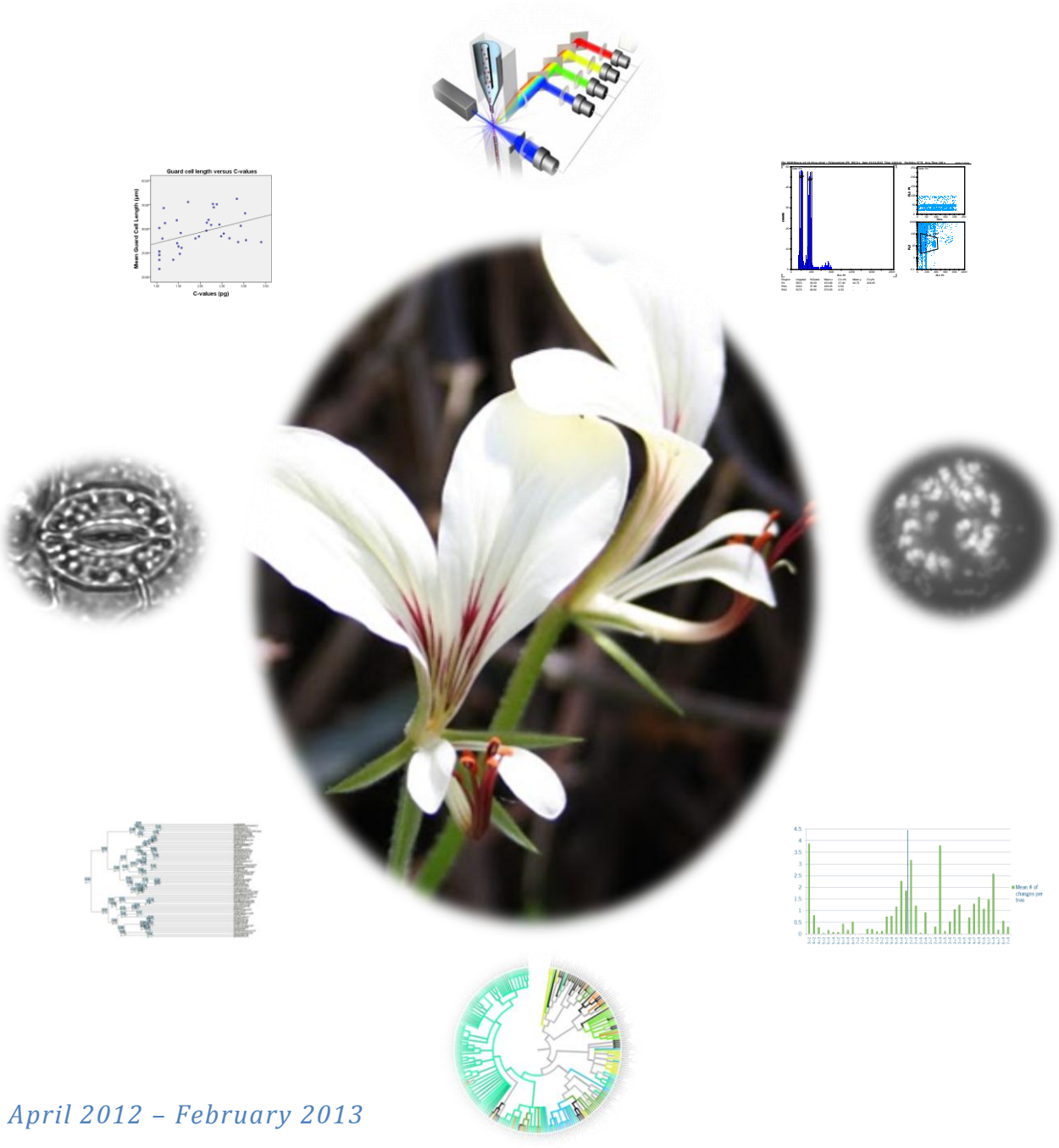
MSc Thesis

Evolutionary trends in genome size and polyploidy in *Pelargonium* (Geraniaceae)

