinations. These 'TYE' motive insertions are the largest in *P. aridum* (14 aa's in total) and are absent in *P. inquinans*, *P. acetosum*, *P.* x *hortorum*, *P.* x *salmoneum* and *P. frutetorum*. Of the seven physico-chemical properties assessed across 19 *rpoB*-vind types (Table 6) iso electric point of zero charge '('IEPoZC') and the 'net-charge at pH 7', show the most striking differences. The highest net charge was for *P. acetosum* with a value of 21.5 C and the lowest is for *P. aridum* (4.5 C), a difference of 17.0 C. The (corresponding) iso-electric points differ the most for these two accessions, as well with highest pH for *P. acetosum* (pH 9.3) and the lowest for *P. aridum* (pH 8.0). When zooming in on the *rpoB*-vind region, the contrast is enhanced with a negative net charge for all accessions except for those of *P.* x *hortorum*, *P. inquinans*, *P. acetosum* and *P. salmoneum* (collectively referred to as the HORT group). The heatmaps of polarity distribution shows that within the 100 aa window spanning the *rpoB*-vind region for the HORT group three aa positions (218, 224 and 231) have 'flipped' polarity from high to low. *Pelargonium acetosum* contains two more positions that 'flipped'. These are: 207 (increased polarity) and 276 (from high to low). The charge distribution heatmap (Fig. 6B) shows that the accessions of the HORT group have five 'charge' flips (215, 223, 257 and 299, all from negative to positive) in the 100 aa window spanning the *rpoB*-vind region. Position 218 also shows this 'charge flip for *P. acetosum*, but for the other HORT group members the charge changed to neutral here.

The length differences between the *rpoB* sequences consist primarily of a threonine (T), Tyrosine (Y) and glutamic acid (E) ('TYE') inserts (Fig. 2). Other aa- are also present, especially in the first part of the *rpoB*-vind region of *P. articulatum* and *P. barklyi*. The 'TYE' inserts are virtually absent from the HORT group and *P. acetosum* but present in varying frequencies and combinations in the other accessions. The full overview of aa variation and length differences are presented in Fig. 2 and Table 3.

The model of the *rpoB* sequence for *P. acetosum* (Fig. 7A) yielded a Qmean score (indicating relatively high homology) of -2.74 which is considered 'good' (Benkert et al. 2011). As a template the *Thermus thermophilus* (template 4g7h.1.C) was selected as 'best match' by Swiss-Model (Zhang et al. 2012). Sequence identity was 37.86% and sequence coverage was from position 1063 aa positions (based on *T. thermophilus*) of the 1079 contained in the *P. acetosum rpoB* aa-sequence. This model shows that the variable *rpoB*-vind region is located on the outside of the operon and mainly contains alpha helices. Further analysis of the structure model of the entire bacterial PEP homologue (Fig. 7B) shows that *rpoB*-vind is not in contact with the other PEP operons; rather, it is in contact with the included sigma factor as well as the template DNA.

TABLE 6 | Physio-chemical properties of rpoB. A) Pelargonium section Ciconium rpoB and B) the rpoB-vind region including ~30 flanking aa positions.

Accession P.	length (aa)	weight (Da)	gravy	aliphatic index	Iso-electric point of zero charge (pH)	net charge (C) (at pH 7)	polarity sum
A *rpoB* peptide properties across the full length of the aa sequence							
HOSA	1079.000	122558.304	-0.278	91.536	9.197	20.624	
inquinans	1079.000	122516.223	-0.282	91.273	9.197	20.624	
frut	1079.000	122558.261	-0.277	91.536	9.140	19.625	
acet	1079.000	122602.228	-0.298	90.209	9.258	21.536	
acra	1083.000	123144.691	-0.299	91.445	8.429	9.639	
zona	1081.000	122795.291	-0.283	91.982	8.510	10.636	
tong	1085.000	123420.939	-0.293	91.891	8.165	6.641	
MUYE	1085.000	123330.699	-0.294	91.627	7.958	4.546	
oman	1085.000	123381.859	-0.288	91.627	8.253	7.544	
insu	1085.000	123333.771	-0.291	91.627	8.162	6.546	
QUSO	1085.000	123333.771	-0.291	91.627	8.162	6.546	
pelt	1081.000	122735.350	-0.293	90.827	9.010	17.538	
ranu	1085.000	123338.707	-0.287	91.364	8.064	5.548	
alch4x	1087.000	123610.067	-0.293	91.364	8.162	6.549	
alch2x	1089.000	123818.194	-0.293	91.364	7.957	4.549	
bark	1084.000	123014.678	-0.276	93.309	8.736	13.449	
arti	1081.000	122640.197	-0.267	92.691	8.587	11.632	
arid	1088.000	123766.202	-0.288	91.536	7.956	4.548	
elon	1077.000	122135.299	-0.284	91.536	8.378	8.601	
B *rpoB* peptide properties across 100 aa containing the *rpoB*-vind region							
HOSA	83.000	9941.094	-0.513	65.400	8.728	3.877	719.6
inqu	83.000	9941.094	-0.513	65.400	8.728	3.877	719.6
frut	83.000	9941.057	-0.509	65.400	8.475	2.877	718.8
acet	83.000	10011.136	-0.582	61.500	9.217	5.785	720.3
acra	87.000	10468.082	-0.703	67.300	4.424	-8.108	778.7
zona	85.000	10250.039	-0.645	68.300	4.625	-6.110	756.7
tong	89.000	10772.235	-0.735	68.300	4.456	-8.107	796.1
MUYE	89.000	10728.157	-0.748	72.200	4.187	-10.197	798.4
oman	89.000	10749.183	-0.734	68.300	4.271	-9.198	796.5
insu	89.000	10749.183	-0.734	68.300	4.271	-9.198	796.5
QUSO	89.000	10749.183	-0.734	68.300	4.271	-9.198	796.5
pelt	85.000	10137.014	-0.731	61.500	5.220	-2.204	757
ranu	89.000	10797.194	-0.729	64.400	4.336	-8.198	791.3
alch4x	91.000	11054.332	-0.803	64.400	4.369	-8.196	814.9
alch2x	93.000	11277.368	-0.814	64.400	4.231	-10.195	832.1
bark	88.000	10546.308	-0.669	71.200	5.251	-1.297	779.4
arti	88.000	10531.140	-0.620	71.200	4.477	-6.202	777.4
arid	95.000	11442.400	-0.785	64.400	4.143	-11.198	845.8
elon	85.000	10191.917	-0.552	67.300	4.314	-7.205	742.7

Discussion

We have obtained over four crossing seasons a total of 30 verified F_1 interspecific hybrids, allowing us to study CNI in *Ciconium* in great detail. We show that biparental inheritance of chloroplasts occurs in *Ciconium* and that is not an artifact of hybridization (Tilney-Basset et al. 1984, 1989a, Breman et al. 2020). More remarkable is the rare biparental or paternal inheritance of mitochondria in *Pelargonium* section *Ciconium* (confirming earlier findings by Weihe et al. 2009 and Apitz et al. 2013). We also explored the highly variable *rpoB* gene and PEP protein structure. We hypothesize that this variability might explain CNI as measured by chlorosis in *Pelargonium* F_1 interspecific hybrids. Although we do not have conclusive evidence, we find that we cannot reject the hypothesis that *rpoB* type determines the level of chlorosis observed in F_1 offspring.

Organelles and speciation

Biparental inheritance

The exact causes of biparental inheritance of organelles remains unknown. But we conclude, based on our experiments, that plastids are more prone than mitochondria to inherit from either parent to the offspring. This can be explained by the fact that mCNI seems to result in more pronounced selection and lethality than pCNI. This could be true as plastids are undeveloped during and directly after fertilization and seed development, whereas the mitochondria are active during these phases. Thus, any mCNI effect would be stronger than pCNI at crucial early developmental stages. This would explain the high number of aborted embryos and empty seeds found on all our F_1 plants. Given that the mother plant is 'responsible' for supplying energy to the development of the seeds, it is logical that there is a strong maternal bias. However, plastids which are introduced to the embryo via pollen (Kuroiwa et al. 1992, 1993) and are sorted out, incompletely in *Pelargonium*, early in development (Kirk and Tilney-Bassett 1967, Weihe et al. 2009). Thus, they can be present in all tissue early in development (Guo and Hu 1995). mCNI hardly plays a role in the post-seedling phase of the plant as evidenced by our KASP genotyping of offspring whereby we found almost only maternally inherited mitome types. It is rather the pCNI that determines the further survival during the vegetative phase of life. Our results lend support for the idea that biparental inheritance of organelles could provide an escape from CNI (Barnard-Kubow et al. 2017). They also support the hypothesis that organellar changes, resulting in CNI, have a profound influence on speciation (Greiner et al. 2013, Barnard Kubow et al. 2016). Further support for these two hypotheses comes from the fact that second generation of plants segregate again for chlorosis with only one plastid type present, showing that selection for organelle management and expression genes acts immediately after the first generation of hybridization (Barnard Kubow et al. 2016, Breman et al. 2020).

Asymmetric inheritance due to lethal and non-lethal CNI

Chlorotic effects of different plastids on offspring are often asymmetric. Different chloroplast types induce different pCNI in crosses with equal nuclear genomic backgrounds (e.g. most HORT crosses are a prime example of this). The preference for one type, as well as preferentially backcrossing with one of the parents (introgression) after a historical hybridization event could explain the problematic position of taxa in phylogenetic trees due to conflict between plastid and nuclear genomic markers. For instance, the four-petalled Clade A species *P. nanum*, which is currently not assigned to a section (Röschenbleck et al. 2014), was suspected to be an ancient hybrid species because of the unique floral morphology and its 'single branch' status in current phylogenies (van de Kerke et al. 2019). It was sister to clade A2 (based on the clade-based systematics used by Bakker et al. 1999). Other cases can be seen in *P.* sect. *Hoarea* where the occurrence of non-monophyletic species has been attributed to 'chloroplast capture' (Bakker et al. 2005). Such taxa would have retained the chloroplast of one species, while displaying the morphology and nuclear genomic type of another. Further testing of such incongruencies could be done by using more markers from the nuclear genomes. For instance, the repeatome appears promising as a source of phylogenetic markers (Dodsworth et al. 2018, Breman et al. submitted) as it provides resolution at a low taxonomic level and provides a genome-wide overview represented by the most abundant parts of the non-coding DNA (repeats).

Naturally occurring hybrids in *Pelargonium* are rarely found (pers. comm. Powrie, Kirstenbosch RSA), but not unheard off (Knuth et al. 1912, van der Walt et al. 1990). This is logical given the reduced fitness characteristic of most hybrid offspring which will result in lower chances of surviving to the reproductive life stage, as is supported by results from our experiments. However, our study also shows that species are highly compatible as we obtained many (~30) interspecific crosses, some of which are fully green and fertile. We therefore, do not exclude that hybridization plays an additional, minor role in *Pelargonium* evolution (see above for *P. nanum*, and the allopolyploids *P. quercetorum, P. endlicherianum* and possibly *P. cayli* from Madagascar). Two cases of possible natural hybrids from section *Ciconium* are known. The first is an herbarium specimen of a wild hybrid between *P. peltatum* and *P. alchemilloides* at RBGE (M. Gibby pers. Comm.). The second case is *P.* x *salmoneum* (from our own collections). We have analyzed the plastome and mitome of *P.* x *salmoneum* and found that it carries the *P. inquinans* plastome and a mix of *P. inquinans* and *P. acetosum* mitochondrial genotypes (supplementary material 5), potentially due to mitome recombination (Apitz et al. 2013). The morphology of *P.* x *salmoneum* is intermediate between the hypothesized parental species. *Pelargonium* x *salmoneum* is a fully fertile, green plant which segregates for numerous traits such as plants size, flower shape, leaf shape indicating it is not a 'stable' species (yet), but a hybrid. We propose that *P.* x *salmoneum*, irrespective of whether it arose naturally or was the result of human crossing activities, is a genuine interspecific hybrid with equal fitness to either of its proposed parents and that it contains possible traces of mitochondrial recombination.

Ciconium rpoB variation and effects on chlorosis

We can explain the occurrence of chlorosis in F_1 hybrids by characterizing the variation in *rpoB*. There are striking differences in length of *rpoB*, largely caused the absence or presence of a 'TYE' aa motif in the *rpoB*-vind region. The structural consequences of the absence or presence of this insert are not completely clear, but there may be some effect on PEP-DNA interaction or interaction with nuclear-encoded σ-factors (as discussed below). Overall charge and polarity changes and absence of the 'TYE' motif in the HORT group rpoB-vind aa sequence correlates with the chlorotic effect of this plastid type in crosses with other *Ciconium* accessions. Variation in polarity and charge are less pronounced between non HORT-group crosses and observations of chlorosis support this (Fig. 4). The *P. barklyi* sequence varies more and it was shown to have a different d_N/d_S ratio (ω) from other accessions, indicating other selective pressures acting on *P. barklyi* evolution when compared to the rest of *Ciconium* (Chapter four). The unique sequence make-up and independent patterns of positive selection of *P. barklyi* may explain the occurrence of (severe) chlorosis (types IV and V) in most of our diploid crosses. The only *P. barklyi* cross where we observed less chlorosis (Type III) than in the other crosses was with *P. multibracteatum*. The difference between the aliphatic indexes (ΔaI for *rpoB*-vind MULT vs BARK=1, while e.g., ΔaI for the *rpoB*-vind region of MULT vs QUIN = 3.9) are much less pronounced between these two accessions even though *P. multibracteatum* and *P. quinquelobatum* are phylogenetically much closer related (van de Kerke et al. 2019). This suggests some mitigating effect of the aliphatic nature of the peptide (Tables 6A & B) on overall compatibility. This is in contrast with the net-charge difference at pH 7 or the iso-electric point differences which are both in the mid-range of pairwise difference sizes (and Δnet charge ~9 C, 'IEPoZC' ΔpH = 1.1 respectively for the *rpoB*-vind region). Another accession which contains a sequence which is quite different from all other species and accessions is *P. peltatum*. However, we were only able to establish a few crosses due to lack of flowering of this accession. The physico-chemical properties for the *rpoB*-vind type of accession *P. peltatum* are also quite extreme (although not as extreme as for the HORT group) suggesting to us that any cross with this species will be chlorotic. From the few crosses we have the F_1 PELT x ACET is the least chlorotic. The others (F_1 PELT x QUIN, F_1 MULT x PELT) are all severely chlorotic. The former also displays dwarf growth indicative of mCNI. Therefore, we conclude that the physico-chemical properties of the *rpoB*-vind type aa-sequence seems to be a good indicator for occurrence of chlorosis in interspecific hybrids in *Ciconium*. We know other factors also influence the expression and regulation of PEP and thereby also the chloroplast (see the review by Siniauskaya et al. 2015), such as sigma (σ) factors (Zhang et al. 2015), *PPR* genes (Wang et al. 2021) and Whirly genes (Maréchal et al. 2009, Isemer et al. 2012). The *PPR* genes seem to be reduced in number in *P.* x *hortorum* when compared with other Angiosperms (Zhang et al. 2013) and may be interesting candidates for further research in *Pelargonium* and CNI. Previous crossing experiments demonstrated clearly that nuclear genomic factors play a role in expression and regulation of chloroplasts (Tilney-Basset et al. 1989b, 1992, Breman et al. 2020) and these must be taken into consideration. The 'interactome' (Westrich et al. 2012), which is the set of mainly nuclear-encoded proteins that fulfill numerous roles during assembly and maturation of PEP (Shikanai and Fujii 2013), should also be studied in

detail. Mitochondrial expression and regulation are managed by the nucleus as well, given the presence of mitochondrially target nuclear-encoded polymerases, but the extent of coevolution remains untested, at least for *Pelargonium*.

Our protein model of *rpoB* with the homologous bacterial model (Zhang et al. 2012), shows the rpoβ operon in its context in the entire protein. Our model demonstrates that the *rpoB*-vind region is located on the surface of the protein and that it interacts directly with target DNA as well as with the σ-factor included in the model (Fig. 7B). Interestingly these changes are reminiscent of what has been recorded for *E. coli* (Nudler 2012). These may be involved in cross talk between DNA repair, replication and transcription. Changes in cross talk in addition to the previously detected changes in the replication, recombination and repair (RRR) machinery in Geraniaceae (Zhang et al. 2016) may, partially, explain the numerous indels and rearrangements found in the *Pelargonium* (and Geraniaceae) plastomes (Ruhlman et al. 2018). Especially changes that affect the transcriptional process are implicated in genomic disruption (Kim and Robertson 2012, Sebastian and Oberdoerffer 2017) and should be considered in future studies of plastome evolution.

Other examples of chlorosis linked to *rpoB* type

The question could be raised if pCNI could be predicted using the *rpoB* genotypes and their respective physio-chemical properties. While it must be realized that nuclear genomic factors play a major role in expression and management of the organelles, we nevertheless speculate here on possible pCNI phenotypes in, as yet untested, F_1 *Ciconium* interspecific hybrids. We predict that a cross of PELT with any HORT clade accession would yield a chlorotic plant (class III) when containing the PELT plastid and class V when the HORT plastid is present. We do not exclude that the HORT plastid may also yield a class IV chlorotic plant. This is because the physico-chemical (charge distribution, gravy, aliphatic index and iso-electric point) properties of *rpoB* from PELT are closest to HORT and the total length of deletions in *rpoB* is comparable in length to the one in the HORT clade even though the 'TYE' insertions are different. For this claim we do have some evidence. The so called 'hanging basket geraniums' are *P. peltatum* hybrids with a partial *P.* x *hortorum* background (James et al. 2004) as evidenced by flower morphology. At some point a viable and successful hybridization between these two taxa must have taken place. Of course, we do not know the plasto-type of each ancestor, so it remains speculative.

We would further predict that a cross of ARTI with BARK will yield green plants. Both their *rpoB*-vind genotypes are highly similar as well as the physio-chemical properties. Only the 'net charge at pH7' differs 5 C at the *rpoB*-vind region. We would predict that the BARK plastome induces worst chlorosis in such a cross because it contains a number of charge (- to +) and polarity (four from high to low/neutral and one from low to high) 'flips' relative to ARTI and these types of 'flips' seem to have a stronger effect (based on the performance of the HORT types in crosses).

An example from our crosses where other factors must be considered is for the cross of F_1 TONG x ACET. The physico-chemical differences for this cross are slightly larger than those for F_1 ACET x ZONA. The plants of the latter cross, have a pCNI phenotype of IV or V. Whereas the F_1 TONG x ACET cross displayed a pCNI phenotype of category II. That said for the F_1 TONG x ACET cross we obtained only one plant and none for the reciprocal. The F_1 ACET x ZONA cross yielded >10 plants. It could be that the nuclear genomic factors in *P. tongaense* and *P. acetosum* are more similar to each other (or more tolerant) than those of *P. zonale* and *P. acetosum*. *Pelargonium acetosum* and *P. tongaense* have ecological preferences (both are heat tolerant) that are much more similar than those between *P. acetosum* and *P. zonale* (prefers more shade and cooler temperatures) (van der Walt and Vorster 1988). This 'similar ecologically adjusted wiring' may have been enough to overcome the, still rather small, differences between the physico-chemical properties of both parents. The F_1 TONG x ACET plants contains the TONG plastome and the 'TYE' motif, absent from ACET was shown previously result in chlorotic phenotypes also in plants with a TONG nuclear genomic background (Breman et al. 2020). Another interesting feature of this plant is that it contained mitochondria from both parents. The effect of this is uncertain, but this may contribute to the overall fitness of the plant. Finally, this plant may simply possess the right combination genes in terms of organelle management and expression. This is possible as the expression of CNI in F_1 interspecific hybrids is controlled by the nuclear genome and if enough crosses are made even rare combination between phylogenetically distant crosses will occasionally yield plants with less severe chlorosis displaying a 'fortunate combination' (See Breman et al. 2020 for the example of this with HORT x ARID). All the above then leads us to conclude that in this case nuclear genomic factors, or ecologically more similar adaptations, may play a role to explain the discrepancy between the F_1 phenotypes of F_1 ACET x ZONA and F_1 TONG x ACET. Despite this one example the overall differences between the physico-chemical properties seem to be a reasonable proxy to predict F_1 interspecific level of chlorosis with. For other predictions we refer to supplementary material 10.

We know of few other examples where *rpoB* genotypes could be used to explain chlorosis in F_1 interspecific hybrids, but one example we do know of occurs in genus *Zantedeschia*. Recently four *Zantedeschia* plastomes became available (He et al. 2020). When comparing the *rpoB* types and known occurrence of chlorosis of interspecific crosses (RCS pers. obs.) (data not shown) we find that in this case there is also an increase of chlorotic phenotypes. Interestingly *Zantedesschia* shares a number of traits with *Pelargonium*. Its species are relatively easy to cross (RCS pers. comm.), it has biparental inheritance of plastids (Yao et al. 1994) and nuclear genomic alleles have been implicated in explaining the observed patterns of chlorosis in the interspecific crosses as well (Yao et al. 2000, Snijder et al. 2007).

Concluding remarks

Genomes evolve 'in concert' (Greiner et al. 2013) but the in the case of *Ciconium* (and probably the entire family Geraniaceae) the orchestra seems to be out of tune leaving the 'organellar' part of the orchestra to 'wonder' which way to evolve. This disharmony however, allows us

to study and discover mechanisms of CNI. This, combined with the relative ease with which these plants can be grown and crossed, propels *Pelargonium* forward as the ideal model to study CNI mechanisms and organellar driven evolution.

Acknowledgements

We would like to thank the crews at Syngenta Enkhuizen (Nl) and Angers (F) for caring for the plants. Tony Lokkers (Syngenta) for pollinations. Mary Gibby, Sabrina Knees (RBGE) for valuable discussion on *Ciconium* species and sharing herbarium specimens and potentially new species. Justin Hooft (Wageningen UR) for help with interpreting the structural peptide models and understanding the physiochemical properties.

Author contributions

Conceived the study: FCB, FTB, MES, RCS. Carried out the analysis: FCB. Wrote the manuscript: FCB. Experimental design PCR: FCB, FTB, *rpoB* typing: FCB, FTB. Experimental design crossing and embryo rescue: FCB, RS, MES, MS-S. Experimental design KASP: FCB, JWK. Laboratory work: FCB, JWK. All authors read the draft and gave feedback.

Funding

This research was funded by the Dutch Foundations for applied scientific research (TTW). Grant number: 14531 "*Pelargonium* genomics for overcoming cytonuclear incompatibility and bridging species barriers" of the Green Genetics program.

Supplementary material

SUPPLEMENTARY MATERIAL 1 | Base and kmer coverage for mito-markers. Base and kmer coverage of each individual fragment for each mitochondrial marker, including reliability scores. Base coverage of >5<10 is moderate, base coverage >10 is deemed reliable.