Mathijs Nieuwenhuis
MBI 900326603020
Supervised by: Freek T. Bakker



April 2012 – February 2013

Contents

Summary **5**

Chapter 1 – Introduction **6**

1.1 - Genome size **6**

Genome size variation 6

Biological consequences 6

Box 1 - The C-value *7*

Mechanisms 9

1.2 – Ploidy **10**

Mechanisms 10

1.3 - Genome Size Evolution **11**

1.4 – Pelargonium **12**

Phylogeny 12

Biogeography 14

Chapter 2 - Scientific Problems and Aims of this Thesis **15**

Genome Size 15

Hypotheses 15

Chapter 3 - Phylogenetic Tree **17**

Box 2 – Bayesian model specifications (adapted from Ringelberg et al. (in prep.) *17*

Chapter 4 - Karyotyping **20**

Technique 20

Materials and methods 20

Results 21

Discussion 22

Chapter 5 - DNA Barcoding **24**

DNA Barcoding 24

CTAB extraction 24

Results and discussion 24

Conclusions 26

Chapter 6 - Flow Cytometry **28**

Technique 28

Sampling 30

Results 30

Weng *et al.* data integration 32

Results and discussion 34

Chapter 7 - Stomata measurements **36**

Results and discussion 37
Conclusions 38
Chapter 8 - Evolutionary inferences 39
8.1 - Pagel's transformations 39
8.2 - Character Optimizations 41
Technique 41
Discrete 41
Continuous 47
Discussion 49
8.3 - Trait correlations 51
Genome size and polyploidy 51
Genome size and guard cell length 53
8.4 – Discussion 55
Chapter 9 – Conclusions 56
9.1 - Future prospects 57
Acknowledgements 58
References 59
Appendices 64
Appendix 1 – Concept Map 65
Appendix 2 – PCR & Cycle Sequencing profiles 66
PCR 66
PCR fragment purification 66
Cycle Sequencing 66
Appendix 3 – Flow Cytometry samples 67
Appendix 4.1 – Flow Cytometry Protocol at Plant Cytometry Services (DAPI) 68
Isolation of Nuclei 68
Flow Cytometry 68
Material 68
Appendix 4.2 – Flow Cytometry Protocol at Plant Cytometry Services (PI) 69
Isolation of Nuclei 69
Flow Cytometry 69
Material 69
Appendix 5 – Flow Cytometry Histograms 70
Appendix 5 – Statistical Tests ..... Error! Bookmark not defined.
Appendix 6 – Discrete Character Optimizations (Square Trees) 77
Appendix 7 - R Scripts 78
Continuous optimization and randomized control of node values: 78
Parsimony and Neighbour Joining barcoding tree analysis: 80
Appendix 8 – Online Resources 81
Accesion photos 81

Large tables..81
Statistical data 81

Summary

Genome size studies are becoming increasingly important with the rapid advance of Next Generation Sequencing technology. The genome size of an organism is an important trait that influences many adaptive characteristics. By estimating genome sizes for entire genera evolutionary patterns in genome size variation can be inferred. The genus *Pelargonium* is known for its morphological diversity, genomic instability and horticultural importance. There seems to be a basal division in the phylogeny of *Pelargonium* based on chromosome length. In addition, the genus shows several independent polyploidization events in all three main clades. Whether the genome size has changed as a result of these chromosomal variations is not known. In this thesis, the genome size of 46 *Pelargonium* species was estimated using flow cytometry. In addition, the ploidy level of these species were determined through karyotyping and data obtained from literature. Genome size changes were mapped across the phylogeny of *Pelargonium* and were related to several factors, including guard cell length and polyploidization to identify possible correlations.

Chapter 1 – Introduction

The genome of an organism is not just a sequence of base pairs that encodes various proteins, it is also a physical structure with a specific size and weight. Like the genetic code contained within, the size of the genome varies among organisms, and is imperfectly transferred from one generation to the next. Deletions, insertions or wholesale duplications alter the DNA content parents pass down to their offspring.

While the combination of nucleotides determines the identity of the individual organism, the raw amount of genetic material plays no role in this. Rather, the effects of varying genome size affect organisms more indirectly, by increasing cell size and cell cycle times (e.g. Bennett 1972). As a varying genome size ultimately affects the whole organism, it becomes a trait that may be selected for like any other, driving genome size changes in species over time.

In this thesis, I tracked the evolutionary history of genome size variation in the angiosperm genus *Pelargonium*, investigating how genome size change correlates to chromosome numbers and cell size in this morphologically diverse clade.

1.1 - Genome size

Genome size variation

The genome size of a species can be highly variable. Within flowering plants alone there is over a 2000-fold difference between the smallest genome (belonging to *Genlisea aurena* at 1C = 0.065 pg, Greilhuber *et al.* 2006; in Leitch & Leitch 2013) and the largest (*Paris japonica*, 1C = 152.23 pg, Pellicer *et al.* 2010). The underlying mechanisms of genome size evolution are unclear and many theories exist (Gregory 2001, Weiss-Schneeweiss *et al.* 2006).

A common misconception is that a larger genome indicates a larger organismal complexity. This has led to the 'C-value paradox', an apparent contradiction in evolutionary science based on the inexplicably large genomes of apparently simple organisms. For instance, the human genome size is calculated as being 3.5 picograms, whereas the amphipod *Ampelisca macrocephala* has a genome size of roughly 64 picograms (Gregory 2005). Genetically, the amphipod therefore seems to be much more complex than a human. The reason this is false is that a higher DNA content does not implicate possession of more functional genes. In fact, most DNA consists of non-coding, repetitive sequences, and the ratio of this non-coding DNA to functional genes is correlated with genome size. This then raises a new question: why does non-coding DNA accumulate in certain species and not in others?

Biological consequences

While non-coding DNA has no direct function as a transcription code for protein synthesis, it does have indirect effects on cellular properties. For instance, increased DNA content enlarges the nucleus and the cell. Bennett (1972) proposed the term 'nucleotype' to refer to properties of DNA that affect the phenotype apart from the genetic code, such as mass and volume. These properties and their consequences for the organism's phenotype can be expected to translate into traits that are beneficial or detrimental to reproductive fitness. As

such, natural selection can act upon the nucleotype as it does on the genetic code, resulting in evolution of genome size. In both plants and animals, genome size might be an important factor driving species formation (Vinogradov 2004, Knight *et al.* 2005, Kraaijeveld 2010).

Several publications highlight the correlations between genome size and life history traits in plants (Matzk *et al.* 2003, Grotkopp *et al.* 2004, Jakob *et al.* 2004). Perennial species are more likely to have large genome sizes than annuals. Bennett (1972) proposed that species with a genome size exceeding a C-value of 20 pg are obligate perennials. In addition, they may be restricted to a terrestrial habitat (Knight & Ackerley 2002). The importance of a small genome size for annual species may be an increased growth rate and more rapid cycling of nutrients (Chase *et al.* 2005).

Box 1 - The C-value

The genome size is usually calculated as the amount of DNA in a haploid nuclear genome. This value, normally represented in picograms, is known as the C-value. The term C-value stems from a publication by Swift (1950), in which the 1C-value represented the DNA content of a haploid cell. 2C, then, represented DNA content of a diploid cell, 4C of a tetraploid cell, and so forth.