SUPPLEMENTARY MATERIAL 2 | Flow-cytometry values for F1 hybrids. Overview of flow-cytometry values obtained for our F_1 hybrids.

SUPPLEMENTARY MATERIAL 3 | Plant list plasto-typing. List of all plants evaluated for chlorosis phenotype and inherited plastid.

SUPPLEMENTARY MATERIAL 4 | Plant list mito-typing. List of all plants evaluated for inherited mitochondrion. Available upon request from FCB.

SUPPLEMENTARY MATERIAL 5 | Mitome alignment. mitome exon alignment.

SUPPLEMENTARY MATERIAL 6 | Spontaneous chlorosis in wildtypes. Photo series of spontaneous chlorosis in wildtype species. *P. multibracteatum*, P. *tongaense*, and *P. yemenense*.

SUPPLEMENTARY MATERIAL 7 | KASP assay protocol. The full mito-typing using a KASP assay protocol.

SUPPLEMENTARY MATERIAL 8 | Tables with rpoB physiochemical properties. A) table showing 11 physiochemical properties of the entire *rpoB* alignment. B) table showing 11 physiochemical properties of the 100 aa regions spanning the *rpoB*-vind region.

SUPPLEMENTARY MATERIAL 9 | Data quality assessment. Discussion on evaluation of F_1 hybrid status of plants and KASP assay issues.

SUPPLEMENTARY MATERIAL 10 | Chlorosis predictions based on rpoB physico-chemical properties.

Online data
All supporting materials can also be viewed at: https://figshare.com/s/556a1b86c083910ca2f3

Chapter 6

General discussion

General discussion

Plant cells contain three genomic compartments, the result of ancient endosymbiosis events whereby a eukaryote-like ancestor engulfed an α-proteobacterial-like organism (becoming the mitochondrion) and a cyanobacterial like organism (yielding the chloroplast). For the chloroplast, the timing of endosymbiosis is tentatively put at 2.1 Gya and it took ~200 MY before the first branching of extant lineages occurred (Sánchez-Baracaldo et al. 2017). the scientific consensus is that the inclusion of these organelles made more complex life possible (Medini et al. 2020). The stable inclusion of the mitochondria and chloroplasts was further achieved by the transfer of many organellar genes (800-2000), not directly involved in photosynthesis or respiration (the oxidative phosphorylation pathway), to the nucleus (Martin et al. 2002, Ku et al. 2015). While the organelles are well established and interactions between the three genomic compartments are generally stable, continued evolution has occasionally upset the delicate balance between them. Whenever a situation changes (biotic or a-biotic) natural selection can act to select otherwise neutral substitutions. Given the important role that each organelle plays in the plant's metabolism any mutation that changes the interactions between nucleus and organelle is likely to have a direct effect on the organism as a whole. It is on these changes and the effects on the organism that this thesis focusses. In my experiments, cyto-nuclear interactions were intentionally disrupted, precisely inducing those responses typical for CNI (i.e., chlorosis). When disruption occurs, it can then be studied within the evolutionary framework of, in this case, an entire clade of species (*P.* sect. *Ciconium*).

Main findings

Chapter two

Inheritance of chloroplasts in *Pelargonium* is considered to be controlled by nuclear genomic loci. This assertion is based on studies using segregation patterns of different cultivars and deducing the number of alleles involved in inheritance of different plastids (Tilney-Basset 1992). In Chapter two we undertook interspecific crosses and determined segregation patterns of chlorosis in F_1 and F_2 offspring. We inferred that one or more nuclear genomic alleles were involved in the expression and management of chloroplasts. Although we could not characterize the actual genes involved, they likely come from one of three major groups involved in organelle management and expression, namely: sigma factors (Wei et al. 2015); PPR genes (Wang et al. 2021) and Whirly genes (Krupinska et al. 2019). We found, in our segregation analysis, that the number of genes needed to explain CNI in different F_1 interspecific hybrids depends on the specific parental combinations. In other words, the more phylogenetically different two hybridizing species are, the more genes were needed to explain segregation patterns of pCNI related phenotypes. This is not surprising, as phylogenetically more distant species have had more time to acquire independent changes as predicted under the BDM model of speciation (Fig. 1). We also found that when we studied CNI at multiple life

stages, the total number of genes needed to explain the observed segregation patterns further increased showing, that CNI acts differently at multiple life stages. Finally, we established that biparental inheritance of chloroplasts is widespread and common across section *Ciconium*. All species have the potential to transmit their plastids either paternally or maternally.

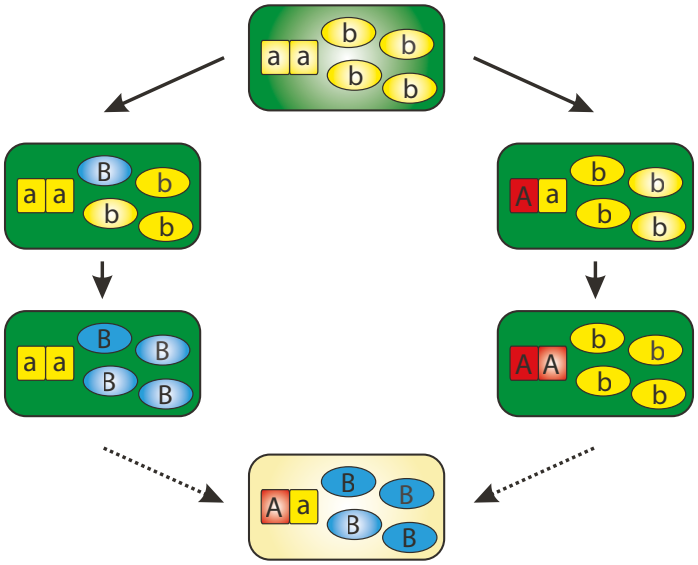


FIGURE 1 | BDM model of speciation by CNI. Green rectangles indicate green plant phenotypes, Aa and Bb denote nuclear genomic or plastome alleles. Colours indicate changes relative to the previous phase, red for the nuclear genome and blue for the chloroplasts. Yellow rectangle indicates (CNI); reproduced with permission from Greiner et al. 2011.

Chapter three

Due to its size and complexity so far relatively few full angiosperm genomes have been annotated. Organelle genomes are much easier to assemble and with high coverage, primarily because they occur in multiple copies in the cell. We used the repetitive part of the genome (the non-coding DNA collectively known as the 'repeatome') as a proxy for the nuclear genotype in our crossing experiments (Chapters two and five). Recently, it was shown that repeat classes can probably be considered homologous across species and hence could be used as phylogenetic markers (Weiss-Schneeweiss et al. 2015, Dodsworth et al. 2015, Vitales et al. 2019). Abundancies of repeat classes as well as sequence similarities between homologous repeats derived from the repeatome were indeed demonstrated to contain phylogenetic signal (Dodsworth et al. 2015, Vitales et al. 2019). Using a comparison of sequence read similarities is a good way to circumvent possible artefacts, such as 'polyploid branch attraction' associated with recently evolved, naturally occurring polyploids. We conclude that these autopolyploids 'attract' each other, when analyzing the repeat abundancy patterns, by virtue of their absolute genomic size being larger. We found that estimated phylogenetic trees, based on

sequence read similarity comparisons, were congruent with plastome-, mitome- or rDNA ITS sequence-based trees. This congruence underscores the fact that the three genomic compartments evolve in concert and may highlight the importance of the organelles on overall *Ciconium* evolution. We found unique repeats that define clades and even species in *Ciconium*. Expanded synapomorphic repeat families were found for the 'Core *Ciconium*' clade (*Ciconium* minus *P. elongatum*) as well as especially leading to the clade containing the species that dispersed outside the Cape Floristic Region (CFR) into the Arabian Peninsula.

Chapter four

There is a large body of literature studying and describing CNI in *Pelargonium*, especially plastid induced CNI (pCNI). There is an equally large volume of literature available (reviewed by Ruhlman and Jansen 2018, 2021) on plastome evolution in Geraniaceae. Rather than approaching plastome evolution at the family and generic level, we approached the study of plastome evolution at the species level using a near-complete coverage of currently described species from section *Ciconium* (Röschenbleck et al. 2014). In this Chapter we focused on concerted evolution of plastid encoded genes and structural changes in the 23S ribosomal subunit. Starting point was the observation of length variation in plastome protein coding genes and especially *rrn23* which was found to contain more length variation among *P.* sect. *Ciconium* species than that observed between plant orders (including *A. trichopoda*) across flowering plants. Based on branch, site and branch-site tests we found that some protein coding genes are under positive selection in *Ciconium*. And that they co-evolve was also found in Geraniaceae (Guisinger et al. 2008 Weng et al. 2016). We do find differences with respect to which genes are under positive selection when comparing *Ciconium* with Geraniaceae: two (*rpoB* and *rpoC1*) of the four *rpo* encoding subunits are under positive selection in *Ciconium*, *rpoC2* is not but shows elevated rate of aa-changes, and the *rpoA* subunit could not be assembled. One gene encoding a ribosomal protein (*uS19c*) was found to be under positive selection but none of the genes involved in photosynthesis. This Chapter provides a first step towards illuminating 'local' patterns of plastome evolution and what drove it. We also addressed the exceptional structural variation encountered in the *Ciconium rrn23* gene, encoding 23S ribosomal RNA in the chloroplast ribosome. We found that this length variation results in possible changed 23S structure. Some of the ribosomal proteins also have a markedly different structure when compared to the *Spinacia oleracea* ribosome model. The implication of this is that the ribosome in *Pelargonium*, while functional, is structurally variable when compared with other plants, probably as a result of overall plastome coevolution (Weng et al. 2016, Zhang et al. 2016), and maybe as a response to changes in the way plastid encoded polymerase (PEP) transcribes genes (Kim and Robertson 2012, Sebastian and Oberdoerffer 2017, Chapter five).

Chapter five

In Chapter five we evaluated F_1 interspecific hybrids for chloroplast and mitochondrial inheritance for three more crossing series. I confirmed our findings for the inheritance of chloroplasts found in Chapter two, i.e., that two nuclear genes are involved in segregation of

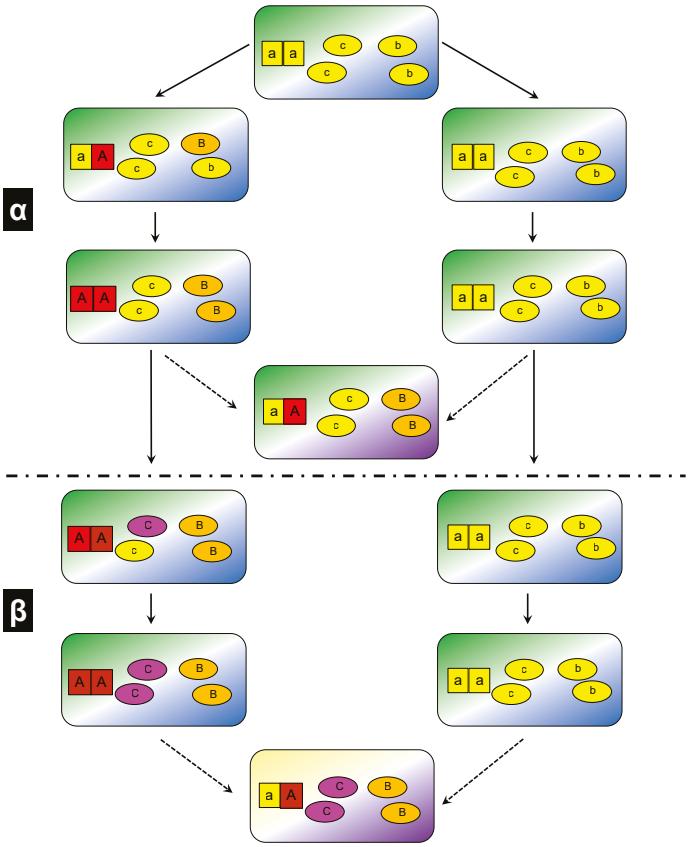


FIGURE 2 | Adjusted BMD model of speciation assuming uniparental, maternal organelle inheritance. 'Aa' represents the nuclear genomic compartment, 'B/b' the mitochondrial compartment and 'C/c' the chloroplast compartment. Background colors Blue and green indicate no mCNI or pCNI respectively. Purple and yellow indicate mCNI and pCNI respectively; α and β indicate mitochondrial and chloroplast phase in BDMI speciation.

chlorosis. I found both maternal and paternal inheritance to be common for chloroplasts in *Ciconium*. In contrast, paternal inheritance of mitochondria is much rarer in the *Ciconium* F_1 interspecific hybrids. Nevertheless, it seems to occur and would settle any previous questions with regards to this phenomenon in *Ciconium* (Weihe et al. 2009, Apitz et al. 2013). The evolutionary implication is that it provides *Pelargonium* with another possible escape from CNI, in addition to the 'escape' that biparental inheritance of chloroplasts offers (Barnard Kubow et al. 2017). Whereas, presumed underlying causes of mCNI are often described in literature (Chase 2007) very little is known about what causes pCNI. In Chapter two we demonstrated that nuclear genomic alleles likely control the expression of the chloroplast. Therefore, we focused on *rpoB* as a plastome encoded gene to explain chlorosis, as it is under direct control of the nuclear genome and central to chloroplast expression and development (Börner et al. 2015). We discovered a ~30 aa *rpoB* region that showed high length variation in most *Ciconium* species. We modelled six physico-chemical properties of the (partial) peptide

encoded by this region and compared it with the chlorosis phenotypes for each variant. We found strong correlation between peptide physico-chemical properties and the occurrence of chlorosis in F_1 interspecific hybrids. Therefore, we propose that, for *Ciconium*, probably for *Pelargonium* and maybe for Geraniaceae, it may be possible to predict the level of CNI in F_1 interspecific hybrids, based on (physico-chemical properties of) *rpoB* type.

Cyto-Nuclear-Interactions and speciation

Cytonuclear incompatibility plays an important role in speciation (Postel and Touzet 2020) whereby mCNI impacts plant reproductive capabilities directly by reducing male fertility, providing a strong bottleneck in the establishment of subsequent generations. In *Ciconium*, F_1 interspecific hybrids showed little fertility and when crossing species from different *Ciconium* clades fertility disappeared completely (except in auto-polyploids). This is in contrast with F_1 interspecific pCNI patterns where crosses across *Ciconium* showed varying levels of chlorosis, but often plants were 'fit enough' to survive. Chlorosis does become stronger in hybrids from phylogenetically more distant parents but, even such crosses sporadically yield virtually unaffected plants. Chlorosis can be asymmetrically expressed in F_1 interspecific hybrids, with one parental type resulting in mild chlorosis whereas the other type induces severe or lethal chlorosis, all in the same nuclear genomic background (Turelli and Moyle 2007). This may be because the nuclear genome has evolved one or two alleles that adapted to, or induced, the specific physico-chemical changes in the rpoB peptide. If other genes remained conserved that would imply that, under this scenario the 'lethal donor' genome retained the ability to manage the non-lethal plastid as well whereas the non-lethal donor genome does not have the modification needed to express the lethal plastome. It could even be that particular changes in physico-chemical properties of the rpoB peptide are less easy to overcome by plants in general. In the case of the lethal *P. inquinans* type plastids (see Weihe et al. 2009 and Chapters two and five) it could be that in *P. inquinans* (and in *P. frutetorum*) a single sigma factor or PPR gene has evolved to precisely overcome those properties that are so different in the plastids of these taxa (distribution and type of charge, optimal pH, Chapter four). Overall, this leads to the conclusion that, not only do mCNI and pCNI play an important role in *Pelargonium* speciation, but it is also possible to discern a sequence of events in the establishment of BDM incompatibilities (see Chapter one, introduction). Based on the findings in Chapters two and five, I would propose that, for *Ciconium*, mCNI evolves first, followed by pCNI (see Fig. 2). Whether this is true in other plant lineages remains to be tested but there are indications from animals that mCNI plays a large and fast acting role in the establishment of BDM incompatibilities (reviewed in Burton and Barreto 2012). I divided the model in an α and β phase, with mCNI dominating the process in the first, and in the second, pCNI would 'finish the job' of establishing full cytonuclear incompatibility between the populations. This model would need further adjustment however in the case of groups that display biparental inheritance of organelles. As we showed in Chapters two and five biparental inheritance of plastids is common in *Ciconium* and, although much less frequent, mitochondria can be transmitted via the paternal line as well. I speculate that the effect of mCNI on fertility acts

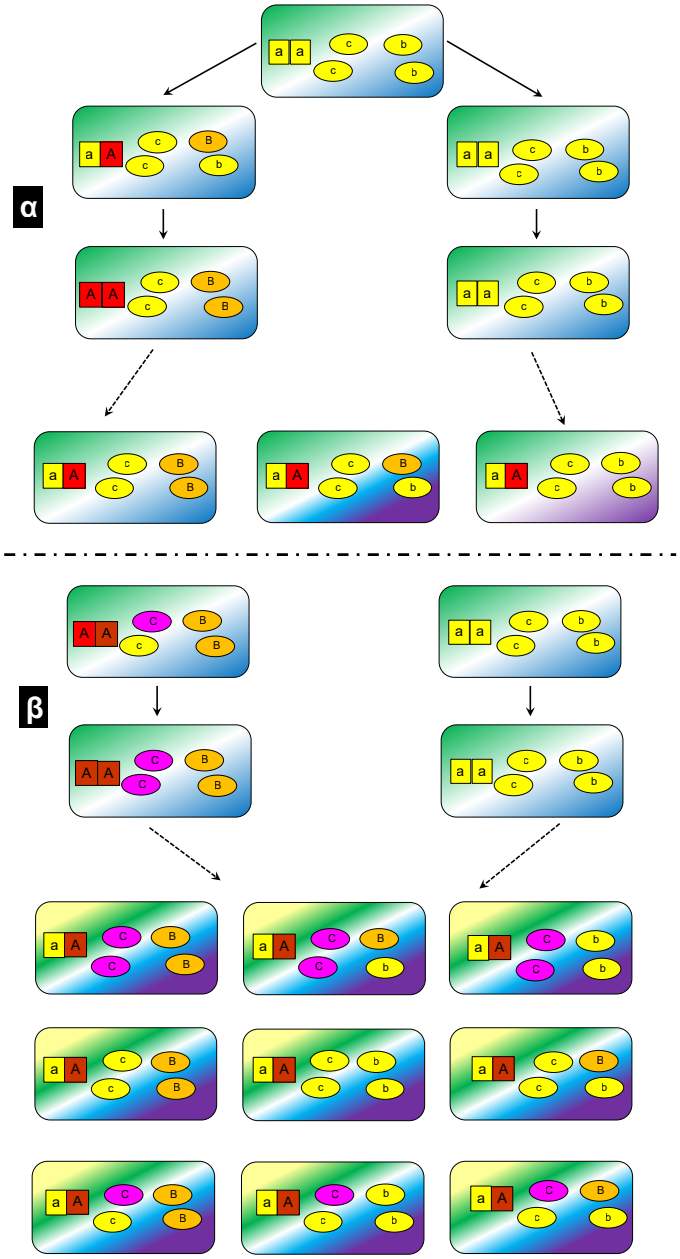


FIGURE 3 | Adjusted BDM model of speciation assuming biparental, organelle inheritance. As in Fig. 2, but now multiple genotypes and phenotypes are possible, some may lead to rescue from CNI.

stronger as it directly impacts (male) fertility while mild chlorosis (especially green virescent plants) does not affect overall fitness per se, and these plants will retain normal flower functionality. Observations on variegated plants (F_1 *P.* x *hortorum* x *P. zonale* and *P.* x *hortorum*

x *P. acetosum*) showed that even white, parasitic, parts of the plant will develop normal flowers with viable pollen. Development is slower on white parts than on the green parts of these plants, but eventually flower, anther and pollen will develop.

Evolutionary consequences of heteroplasmy

It was recently proposed that heteroplasmy, and thus biparental inheritance, of organelles may be widespread across angiosperms (Gonçalves et al. 2020) with the two organelle types being present in varying ratios in offspring. Heteroplasmy could arise after locally acquired structural and sequence divergence or because of a biparental inheritance event during evolutionary history. The question arises then what the evolutionary benefits of this trait may be and why is it particularly strong in *Pelargonium*? Mitochondrial biparental inheritance may confer an advantage by being able to adapt quickly to demanding metabolic requirements through recombination between both parental mitomes (Barr et al. 2005, Woloszynska, 2009, Arrieta-Montiel and Mackenzie, 2011. Apitz et al. 2013, Johnston 2019). This could be advantageous in a region where and habitats are highly compartmentalized and variable steep rainfall clines, fire, and soil types across short geographical distances per growing season see (Fig. 5B) such as is typical in the CFR (Cowling and Lombard 2002, Linder 2003; Martínez-Cabrera 2012). Ostensibly, chloroplasts do not recombine (but see Medgyesy et al. 1985) and they would, therefore, adapt to rapid environmental changes slower than mitochondria would. Nevertheless, being able to maintain two (or more?) types of chloroplasts or mitochondria (heteroplasmy), implies that, should the circumstances require, such a plant could 'switch'. Thus, having the ability to maintain more than two types of an organelle would also increase adaptability from an evolutionary point of view even if one type is less optimal than the other (Woloszynska et al. 2009). However, this could change if circumstances change, when the effect on fitness conferred by one type changes with those circumstances. In the most extreme case, a plant could even switch within the same season where one plastid would be active if there is low rainfall, increased heat and radiation. The other could become active when rainfall increases, and temperature and radiation drop. While this may seem far-fetched, I think it is not impossible. Over the course of this project, I observed spontaneous chlorosis occurring within individuals, during ontogenesis, in at least four species (*P. zonale*, *P. inquinans*, *P. multibracteatum*, *P. yemensene* and *P. tongaense* see Figs. 6A-C). For this to work the plant must have the ability to retain both types and increase or decrease expression of either, as the circumstances demand. While I did not test this in my crossings, the observation that chlorosis fluctuates with green house temperatures, in combination with the differences in ecological preference of the species involved, is intriguing. Evidence for such a seasonal 'switch' is unfortunately not available, but the introduction of different plastid types in a homozygotic nuclear genomic background ('cytoplasmic swapping') in *A. thaliana* were shown to result in changes in photosynthetic rates (Flood et al. 2020). This suggests that it is indeed possible that slightly different chloroplasts can be expressed by one nuclear genome and these result in variation of photosynthetic efficiency (Flood et al. 2020).

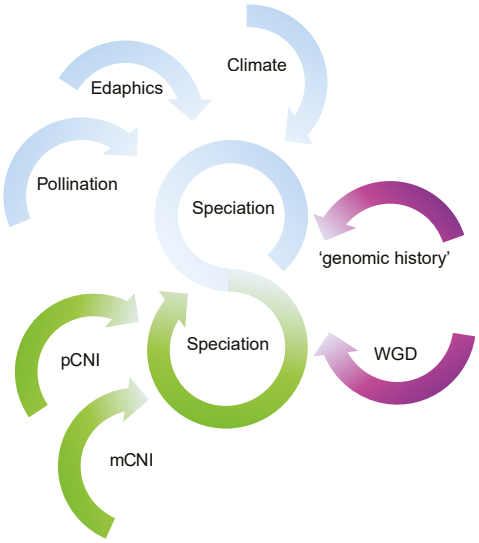


FIGURE 4 | Expanded 'cyclical' model for *Pelargonium* speciation. Blue indicates pre-zygotic or environmental pressures influencings speciation, green indicates post-zygotic pressures, purple indicate inherent genomic properties influencing speciation. Adjusted with permission from van de Kerke 2019.

Compatibility of nuclear genomes

In addition to CNI, nuclear genomes themselves can be incompatible as well. This manifests itself mostly by the occurrence of 'hybrid necrosis' caused by insurmountable differences in the plant immune systems (Wan et al. 2021). We have not observed any signs of this in our work on *P.* section *Ciconium*. Even crosses between the most phylogenetically distant parents, showed normal F_1 plant phenotypes (apart from severe CNI phenotypes) when grown under controlled conditions in a laboratory. Nuclear genomic incompatibility (NGI) does not seem to play a role (yet) at least at the sectional level in *Pelargonium*. A much stronger nuclear genomic barrier to crossing are differences in base chromosome numbers. For instance, we have attempted to cross *P. elongatum* (x=4), with other *Ciconium* (x=9) species (Gibby et al. 1989), but no development on the plant was ever observed. We did not observe specific repeatome types and the repeatome patterns in *Ciconium* (Chapter three), indicating that the nuclear genomes are quite similar across the section (as compared to *P. karooicum*, unassigned within subgenus *Paucisignata* sensu Röschenbleck et al. 2014). The species analyzed do share >25 *Ciconium* specific synapomorphic repeats (representing ~25% of the repeatome). If this is indicative of nuclear genomic compatibility remains to be tested. Repeat profiles should be made for species from different *Pelargonium* sections and intersectional crosses could be attempted. If successful in establishing hybrid plants these plants could then be evaluated for signs of nuclear genomic incompatibility such as hybrid necrosis and this could be compared to repeat profiles. Overall, though, I would suspect that incompatibilities between nuclear genomes (NGI) across the genus are not the primary crossing barrier as intersectional crosses are well-known from amateur breeder sources for instance with *P. fulgidum*, https://www.geraniaceae-group.org/society/links/) and do result in viable plants (Horn 1994, James et al. 2004). I suspect that CNI is a much stronger barrier between *Pelargonium* species than NGI.

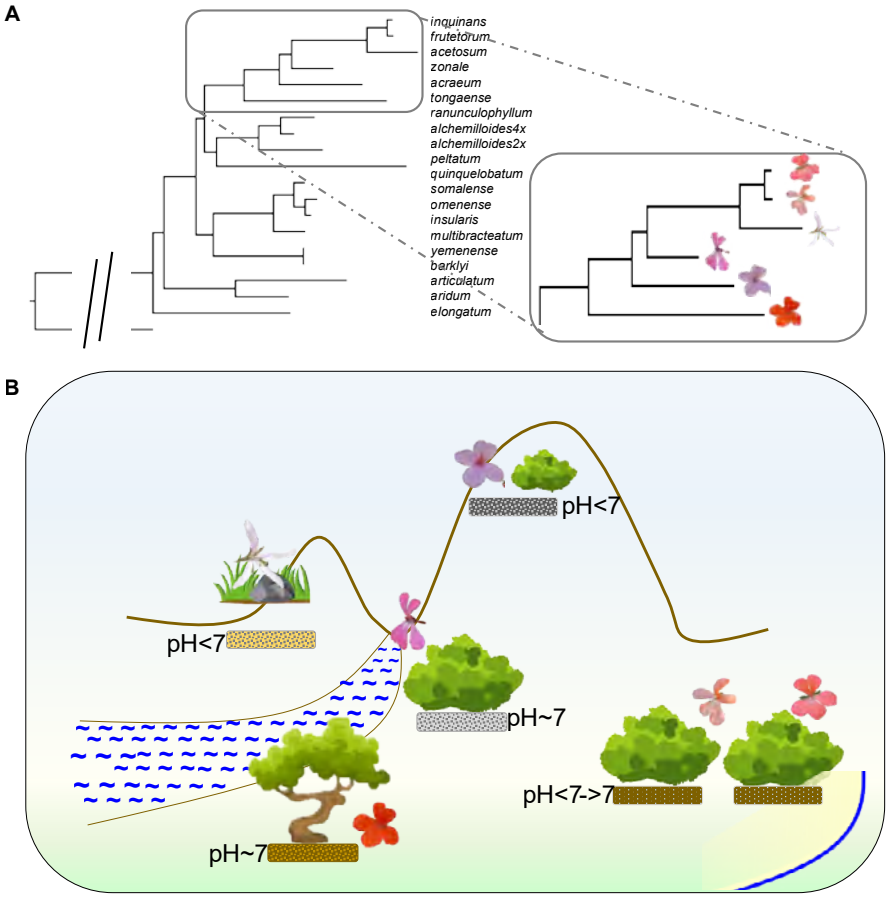


FIGURE 5 | A) Plastome based *Ciconium* phylogeny The red flowered clade has been cut out and the flowers represent each species. B) Schematic overview of ecological preferences of the species in the red flowered clade. (source: http://pza.sanbi.org/, Van der Walt and Vorster 1988, appendix 1 this thesis)

P. tongaense: Loose sandy soil, cool, forests-bushy trees.

P. acraeum: Heavy soil, sand, clay loam, dolerite, exposed-semi shade, montane.

P. zonale: rocky soil, sand, clay loam, shade, valleys.

P. acetosum: Stony, sandy soil, dry grassland, hillsides, exposed.

P. inquinans and *P. frutetorum*: Shale, sandy loam shade, coastal bush.

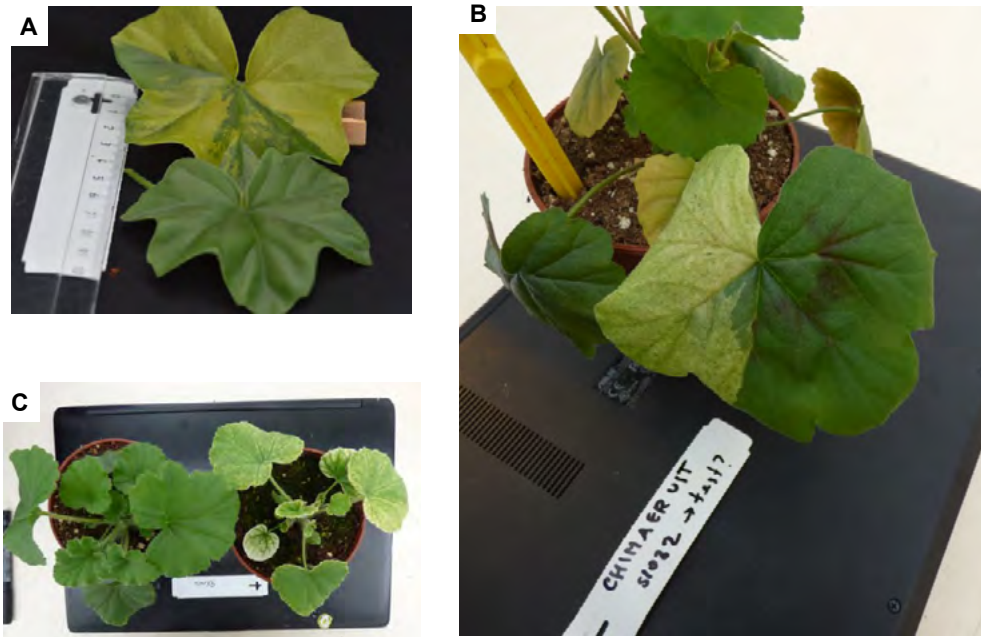


FIGURE 6 | Spontaneous chlorosis in *Ciconium* wildtype accessions. A) *P. tongaense*, B) *P. multibracteratum*, *P. yemenense* sp. nov.

Speciation and genomic changes in *Pelargonium*

As in most cases of plant speciation there is a pre-zygotic and post-zygotic component, signifying the importance of external and internal processes for evolution. Pre-zygotic barriers have been well investigated in *Pelargonium* (see Fig. 5 for a model of *Pelargonium* speciation), especially pollination (focusing on floral aspects such as nectar tube as well as corolla shape evolution) (Bakker et al. 2005, Moore et al. 2018, van de Kerke 2019) as well as other factors such as niche variation, edaphic shifts, water availability, or the occurrence of fire, (Jones et al. 2003, 2009, Martínez-Cabrera et al. 2012). Of these, nectar tube length (Ringelberg 2012, van der Niet et al. 2006; van de Kerke 2019) and high CFR niche diversity (Martínez-Cabrera et al. 2012) have been shown to correlate strongest with speciation in *Pelargonium*. For the first it appears that with increasing nectar-tube length speciation rates decrease. For the latter, the authors conclude that "in most instances, niche similarity was independent of phylogenetic relatedness, the *Pelargonium* radiation being characterized neither by high pronounced niche conservatism nor lability."

This thesis demonstrates that post-zygotic barriers too, although often not absolute, play a major role in *Ciconium* evolution as is seen in the CNI patterns encountered in the F_1 interspecific hybrids. Therefore, our model of speciation for *Pelargonium* must be refined (Fig. 3). We propose that effects of pre-zygotic barriers act after the mitochondrial phase has been concluded (the α-phase, see Fig. 2), i.e., after partial reproductive barriers have been

established as this isolation occurs restively fast. These reproductively isolated populations start to coevolve their plastomes within rather than between population (phase β). Edaphic and climatic variability would start to kick in at around the same time reinforcing selection for adaptive traits, presumably in the plastome as well, that perform best under these ecological selective pressures (van der Niet et al. 2006), thus further reinforcing reproductive isolation and speciation. This scenario supports the findings by van der Niet and Johnson (2009) who state that: "ecological shifts between sister species were a key part of their speciation. The alternative scenario would be some unknown process of speciation followed by the divergence of ecologically linked traits through anagenesis." This unknown process of speciation could be CNI, based on the evidence presented in this thesis. Even though it is a "less parsimonious" and therefore a "less likely explanation" (van der Niet and Johnson 2009), it may be exactly what is happening. Once metabolism is optimized to best 'fit' the a-biotic ecological situation or niche, the floral display apparatus may start to adapt to pollinators. The adaptation of flowers to pollinators then starts with the nectar tube, the variation in which was shown to be positively correlated with *Pelargonium* speciation rate (Ringelberg 2012, van de Kerke 2019) but not corolla shape per se. As a final step, remaining floral structures (petal curvature, petal markings, stigma—anthers distance; Steve Johnson, pers. comm.), may become more important and refined in attracting more or specialized pollinators. Section *Ciconium* flowers are characterized by an absence of petal markings (otherwise ubiquitous across genus *Pelargonium*) and hence the subgeneric name '*Paucisignata*' (Röschenbleck et al. 2014). This would suggest that pollinator-driven evolution is of minor importance; and indeed we don't see *Ciconium* species aggregating in the PCA's summarizing corolla shape variation (van de Kerke 2019, Fig. S4, S5 and 6). However at least four species do display clear petal markings (*P. peltatum*, *P. quinquelobatum*, *P. acraeum*, *P. articulatum*, *P. yemenense*) and *Ciconium* hybrids, F_1 and complex ones, display them regularly. First generation hybrids always have petal markings if at least one of the parents displays them (e.g., F_1 *P. multibracteatum* x *P. quinquelobatum*) indicating dominance of the gene(s) involved. I would propose that the 'absence of petal markings' is a null mutation (silencing or absence of the allele). When for instance a species with a pale flower and no petal markings (e.g., white-yellow *P. barklyi*) is crossed with another *P.* subg. *Paucisignata* species that does not show petal markings (e.g., the tetraploid *P. alchemilloides*) the resulting plant displays darkened veins in the flower that act as petal marking. The genes for 'color filling' the petal markings region (Nasser and Tilney-Basset 1992) may still be there. Indeed, the species that do not display them either have a uniform petal color across the petals, or contain almost white veins at the markings position. Null mutations may also be responsible for the white color occasionally observed in *P. inquinans*, *P. zonale* and the derived hybrid *P.* x *hortorum* (Ronald Snijder pers. comm). The genetic machinery underlying petal markings may therefore be silenced, or dormant, or absent in some species, but it could easily be activated or reintroduced. I am therefore of the opinion that the subgeneric name '*Paucisignata*' is somewhat misleading. Finally, I hypothesize that, at least in some species, the petal markings are invisible to humans, but possibly visible to invertebrates. *Pelargonium acraeum* is the only *Ciconium* species do display white surrounding pink, linear petal markings on the posterior two petals of the corolla (Fig.

7). This species occurs in montane regions where UV reflection may play an important role in marking visibility (Burr and Barthlott 1993).

Another explanation could be that the absence of petal markings is a *Ciconium*-unique adaptation indicative of early phase floral adaptation (late phase ecological and organellar adaptations). It could be advantageous to have a general pollination syndrome rather than a specialized one. As long as shapes do not confer obvious disadvantages, the corolla is shape is 'free to vary' (or in state of "morphological drift", Bas Zwaan, pers. comm.). This was suggested to be the case by van de Kerke 2019 who showed that: "General floral shape does not correlate with speciation rates in *Pelargonium* indicating that this part of the 'display' system is a general attractant and that the wide variety in shapes may be the result of genetic drift".

FIGURE 7 | Flower of *P. acraeum*. Two posterior petals show linear petal markings surrounded by white. Base of anterior petals darkened.

Interestingly, neither ecological parameters, nor flower shape adaptations, nor organellar induced barriers are necessarily absolute. Pollinators can still transfer pollen between different populations because they are not restricted in moving between habitats and in that way potentially (re)-introduce genetic information (including organelles) into other populations. Though unlikely to result in new species or a real shift in organelle type, the results from this thesis show that it is not impossible. This scenario could for instance explain the reported 'chloroplast captures' in sect. *Hoarea* and *Ligularia* species (Bakker et al. 2005), invoking a possible new cycle of CNI driven speciation as the genomic compartments adapt. Speciation in *P.* section *Ciconium* is probably best viewed as an ongoing phenomenon with different selective forces acting simultaneously but each one gaining the 'upper hand' at a different stage. Each time when adaptation to one force has taken place another can take over again (Fig. 4). Evolution in *Ciconium* currently seems to be driven by CNI, whereas in other sections (e.g., *Hoarea*, Bakker et al. 2005; Burger et al. in prep.) pre-zygotic barriers seems to be dominant.

The repeatome as proxy for nuclear genomic types?

The repeatome was shown to be highly discriminative between species (Dodsworth et al. 2017, Vitales et al. 2019, Chapter three) and the large amount of data and information contained therein could be used in identifying species alongside current mitochondrial and plastid

markers (Hebert et al. 2003, de Vere et al. 2015). When we included *P. karooicum* (clade C2 and sister to the section *Ciconium* sensu Röschenbleck et al. 2014 clade) it became obvious that few repeats were shared between *P. karooicum* and *Ciconium*, many repeats were unique to *Ciconium*, suggesting there is not a universal homologous repeat, not even at the intrageneric level. Obtaining repeat profiles, for instance across a family (e.g., Geraniaceae) would allow us the determine a hierarchy in evolution of repeats. More conserved repeats will connect the repeat profiles, while repeats synapomorphic for a clade (at any taxonomic level) may provide interesting clues with respect to speciation or the evolutionary history of the clade. The use of the repeatome may also allow us to test the hypothesis of historical plastome capture (see above) and trace historical hybridizations. Finally, it would allow us to circumvent the effects of heteroplasmy on phylogeny reconstruction, by providing a non-organellar perspective.

Speciation in *Ciconium*: from autecology of species to changes in rpo subunits?

As we saw in Chapter five, rpoB peptide physico-chemical properties can, partially, explain the occurrence of chlorosis in F1 interspecific *Ciconium* hybrids. But what could be the mechanism behind these changes? So far, we saw that speciation in *Pelargonium* is driven by ecological factors (Bakker et al. 2005, van der Niet et al. 2006, van der Niet and Johnson 2009, van de Kerke 2019) as well as by CNI (this thesis). And correlation of morphological traits such as changes in leaf surface area with pre-zygotic barriers, while weak, do exist especially for adaptations to drought and heat (Nicotra et al. 2008, Jones et al. 2009, Nicotra et al. 2011). But how could external ecological factors influence speciation. Soil parameters (pH, nutrient content, water retention) have a direct impact on plant well-being. Acidic sandy soils will leech out nutrient and have low overall nutrient supply. Sandy, or rocky soils will retain less water than heavy soils with high loam and clay content and. Rainfall, radiation intensity further influences well-being directly and, high or low, will change the dynamics of evapotranspiration. Changes in radiation and water availability will further affect photosynthesis and have been shown to do so in *Pelargonium*, where in some lineages CAM-like metabolism and closure of stomata was demonstrated by Jones et al. 2003). Low rainfall will also necessitate retention of water or synchronization of the lifecycle with water availability (e.g., in section Hoarea, Bakker et al. 2005). In section *Ciconium* the red-flowered clade (cutout Fig. 5A) presents a group of species that evolved in different environments and their *rpo* genotypes are also highly variable, possibly presenting us with an example where environmental differences, (e.g., edaphic shifts, exposed or more sheltered growth environment, water availability) may have affected the chloroplast gene expression. Temperature, radiation, water, and nutrients may affect the overall environment in the plant possibly resulting in pH changes and nutrient concentrations. Both these factors require adaptation from the metabolic processes as these operate at specific optima (temperature, pH). I would propose that the signaling proteins (passing through media with changed pH) may have changed to adjust and stay functional, resulting in changes in *rpo* as it is central in expression plastid encoded genes. These genes must have evolved themselves as well do deal with the changed radiation and water and nutrient availability. Take for example the sister species *P. zonale* and *P. acetosum* (Figs. 5A

and B), originating from contrasting ecosystems. Variation in rpoB peptide physico-chemical properties between these two species is quite pronounced (Chapter five). If we zoom in on *P. acetosum* we can see that it has acquired numerous unique amino acid (aa) changes in its *rpo* sequences. So much so that sequences for this species had one of the longest branches (Fig. 5A and Chapter four), which indicates many unique changes. Tests for positive selection showed that the optimal explanation for observed patterns of aa changes was to assume a separate ratio of silent vs non-silent substitutions d_N/d_S (ω) for *P. acetosum*. These tests showed that alongside *P. acetosum*, separate ω's should be assumed for the other long branched species (see also Fig. 5A and Chapter four) *P. peltatum*, *P. barklyi* and *P. aridum*. Coincidentally or not, these species are all adapted to hotter, drier, more exposed and nutrient poor environments, and represent extremes as they have had to cope with a multiple of ecological pressures. Whereas, for instance, *P. zonale* and *P. tongaense* do not differ much in their ecological requirements and differences between their respective physico-chemical properties are smaller than between these two species and *P. acetosum*. Interestingly crosses between *P. tongaense* and *P. acetosum* yielded a green plant, whereas those between *P. zonale* and *P. acetosum* yielded severely chlorotic and lethal plants only. This suggests that, while *P. tongaense* prefers the same habitat as *P. zonale* it may have retained genetic properties allowing it to tolerate heat. Indeed, this species still shows heat-related adaptations because it leaves are somewhat more succulent and thicker and despite its preferred habitat it more heat tolerant than *P. zonale* and has been used in the past to introduce heat resistant in *P.* x *hortorum* cultivars (Ronald Snijder pers. comm). Testing for correlation between adaptations to various, extreme, environments and *rpo* genotypes should be undertaken to test the hypothesis posed here that 'changes in ecological circumstances lead to physico-chemical changes in the plant's transcriptional and metabolic pathways, which in turn lead to genomic adaptations, I would predict that in other *Pelargonium* sections species adjusted to more extreme (non-standard for most species in a section) environments will show longer branches in all or some *rpo* subunit gene trees and possibly positive selection.

The rpoB peptides evaluated by us, display changes in optimal pH, 3D shape parameters, sequence length and distribution of charge. Presumably, these play a role when peptides or indeed entire proteins need to interact. The most important properties in *Ciconium* seem to be the 'iso-electric point of zero charge' and the 'net charge' at pH7. When each increased, chlorosis becomes more severe. In some cases, this was mitigated when, for instance, the aliphatic index (as a measure of 3D shape parameter) differences were small. Testing of this should be undertaken with crossing series not (yet) performed by us or in other sections of *Pelargonium* after analyzing their plastomes with regards to *rpoB* variants. The practical implications could be large, because the occurrence of CNI is major hurdle in plant breeding when attempting to introduce traits from 'crop wild relatives' (CWR) in cultivars (Flint-Garcia 2013, Kuligowska et al. 2012).

Possible applications of the altered *P.* sect. *Ciconium* plastid ribosome

We found that the *Ciconium* plastid ribosomes are structurally highly diverse, because of the unusually high sequence divergence in the *rrn23* gene (Chapter four). Gaining insight into what drives ribosomal variations can have important implications for diverse fields such as creating artificial ribosomes (Levy et al. 2020), but also microbial antibiotic resistance. Growing 'ribosomes on a chip' is an important part of the effort to develop artificial cells. A ribosome with a different design, for instance a differently shaped 'peptide synthesizing tunnel' could perhaps produce compounds which are otherwise hard or impossible to synthesize (Schmied et al. 2018). But even a larger (such as the one in *Ciconium*) or smaller functioning ribosome can already be an important step forward. Perhaps a larger one can produce larger volumes or perhaps a smaller one will be easier to synthesize as it may require fewer steps to assemble if non-essential parts are stripped away. Before this can be undertaken with any changes of success we must know, what changes are needed and more importantly how the other ribosomal components (the ribo-proteome) must adapt to still retain a function. As we have no indications that the *Ciconium* ribosome functions differently from others it provides an example of a functioning yet altered configuration. The results discovered and analyzed in this thesis provide a hint as to 'what is possible' and what needs to change with what. Ribo-proteins must evolve if the rRNA backbone is changed. We found that ribo-proteins involved in ribosomal stability show, in *Ciconium*, structural flexibility and have evolved extra alpha helices in some instances (e.g., *bl32c* see Chapter four), likely to maintain a stable connection to between 23S the large subunit and 4.5S rRNA subunit.