Since then, however, the term C-value has started a life of its own in scientific literature and several definitions have been in use. Greilhuber *et al.* (2005) proposed a uniform terminology which will be followed in this thesis. Greilhuber's definition of a 1C-value is as follows:

"DNA content of one non-replicated reduced holoploid genome with the chromosome number n. Also the half of a non-replicated holoploid non-reduced genome with the chromosome number 2n." (Greilhuber *et al*. 2005)

With a 'holoploid' genome consisting of the DNA content of the full chromosome complement n, regardless of ploidy level. For example, an octaploid species with basic chromosome number x=11 would have 88 chromosomes in a meiotically unreduced cell (2n=88). The 1C-value of such a species, then, would be the DNA content for a reduced cell with n=44 chromosomes.

In addition, Greilhuber defined DNA content of a cell nucleus containing *x* chromosomes as the 1Cx-value. In the previous example, the 1Cx-value of the octaploid species would therefore be equal to the DNA content of its basic chromosome number, x=11.

One important consequence of increased genome size is a higher cell volume. For plants, smaller guard cells are a common adaptation to cope with water stress. It can therefore be expected that plants suffering from water stress have a reduced genome size. Chase *et al.* (2005) found support for this hypothesis in the *Oncidiinae (Orchidaceae)*, showing epiphytic orchids had smaller genomes than terrestrial species, probably due to selection on smaller cells to alleviate water stress. Beaulieu *et al.* (2008) confirmed that genome size is a strong predictor of cell size based on epidermal cell area and guard cell length. In addition, they found a negative correlation between genome size and stomata density. Veselý *et al.* (2012) also found strong positive correlations between genome size and stomata length. The link between genome size and guard cell size has in fact been used to estimate genome sizes of fossilized plant remains (Masterson *et al.* 1994, Franters *et al.* 2012). In addition, the genome size shows a strong, positive correlation with cell cycle time (e.g. Vinogradov 1999, Francis *et al.* 2008).

Several publications highlight the correlations between genome size and life history traits in plants (Matzk *et al.* 2003, Grotkopp *et al.* 2004, Jakob *et al.* 2004). Perennial species are more likely to have large genome sizes than annuals. Bennett (1972) proposed that species with a genome size exceeding a C-value of 20 pg are obligate perennials. In addition, they may be restricted to a terrestrial habitat (Knight & Ackerley 2002). The importance of a small genome size for annual species may be an increased growth rate and more rapid cycling of nutrients (Chase *et al.* 2005).

Knight and Ackerley (2002) investigated the occurrence of angiosperm species with small or large genomes along an environmental gradient. For their hypothesis they summarized previous studies on genome size/environment correlations, which often contradicted each other, and proposed species with a large genome size are rare in arid regions with extreme temperatures (both cold and warm). Applying quantile regression methods, they found species with large genomes to indeed be much more sensitive to environmental change. Species with small chromosomes were equally abundant across the entire gradient while large genome species became increasingly rare in extreme, stressful environments.

This does not mean that species living in dry, stressful environments will always have small genomes. For instance, Jakob *et al.* (2004) found species in the genus *Hordeum* with genome sizes varying from small to large living in both arid conditions as well as more moderate climates. In this genus, genome size seemed uncorrelated with water stress. However, the genome size did show a strong correlation to annual or perennial life form. Perennials were found to possess consistently larger genomes than annual species.

Another possible consequence of a larger genome is a decreased speciation rate and a more restricted ecological distribution. Knight *et al.* (2005) assembled C-values of many different plant lineages and found a correlation between their genome size and speciation rate. Clades with large genomes were overall significantly less specious than clades with smaller genomes. They also found that species with a small genome size often had a wider habitat range. The restriction of speciation and distribution of plants with larger genomes may be caused by their lower morphological and physiological variety. For instance, the authors found a significant, negative correlation between genome size and photosynthetic rate, whereas small genome species had a wide range of photosynthetic rates. In addition, large genome species always produce relatively large seeds, hampering their dispersal capacity (Knight & Beaulieu 2008).