Another important field where insights in ribosomal evolution are valuable is the study of antibiotic resistance. More than half of antibiotic in use today inhibit the bacterial ribosome (Charon et al. 2019) and these compounds act by directly upsetting physico-chemical interactions in the ribosome inhibiting aa translation to functional peptides. Antibiotic resistance often arises from structural ribosomal changes that evolved as an adaptational response to neutralize and counteract the disrupting effects of antibiotics (Davies et al. 2010, Charon et al. 2019).

Whole genome duplications in *Pelargonium*?

Whole Genome Duplications (WGD) should not be overlooked when discussing speciation in general and in plants in particular. WGD's are widespread in Angiosperms, and these are considered to have had a profound effect on the evolution of flowering plants (Wendel et al. 2015, Soltis et al. 2009, 2015; Schranz et al. 2012). One of the resulting effects is an increased speciation rate (Landis et al. 2018) in angiosperms. Previously (Ren et al. 2018) deduced that there may have been a WGD at +/- 11 Mya, leading to the lineage *Pelargonium*. Interestingly, this time estimate is close to the estimate given by van de Kerke et al. 2019 for the proliferation of *Pelargonium* (9.7 Mya with an HPD: 9.0-10.5 Mya). While we have no direct evidence, we speculate there have been 12-13 independent polyploidizations (Nieuwenhuis 2013) in species scattered throughout the *Pelargonium* phylogenetic tree, but ancient WGD's may have led to one or more sections. We have some circumstantial evidence for this: F_1 interspecific

hybrids readily form auto polyploids as evidenced from our studies (Chapters two and five). Some wildtype species from *Ciconium* are autopolyploid (*P. articulatum*, *P. yemenense*, *P. alchemilloides*), based on flow-cytometry values (Chapter two) and on previously published karyology studies (Gibby et al. 1990, 1996). A final piece of evidence is the occurrence of varying chromosome numbers indicative of incomplete genome reductions which occur across *Pelargonium* (Bakker et al. 2005) indicative of incomplete genome reduction which occurs post-polyploidization (Sankoff et al. 2010, Buggs et al. 2012).

Hyper-diverse organellar genomes and possible consequences

The chloroplast perspective

Pelargonium plastomes are, when compared to other angiosperm groups, highly variable structurally (Ruhlman and Jansen 2018). Knowing what causes the cause can help when determining further avenues of research and would also allow speculation as to determine the cause of specific changes per clade. Did different clades adapt in different ways to the underlying causes? To address this the nuclear genome of Geraniaceae must be characterized, specifically for the nuclear encoded plastid targeting genes ('neptag's). One such group is involved in repair, replication and recombination (so called RRR genes) of chloroplasts. Four of these genes have been found to be under positive selection in Geraniaceae and it was suggested that dysfunctions in these genes explains aberrant Geraniaceae plastomes (Zhang et al. 2016). However not only repair, replication and recombination are at work. For the organelles to function properly the signaling between nucleus and organelles must function properly as well. It was shown that PEP encoding subunits are all affected to some degree, with *rpoB* and *rpoC1* being under positive selection and *rpoA* being broken up in multiple orfs across the *Pelargonium* plastome (Chumley et al. 2006). This thesis shows that, by analyzing the segregation ratios of CNI in F_1 interspecific crosses, there is variation in neptag's involved in the expression of chloroplasts. We speculate that genes, such as PPR (Wang et al. 2021), Whirly (Maréchal et al. 2009, Isemer et al. 2012) and sigma factors (Lerbs-Mache 2011, Zhang et al. 2013, Wei et al. 2015) could also be highly variable between *Pelargonium* species. Especially those genes that target *rpoB*, which is the initial recipient of anterograde signals, may have been affected. Characterizing these genes and subsequently plotting observed changes over currently available, dated phylogenetic trees (e.g., van de Kerke et al. 2019) would be an interesting way to piece to together the chain of events after the RRR induced plastome changes. A question that arises is, to what extent have different groups in Geraniaceae dealt with the RRR induced changes? Do all genera and section show a similar response or does one genus evolved a different response than the others.

The mitochondrial perspective

The preceding paragraph mainly discussed the plastome. But mitomes appear to be similarly affected by a genomic upheaval as the plastomes in Geraniaceae show increased substitution, rates have rearrangements, lack introns (Bakker et al. 2006, Park et al. 2017, Cole et al. 2018, Choi et al. 2021). Mitochondria do not, like chloroplast, carry their own RNA polymerase

but its genes are expressed only by nuclear encoded polymerases (NEP'S). More specifically the one NEP that targets chloroplast, targets mitochondria as well (referred to as 'mpNEP's, Courtois et al. 2007). I would propose that the same nuclear genomic changes (such as those to RRR genes and perhaps PPR, whirly and sigma factors as well) are also explain the high mitochondrial diversity encountered in Geraniaceae. Mitome disruption seems to be even worse than the plastome disruptions and this leads me to hypothesize that the fact that the plastome carries its own polymerase may actually serve as a 'buffer' to counteract the disrupting effect of the changed nuclear genomic alleles mentioned above. If so this 'lack of a mitochondrially encoded polymerase (MEP) buffer' may explain also why mCNI seems to act stronger and faster in *Ciconium* speciation as discussed in the preceding paragraphs (Figs. 1 & 2). The mitochondria acquire changes quicker, leading to faster CMS and thus faster establishment of BMD incompatibilities. Finally, in Chapter five we showed that, just as plastids, mitochondria can be inherited both paternally and maternally as well as biparentally. This confirms earlier findings by Weihe et al. (2009) and Apitz et al. (2013) and points to a, partially, shared mechanism controlling inheritance (see e.g., Tilney -Basset et al. 1992 for supporting information on the genetic control of plastid inheritance). As speculated above for chloroplasts, heteroplasmy could perhaps counteract some effects of CNI and increase evolutionary adaptability. However, the level of heteroplasmy, at least in interspecific crosses seems insufficient to counteract the fast rate of mitome changes in *Ciconium*. What the role of the mitochondria is in speciation in other plant groups and whether heteroplasmy for mitochondria is also more common than suspected remains to be investigated.

Reflections

For this thesis I was able to work in several arena's but as is always the case, certain topics remain unexplored, incomplete, or overlooked. This is normal, no person can do all, but I would still like to reflect on 'what if' or 'what else'. In this thesis four complete series of interspecific crosses were created, many of which never created before. However, *P.* section *Ciconium* contains more species and having one or two series would have been nice if only to underpin the predictions we make about chlorosis in Chapter five. I think studying one crossing series over more than two generations would have illuminated the nuclear genomic alleles underlying pCNI better and would have also gained insights into the process of selection for or against certain alleles.

I was not able to assemble full plastomes, which was not unexpected. However, the attempts to optimize the data that I did have, the great number of studies already done on the *Pelargonium* plastome and the desire to filter out as much information as possible, blinded me for the fact that the data that we did have already provided valuable insights (Chapters four and five). Still, it would have been nice to have full plastomes available as we likely miss a relevant part of the plastome. I think more genes are under positive selection or are relevant for explaining co-evolution. What I would have liked to explore as well in more detail is the mitochondrion. I feel the mitochondrion remains understudied here, but it is just as vital to the plant and to its evolution as the plastome.

I would also like to reflect on the overall thesis. The research played itself out in three arenas. In the greenhouse, in the laboratory and behind the desk. This work to me shows the importance of an integrated approach studying evolutionary biology. I would have loved to add the 'field' arena to this as well. But I do not believe that the fact that I have not studied *Pelargonium* in the wild weighs negatively on this work. That said, field work by other resulted in the discovery of a new species from Oman (*P. omenense* sp. Nov Knees et al. in prep) and one from Yemen (*P. yemenense* sp. Nov. Gibby et al. in prep). *Pelargonium yemenense* was also discovered by Ronald Snijder (co-promotor) in the seed collections of Syngenta and it was used in this thesis to cross with and for study.

Ideal world

As stated above issues remain. What would I like to do if I had unlimited resources and time? First of all, I would have performed more crossing series, preferably all. Having the full overview of crossing combinations would place us in the unique position to attempt to determine for each branch in the *Ciconium* phylogenetic tree a hypothesis with regards to the impact that plastome differences have on hybridization and speciation. I would have loved to sequence one or more nuclear genomic genes, such as the PPR and whirly genes or the sigma factors involved in plastome expression and management (Wang et al. 2021, Maréchal et al. 2009, Isemer et al. 2012, Börner et al. 2015) for *Ciconium*. This would allow us to study interactions between the nucleus and organelles in even more detail. That brings us to the lack of a fully annotated nuclear genome for *Pelargonium*. Generating a detailed fully annotated overview of the nuclear genomic compartment is long overdue. The same could be said for the mitochondrion though. As few are described for plants in general and they remain hard to study (Christensen 2021) but having one for *Pelargonium* would be great. Another interesting avenue would be to perform comparative transcriptomics to investigate the cyto-nuclear interactions in real time. It was demonstrated to yield interesting results when applied to Geraniaceae (Zhang et al. 2013). Performing a detailed study of the altered *Pelargonium* ribosomes (Chapter four) would be insightful and have implications that reach further than plants. Plant ribosomal structural models are becoming available, but so far mainly for the model organism *S. oleraceae* (Bieri et al. 2017, Perez-Boerema et al. 2018). Doing this would help us gain insight into practical topics such as protein synthesis (Levy et al. 2020) and antibiotic resistance (Davies 2010, Charon et al. 2019), but also into the fundamental evolutionary questions on life as we know it (Bowman et al. 2020, Moelling and Broecker 2021).

Significance of this study

This research was funded by the TTW Green genetics program. This project set out to develop markers to detect organelle types in hybrid plants, establish limits to crossing possibilities and generate possible mechanisms for explaining the occurrence of CNI.

In this thesis CNI was studied in detail and an attempt was made to integrate information from all three interacting genomes. The results from crossings will be useful not only to the scientific community but also to *Pelargonium* breeders. The insights in plastome evolution fill in gaps and opened new avenues of research (ribosome studies, low level taxonomic comparisons). The markers developed in this thesis, especially the KASP markers for mitochondrial detection can be used to study other *Pelargonium* material. Insights from this thesis can be useful when studying other groups of plants especially with the intent to incorporate crop wild relatives (CWR) with desirable traits in breeding programs. The observations and discoveries made in the course of this work warrant further detailed investigation, especially the altered ribosomes. The role of plastic encoded polymerase in explaining chlorosis may be of great importance in future breeding efforts.

Appendix

Comparative observations on variation between *Pelargonium* species with respect to their reproductive biology

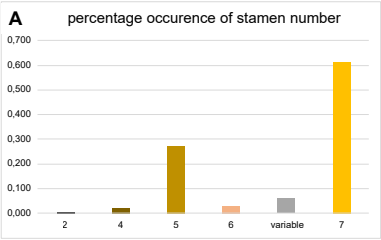
For five seasons (~9-11 months each) from 2016 to 2020 nearly all currently described *Ciconium* species were grown in semi-open greenhouse in NW Europe. Per species we often had dozens of individual plants growing. Because the greenhouse was semi-open our populations were partly exposed to the variable climate. This allowed us to observe behavior of each species under different environmental conditions. A study of the species-specific ecological preferences was not part of this thesis, but I feel it is important to document relevant biological observations, perhaps for use in forthcoming studies. I therefore discuss these observations here, ordered according to life stage, physiology, ontogeny etc.

Flowering strategies

All *Ciconium* species show different flowering strategies as the seasons progressed and they did so every year. The first flowering of most species was characterized by 'stamen first'. This means that the anthers would ripen and open, after which the style would stretch and in the process of doing so, the stigma would strike past the freshly opened anthers, thus achieving effective self-pollination. Seed set after this event was abundant. In the second, third and subsequent flowering series of a plant the sequence of events differed, even for species which appeared to be cleistogamous (*P. barklyi*, *aridum*, *elongatum* and *alchemilloides*). This could indicate an adaptation to the highly variable environment of the CFR where it pays off to quickly secure the next generation and if more than one series of flowering are possible the option for outcrossing is then pursued. This observation is further reinforced by the fact that the species from semi-deserts display this behavior the strongest (*P. acetosum*, *P. barklyi*, *P. aridum*, and *P. yemenense*).

Pollination requirements

The majority of phylogenetically distant, interspecific crosses in our experiments could be traced to one or two pollination events per season. The environmental conditions were very specific on these dates. The sky was semi-clouded, the temperature ranged between 18 and 22 °C and most importantly, the humidity was <40%. These conditions are rare in the Netherlands which has a prevailing sea-climate with relative humidity nearly always >50%. Further observations shows that it is difficult for two species (*P. acetosum* and *P. acraeum*) to obtain seeds from selfings. This could be by a preference of outcrossing rather than selfings (as they are excellent pollen donors) and it could indicate the presence of S-alleles. Another explanation is that these two species have narrower 'window of tolerance' with respect to conditions required for successful pollinations. It could also be that the requirements are different from those of all other *Ciconium* species. I believe both to be the case as they seem to require a higher temperature and a much lower humidity than the average *Ciconium*. One species (*P. yemenense*) stood out because it appeared as if it would not set seed at all under the greenhouse conditions we used. However, when the drip-irrigation was removed and watering was reduced to once a week, seed set suddenly exploded, suggesting adaptation to its original semi-desert environment (see also the Autecology paragraph in General Discussion).



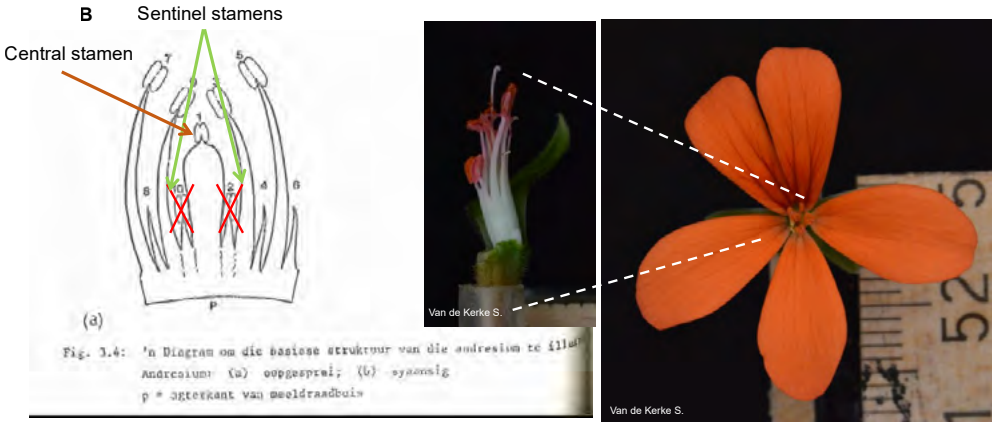


FIGURE 8 | Stamens in *Pelargonium*. A) Percentages of stamen numbers counts for 70 species. (van der Walt et al. 1988 vol I-II). B) stamen 'layout' of *Pelargonium*. *Pelargonium tongaense* photos from van de Kerke with permission and drawing is after Marais 1994, with modifications.

Stamen numbers across *Pelargonium* and section *Ciconium*

Pelargonium species contain varying number of stamens, based on 70 species discussed by van der Walt and Vorster (vols. I and II 1988) 60% of species has seven stamens, but a considerable minority has five (see fig. 8A). When this is the case the two stamens closest to the nectar tube opening (called 'sentinel stamens' here see Fig. 8B) are missing. The possible reasons for the evolution of this trait are beyond the scope of this thesis, but one species from section *Ciconium* (*P. acetosum*) also has five rather than the seven stamens that all other *Ciconium* species have (including the newly discovered species). This trait is stable within the *P. acetosum* population I used, and I never observed plants with a different number over four seasons. But when crossed with other species this trait disappears and all offspring of crosses from *P. acetosum* x *Ciconium* series displayed seven stamens. For one cross we obtained an F_2 population (F_2 HORT x ACET), and numerous 'grand parental' traits segregated in these populations (flower color and shape, leaf shape, plant form and more), but stamen number did not. The second generation also always contained seven stamens. The 'gene(s)' appear to have been silenced. What causes this is unknow to me, but it could be that epigenetic effects play a role. Subsequent generations (F_3, F_4 and further) should be created to evaluate this.

Flowering requirements

The majority of species flowered easily and as the seasons progressed nearly all would flower, even those that required heat. Still a few species stood out here by virtue of 'extreme' requirements when compared to other *Ciconium* species. The first one is *P. acetosum*, which requires abundant heat, or its flowering will be limited. Season 2018 was extremely hot (with temperatures >40°C in and even outside the greenhouse setting temperature records across NW Europe), but *P. acetosum* thrived and flowered abundantly allowing us to make many crossing attempts with this species. The second species that stands out is *P. articulatum*. We were never able to bring this species to continued flowering under normal greenhouse conditions. Experiments with this species showed that a cold period of at least two months is required for it to start flowering. The third species worth mentioning is *P. acraeum*. This little-known species would also not flower in the greenhouse. Suspicions by breeders and a search in online repositories (http://pza.sanbi.org/pelargonium-acraeum) suggested it may need bright light (and higher temperatures than others, but not as high as for *P. acetosum*) as it originates from montane environments (van der Walt and Vorster 1988). Experiments with this species confirmed this. *Pelargonium acraeum* needs bright light to flower. Interestingly, it does not grow well when it is just exposed to the sun and requires shade for part of the day to develop healthy.

References

A

Alix K, Gérard PR, Schwarzacher T, Heslop-Harrison JSP. (2017). Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. Ann. Bot, 120(2):183-194. Doi: 10.1093/aob/mcx079.

Allen JF. (2015). CoRR hypothesis for genes in organelles. Proc. Nat. Acad. Sci. USA: 112 (33) 10231-10238. Doi: 10.1073/pnas.1500012112.

Apitz J, Weihe A, Pohlheim F, Börner T. (2013). Biparental inheritance of organelles in *Pelargonium*: evidence for intergenomic recombination of mitochondrial DNA. *Planta* 237, 509–515. Doi: 10.1007/s00425-012-1768-x.

Arrieta-Montiel M.P., Mackenzie S.A. (2011) Plant mitochondrial genomes and recombination. In: Kempken F. (eds) Plant Mitochondria. Advances in Plant Biology, vol 1. Springer, New York, NY. Doi: 10.1007/978-0-387-89781-3_3.

B

Bakker FT, Hellbrügge D, Culjam A, Gibby M, (1998). Phylogenetic relationships within *Pelargonium* sect. *Peristera* (Geraniaceae) inferred from nrDNA and cpDNA sequence comparisons. Syst. Evol. 211, 273-287. Doi: 10.1007/BF00985364.

Bakker FT, Culham A, Daugherty LC, Gibby M. (1999a). A trnL- F based phylogeny for species of *Pelargonium* (Geraniaceae) with small chromosomes. Plant Syst. Evol. 216:309–324. Doi: 10.1007/BF01084405.

Bakker FT, Culham A, Gibby M. (1999b). Phylogenetics and diversification in *Pelargonium*. In: Hollingsworth P, Bateman R, Gornall R editors. Molecular systematics and plant evolution. Chapman and Hall, London, pp 353–374.

Bakker FT, Culham A, Pankhurst CE, Gibby M (2000) Mitochondrial and chloroplast DNA based phylogeny of *Pelargonium* (Geraniaceae). Am. J. Bot. 87:727–734. Doi: 10.2307/2656859.

Bakker FT, Culham A, Hettiarachi P, Touloumenidou T, Gibby M. (2004). Phylogeny of *Pelargonium* (Geraniaceae) based on DNA sequences from three genomes. Taxon 53: 17–28. Doi: 10.2307/4135485.

Bakker FT, Culham A, Marais EM, Gibby M (2005) Nested radiation in Cape *Pelargonium*. In: Bakker FT, Chatrou LW (eds) Plant species-level systematics: new perspectives on pattern and process. Koeltz Scientific Books, Koenig- stein, pp 75–100.

Bakker FT, Breman FC, Merckx V. (2006). DNA sequence evolution in fast evolving mitochondrial DNA nad1 exons in Geraniaceae and Plantaginaceae. Taxon 55:887–896. Doi: 10.2307/25065683.

Baluška F, Lyons S. (2021). Archaeal origins of eukaryotic cell and nucleus. BioSystems 203: 104375. Doi: 10.1016/j.biosystems.2021.104375.

Barkan A, Small I. (2014). Pentatricopeptide repeat proteins in plants. Annu. Rev. Plant Biol. 65, 415–42. Doi: 10.11646/phytotaxa.159.2.1

Barnard-Kubow KB, So N, Galloway LF. (2016). Cytonuclear incompatibility contributes to the early stages of speciation. *Evolution* 70 (12), 2752-2766. Doi: 10.1111/evo.13075.

Barnard-Kubow KB, McCoy MA, Galloway LF. (2017). Biparental chloroplast inheritance leads to rescue from cytonuclear incompatibility. New Phytol. 213(3), 1466-1476. Doi: 10.1111/nph.14222

Barr CM., Neiman M, Taylor DR. (2005). Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. New Phytol. 168:39–50. Doi: 10.1111/j.1469-8137.2005.01492.x.

Barr CM, Fishman L. (2010). The nuclear component of a cytonuclear hybrid incompatibility in *Mimulus* maps to a cluster of pentatricopeptide repeat genes. Genetics 184, 455– 465. Doi: 10.1534/genetics.109.108175.

Bateson W. (1909). Heredity and variation in modern lights. In Darwin and modern science (ed. A. C. Seward), pp. 85– 101. Cambridge, UK: Cambridge University Press.

Baur E. (1909). Das Wesen und die Earblichkeitsverhältnisse der 'arietates albomarginatae hort' von *Pelargonium zonale*. Z. Indukt. Abstammungs-Vererbungsl. 1, 330–351.

Benkert, P., Biasini, M., Schwede, T. (2011) Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27, 343-350. Doi: 10.1093/bioinformatics/btq662.

Berlin, B. (1992). Ethnobiological classification: principles of categorization of plants and animals in traditional societies. Princeton, New Jersey: Princeton University Press. Retrieved May 22, 2021, from http://www.jstor.org/stable/j.ctt7ztq5q.

Bertoni M, Kiefer F, Biasini M, Bordoli L, Schwede T. (2017). Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. Scientific Reports 7 10480. Doi: 10.1038/s41598-017-09654-8.

Bhattacharya D, Archibald, JM, Weber AP, Reyes-Prieto A. (2007). How do endosymbionts become organelles? Understanding early events in plastid evolution. Bioessays 29: 1239–1246. Doi: 10.1002/bies.20671.

Bieri P, Leibundgut M, Saurer M, Boehringer D, Ban N. (2017). The complete structure of the chloroplast 70S ribosome in complex with translation factor pY. The EMBO Journal 36(4): 475-486. Doi: 10.15252/embj.201695959.

Blanchard JL, Lynch M. (2000). Organellar genes - why do they end up in the nucleus? Trends Genet., 16: 315-320. Doi: 10.1016/S0168-9525(00)02053-9.

Blazier JC, Jansen RK, Mower JP, Govindu M, Zhang J. et al. (2016a). Variable presence of the inverted repeat and plastome stability in *Erodium*. Ann. of Bot. 117, 1209–1220. 10.1093/aob/mcw065.

Blazier JC, Ruhlman TA, Weng M-L, Rehman SK, Sabi JSM, et al. (2016b). Divergence of RNA polymerase a subunits in angiosperm plastid genomes is mediated by genomic rearrangement. Sci. Rep. 6, 24595. Doi : 10.1038/srep24595.

Börner T, Aleynikova AY, Zubo YO, Kusnetsov VV. (2015). Chloroplast RNA polymerases: role in chloroplast biogenesis. Biochimica et Biophysica Acta 1847: 761–769

Britten RJ, Graham DE, Neufeld BR. (1974). Analysis of repeating DNA sequences by reassociation methods. Enzymology 29:363–418. Doi: 10.1016/0076-6879(74)29033-5.

Bowman JC, Hud NV, Williams LD. (2015). The ribosome challenge to the RNA world. J. Mol. Evol.:143-61. Doi: 10.1007/s00239-015-9669-9.

Bowman JC, Petrov AS, Frenkel-Pinter M, Penev PI, Williams LD. (2020) Root of the tree: the significance, evolution, and origins of the ribosome. Chem. Rev. 10;120(11):4848-4878. Doi: 10.1021/acs.chemrev.9b00742.

Breman FC, Snijder RC, Korver JW, Pelzer S, Sancho-Such M, Schranz ME and Bakker FT (2020). Interspecific hybrids between *Pelargonium* × *hortorum* and species From *P.* Section *Ciconium* reveal biparental plastid inheritance and multi-locus cyto-nuclear incompatibility. Front. Plant Sci. 11:614871. doi: 10.3389/fpls.2020.614871.

Buggs RJA, Chamala S, Wu W, Tate JA, Schnable PS, et al. (2012). Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. Curr. Biol. 22: 248–52. Doi: 10.1016/j.cub.2011.12.027.

Burr B, Barthlott W. (1993). Untersuchungen zur Ultrareflexion von Angiospermenblüten II. Magnoliidae, Ranunculidae, Hamamelididae, Caryophyllidae, Rosidae. Tropische und subtropische Pflanzenwelt 87: 1–193.

Burton RS, Barreto FS. (2012). A disproportionate role for mtDNA in Dobzhansky-Müller incompatibilities? Mol Ecol.;21(20):4942-57. Doi: 10.1111/mec.12006.

Burton RS, Pereira RJ, Barreto FS. (2013). Cytonuclear genomic interactions and hybrid breakdown. Annu. Rev. Ecol. Evol. Syst. 44:281–302. Doi: 10.1146/annurev-ecolsys-110512-135758.

Bychkov IA., Andreeva, A.A., Kudryakova, NV, Pojidaeva ES. (2021). Nuclear-encoded plastid RNA polymerases are components of anterograde control in hormonal regulation of chloroplast gene expression. Russ. J. Plant Physiol. 68, 228–237. doi: 10.1134/S1021443721020023.

C

de Candolle AP. (1824). Geraniaceae. Prodromus systematis naturalis regni vegetabilis, vol. 1. Treutel & Wuertz, Paris, pp. 637–682

Cappadocia L, Maréchal A, Parent JS, Lepage E, Sygusch J, Brisson N.(2010). Crystal structures of DNA-Whirly complexes and their role in *Arabidopsis* organelle genome repair. Plant Cell.;22(6):1849-67. doi: 10.1105/tpc.109.071399.

Charon J, Manteca A, Innis CA. (2019). Using the bacterial ribosome as a discovery platform for peptide-based antibiotics. Biochemistry 15;58(2):75-84. Doi: 10.1021/acs.biochem.8b00927.

Chen C, Zhiguo E, Hong-Xuan L. (2016). Evolution and molecular control of hybrid incompatibility in plants. Fr. Pl. Sci. 7:1208. Doi: 10.3389/fpls.2016.01208.

Chi W, Ma J, Zhang D, Guo J, Chen F, Lu C, Zhang L. (2008). The Pentratricopeptide Repeat Protein DELAYED GREENING1 Is involved in the regulation of early chloroplast development and chloroplast gene expression in *Arabidopsis*. Plant Physiology. 147, 573–584. Doi: 10.1104/pp.108.116194.

Choi K, Weng M-L, Ruhlman TA, Jansen RK. (2021). Extensive variation in nucleotide substitution rate and gene/intron loss in mitochondrial genomes of *Pelargonium*. Mol. Phyl. Evol. 155, (106986). Doi: 10.1016/j.ympev.2020.106986.

Christensen AC. (2021). Plant mitochondria are a riddle wrapped in a mystery inside an enigma. J Mol. Evol. 89, 151–156. Doi: 10.1007/s00239-020-09980-y.

Chumley TW, Palmer JD, Mower JP, Fourcade MH, Calie PJ, et al. (2006). The complete chloroplast genome sequence of *Pelargonium* x *hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol. Biol. Evol. 23(11), 2175–2190. Doi: 10.1093/molbev/msl089.

Chateigner-Boutin AL, Ramos-Vega M, Guevara-Garcia A, Andres C, de la Luz G-NM, et al. (2008). CLB19, a pentatricopeptide repeat protein required for editing of rpoA and clpP chloroplast transcripts. Plant J. 56, 590–602. 10.1111/j.1365-313X.2008.03634.x

Cortés AJ, Blair MW. (2018). Genotyping by sequencing and genome–environment associations in wild common bean predict widespread divergent adaptation to drought. Front. Plant Sci. 9:128. Doi: 10.3389/fpls.2018.00128.

Cowling RM, Lombard AT. (2002). Heterogeneity, speciation/ extinction history and climate: explaining regional plant diversity patterns in the Cape Floristic region. Diversity & Distributions 8: 163 – 179. Doi: 10.1046/j.1472-4642.2002.00143.x.

Craig NL, Chandler M, Gellert M, Lambowitz AM, Rice PA, et al. (2015). Mobile DNA III. 3rd ed. Washington, DC: American Society for Microbiology (ASM).

Cruz JA, Savage LJ, Zegarac R, Hall CC, Satoh-Cruz M, et al. (2016). Dynamic environmental photosynthetic imaging reveals emergent phenotypes. Cell systems 2(6), 365-377. Doi: 10.1016/j.cels.2016.06.001.

D

Darling ACE, Mau B, Blattner FR, Perna NT. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Research 14: 1394–1403. Doi: 10.1101/gr.2289704.

Darwin, C. (1859). On the origin of species by means of natural selection. London: Murray. p. 48. ISBN 978-84-206-5607-6.

Davies J, Davies L. (2010). Origins and Evolution of Antibiotic Resistance. Microbiol. Mol. Biol. Rev.;74(3):417-33. Doi: 10.1128/MMBR.00016-10.

Dawood MAO, El Basuini MF, Zaineldin AI, Yilmaz S, Hasan MT, et al. (2021). Antiparasitic and antibacterial functionality of essential oils: an alternative approach for sustainable aquaculture. Pathogens; 10(2):185. Doi: 10.3390/pathogens10020185.

Demarsy E, Courtois F, Azevedo J, Buhot L, Lerbs-Mache S. (2006). Building up of the plastid transcriptional machinery during germination and early plant development. Pl. Physiol. 142: 993–1003. Doi: 10.1104/pp.106.085043.

Deng J, Hondo K, Kakihara F, Masahira K. (2004). Seasonal variation of flowering characteristics and high-temperature stress tolerance in Geraniums. J. of high technology in agriculture 16(4), 173-182. doi.org/10.2525/jshita.16.173.

Deniz O, Frost JM, Branco MR. 2019. Regulation of transposable elements by DNA modifications. Nat. Rev. Genet. 20, 417–431. Doi: 10.1038/s41576-019-0106-6.

Dobzhansky T. (1936) Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. Genetics 21, 113–135.

Dodsworth S, Chase MW, Kelly LJ, Leitch IJ, Macas J, et al. (2015). Genomic repeat abundances contain phylogenetic signal. Syst. Biol. Doi: 10.1093/sysbio/syu080.

Dodsworth S, Jang TS, Struebig M, Chase MW, Weiss-Schneeweiss H, et al. (2017). Genome-wide repeat dynamics reflect phylogenetic distance in closely related allotetraploid *Nicotiana* (Solanaceae). Pl. Syst. and Evol. Doi: 10.1007/s00606-016-1356-9.

Du L, Zhang J, Qu S, Zhao Y, Su B, et al. (2017). The pentatricopeptide repeat protein pigment-defective mutant2 is Involved in the regulation of chloroplast development and chloroplast gene expression in *Arabidopsis*. *Plant and Cell Physiology*. 58(4), pp. 747–759. Doi: 10.1093/pcp/pcx004.

E

Eckardt NA. (2001). A sense of self: the role of DNA sequence elimination in allopolyploidization. Plant Cell 13:1699–1704. Doi: 10.1105/tpc.13.8.1699.

Ecklon CF, Zeyher KL, (1835). Enumeratio plantarum africae australis extratropicae 1. Perth & Besser, Hamburg.

Elliott TA, Gregory TR. (2015). What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. Phil. Trans. Roy. Soc. B 26;370(1678):20140331. Doi: 10.1098/rstb.2014.0331.

Enriquez-Gasca R, Gould PA, Rowe HM. (2020). Host gene regulation by transposable elements: the new, the old and the ugly. Viruses 12, 1089. Doi: 10.3390/v12101089.

F

Freeling M, Woodhouse MR, Subramaniam S, Turco G, Lisch D, et al. (2012). Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. Curr. Opin. Plant Biol. 15:131–139. Doi: 10.1016/j.pbi.2012.01.015.

Flavell RB. Bennett MD, Smith JB, Smith DB. (1974). Genome size and the proportion of repeated nucleotide sequence DNA in Plants. Biochem. Genet 12:257–269. Doi: 10.1007/BF00485947.

Flint-Garcia SA. (2013). Genetics and consequences of crop domestication. J Agric. Food Chem.;61(35):8267-76. Doi: 10.1021/jf305511d.

Flood PJ, Harbinson J, Aarts MGM. (2011). Natural genetic variation in plant photosynthesis. *Trends Plant Sci.* 16, 327–335. Doi: 10.1016/j.tplants.2011.02.005

Flood PJ, Theeuwen TPJM, Schneeberger K, Keizer P, Kruijer W, Severing et al. (2020). Reciprocal cybrids reveal how organellar genomes affect plant phenotypes. *Nature Plants* 6, 13–21. Doi: 10.1038/s41477-019-0575-9.

Forsythe ES, Williams AM, Sloan DB. (2021). Genome-wide signatures of plastid-nuclear coevolution point to repeated perturbations of plastid proteostasis systems across angiosperms. The Plant Cell: 00: 1–18. Doi: 10.1093/plcell/koab021.

G

Galinsky KJ, Bhatia G, Loh PR, Georgiev S, Mukherjee S, et al. 2016. Fast principal-component analysis reveals convergent evolution of ADH1B in Europe and East Asia. Am J Hum Genet. 98:456-472. doi: 10.1016/j.ajhg.2015.12.022.

GBIF.org (2020), GBIF Home Page. Available from: https://www.gbif.org, https://www.gbif.org/occurrence/search?q=Pelargonium (accessed September 2020)

Gibby M, Westfold J. (1986). A cytological study of *Pelargonium* sect. *Eumorpha*. *Syst. Evol.* 153, 205 - 222. 10.1007/BF00983688.

Gibby M, Albers F, Prinsloo B. (1990). Karyological studies in *Pelargonium* sectt. *Ciconium*, *Dibrachya*, and *Jenkinsonia* (Geraniaceae). *Plant Systematics and Evolution* 170, 151–159. Doi: 10.1007/BF00937700.

Gibby M, Hinnah S, Marais EM, Albers F. (1996). Cytological variation and evolution within *Pelargonium* Section *Hoarea* (Geraniaceae). P1. Syst. Evol. 203:111-14. Doi: 10.1007/BF00985241.

Goloboff PA, Mattoni CI, QuinterosAS. 2006. Continuous characters analyzed as such. Cladistics 22:589–601. Doi: 10.1111/j.1096-0031.2006.00122.x

Goloboff PA, Farris JS, Nixon KC. 2008. TNT, a free program for phylocladistics 22:589–601. Doi: 10.1111/j.1096-0031.2008.00217.x.

Goloboff P, Catalano S. (2016) TNT version 1.5 with a full implementation of phylogenetic morphometrics. Cladistics 32(3): 221-238. Doi: 10.1111/cla.12160.

Gonçalves DJP, Jansen RK, Ruhlman TA, Mandel JR. (2020). Under the rug: Abandoning persistent misconceptions that obfuscate organelle evolution. Mol. Phyl. Evol., 151;106903. Doi: 10.1016/j.ympev.2020.106903.

Gray M.W., Archibald J.M. (2012) Origins of Mitochondria and Plastids. In: Bock R., Knoop V. (eds) Genomics of chloroplasts and mitochondria. Advances in photosynthesis and respiration (Including bioenergy and related processes), vol 35. Springer, Dordrecht. doi: 10.1007/978-94-007-2920-9_1

Greiner S, Wang X, Herrmann RG, Mayer K, Haberer G, et al. (2008a). The complete nucleotide sequences of the 5 genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: II. A microevolutionary view using bioinformatics and formal genetic data. *Molecular Biology and Evolution*. 25(9), 2019–2030. Doi: 10.1093/molbev/msn149.

Greiner S, Wang X, Rauwolf U, Silber MV, Meurer J, et al. (2008b). The complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: I. Sequence evaluation and plastome evolution. *Nucleic Acids Research*. 36(7), 2366–2378. Doi: 10.1093/nar/gkn081

Greiner S, Rauwulf, U, Meurer J, Hermann RG. (2011). The role of plastids in plant speciation. Mol. Ecol. 20, 671–691. Doi: 10.1111/j.1365-294X.2010.04984.x

Greiner S, Bock R. (2013). Tuning a ménage à trois: Co-evolution and co-adaptation of nuclear and organellar genomes in plants. *Bioessays* 35, 354–365. Doi: 10.1002/bies.201200137

Greiner S, Sobanski J, Bock R. (2015). Why are most organelle genomes transmitted maternally? Bioessays 37: 80–94. 10.1002/bies.201400110

Grewe F, Zhu A, Mower JP. (2016). Loss of a Trans-Splicing nad1 intron from Geraniaceae and transfer of the maturase gene *matR* to the nucleus in *Pelargonium*. Genome Biol Evol. 30;8(10):3193-3201. doi: 10.1093/gbe/evw233.

Guisinger MM, Kuehl JV, Boore JL, Jansen RK. (2008). Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. *Proc. Natl. Acad. Sci. U S A.* 105(47), 18424–18429. Doi: 10.1073/pnas.0806759105

Gualberto JM, Mileshina D, Wallet DC, Niazi AK, Weber-Lotfi F, et al. (2014). The plant mitochondrial genome: dynamics and maintenance. Biochimie, 100: 107-120. Doi: 10.1016/j.biochi.2013.09.016.

Guisinger MM, Kuehl JV, Boore JL, Jansen RK. (2011). Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol Biol Evol. 28(1): 583–600. 10.1093/molbev/msq229

Guo FL, Hu SY. (1995). Cytological evidence of biparental inheritance of plastids and mitochondria in *Pelargonium*. Protoplasma. 186:201-207. Doi: 10.1007/BF01281330.

Gupta N, Van der Heiden J, Uduman M, Gadala-Maria D, Yaari G, et al. (2015). "Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data." Bioinformatics, 1-3. doi: 10.1093/bioinformatics/btv359.

H

Hanada K, Vallejo V, Nobuta K, Slotkin RK, Lisch, D, et al. (2019). The functional role of Pack-MULEs in rice inferred from purifying selection and expression profile. The Plant Cell, 21: 25–38. Doi: 10.1105/tpc.108.063206Hartley G, O'Neill RJ. 2019. Centromere repeats: hidden gems of the genome. Genes, 10, 223. Doi: 10.3390/genes10030223.

Hansen AK, Escobar LK, Gilbert LE, Jansen RK. (2007). Paternal, maternal, and biparental inheritance of the chloroplast genome in *Passiflora* (Passifloraceae): implications for phylogenetic studies. Am. J. of Bot. 94(1), 42–46. 10.3732/ajb.94.1.42.

Harvey WH. (1860). Geraniaceae. In: Harvey, WH. & Sonder OW. (eds.), Flora capensis, Vol. 1. Hodges, Smith & Co, Dublin, pp. 259–308

Havey MJ. (1997). Predominant paternal transmission of the mitochondrial genome in cucumber. J. Hered. 88:232–235. Doi: 10.1093/oxfordjournals.jhered.a023094.

He S, Yang Y, Li Z, Wang X, Guo Y, Wu H. (2020). Comparative analysis of four *Zantedeschia* chloroplast genomes: expansion and contraction of the IR region, phylogenetic analyses and SSR genetic diversity assessment. PeerJ 8:e9132. Doi: 10.7717/peerj.9132.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR. (2003). Biological identifications through DNA barcodes. Proc R Soc Lond B 270:313–32. Doi: 10.1098/rspb.2002.2218.

Hey J. (2006). On the failure of modern species concepts. Trends in Ecology & Evolution 21(8): 447-450. Doi: 10.1016/j.tree.2006.05.011.

Hiremath PJ, Kumar A, Penmetsa RV, Farmer A, Schlueter JA, et al. (2012). Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. Plant Biotechnol. J. 10:716–32. 10.1111/j.1467-7652.2012.00710.x.

Hondo K, Iwasaki C, Kakihara F. (2014). Characteristics of progeny obtained from the cross between Geraniums (*Pelargonium* x *hortorum*) and wild species in section *Ciconium*, Including heat tolerance and odor of leaves. *Environ. Control Biol*. 52 (1), 37-43. ttps://doi/irg. /10.2525/ecb.52.37.

Hondo K, Sukhumpinij P, Kakihara F. (2015). Flower color and pigments in yellow-flowered hybrid progeny raised from the interspecific cross *Pelargonium quinquelobatum* × white-flowered geraniums. Scientia Horticulturae 195: 145-153. Doi: 10.1016/j.scienta.2015.09.014

Horn W. (1994). Interspecific crossability and inheritance in *Pelargonium*. *Plant Breeding*. 113(1) 3–17. Doi: 10.1111/j.1439-0523.1994.tb00696.x.

Horn R, Gupta KJ, Colombo N. (2014) Mitochondrion role in molecular basis of cytoplasmic male sterility. Mitochondrion. 19:198-205. doi: 10.1016/j.mito.2014.04.004.

Hu YC, Zhang Q, Rao GY, Sodmergen. (2008). Occurrence of plastids in the sperm cells of Caprifoliaceae: biparental plastid inheritance in angiosperms is unilaterally derived from maternal inheritance. *Plant and Cell Physiology*. 49(6), 958–968. Doi: 10.1093/pcp/pcn069.

Huson DH, Bryant D. (2006). Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23(2):254-267. Doi: 10.1093/molbev/msj030.

I

Isemer R, Mulisch M, Schäfer A, Kirchner S, Koop H-U, et al. (2012). Recombinant Whirly1 translocates from transplastomic chloroplasts to the nucleus. *FEBS Letters*. 586() 85–88. doi:10.1016/j.febslet.2011.11.029.

Islam MS, Thyssen GN, Jenkins JN, Fang DD. (2015). Detection, validation, and application of genotyping-by-sequencing based single nucleotide polymorphisms in upland Cotton. Plant Genome 8:1–10. Doi: 10.3835/plantgenome2014.07.0034.

J

James, CM, Gibby M, Barrett JA. (2004). Molecular studies in *Pelargonium* (Geraniaceae). A taxonomic appraisal of section *Ciconium* and the origin of the "Zonal" and "Ivy-leaved" cultivars. *Plant Syst. Evol.* 243, 131–146. 10.1007/s00606-003-0074-2.

Jiang M, Chen H, He S, Wang L, Chen AJ, et al. (2018). Sequencing, characterization, and comparative analyses of the plastome of *Caragana rosea* var. *rosea* Int. J. Mol. Sci. 19, 1419. Doi:10.3390/ijms19051419.

Jin JJ, Yu W-B, Yang J-B, Song Y, dePamphilis CW, et al. (2019). GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. doi: 10.1101/256479.

Johnston JS, Pepper AE, Hall AE, Chen ZJ, Hodnett G, et al. (2005). Evolution of genome size in Brassicaceae. Ann. Bot. 95:229-235. Doi: 10.1093/aob/mci016.

Johnston IG. (2019). Tension and resolution: dynamic, evolving populations of organelle genomes within plant cells. Mol. Plant. 12, 764–783. Doi: 10.1016/j.molp.2018.11.002.

Jones CS, Cardon ZG, Czaja AD. (2003). A phylogenetic view of low-level CAM in *Pelargonium* (Geraniaceae). Am. J. Bot. 90(1):135-42. Doi: 10.3732/ajb.90.1.135.

Jones C, Bakker FT, Schlichting CD, Nicotra AB. (2009). Leaf shape evolution in the south African genus *Pelargonium* l'Hér. (Geraniaceae). Evolution 63-2:479–497. Doi: 10.1111/j.1558-5646.2008.00552.x.

Jurka J, Bao W, Kojima KK. (2011). Families of transposable elements, population structure and the origin of species. Biol. Direct 6, 44. Doi: 10.1186/1745-6150-6-44.

Jurka J, Bao W, Kojima KK, Kohany O, Yurka MG. (2012). Distinct groups of repetitive families preserved in mammals correspond to different periods of regulatory innovations in vertebrates. Biol. Direct 7, 36. Doi: 10.1186/1745-6150-7-36.

K

Ka S, Quinton-Tulloch MJ, Amgai RB, et al. (2018). Accelerating public sector rice breeding with high-density KASP markers derived from whole genome sequencing of indica rice. Mol Breed 38:38. Doi:10.1007/ s11032-018-0777-2. PMid: 29563850

Kamlah R, Pinker I, Plaschil S, Olbricht K. (2019). Hybridization between *Pelargonium acetosum* L'Hér. and *Pelargonium × peltatum*. *J. of Applied Botany and Food Quality*. 92, 49 – 56, doi 10.5073/JABFQ.2019.092.007.