Mechanisms

In plants, the most straightforward mechanism of genome enlargement is probably polyploidization. Either by duplicating the plant's own genome (autoploidy), or by incorporating another plant's genome with its own (alloploidy), polyploidization is known to occur constantly across all plant families (Ramsey and Schemske 1998, Wendel 2000).

Another common mechanism for genome enlargement is retrotransposon amplification (Park *et al.* 2011). Transposons are non-coding, mobile DNA strands that make up large parts of the genome. In humans, in fact, they may make up as much as 50% of our total DNA, while in plants they account for as much as 90% (Kazazian *et al.* 2004) In plants, proliferation of transposable elements is perhaps the most important factor in genome size enlargement. The timescale at which DNA content is added through retrotransposon amplification is relatively short; in various studies the doubling of genome size by these means is shown to range from 3-5 million years (Hawkins *et al.* 2009). Plants with small genomes generally have only a small fraction of their genome made up of retrotransposons. In maize and other plants with large genomes, however, retrotransposons usually make up over half of the entire genetic code (San Miguel *et al.* 1996).

For a time, only mechanisms causing increased genome size were known, such as polyploidy and retrotransposon amplification. If no mechanism for genome size reduction existed, than this would inevitably lead to irreversible accumulation of genetic material in the cell, or 'genetic obesity' (Bennetzen & Kellogg 1997). However, such mechanisms do exist, and include unequal crossing-over events, illegitimate recombination (Devos *et al.* 2002), enhanced deletion rate, and selection against transposable elements. In fact, Leitch and Bennett (2004) showed that polyploid plants usually had smaller Cx-values than their diploid relatives, proposing that 'genome downsizing' may be much more common than previously thought. For these reasons, the number of chromosomes in a cell is not directly correlated to genome size (Chase *et al.* 2005). An organism may have twice as many chromosomes, but they may be twice as small individually as the result of genome downsizing. See Figure 1 for a schematic outline of the major processes involved in genome size variation, as well as some of the possible consequences of altered DNA content.