Kassambara A, Mundt F. (2020). factoextra: Extract and visualize the results of multivariate data analyses. R package version 1.0.7. https://CRAN.R-project.org/package=factoextra.

Katoh K, Rozewicki J, Yamada KD. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in Bioinformatics 20:1160-1166. Doi: 10.1093/bib/bbx108.

van de Kerke, SJ, Shrestha B, Ruhlman TA, Weng M-L, Jansen RK, et al. (2019). Plastome based phylogenetics and younger crown node age in *Pelargonium*. *Molecular Phylogenetics and Evolution* 137, 33–43. 10.1016/j.ympev.2019.03.021.

van de Kerke SJ. (2019). Exploring floral evolution in *Pelargonium* (Geraniaceae) linking shapes and macro-evolution. PhD thesis Wageningen UR, the Netherlands. Doi: 10.18174/497382.

van de Kerke SJ, van Engelenhoven T, van Es AL, Schat L, van Son LM, et al. (2020). Capturing variation in floral shape: a virtual3D based morphospace for *Pelargonium*. PeerJ 8:e8823 doi: 10.7717/peerj.8823.

Kim N, Jinks-Robertson S. (2012). Transcription as a source of genome instability. Nat Rev Genet 13, 204–214. Doi: 10.1038/nrg3152.

Kirk JTO, Tilney-Bassett RAE. (1967). The plastids. Freeman and Co., London.

Kirk JTO, Tilney-Bassett RAE. (1978). The plastids. Their chemistry, structure, growth and inheritance, 2nd edn, pp. 1–960. Elsevier, Amsterdam.

Knuth R. (1912). Geraniaceae. In: Engler, A. (eds.), Das Pflanzenreich IV.129. Engelmann, Leibzig, pp. 640

Koo D-H, Hong CP, Batley J, Chung YS, Edwards D, et al. (2011). Rapid divergence of repetitive DNAs in *Brassica* relatives. Genomics 97:173–185. Doi: 10.1016/j.ygeno.2010.12.002.

Koonin EV, Novozhilov AS. (2009). Origin and evolution of the genetic code: the universal enigma. IUBMB Life;61(2):99-111. doi: 10.1002/iub.146.

Krausea K, Kilbienski I, Mulisch M, Rödiger A, Schäfer A, Krupinska K. (2005). DNA-binding proteins of the Whirly family in *Arabidopsis thaliana* are targeted to the organelles. *FEBS Letters* 579(17), 3707-3712. 10.1016/j.febslet.2005.05.059

Krupinska K, Oetke S, Desel C, Mulisch M, Schäfer A, Hollmann J, Kumlehn J, Hensel G. (2014). WHIRLY1 is a major organizer of chloroplast nucleoids. Front. Pl. Sci. (5): 432. 10.3389/fpls.2014.00432.

Krupinska K, Braun S, Nia MS. Schäfer A, Hensel G, et al. (2019). The nucleoid-associated protein WHIRLY1 is required for the coordinate assembly of plastid and nucleus-encoded proteins during chloroplast development. Planta 249, 1337–1347doi:. 10.1007/s00425-018-03085-zde Laat, A, Godhe, W, Vogelzang, M, (1987). Determination of ploidy of single plants and populations by flow cytometry. Plant Breeding 99(4), 303-307. 10.1111/j.1439-0523.1987.tb01186.x.

Ku C, Nelson-Sathi S, Roettger M, Sousa FL, Lockhart PJ, et al. (2015). Endosymbiotic origin and differential loss of eukaryotic genes. Nature. 27;524(7566):427-32. Doi: 10.1038/nature14963.

Kubo, T., Arakawa, T., Honma, Y., & Kitazaki, K. (2020). What does the molecular genetics of different types of restorer-of-fertility genes imply? Plants, 9(3), 361. doi: 10.3390/plants9030361.

Kuligowska K, Lütken H, Müller R. (2016). Towards development of new ornamental plants: status and progress in wide hybridization. Planta.;244(1):1-17. Doi: 10.1007/s00425-016-2493-7.

Kumar S., Stecher G., and Tamura K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 33:1870-1874.

Kuroiwa, H., Kuroiwa, T. (1992). Giant mitochondria in the mature egg cell of *Pelargonium zonale*. Protoplasma 168, 184–188. Doi: 10.1007/BF01666264.

Kuroiwa T, Kawazu T, Ushida H, Ohta T, Kuroiwa H. (1993). Direct evidence of plastid DNA and mitochondrial DNA in sperm cells in relation to biparental inheritance of organelle DNA in *Pelargonium zonale* by fluorescence/electron microscopy. Eur. J. Cell Biol. 62:307–313

L

Landis JB, Soltis DE, Li Z, Marx HE, Barker MS, et al. (2018). Impact of whole-genome duplication events on diversification rates in angiosperms. Am. J. Bot. 105(3): 348–363. doi:10.1002/ajb2.1060.

Lerbs-Mache S. (2011). Function of plastid sigma factors in higher plants: regulation of gene expression or just preservation of constitutive transcription? Plant Molecular Biology 76, 235–249. Doi: 10.3389/fpls.2014.00316.

Le S, Josse J, Husson F. 2008. FactoMineR: An R package for multivariate analysis. J. Statistical Software, 25:1-18. Doi: 10.18637/jss.v025.i01.

Leister D. (2005). Genomics-based dissection of the cross-talk of chloroplasts with the nucleus and mitochondria in *Arabidopsis*. Gene. 18;354:110-6. doi: 10.1016/j.gene.2005.03.039.

Leitch AR, Leitch IJ. 2008. Genomic plasticity and the diversity of polyploid plants. Science 320:481–483. doi:10.1126/science.1153585.

Leitch IJ, Beaulieu JM, Chase MW, Leitch AR, Fay MF. 2010. Genome size dynamics and evolution in monocots. J. Bot: 1–18. doi: 10.1155/2010/862516.

Leitch IJ, Johnston E, Pellicer J, Hidalgo O, Bennett MD. (2019). Plant DNA C-values Database (release 7.1, April 2019).

Lerbs-Mache S. (2011). Function of plastid sigma factors in higher plants: regulation of gene expression or just preservation of constitutive transcription? Plant Mol. Biol., 76 :235. Doi: 10.1007/s11103-010-9714-4.

Levy M, Falkovich R, Daube SS, Bar-Ziv RH. (2020). Autonomous synthesis and assembly of a ribosomal subunit on a chip. Science Advances 6(16): eaaz6020. Doi: 10.1126/sciadv.aaz6020.

LGC Genomics. (2013). KASP genotyping chemistry: User guide and manual. 9. http://www.lgcgenomics.com/genotyping/kasp-genotyping-reagents/how-does-kasp-work/

Li D, Qi X, Li X, Li L, Zhong C, et al. (2013). Maternal inheritance of mitochondrial genomes and complex inheritance of chloroplast genomes in *Actinidia* Lind.: evidences from interspecific crosses. *Mol. Genet. Genomics*. 288, 101–110. 10.1007/s00438-012-0732-6

Linder HP. (1991). Environmental correlates of patterns of species richness in the south-western Cape province of South Africa. Journal of Biogeography 18: 509 – 518.

Linder R, Meadows ME, Cowlino RM. (1992). History of the Cape flora. - In Cowlino R. editor. The ecology of fynbos. - Cape Town: Oxford University Press.

Linder HP, (2003). The radiation of the Cape flora, southern Africa. Biol. Rev. 78, 597–638. Doi: 10.1017/S1464793103006171.

Linnaeus C. (1758). Systema naturae vol I. Species plantarum, exhibentes plantas rite cognitas ad genera relatas cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas. Published by Vindobonae: Typis Joannis Thomae de Trattner.

Lis-Balchin MT. (1996). A chemotaxonomic reappraisal of the section *Ciconium Pelargonium* (Geraniaceae)." S. Afr. J. Bot. Doi: 10.1016/S0254-6299(15)30657-8.

Lis-Balchin M. (1997). A chemotaxonomic study of the *Pelargonium* (Geraniaceae) species and their modern cultivars. Journal of Horticultural Science 72: 791–95. Doi: 10.1080/14620316.1997.11515572.

Llorens C, Futami R, Covelli L, Domínguez-Escribá L, Viu JM, et al. (2011). The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. Nucleic Acids Res. 39:70–74. Doi: 10.1093/nar/gkq1061.

Livingstone K, Rieseberg L. (2003). Chromosomal evolution and speciation: a recombination-based approach. New Phytologist 161: 107–112. Doi: 10.1046/j.1469-8137.2003.00942.x.

Lloyd AH, Rousseau-Gueutin M, Timmis JN, Sheppard AE, Ayliffe MA. (2012) Promiscuous organellar DNA. In: Bock R., Knoop V. (eds) Genomics of chloroplasts and mitochondria. Advances in Photosynthesis and Respiration (Including Bioenergy and Related Processes), vol 35. Springer, Dordrecht. Doi: 10.1007/978-94-007-2920-9_9

Lohani N, Singh MB, Bhalla PL. (2020). High temperature susceptibility of sexual reproduction in crop plants. J. of Experimental Botany. 71(2), 555–568. doi:10.1093/jxb/erz426.

Louzada S, Lopes M, Ferreira D, Adega F, Escudeiro A, et al. (2020). Decoding the role of satellite DNA in genome architecture and plasticity—An evolutionary and clinical affair. Genes, 11. doi: 10.3390/genes11010072.

M

Mackenzie SA, Kundariya H. (2019). Organellar protein multi-functionality and phenotypic plasticity in plants. *Phil. Trans. R. Soc. B* 375: 20190182. doi: 10.1098/rstb.2019.0182

Maddison, WP, Maddison DR. (2019). Mesquite: a modular system for evolutionary analysis. Version 3.61 http://www.mesquiteproject.org.

Marais, EM. (1994). Taxonomic studies in *Pelargonium* section *Hoarea* (Geraniaceae). Dissertation Presented for the degree of Doctor of Philosophy at the University of Stellenbosch, South Africa.

Marais, EM. (2005). Differences between *Pelargonium moniliforme* (Geraniaceae) and the closely related *P. vinaceum*. S. Afr. J. Bot. 71:221–227. Doi: 10.1016/S0254-6299(15)30136-8.

Marais EM. (2014). One name change and three new species of *Pelargonium* section *Hoarea* (Geraniaceae) from Western Cape Province. S. Afr. J. Bot. 90:118–127. Doi: 10.1016/j.sajb.2013.10.013.

Maréchal A, Parent J-S, Véronneau-Lafortune F, Joyeux A, Lang BF, et al. (2009). Whirly proteins maintain plastid genome stability in *Arabidopsis*. *P.N.A.S*. 106(34), 14693–14698. Doi: 10.1073/pnas.0901710106.

Metzlaff M, Börner T., Hagemann R. (1981). Variations of chloroplast DNAs in the genus *Pelargonium* and their biparental inheritance. Theor. Appl. Genet. 60, 37-41. Doi: doi.org/10.1007/BF00275175.

Martin W, Herrmann RG. (1998). Gene transfer from organelles to the nucleus: how much, what happens, and Why? Plant Physiol.;118(1):9-17. Doi: 10.1104/pp.118.1.9.

Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, et al. (2002). Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveal plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc. Natl. Acad. Sci. USA 99:12246–12251.doi: 10.1073/pnas.182432999

Martínez-Cabrera HI, Schlichting CD, Silander JA, Jones CS. (2012). Low levels of climate niche conservatism may explain clade diversity patterns in the South African genus *Pelargonium* (Geraniaceae). Am J Bot.;99(5):954-60. Doi: 10.3732/ajb.

Masutani B, Arimura S-i, Morishita S (2021) Investigating the mitochondrial genomic landscape of Arabidopsis thaliana by long-read sequencing. PLoS Comput. Biol. 17(1): e1008597. Doi: 10.1371/journal.pcbi.1008597

Maumus F, Quesneville H. (2014). Deep investigation of *Arabidopsis thaliana* junk DNA reveals a continuum between repetitive elements and genomic dark matter. PLoS ONE 9:e94101. Doi: 10.1371/journal.pone.0094101.

Mauro VP, Edelman GM. (2002). The ribosome filter hypothesis. Proc. Natl. Acad. Sci. U S A. 17;99(19):12031-6. Doi: 10.1073/pnas.192442499.

Medgyesy P, Fejes E, Maliga P. (1985). Interspecific chloroplast recombination in a *Nicotiana* somatic hybrid. Proc. Natl. Acad. Sci. U S A.;82(20):6960-6964. doi:10.1073/pnas.82.20.6960.

Medini H, Cohen T, Mishmar D. (2020). Mitochondria are fundamental for the emergence of metazoans: on metabolism, genomic regulation, and the birth of complex organisms. Annu. Rev. Genet. 2020. 54:151–66. Doi: 10.1146/annurev-genet-021920-105545.

Metzlaff M, Pohlheim F, Börner T, Hagemann R. (1981). Variations of Chloroplast DNAs in the genus *Pelargonium* and their biparental inheritance. Theor. Appl. Genet. 5: 245-49. Doi: 10.1007/BF00275175.

Metzlaff M, Pohlheim F, Börner T, Hagemann R. (1982). Hybrid variegation in the genus *Pelargonium*. Curr Genet 5: 245-49. Doi: 10.1007/BF00391813.

Mezghani N, Khoury C-K, Carver D, Achicanoy HA, Simon P, et al. (2019). Distributions and conservation status of carrot wild relatives in Tunisia: a case study in the western Mediterranean basin. Crop Sci. 59, 2317–2328 10.2135/cropsci2019.05.0333.

Migicovsky Z, Myles S (2017). Exploiting wild relatives for genomics-assisted breeding of perennial crops. Front. Plant Sci. 8:460. doi: 10.3389/fpls.2017.00460.

Miller AJ, Matasci N, Schwaninger H, Aradhya MK, Prins B, et al. (2013). *Vitis* phylogenomics: hybridization intensities from a SNP array outperform genotype calls. PloS one, 8(11), e78680. Doi: 10.1371/journal.pone.0078680.

Moelling K, Broecker F. (2021). Viroids and the origin of life. Int. J. Mol. Sci. 28;22(7):3476. Doi: 10.3390/ijms22073476.

Montoliu L. (2012). Mendel: a simple excel workbook to compare the observed and expected distributions of genotypes/phenotypes in transgenic and knockout mouse crosses involving up to three unlinked loci by means of a χ2 test. *Transgenic res*. 21(3), 677-681. Doi: 10.1007/s11248-011-9544-4.

Moore TE, Schlichting CD, Aiello-Lammens ME, Mocko K, Jones CS, (2018). Divergent trait and environment relationships among parallel radiations in *Pelargonium* (Geraniaceae): a role for evolutionary legacy? New Phytol. Doi: https://doi. org/10.1111/nph.15196.

Morgante M, De Paoli E, Radovic S. (2007). Transposable elements and the plant pan-genomes. Curr. Opin. Plant. Biol. 10:149–55. Doi: 10.1016/j.pbi.2007.02.001.

Mower JP, Touzet P, Gummow JS, Delph LF, Palmer JD. (2007). Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. BMC Evolutionary Biology 7:135. Doi: 10.1186/1471-2148-7-135.

Müller, HJ. (1942). Isolating mechanisms, evolution, and temperature. Biol. Symp. 6, 71–125.

N

Nasser NS, Tilney Basset RAE (1992). The inheritance of patterned white flowers in zonal Pelargoniums. Euphytica 41: 23-29.

Nicotra AB, Cosgrove MJ, Cowling A, Schlichting CD, Jones CS, (2008). Leaf shape linked to photosynthetic rates and temperature optima in South African *Pelargonium* species. Oecologia 154, 625–635. Doi: 10.1007/s00442-007-0865-1

Nicotra A, Leigh A, Boyce CK, Jones C, Niklas K, Royer D, Tsukaya H. (2011). The evolution and functional significance of leaf shape in the angiosperms. Functional plant biology: FPB, 38 7, 535-552. Doi: NA

van der Niet T, Johnson SD, Linder HP. (2006). Macroevolutionary data suggest a role for reinforcement in pollination system shifts. Evolution60:1596–1601. Doi: 10.1111/j.0014-3820.2006.tb00504.x.

van der Niet T, Johnson SD. (2009). Patterns of plant speciation in the Cape floristic region. Molecular Phylogenetics and Evolution 51 (2009) 85–93. Doi: 10.1016/j.ympev.2008.11.027.

Nieuwenhuis M. (2013). Evolutionary trends in genome size and polyploidy in *Pelargonium* (Geraniaceae). MSc thesis Wageningen UR. DOI: 10.13140/RG.2.2.25087.36005.

Novák P, Neumann P, Macas J. (2010). Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinformatics 11:378. Doi: 10.1186/1471-2105-11-378.

Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. (2013). RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 29:792–93. Doi: 10.1093/bioinformatics/btt054.

Novák P, Guignard MS, Neumann P, Kelly LJ, Mlinarec J. et al. (2020). Repeat-sequence turnover shifts fundamentally in species with large genomes. Nat. Plants 6:1325–1329 Doi: 10.1038/s41477-020-00785-x.

Nudler E. (2012). Linking RNA polymerase backtracking to genome instability in *E. coli*. Cell 146, 533–543. Doi: 10.1016/j.cell.2011.07.034.

O

Oliver KR, McComb JA, Greene WK. (2013). Transposable elements: powerful contributors to angiosperm evolution and diversity, Genome Biol. Evol. 5:1886–1901. Doi: 10.1093/gbe/evt141.

Ollitrault P, Terol J, Garcia-Lor A, Bérard A, Chauveau A, et al. (2012). SNP mining in *C. clementina* BAC end sequences; transferability in the *Citrus* genus (Rutaceae), phylogenetic inferences and perspectives for genetic mapping. BMC Genomics 10:13:13. doi: 10.1186/1471-2164-13-13.

Osorio D, Rondon-Villarreal P, Torres R. (2015). Peptides: A package for data mining of antimicrobial peptides. The R Journal. 7(1): 4-14. Doi: 10.32614/RJ-2015-001.

P

Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu YL, et al. (2000). Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. Proc. Nat. Acad. Sci. U.S.A. 97: 6960–6966. Doi: 10.1073/pnas.97.13.6960.

Paço A, Freitas R, Vieira-da-Silva A. (2019). Conversion of DNA sequences: from a transposable element to a tandem repeat or to a gene. Genes, 10:1014. Doi: 10.3390/genes10121014.

Paradis E, Schliep K. (2018). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528. Doi: 10.1093/bioinformatics/bty633.

Perez Boerema A, Aibara S, Paul B, Tobiasson B, Kimanius et al. (2018). Structure of the chloroplast ribosome with chl-RRF and hibernation-promoting factor. Nature Plants 4, 212–217 Doi: 10.1038/s41477-018-0129-6.

Pfannschmidt T, Terry MJ, Van Aken O, Quiros PM. (2020). Retrograde signals from endosymbiotic organelles: a common control principle in eukaryotic cells. Phil. Trans. R. Soc. B 375: 20190396. Doi: 10.1098/rstb.2019.0396.

Plaschil S, Budahn H, Wiedemann M, Olbricht K. (2017). Genetic characterization of *Pelargonium* L'Hér. germplasm. *Genet. Resour. Crop Evol.* 64, 1051–1059. 10.1007/s10722-016-0424-x.

Poole AM, Gribaldo S. (2014). Eukaryotic origins: how and when was the mitochondrion acquired? Cold Spring Harb Perspect; 6:a015990. Doi: 10.1101/cshperspect.a015990.

Postel Z, Touzet P. (2020). Cytonuclear genetic incompatibilities in plant speciation. Plants, 9, 487; doi:10.3390/plants9040487.

R

Ramsey AJ, Mandel JR. (2019). When one genome is not enough: organellar heteroplasmy in plants. Ann. Pl. Rev. 2, 1–40 doi: 10.1002/9781119312994.apr0616.

Ren R, Wang H, Guo C, Zhang N, Zeng L, et al. (2018). Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. Molecular Plant 11:414–28. Doi: 10.1016/j.molp.2018.01.002.

Ribera A, Bai Y, Wolters A-MA, van Treuren R, Kik C. (2020). A review on the genetic resources, domestication and breeding history of spinach (*Spinacia oleracea* L.). *Euphytica* 216. Article 48. 10.1007/s10681-020-02585-y.

Ringelberg J. (2012). Floral evolution in Cape *Pelargonium* (Geraniaceae). Inferring shifts in nectar spur length and pollination. MSc thesis Wageningen UR. https://edepot.wur.nl/240312.

Rojas M, Yu Q, Williams-Carrier R, Maliga P, Barkan A. (2019). Engineered PPR proteins as inducible switches to activate the expression of chloroplast transgenes. *Nature Plants*. 5: 505–511. 10.1038/s41477-019-0412-1.

Röschenbleck J, Albers F, Müller K, Weinl S, Kudla J. (2014). Phylogenetics, character evolution and a subgeneric revision of the genus *Pelargonium* (Geraniaceae). Phytotaxa. 159(2), 31–76. 10.11646/phytotaxa.159.2.1

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. (2017). DnaSP v6: DNA Sequence Polymorphism Analysis of Large Datasets Mol. Biol. Evol. 34: 3299-3302.

Ruhlman TA, Jansen RK. (2018). Aberration or analogy? The atypical plastomes of Geraniaceae. in: Advances in Botanical Research 85: Plastid genome evolution. 10.1016/bs.abr.2017.11.017. Eds S-M. Chaw, R. K. Jansen (Amsterdam NL, Elsevier). 223–262.

Ruhlman TA, Jansen RK. (2021). Plastid genomes of flowering plants: essential principles. In: Maliga P. (eds) Chloroplast Biotechnology. Methods in molecular biology, vol 2317. Humana, New York, NY. Doi: 10.1007/978-1-0716-1472-3_1.

S

Salim D, Gerton JL. (2019). Ribosomal DNA instability and genome adaptability. Chromosome Res. 27:73–87. Doi: 10.1007/s10577-018-9599-7.

Sánchez-Baracaldo P, Raven JA, Pisani D, Knoll AH. (2017). Early photosynthetic eukaryotes inhabited low-salinity habitats. PNAS September 12,114 (37) E7737-E7745. Doi: 10.1073/pnas.1620089114.

Sankoff D, Zhang C, Zhu Q. (2010). The collapse of gene complement following whole genome duplication. BMC Genomics, 11:313. Doi: 10.1186/1471-2164-11-313.

Schmied WH, Tnimov Z, Uttamapinant C, Rae CD, Fried SD, et al. (2018). Controlling orthogonal ribosome subunit interactions enables evolution of new function. Nature ;564(7736):444-448. Doi: 10.1038/s41586-018-0773-z.

Schnable PS, Wise RP. (1998). The molecular basis of cytoplasmic male sterility and fertility restoration. Trends Plant Sci 3:175–180. Doi: 10.1016/S1360-1385(98)01235-7.

Schranz ME, Mohammadin S, Edger PP. (2012) Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. Curr. Opin; Plant Biol. 15:147-53. Doi: 10.1016/j.pbi.2012.03.011

Sebastian R, Oberdoerffer P. (2017). Transcription-associated events affecting genomic integrity. Philos. Trans. R. Soc. Lond. B Biol. Sci. 372(1731):20160288. Doi: 10.1098/rstb.2016.0288.

Semagn K, Babu R, Hearne S. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop improvement. Mol. Breed. 33:1–14. Doi: 10.1007/s11032-013-9917-x.

Sharma, G., Sharma, N. (2014). Cytology as an important tool for solving evolutionary problems in angiosperms. Proc. Natl. Acad. Sci., India, Sect. B Biol. Sci. 84, 1–7 (2014). Doi: /10.1007/s40011-013-0203-9.

Shikanai T, Shimizu K, Ueda K, Nishimura Y, Kuroiwa T, et al. (2001). The chloroplast clpP gene, encoding a proteolytic subunit of ATP-dependent protease, is indispensable for chloroplast development in Tobacco. *Plant Cell Physiol*. 42(3): 264–273. Doi: 10.1093/pcp/pce031.

Singh R, Singh S, Parihar P, Singh VP, Prasad SM. (2015). Retrograde signaling between plastid and nucleus: A review. J Plant Physiol. 1;181:55-66. Doi: 10.1016/j.jplph.2015.04.001.

Singh M, Kuma, S, Basandrai AK, Basandrai D, Malhotra Net al. (2020). Evaluation and identification of wild lentil accessions for enhancing genetic gains of cultivated varieties. *PLoS ONE*. 15(3), e0229554. 10.1371/journal.pone.0229554.

Siniauskaya MG, Danilenko NG, Lukhanina NV, Shymkevich AM, Davydenko OG. (2016). Expression of the chloroplast genome: modern concepts and experimental approaches. Russ. J. of Genet.: Applied Research6(5): 491–509. Doi: 10.1134/S2079059716050117.

Sloan DB, Triant DA, Forrester NJ, Bergner LM, Wu M, et al. (2014). A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe Sileneae (Caryophyllaceae). Mol. Phylogenet. Evol. 72:82–89. Doi: 10.1016/j.ympev.2013.12.004.

Small ID, Schallenberg-Rüdinger M, Takenaka M, Mireau H, Ostersetzer-Biran, O. (2020). Plant organellar RNA editing: what 30 years of research has revealed. *The Plant J*. 101(5), 1040–1056. Doi: 10.1111/tpj.14578.

Smith R, Rassmann K, Davies H, King, N. (eds.), (2011). Why taxonomy matters. http://www.bionet-intl.org/why. BioNET-INTERNATIONAL, Egham, UK. (Accessed [22/05/2021]).

Snijder RC, Brown FS, van Tuyl JM. (2007). The role of plastome-genome incompatibility and biparental plastid inheritance in interspecific hybridization in the genus *Zantedeschia* (Araceae). *Floriculture and Ornamental Biotechnology*. 1(2), 150-157.

Sobanski J, Giavalisco P, Fischer A, Kreiner JM, Walther D, et al. (2019). Chloroplast competition is controlled by lipid biosynthesis in evening primroses. *PNAS*. 116(12), 5665-5674. Doi: 10.1073/pnas.1811661116.

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, et al. (2009). Polyploidy and angiosperm diversification. Am. J. Bot 96:336-348. Doi: 10.3732/ajb.0800079.

Soltis PS, Marchant DB, Van de Peer Y, Soltis DE. (2015). Polyploidy and genome evolution in plants. Curr. Op. in Genet. Devel. 35: 119–125. Doi: 10.1016/j.gde.2015.11.003.

Spencer PS, Barral JM. (2012). Genetic code redundancy and its influence on the encoded polypeptides. Comput. Struct. Biotechnol. J. 20;1:e201204006. Doi: 10.5936/csbj.201204006.

Stubbe, W. (1958). Dreifarbenpanschierung bei *Oenothera*: II. Wechselwirkungen zwischen Geweben mit zweierblich verschiedenen Plastidensorten. *Z. für Vererbungslehre*. 89, 189–203.

Stubbe, W. (1959). Genetische Analyse des Zusammenwirkens von Genom und Plastom bei *Oenothera*. *Z. für Vererbungslehre*. 90, 288–298

Stubbe W (1989). *Oenothera*—An ideal system for studying the interaction of genome and plastome. Plant Mol. Biol. Reporter, 7, 245–257. Doi: 10.1007/BF02668633.

Sweet R. (1820). Geraniaceae the natural order of Gerania Vol1. London: Printed for James Ridgway, Piccadilly UK

Sweet R. (1822). Geraniaceae the natural order of Gerania Vol2. London: Printed for James Ridgway, Piccadilly UK

Sweet R. (1824–1826) Geraniaceae the natural order of Gerania vol3. Ridgway, London, Pl. 201–30

Swofford DL. (2002). PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4. Sunderland, Massachusetts.

T

Tilney-Bassett RAE. (1973). The control of plastid inheritance in *Pelargonium* II. *Heredity*. 30, 1-13. 10.1038/hdy.1973.1.

Tilney-Bassett RAE. (1974). The control of plastid inheritance in *Pelargonium* III. *Heredity*. 33, 353-360. 0.1038/hdy.1974.102.

Tilney-Bassett RAE. (1975). Genetics of variegated plants. In: Genetics and biogenesis of mitochondria and chloroplasts. eds Birky C. W. Jr, Perlman P. S, Byers T. J. Ohio State University Press, Columbus. 268-308.

Tilney-Bassett RAE. (1976). The control of plastid inheritance in *Pelargonium* IV. *Heredity*. 37, 95-107. 10.1038/hdy.1976.68.

Tilney-Bassett RAE, Birky CW Jr. (1981). The mechanism of the mixed inheritance of chloroplast genes in *Pelargonium*: Evidence from gene frequency distributions among the progeny of crosses. Theor. Appl. Genet. 60(1), 43-53. 10.1007/BF00275177.

Tilney-Bassett RAE, Abdel-Wahab, OAL. (1982). Irregular segregation at the Pr locus controlling plastid inheritance in *Pelargonium*: Gametophytic lethal or incompatibility system? *Theor. Appl. Genet.* 62, 185-191. 10.1007/BF00293357

Tilney-Bassett RAE. (1984). The genetic evidence for nuclear control of chloroplast biogenesis in higher plants. In: Chloroplast biogenesis. Society for Experimental Biology Seminar Series 21. Ed R. J. Ellis, (London, UK, Cambridge University Press). 13-50.

Tilney-Bassett RAE (1988). Inheritance of plastids in *Pelargonium*. In: The division and segregation of organelles. Society for Experimental Biology Seminar Series 35. eds S. A. Boffey, D. Lloyd (London, UK, Cambridge University Press). 115-129.

Tilney-Bassett RAE, Almouslem AB. (1989a). Variation in plastid inheritance between *Pelargonium* cultivars and their hybrids. *Heredity*. 63, 145-153. 10.1038/hdy.1989.86.

Tilney-Bassett RAE, Almouslem AB, Amoatey HM. (1989b). The complementary gene model for biparental plastid inheritance. In: Physiology, biochemistry, and genetics of non-green plastids. The American Society of Plant Physiologists. eds C. D. Boyer, J. C. Shannon, R. C. Hardison. 265-266.

Tilney-Bassett RAE, Almouslem AB, Amoate HM. (1992). Complementary genes control biparental plastid inheritance in *Pelargonium*. *Theor. Appl. Genet.* 85, 317-324. 10.1007/BF00222876.

Turelli M, Moyle LC. (2007) Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. Genetics, 176, 1059–1088. Doi: 10.1534/genetics.106.065979.

V

de Vere N, Rich TC, Trinder SA, Long C. (2015). DNA barcoding for plants. Methods Mol Biol.;1245:101-18. doi: 10.1007/978-1-4939-1966-6_8.

Verboom, GA, Archibald JK, Bakker FT, Bellstedt DU, Conrad F. et al. (2009). Origin and diversification of the Greater Cape flora: ancient species repository, hot-bed of recent radiation, or both? Mol. Phylogenet. Evol. 51:44–53. Doi: 10.1016/j.ympev.2008.01.037.

Vitales D, Garcia S, Dodsworth S. (2019a). Reconstructing phylogenetic relationships based on repeat sequence similarities. Mol. Phyl. Evol. 147:106766. Doi: 10.1101/624064.

Vitales, D, Fernández P, Garnatje T, Garcia S. (2019b). Progress in the study of genome size evolution in Asteraceae: analysis of the last update. Database Vol. 2019; Doi: 10.1093/database/baz098.

van der Walt, JJA. (1977). Pelargoniums of Southern Africa. Purnell, Cape Town, 51 pp

van der Walt, JJA. (1979). Notes on the nomenclature of *Pelargonium* (Geraniaceae). Journal of South African Botany 45: 378–379

van der Walt JJA, VorsterPJ, (1988). *Pelargonium*s of Southern Africa. Volumes I-III. Annals of Kirstenbosch botanic gardens.

Van der Walt JJA, Albers F, Gibby M. (1990). Delimitation of *Pelargonium* sect. *Glaucophyllum*. PI. Syst. Evol. 171: 15—26. Doi: 10.1007/BF00940594.

W

Wan W-L, Kim S-T, Castel B, Charoennit N, Chae E. (2021). Genetics of autoimmunity in plants: an evolutionary genetics perspective. New Phytologist (2021) 229: 1215–1233 Doi: 10.1111/nph.16947.

Wang W, Wu Y, Messing J. (2016a). Genome-wide analysis of pentatricopeptide-repeat proteins of an aquatic plant. *Planta* 244, 893–899. 10.1007/s00425-016-2555-x.

Wang D, Liu H, Zhai G, Wang L, Shao J, Tao Y. (2016b). OspTAC2 encodes a pentatricopeptide repeat protein and regulates rice chloroplast development. J. of Genetics and Genomics 43. 601e608. 10.1016/j.jgg.2016.09.002

Wang J, Li Y, Li C, Yan C, Zhao X, et al. (2019). Twelve complete chloroplast genomes of wild peanuts: great genetic resources and a better understanding of *Arachis* phylogeny. BMC Plant Biology. 19, 504. 10.1186/s12870-019-2121-3.

Wang X, Yaqi A, Pan X, Jianwei X. (2021). Functioning of PPR proteins in organelle RNA metabolism and chloroplast biogenesis. Fr. in Pl. Sci. 12. Doi: 10.3389/fpls.2021.627501.

Warren BH, Bakker FT, Bellstedt DU, Bytebier B, Claßen-Bockhoff R, et al. (2011). Consistent phenological shifts in the making of a biodiversity hotspot: the Cape flora. BMC Evol. Biol. 11: 10.1186/1471-2148-11-39.

Wei C, He B, Mao J, Jiang J, Zhang L. (2015). Plastid sigma factors: Their individual functions and regulation in transcription. Biochimica et Biophysica Acta 1847 770–778. Doi: /10.1016/j.bbabio.2015.01.001.

Weihe A, Apitz J, Salinas A, Pohlheim F, Börner T. (2009). Biparental inheritance of plastidial and mitochondrial DNA and hybrid variegation in *Pelargonium*. Mol. Genet. Genomics, 282:587–593. Doi: 10.1007/s00438-009-0488-9.

Weiss-Schneeweiss H, Leitch AR, McCann J, Jang T-S, Macas J. (2015). Employing next generation sequencing to explore the repeat landscape of the plant genome. in Hörandl E, Appelhans MS. editors, Next-generation sequencing in plant systematics. Chapter 5, http://dx.doi.org/10.14630/000006.

Wendel JF. (2015). The wondrous cycles of polyploidy in plants. Am. J. Bot. 102:1753–1756. Doi: 10.3732/ajb.1500320.

Weng ML, Ruhlman TA, Gibby M, Jansen RK. (2012). Phylogeny, rate variation, and genome size evolution in *Pelargonium* (Geraniaceae). Mol. Phylogenet. Evol. 64:654–670. Doi: 10.1016/j.ympev.2012.05.026.

Weng ML, Blazier JC, Govindu M, Jansen RK. (2013). Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. Mol Biol Evol. 31:645-59. Doi: 10.1093/molbev/mst257.

Weng M-L, Ruhlman TA, Jansen RK (2016) Plastid-nuclear interaction and accelerated coevolution in plastid ribosomal genes in Geraniaceae. Genome Biol Evol 27;8(6):1824-38 Doi: 10.1093/gbe/evw115.

Westerbergh,A, Lerceteau-Köhler E, Sameri M, Bedada G, Lundquist P-O. (2018). Towards the development of perennial barley for cold temperate climates—evaluation of wild barley relatives as genetic resources. *Sustainability* 10(1969). doi:10.3390/su10061969.

Weng M-L, Ruhlman TA, Jansen RK. (2017). Expansion of inverted repeat does not decrease substitution rates in *Pelargonium* plastid genomes. *New Phytologist*. 214, 842–851 10.1111/nph.14375.

Weng M-L, Ruhlman TA, Jansen RK, (2017). Expansion of inverted repeat does not decrease substitution rates in *Pelargonium* plastid genomes. New Phytol. 214: 842–851. Doi: 10.1111/nph.14375.

Westerich LD, Leon Gotsmann VL, Herkt C, Ries F, Kazek T, et al. (2021). The versatile interactome of chloroplast ribosomes revealed by affinity purification mass spectrometry. Nucl. Acids Res., Vol. 49(1). Doi: 10.1093/nar/gkaa1192.

Wick RR, Schultz MB, Zobel J & Holt KE. (2015). Bandage: interactive visualization of de novo genome assemblies. Bioinformatics, 31(20): 3350–3352. Doi: 10.1093/bioinformatics/btv383.

Wicke S, Schneeweiss GM, Muller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol. Biol. 76:273–297. Doi: 10.1007/s11103-011-9762-4.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, et al. (2007). A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8:973–982. doi: 10.1038/nrg2165.

Wickham H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Wilkins JS. (2003). "How to be a chaste species pluralist-realist: The origins of species modes and the synapomorphic species concept", Biology and Philosophy 18:621-638. Doi: 10.1023/A:1026390327482.

Woloszynska, M. (2009). Heteroplasmy and stoichiometric complexity of plant mitochondrial genomes: though this be madness, yet there's method in't. J. Exp. Bot. 61:657–671. Doi: 10.1093/jxb/erp361.

Y

Yang H, Wei C-L, Liu H-W, Wu J-L, Li Z-G, et al. (2016). Genetic divergence between *Camellia sinensis* and its wild relatives revealed via genome-wide SNPs from RAD sequencing. PLoS ONE 11(3): e0151424. Doi: 10.1371/journal.pone.0151424.

Yao JL, Cohen D (2000). Multiple gene control of plastome-genome incompatibility and plastid DNA inheritance in interspecific hybrids of *Zantedeschia*. Theoretical and Applied Genetics 101, 400-406.

Z

Zhang J, Ruhlman TA, Mower JP, Jansen RK. (2013). Comparative analyses of two Geraniaceae transcriptomes using next-generation sequencing. BMC Plant Biol. 13, 228. Doi: 10.1186/1471-2229-13-228.

Zhang J, Ruhlman T, Sabir J, Blazier J, Jansen R. (2015). Coordinated rates of evolution between interacting plastid and nuclear genes in Geraniaceae. The Plant Cell, 27(3): 563-573. Doi: 10.1105/tpc.114.134353.

Zhang JJ, Ruhlman TA, Sabir JSM, Blazier JC, Weng M-L, et al. (2016). Coevolution between nuclear-encoded DNA replication, recombination, and repair genes and plastid genome complexity. Genome Biology and Evolution, 8,3: 622–634, Doi: 10.1093/gbe/evw033.

Zhang J, Xiao J, Li Y, Su B, Xu H, et al. (2017). *PDM3*, a pentatricopeptide repeat-containing protein, affects chloroplast development. J. of Exp. Bot. 68(20), 5615–5627. doi: 10.1093/jxb/erx360.

Zhang Y, Feng Y, Chatterjee S, Tuske S, Ho MX, et al. (2012). Structural Basis of transcription initiation. Science 338(6110): 1076-1088. Doi: 10.1126/science.

Zerges W, Hauser C. (2009). Chapter 28—Protein synthesis in the chloroplast E. Harris, D.B. Stern, G.B. Witman (Eds.), The Chlamydomonas sourcebook (Second Edition), Elsevier (2009), pp. 967-1025. Doi: 10.1016/j.bbabio.2015.05.008.

Zipper H, Brunner H, Bernhagen J, Vitzthum F. (2004). Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications. Nucl. Acids Res. 12;32(12):e103. Doi: 10.1093/nar/gnh101.

Zoschke R, Bock R. (2018). Chloroplast translation: structural and functional organization, operational control, and regulation. The Plant Cell 30: 745–770. Doi: 10.1105/tpc.18.00016.

R

Nagoya compliancy statement

"All plant material used in this study is fully compliant with the rules layed down in the Nagoya protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilization to the convention on biological diversity. Paperwork is on file to support this."

Summary
Samenvatting
Acknowledgements
About the author
List of publications
Education statement

Summary

In this thesis I studied cause and effect of Cyto Nuclear Incompatibility (CNI) induced speciation in plants. By combining comparative genomics, phylogenetics and bioinformatics with large scale crossing experiments I was able to infer part of the genetic machinery underlying CNI, determine plastid inheritance, study plastome evolution in detail and establish a possible mechanism for explaining chlorosis.

After introducing (Chapter 1) the phenomenon of CNI, its consequences for plant evolution and speciation, and *Pelargonium* section *Ciconium* as a model system, I investigate patterns of CNI in a comprehensive series of *P.* sect. *Ciconium* (Geraniaceae) interspecific crosses with a single parental population of *P.* x *hortorum* in Chapter two. I deduce that one or more nuclear genomic alleles are involved in the expression and management of chloroplasts, and that these act throughout the different life stages. It appears to be the case that the number of alleles needed to explain the observed segregation patters increases with phylogenetic distance between the parent species. Finally, I established that biparental inheritance of chloroplasts is widespread and common across *P.* sect. *Ciconium*. All species have the potential to transmit their plastids either paternally or maternally.

In Chapter three I use the variation on abundancies of the repetitive part of the genome (the non-coding DNA collectively known as the 'repeatome') as a proxy for the nuclear genotype in our crossing experiments (Chapters two and five). In addition, I use genomic repeat abundancy and repeat sequence similarity to reconstruct phylogenetic trees. Hereby the repeat clusters are characters, and their abundancies states. The interpretation of abundancy variation must be done with care, because autopolyploids can cause 'false' sister species relationships to occur in the reconstructed tree, based on their genomic sizes, not on shared history. Nevertheless, when I included sequence read variation in the analysis as well, it became clear that the 'Core-*Ciconium*' species share >20 large repeat clusters when compared to *P. elongatum* and that the species *P. multibracteatum*, *P. quinquelobatum* and *P. yemenense* sp. nov., occurring outside the Cape Floristic Region (CFR), again share several large synapomorphic repeats.

In Chapter four, I use a near complete selection of species from *P.* sect. *Ciconium* to study plastome evolution in detail. I create a detailed picture of the variation that occurs in *P.* sect. *Ciconium* plastomes and present evidence that *rrn* genes (especially *rrn23*), encoding ribosomal RNA, and ribosomal proteins have undergone concerted evolution. *Rrn23* displays more variation within section *P.* sect. *Ciconium* than it does between angiosperm orders. I speculate that one of the consequences of plastome evolution in *P.* sect. *Ciconium* (and presumably *Pelargonium* as well) is the presence of a, structurally, altered plastidial ribosome. In this Chapter I further describe that two of the three available plastid encoded polymerase (PEP) subunits (*rpoB* and *rpoC1*) are under positive selection in *Ciconium* suggesting strong involvement in plastome evolution.

Chapter five draws, in part, on the results reported in the previous Chapters. I continued to determine inheritance of plastids, but report newly developed markers for the mitochondria as well. I confirmed the findings from Chapter two that plastids inherit via both parental

lines, but biparental inheritance of mitochondria turned out to be rarer. In this Chapter I present three more interspecific crossing series and describe more chlorosis patterns and their associated plastid types. I investigate the occurrence of chlorosis from the chloroplast perspective and find that changes in physico-chemical properties of resulting rpoB peptides occurring in *P.* sect. *Ciconium*, by and large, correlate with the observed patterns of chlorosis in F_1 interspecific offspring. There is a strong indication that the *Ciconium*-unique sequence insertions in *rpo* genes encode parts of the polymerase that are in direct contact with template DNA during transcription. If so, then *rpo* variation, or PEP structure variation, must be considered, alongside nuclear genomic changes reported in other studies, when understanding plastome evolution in Geraniaceae.

In chapter six I discuss the results from the previous chapters in the context of the Bateson-Dobzhansky-Müller model of speciation. I propose a refinement of the model, in which different kinds of selective pressures act at different moments in time, for *Pelargonium*. I discuss that *P.* section *Ciconium* is, likely in an evolutionary phase where CNI is the driver for speciation.

Samenvatting

In dit proefschrift heb ik gepoogd 'oorzaak en gevolg' uiteen te rafelen voor wat betreft het voorkomen van soortvorming gedreven door Cyto Nucleaire Incompatibiliteiten (CNI). Door het combineren van vergelijkende genomica, phylogenetica en bioinformatica was ik in staat om een delen van de onderliggende genetica van CNI af te leiden, overerving van chloroplasten te bepalen, plastoom evolutie in detail te bestuderen en een mechanisme voor te stellen dat het voorkomen van chlorose verklaart.

Na de introductie van het fenomeen CNI in hoofdstuk één, de gevolgen voor evolutie en soortvorming en de introductie van *Pelargonium* sectie *Ciconium* als model onderzoek ik in hoofdstuk twee patronen van CNI in een volledige serie interspecifieke kruisingen van *P.* sectie *Ciconium* soorten met één ouderlijke populatie van *P.* x *hortorum*. In hoofdstuk twee heb ik afgeleid dat één of meerdere nucleair genomische allelen betrokken zijn bij de expressie en het beheer van chloroplasten. Ik heb ook gevonden dat deze doorwerken in verschillende levensstadia. Het lijkt erop dat het aantal allelen dat nodig is om de segregatie patronen te verklaren toe neemt met de fylogenetische afstand tussen de twee ouders. Tot slot stelde ik vast dat biparentale overerving van chloroplasten wijdverbreid en algemeen is in *P.* sectie *Ciconium*. Van alle soorten kunnen, potentieel, hun chloroplasten in vaderlijke of moederlijke lijn overerven.