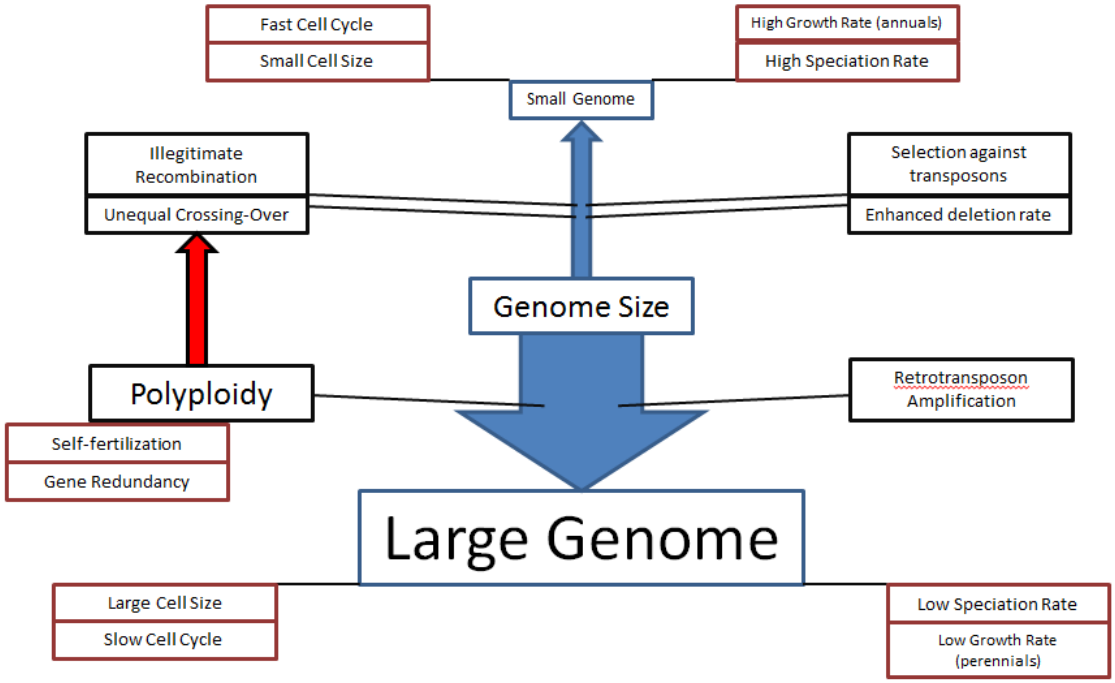


Figure 1 – Overview of the causal mechanisms behind genome size variation (black boxes), as well as some of the possible consequences on a cellular, organismal and species level (red boxes). Polyploidy is responsible for both genome enlargement through direct duplication of genetic material, as well as genome downsizing as subsequent deletion of genetic material is quite common in polyploids.

1.2 – Ploidy

The number of chromosomes in a cell varies strongly among organisms. In plants polyploidy is a particularly common occurrence, and may arise from duplication of the genome (autopolyploidy) or inclusion of another plant's genome during hybridization (allopolyploidy).

Polyploids may occur in a population of diploids as a result of unreduced gamete formation, possibly with an unstable 'triploid bridge' following the union of a reduced and an unreduced gamete. Newly formed polyploid plants are faced with diminished reproductive fitness in their first few generations (Comai 2005). As the genome stabilizes and the species adapts, however, the evolutionary advantages of polyploidization start to tell.

Mechanisms

Polyploidization has played a major part in plant evolution, with as many as 50 to 80 percent of all angiosperm species having undergone polyploidization at some point in their evolutionary history (Masterson 1994, Leitch and Bennett 1997). Polyploidization occurs in individual plants, usually through erroneous production of unreduced gametes. Temperature variations, especially cold weather, can increase the frequency of unreduced gamete production (Bretagnolle and Thompson 1995). This may explain the abundance of polyploid plants in mountainous or arctic regions as the lower temperatures induce formation of unreduced gametes.

The number of chromosomes is not directly correlated to genome size (Chase *et al.* 2005). Due to genome downsizing, C-values do not increase linearly with ploidy levels. Leitch and Bennett (2004) showed that polyploid plants tend to have smaller Cx-values than their diploid ancestors, indicating that loss of DNA is a frequent consequence of genome duplication. They identified homeologous recombination, activation of transposons, and gene elimination as the most likely mechanisms reducing genome size during polyploidization.

1.3 - Genome Size Evolution

Various studies have sought out evolutionary patterns in genome size variation (e.g. Weiss-Schneeweiss et al. 2006, Beaulieu et al. 2008, Leitch et al. 2009). With so many adaptive consequences it is not unlikely that there is a selection on the size of an organism's genome as well as its functional code. To infer evolutionary trends in genome size variation it would be best to select a genus with an instable genomic history to ensure sufficient variation in both C- and Cx-values. For these reasons, the genus that I chose to study evolutionary trends in genome size was *Pelargonium* L'Her.

Grotkopp et al. (2004) inferred evolutionary patterns in genome size in *Pinus* and correlated this to life history traits, environmental characteristics and so forth. They found no correlation between genome size and minimum generation times, but found a strong correlation between genome size and seed mass. In addition, invasiveness was negatively correlated to genome size. *Pinus* shows both increase and decrease of genome size across its taxa. However, genome size increase was more common and Grotkopp *et al.* suggest this may be due to vertebrates selecting larger seeds and therefore exerting selection on pines with larger genome sizes.

A good candidate for an evolutionary reconstruction of genome size variation should have a well-supported phylogenetic consensus tree. In addition, variation in chromosome numbers (including different levels of ploidy), as well as known differences in chromosome size are important to confirm suspected variation in genome size in the first place. Secondary considerations are differences in morphology, physiology and ecology between species to allow for correlations between genome size and various biological traits. The genus chosen for this study is *Pelargonium* (Geraniaceae), which complies to all mentioned criteria.