In hoofdstuk drie heb ik de variatie in hoeveelheid van het 'repeatome'(dat deel van het genoom dat repeterende sequentie motieven bevat) gebruikt als proxy om de nucleaire genomen die we gebruiken bij onze kruisingen (hoofdstukken twee en vijf) te karakteriseren. Daarnaast is er ook mogelijkheid het repeatome te gebruiken als bron voor fylogenetische markers. Hierbij zijn repeat clusters kenmerken, en hun abundanties de states. Fylogenetische bomen kunnen dan gebaseerd worden op abundantie én sequentie variatie verschillen. De interpretatie van de variatie in hoeveelheid moet met de nodige voorzichtigheid gebeuren omdat autopolyploïden 'valse' zuster soort verwantschappen veroorzaken enkel op basis van genoomformaat, niet op gemeenschappelijke afkomst. Desalniettemin, toen ik ook de sequentie variatie meenam in de analyses werd het duidelijk dat de 'Centrale *Ciconium*' soorten > 25 grote repeats delen bij vergelijking met *P. elongatum* én dat de soorten *P. multibracteatum*, *P. quinquelobatum* en *P. yemenense* sp. nov., die allen voorkomen buiten de Cape Floristic Region (CFR), nog een groep unieke repeats bevatten.

In hoofdstuk vier heb ik, een vrijwel complete, selectie van soorten uit *P.* sectie *Ciconium* gebruikt om plastoom evolutie binnen de sectie te bestuderen. Ik heb een gedetailleerd beeld geschetst van de variatie de voorkomt binnen de *Ciconium* plastomen. Ik presenteer aanwijzingen dat *rrn* genen (m.n. *rrn23*), die de code bevatten voor het ribosomaal RNA, en genen die coderen voor ribosomale proteïnen, gezamenlijk en synchroon zijn geëvolueerd. Het *rrn23* gen vertoont meer variatie tussen *P.* sectie *Ciconium* soorten dan tot nu toe is vastgesteld tussen angiosperme ordes. Ik speculeer dat een van de gevolgen van *P.* sectie *Ciconium* plastoom evolutie kan zijn dat er structureel veranderde ribosomen aanwezig zijn in de *P.* sectie *Ciconium* chloroplasten (en mogelijk in die van geheel het genus *Pelargonium*).

Ik beschrijf in dit hoofdstuk verder dat twee van de drie beschikbare gen sequenties voor chloroplast gecodeerde RNA polymerase sub-units (*rpoB* en *rpoC1*) onder positieve selectie staan in *P.* sectie *Ciconium*, wat doet vermoeden dat deze genen een belangrijke rol spelen in de evolutie van het plastoom.

Hoofdstuk vijf bouwt ten dele voort op de resultaten verkregen in de eerdere hoofdstukken. In hoofdstuk vijf ging ik verder met het vaststellen van de chloroplast overerving, maar ik gebruikte nu ook nieuw ontwikkelde merkers om de overerving van mitochondriën te bepalen. Ik bevestigde dat chloroplasten van alles *P.* sect. *Ciconium* soorten zowel maternaal als paternaal overerven, maar opvallend genoeg was biparentale overerving veel zeldzamer voor mitochondriën. In dit hoofdstuk presenteer ik nog drie series van interspecifieke kruisingen en beschrijf ik patronen van chlorose bij de kruisingsseries. Ik heb het voorkomen van chlorose vanuit het perspektief van de chloroplast onderzocht en gevonden dat veranderingen in de fysische en chemische eigenschappen van m.n. de *Ciconium* rpoB peptiden, voor een groot deel correleren met de waargenomen chlorose patronen in F_1 nakomelingen van interspecifieke kruisingen. Er is een sterke aanwijzing dat *P.* sectie *Ciconium* unieke sequentie inserties bevat in delen van het chloroplast gecodeerde RNA polymerase die in direct contact staan met het sjabloon DNA gedurende transcriptie. Indien dit zo is, zullen variatie in *rpo* sequenties of structurele variatie van het chloroplast gecodeerde RNA polymerase complex ook in overweging genomen moeten worden wanneer we plastoom evolutie in Geraniaceae willen begrijpen.

In het laatste hoofdstuk bediscussieer ik de gevonden resultaten in de context van het Bateson-Dobzhansky-Müller model van soortvorming. Ik stel een verfijning voor van het model, waarin verschillende typen selectiedruk actief zijn op verschillende momenten in de tijd, voor *Pelargonium*. Ik bespreek dat *P.* sectie *Ciconium* zich waarschijnlijk in een fase van evolutie bevindt waar CNI soortvorming aandrijft.

Acknowledgements

Being able to do a PhD was a privilege. You get to work on a research project, but in the process you get to deal with yourself as well. Fortunately there are many people who support you or who understand. I am grateful to you all. One of the things that I had to work on a lot was my writing skill. Many people will know that I much preferred to talk instead of engaging in endless email discussions. I think I have improved my scientific writing skills, but writing words of gratitude is hard. To all readers. Please take these words to mean the best thing for you. Forgive my clumsy wording or the occasional omission. I can tell you so much better in person.

First of all I would like to thank Freek Bakker, co-promotor, supervisor, project manager and teacher. But where to start? You invited me in 2016 to apply and thought I could do this. I thought so too, so we started. But, there was so much to learn and do. Throughout the project you never stopped challenging. Whenever I conquered a new concept, analysis, idea (fill in another word) you already had new ones to consider. I appreciated the sophistication of the supervision you displayed and any new candidate will be lucky to have you as a supervisor. You make teaching and supervision seem easy, but I know for a fact how much thought and energy you put into this. Keep it up. Our discussion evolved from how to organize the work, to choices on topics and types of analyses, all the way through deep scientific discussions to…: 'what next?' Come what may, I am looking forward to the 'what next' part. We have some great ideas and I hope these come to fruition. I thank you for your hospitality and having me over for dinners and lunches where we enjoyed lengthy discussion, coffee and original Thai food.

Next, I would like to thank the other co-promotor, (yes you really are) Ronald Snijder. I often got the idea that you thought I was crazy, but I'd like to think we managed to make it work. Your tireless commitment to the project and ever critical views kept me sharp. Your energy and 'work in the background' made this project work and you have my gratitude. Thanks to you I discovered what plant breeding is all about and I feel confident to approach plant breeding in general even though there is still so much to learn about this.

Last but not least of my promotor of course. Dear Eric Schranz, your quick mind and helicopter view combined with enormous experience have helped shape this project and have taught me a lot. You were always fair and your skill in bringing together all of the different types of research under the umbrella of biosystematics sets an example.

I would like to thank the people at Syngenta who suddenly had to put up with research of a different kind. Thank you to Sabine Lorente for managing the earlier phases of this project from the Syngenta side. A big thank you to the greenhouse management also. No longer did we throw away 'bad plants', some needed 'special treatment' and all were important. You all did great. Thank you Dennis, Harald. We also used the considerable experience Syngenta staff have with embryo rescue techniques. Thank you all the project is so much richer because of the miracles you worked. Did I say miracles? I meant skill and experience of course. Thank to these, interspecific crosses were grown that the world has not yet seen. Thank you all, Ewa, Anja,

Maria, Marjon and Mireilla. A special word of thanks needs to go to the 'lady of pollinations' Tony Lokkers. Beste Tony, je werd in het diepe gesmeten toen ze je het werk aan mijn project gaven. En wat voor project. Niks was standaard, zaken veranderden ook nog vaak en ook ik was nogal eens onzeker. Je bent dit echter met vakkundigheid en rust tegemoet getreden. Ik heb echt enorm genoten van onze samenwerking. Het was altijd een plezier om samen met jouw 'door de planten te gaan.' Elke keer leerde ik weer meer over hoe de verschillende soorten zich gedroegen. Dankzij jouw kan ik nu ook 'echt bestuiven' en kennen we de soorten veel beter dan anders het geval was geweest. Jouw inzet en inzicht hebben het werk succesvol gemaakt en ik hoop dat we nog eens samen mogen werken in de toekomst.

Next, I would like to thank the members of the PhD committee: Mary Gibby, Stephen Dodsworth, Sjoukje Heimovaara and Mark Aarts. I am honored that you have taken the time to read and evaluate my thesis in such detail. Thank you for your comments and suggestions.

I would also like to thank the co-authors and other collaborators in my project. Stephan Greiner: thank you for your in depth insights and very good remarks on the work we did. Mireilla Sancho-Such: With all your other responsibilities you and your team managed to create numerous hybrids and show me how it is done as well. We have very nice chapters and papers as a result. Thank you to Mary Gibbing for kindly receiving me in your home during my visit to Edinburhg. Thank you also for the nice discussions and interesting visits to the Edinburgh Herbarium.

I am very grateful to all my students. I was privileged to supervise four. Two students did their bachelor's degree with me and two wrote a master thesis. First: Sieme Pelzer. You did great work with me and I am proud you found work at Wageningen University. Joost Korver. You worked with me during the Covid crisis and you rose to the challenge of my forced absence and took responsibility of a number of experiments. I am confident you will finish your education with a great report and rest assured you will be a fantastic colleague wherever you will be. Guangnan Chen, you tackled one of the more difficult topics and produced a well-rounded thesis on the repeat profiling. Lastly, I would like to mention Jeroen Hofs. We were allowed to pursue an altogether different topic with a beautiful results. It is good to hear from you on occasion and rest assured the *Accipiter* work is in preparation as well.

To all the people at biosystematics. Thank you for bearing with me when I felt talkative. Thank you too for the numerous times we had coffee breaks together. You all provide such a nice environment for growing as a scientist or for simply having nice conversations with. Wilma, thank you for being patient when I did not (want) to understand the mysteries of the university bureaucracy, thank you for the Yoga exercises. They helped a lot. Thank you also for putting things in perspective. Sara, my fellow 'Pelargonista'. We shared more than a topic and a supervisor. We levelled off many times. Thanks for introducing me to IOGA and reflecting on texts. Thanks for helping me stay focused on the science. Finally, thanks for being my paranymph (see below). Dêêdi, take care in Benin. Hope to see you in real life someday again. Thijmen. We were office neighbors and helped each other out with scripts, students, texts and more, but especially during the last phase of our PhD we worked together on nearly every aspect. Thank you and you're a great researcher. Wei and Niccolo, good luck and thanks for

helping with the test defense. Looking forward to yours. To all the new colleagues whom I did not get to know as well as I would have liked. Eva, Wouter, Li, Shi, Liana, Pamela, Nora, Cloë. Good luck with all your research. Cloë and Wouter thanks for helping out with the test defense as well of course. To Ronald, Roel and Nynke, enjoy your retirement. And to Lars. I hope you'll do well in Belgium. It's not such a bad place. A thank you as well to the people from the lab, Frank Beckers, Corry en José. It was nice working with you all.

To all my friends. Thank you for understanding me and listening to my stories. Tijn, Berit, Victor, Pieter, Martijn en Karlijn. Thanks for the great games and times spent throughout all the years. May there be many more years to come. Stanly, thanks for showing me how to use my strengths and keeping me straight. To Franck Theeten. Tu ma's eu appris à gérer les scripts et l'informatique bien avant que j'en ai besoin. Nous avons tous les deux réalisé l'importance de l'informatique en biologie et avons travaillé sur des choses intéressantes au fil des ans. J'attends avec impatience la suite. Tu es devenu un grand ami. Take care. Special mention needs be go here to Louay and Nahed. Ik ben jullie ongelooflijk dankbaar voor alle keren dat jullie ons geholpen hebben om Peter van school te halen en op te vangen. Last but not least in this paragraph, I would like to add a special thanks to Martijn en Karlijn who provided a 'home away from home' for me. You kindly let me stay with you during this project and I cannot express enough gratitude for this. We ate together and discussed numerous topics during the years. I was also privileged to witness the arrival of your two beautiful daughters. Thank you so much.

Finally a big thanks to my family. My parents who always listened to the long stories. Thank to my brother and sisters for interest and support. My special thanks goes to Lina, my wife, and Peter, my son. Lina you realized what this work meant to me and despite the difficulties you supported me throughout and encouraged me to do this. Matur sembah nuwun lan kawulå tresnå panjenengan. Kawulå mboten saket nindakaken punika tanpå panjenengan. Peter, mijn grote jongen. Je moest me vaak missen de afgelopen jaren, maar bent altijd gelukkig als we weer samen zijn. Je doet het fantastisch en ik ben trots op je.

Floris

About the author

Floris Breman was born in Hengelo (Ov) The Netherlands on the 7th of November 1977. After attending the MAVO and HAVO, Floris graduated from het Twickel College for VWO in 1998 and started his study at the University of Wageningen. He graduated in 2006 as MSc in plant sciences with a free program. The focus in this program was on biosystematics and plant ecology. A master thesis was written under supervision of Dr. Freek T. Bakker (Biosystematics Group). An internship was done at the university of Leiden and this took place in Indonesia under the Supervision of Dr. Ferry Slik. After graduation Floris worked as a scientific assistant at the Royal Museum for Central Africa in Tervuren and the Royal Belgian Institute for Natural sciences in Brussel, both in Belgium. After this he started a training to become a fully-licensed teacher, which was cut short when he started his PhD in Wageningen at the Biosystematics Group, in collaboration with Syngenta, under the supervision of Dr. Freek T. Bakker, Prof. Dr. Eric Schranz and Dr. Ronald Snijder.

List of publications

- Tessens B., Monnens M., Backeljau T., Jordaens K., Van Steenkiste N**., Breman FC**., Smeets K., Artois T. (2021). Is 'everything everywhere'? Unprecedented cryptic diversity in the cosmopolitan flatworm *Gyratrix hermaphroditus*. Zoologica Scripta, 00, 1– 15. doi: 10.1111/zsc.12507

- **Breman FC**, Schranz ME, Snijder RC, Bakker FT. (2020). Interspecific Hybrids Between *Pelargonium* × *hortorum* and species From *P.* Section *Ciconium* reveal biparental plastid inheritance and multi-locus cyto-nuclear incompatibility. Front. Plant Sci. 11:614871. doi: 10.3389/fpls.2020.614871.

- Freek T. Bakker, **Floris C. Breman**, Frank Breteler (submitted). Plastome evolution in the newly described genus *Jongkindia*.

- **Floris C. Breman**, M. Eric Schranz, Guangnan Chen, Ronald Snijder, and Freek Bakker (submitted). Repeatome evolution in *Pelargonium* section *Ciconium* (Sweet) Harvey.

- **Floris C. Breman**, Eric Schranz, Stephan Greiner, Ronald Snijder, Freek T. Bakker. Rapid plastome evolution *in Pelargonium* section *Ciconium* linked with altered plastidic ribosomes.

- **Floris C. Breman**, M. Eric Schranz, Ronald Snijder and Freek Bakker (in prep). Possible causes of plastid induced CNI and organellar inheritance in *Pelargonium* section *Ciconium* (Sweet).

- **Floris C. Breman,** M. Eric Schranz, Jeroen Hofs, Kim Peterse, Sara van de Kerke, Freek T. Bakker (in prep). Evolution and speciation in the clade of *Accipiter tachiro* and its sister species.

Education Statement of the Graduate School Experimental Plant Sciences

Issued to: **Floris C. Breman**
Date: **13 September 2021**
Group: **Biosystematics**
University: **Wageningen University & Research**

1) Start-Up Phase	date	cp
► **First presentation of your project**		
Nuclear genomic characterisation via repeat profiling in *Pelargonium* L'Her	16/11/2016	1,5
► **Writing or rewriting a project proposal**		
Pelargonium genomics for overcoming species crossing barriers	04/2016 - 06/2016	3,0
► **MSc courses**		
Subtotal Start-Up Phase		4,5

2) Scientific Exposure	date	cp
► **EPS PhD student days**		
EPS PhD Get2Gether, Soest, NL	09-10/02/2017	0,6
EPS PhD Get2Gether, Soest, NL	15-16/02/2018	0,6
► **EPS theme symposia**		
EPS theme 4 symposium Genome Biology, Amsterdam, NL	25/09/2018	0,3
EPS theme 4 symposium Genome Biology, Wageningen, NL	13/12/2019	0,3
► **Lunteren Days and other national platforms**		
Experimental Plant Sciences meeting, Lunteren, NL	11-12/04/2016	0,6
Experimental Plant Sciences meeting, Lunteren, NL	09-10/04/2018	0,6
► **Seminars (series), workshops and symposia**		
Seminar by Peter Linder at WUR	06/10/2017	0,1
Pelargonium meeting Hortus-Botanicus Amsterdam, NL	20/10/2017	0,3
Meeting (herbarium studies, discussions with scientists, presentation) at the Royal Botanical Gardens Edinburgh, UK	31/03/2019 - 03/04/2019	0,8
Wageningen PhD symposium 2019 "Science with Impact", Wageningen, NL	25/10/2019	0,3
Meeting (herbarium studies, discussions with scientists, presentation) at the Royal Botanical Gardens of Meisse, BE	18/11/2019	0,3
► **Seminar plus**		
► **International symposia and congresses**		
16th Congress of the European Society of Evolutionary Biology (ESEB), Groningen, NL	20-25/08/2017	1,0
3rd UK plant evolution meeting, Jodrell Laboratory, Royal Botanical Gardens Kew, London, UK	03-05/04/2019	0,6
71st National Botany Congress, BR (cancelled due to COVID-19 pandemic)	2020 (cancelled)	0,0
Plant and Animal Genome (PAG) XXVIII conference in San Diego, CA, USA (cancelled due to COVID-19 pandemic)	2020 (cancelled)	0,0
► **Presentations**		
Poster at the 16th ESEB congress entitled: Using repeat abundance to obtain insights into the phylogeny of species in section *Ciconium* (Geraniaceae)	21/08/2017	1,0

Poster at the annual Experimental Plant Sciences meeting entitled: *Pelargonium* genomics for overcoming cytonuclear incompatibility and bridging species barriers	09-10/04/2018	1,0
Presentation at the Royal Botanical Gardens Edinburgh entitled: Towards a *Ciconium* genomic atlas How to infer cytonuclear incompatability	02/04/2018	1,0
Presentation at the 7th workshop on the application of next generation seqeuncing to repetitive DNA in plants	23/05/2018	1,0
Presentation at the Genetics Department of WUR entitled: The first hints of the genetic processes involved in chloroplast inheritance and expression in *Pelargonium* section *Ciconium*	03/09/2019	1,0
Presentation at EPS theme 4 symposium entitled: A first step towards an 'atlas' of all three genomes in *Pelargonium* section *Ciconium* for use in horticulture	25/09/2018	1,0
Presentation at Syngenta laboratories entitled: The first hints of the genetic processes involved in chloroplast inheritance and expression in *Pelargonium* section *Ciconium*	14/10/2019	1,0
Presentation at the Wageningen PhD symposium entitled: Patterns of chloroplast inheritance in *Pelargonium* section *Ciconium* - First hints of the underlying processes	25/10/2019	1,0
Presentation at the Royal Botanical Gardens of Meisse entitled: Cyto-Nuclear incompatibility (CNI) and speciation in *Pelargonium* section *Ciconium*	18/11/2019	1,0
Presentation at EPS theme 4 symposium entitled: Cytonuclear incompatibility and speciation in *Pelargonium* section *Ciconium*	13/12/2019	1,0
► **3rd year interview**		
► **Excursions**		
Visit to the Syngenta *Pelargonium* collection in Angers, FR	18-19/10/2016	0,6
Subtotal Scientific Exposure		17,0

3) In-Depth Studies	*date*	*cp*
► **Advanced scientific courses & workshops**		
Course: Phylogenetics, principles and methods, EPS & WUR, Wageningen, NL	17-19/05/2016	1,0
Course: Introduction to next generation sequencing, EMBL, Hinxton, Cambridge, UK	03-06/10/2016	1,0
Workshop: 7th workshop on the application of next generation sequencing to repetative DNA analyses in plants, Ceske Budejovice, CZ	22-24/05/2018	0,6
Workshop: Computing skills for reproducible research, Software Carpentry/ ELIXIR, Ghent, BE	24-25/10/2018	0,4
► **Journal club**		
Journal club of the Biosystematics group	29-02-2018	0,2
► **Individual research training**		
Overal training in managing and executing breeding experiments in an industrial context at Syngenta NV, NL	2016-2021	3,0
Subtotal In-Depth Studies		6,2

4) Personal Development	date	cp
► **General skill training courses**		
Course: EPS introduction course, Wageningen, NL	29/09/2016	0,3
Course: Braintraining, Wageningen, NL	20/09/2017	0,2
Course: Bridging across cultural differences, Wageningen, NL	28/11/2018 - 12/12/2018	0,7
Course: Effective behaviour in your professional surroundings, Wageningen, NL	23/10/2018 - 13/11/2018	1,3
Course: Efficient writing strategies, Wageningen, NL	03/2019 - 05/2019	1,3
Workshop: From emotional to cultural intelligence, Wageningen, NL	21/09/2018	0,2
Workshop: Last Stretch of the PhD Programme, Wageningen, NL	02/10/2020	0,0
Workshop: Applying for a Marie Sklodowska-Curie Fellowship: from proposal to project, Wageningen, NL	10/12/2020	0,2
Workshop: Writing propositions for your PhD, Wageningen, NL	11/12/2020	0,0
► **Organisation of meetings, PhD courses or outreach activities**		
► **Membership of EPS PhD Council**		
Subtotal Personal Development		4,2

5) Teaching & Supervision Duties	date	cp
► **Courses**		
Advanced Biosystematics	2016,2017,2018,2019	3,0
Evolution and Systematics	2017,2018,2019	
Biosystematics, Evolution and Agrobiodiversity	2018	
Introduction Bioinformatics	2019,2020	
Ethnobotany	2019	
► **Supervision of BSc/MSc students**		
Jeroen Hofs, MSc student	2018-2019	3,0
Guangnan Chen, MSc student	2019	
Sieme Pelzer, BSc student	2018	
Joost Korver, BSc student	2020-2021	
Subtotal Teaching & Supervision Duties		6,0

TOTAL NUMBER OF CREDIT POINTS*	**37,9**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS with a minimum total of 30 ECTS credits.

** A credit represents a normative study load of 28 hours of study.*

This thesis was funded by the Dutch Foundations for applied scientific research (TTW). Grant number: 14531 “Pelargonium genomics for overcoming cytonuclear incompatibility and bridging species barriers” of the Green Genetics program.

This project was also financially supported by the Alberta Mennega foundation Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design: Floris C. Breman & Iliana Boshoven-Gkini | AgileColor.com
Layout design: Iliana Boshoven-Gkini | AgileColor.com
Printed by: GVO drukkers & vormgevers | gvo.nl