1.4 – Pelargonium

Pelargonium L'Her is a genus of angiosperm plants mostly found in southern Africa, specifically the Cape Floristic Region where it is the sixth largest angiosperm genus. The CFR is geographically isolated, with a level of endemism normally found on islands (>70%,), and the region is a model system for angiosperm evolution (Linder 2003, Bakker *et al.* 2005, Verboom *et al.* 2009). As a result, *Pelargonium* is one of the genera studied here. It comprises a high diversity of vegetative and floral morphologies, ranging from stem succulents to woody geophytes. The genus is primarily of economic importance for it contains the popular garden geranium, *P. x hortorum,* which itself is a cross between the species *P.inquinans* and *P. zonale.* In addition, many *Pelargonium* species build up secondary metabolites in their leaves, which are extracted for use in the pharmaceutical industry as essential oils or antimicrobial compounds (Kolodziej *et al.* 2006).

From an evolutionary perspective, *Pelargonium* is of particular interest due to genomic instability and an abnormally high rate of substitution (Mower *et al.* 2007**,** Bakker *et al*. 2007 Chumley *et al.* 2006), indicated by its long branch in phylogenetic trees. The mtDNA nucleotide substitution rate appears to be roughly 600 times as high as in other angiosperm groups. Another aspect of the perceived genomic instability in this genus is the broad range of karyotypes, showing interspecific differences in chromosome numbers and size.

Phylogeny

The *Pelargonium* genus is divided in two major clades based on chromosome size (Gibby et al. 1996). The 'small chromosome' clade contains species with a chromosome length of less than 1.5 µm while the 'large chromosome' clade has chromosomes between 1.5 and 3 µm in length. So far no other morphological distinction between these two clades has been identified.

The small chromosome clade comprises two smaller clades, simply known as clade A and B (Bakker *et al.* 2004, Figure 2). Together they comprise approximately 80% of the 280 known Pelargonium species. Clade A is morphologically the more diverse of these, featuring 11 out of 16 described morphological varieties, most of them found in the winter rainfall region of the South African Cape. Karyological differentiation is rare, but a few polyploids in the Magnistipulaceae clade have settled in areas with summer rain (Bakker et al. 1999). Two subclades can be identified within clade A. Clade A1 mostly consists of woody shrubs, whereas clade A2 features a diverse and specious range of xerophytic plants which may have arisen through adaptive radiation. Clade A1 contains sections *Campylia* and Pelargonium. A2 contains sections *Cortusina, Polyactium, Otidia, Magnistipulacea, Ligularia s.s.*, and *Hoarea*.

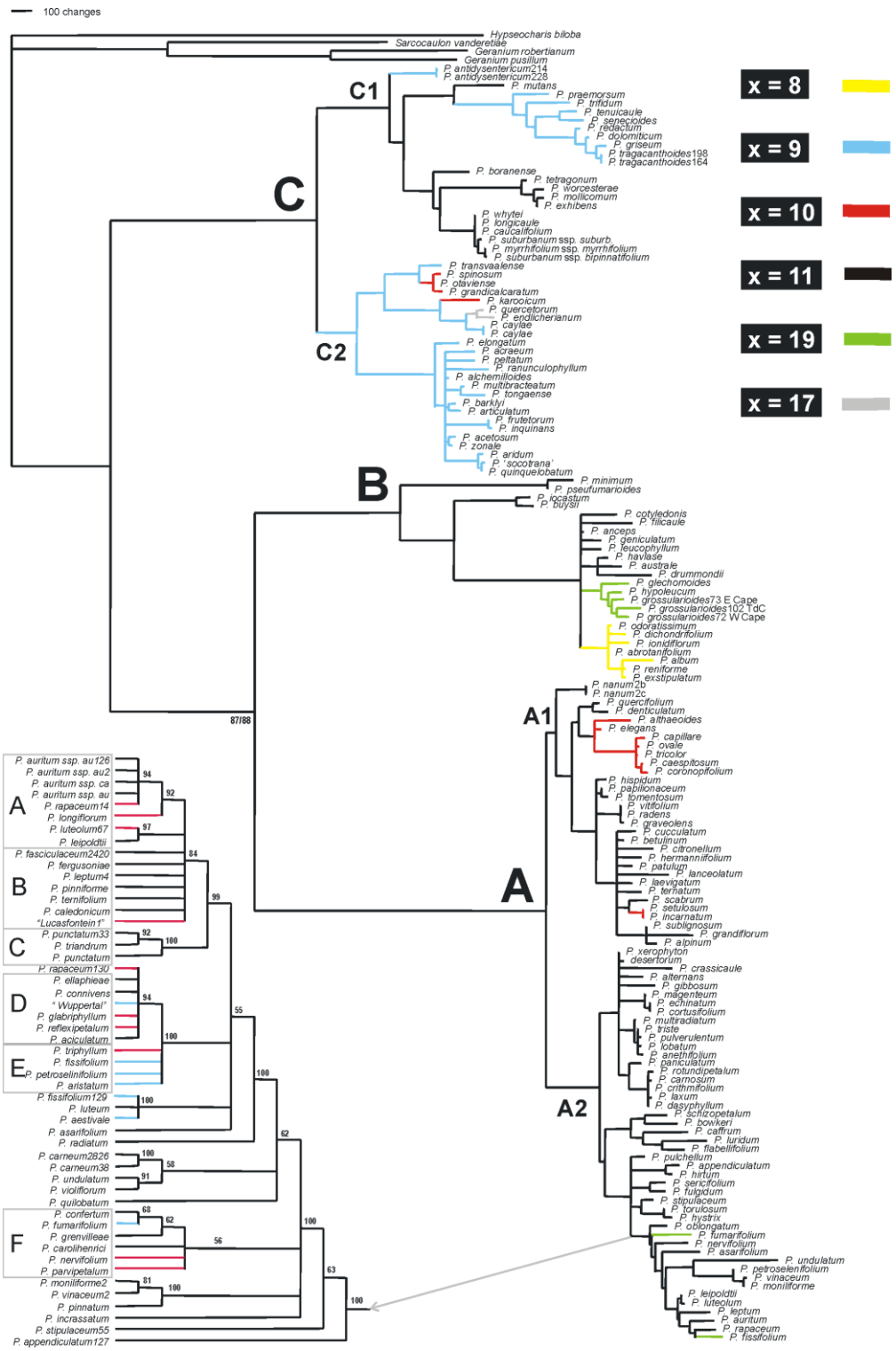


Figure 2 – *Phylogeny of* Pelargonium *from Bakker et al. (2004). Basic chromosome numbers are indicated by coloured branches. Shown left is the phylogeny of* Hoarea*, a subclade of clade A2, noted for its nested radiation. Adaptive radiation and species proliferation like this is a trademark of clade A.*

Clade B contains many annuals and has a wide distribution, ranging from Australia to St. Helena island, though most species are found in South Africa. Its high dispersal capacity and

low pollinator dependence stands in stark contrast with its sister clade (A), which contains plants with low dispersal capacity and high pollinator dependence. Clade B contains sections *Peristera* and *Reniformia*.

The large chromosome clade comprises two subclades, C1 and C2 (Bakker *et al.* 2004). Clade C1 consists of sections *Jenkinsonia, Chorisma* and *Myrrhydium*. C2 contains sections *Subsucculentia* and *Ciconium*. Both clades have a wide distribution range, from Turkey to Madagascar. Interesting to note is that many sister species are separated geographically by large distances, indicating they only recently dispersed and became isolated from each other

Biogeography

Pelargonium L'Hérit is mostly found in the South African Cape region, with some species settling in distant regions such as St. Helena island, Australia, Madagascar or the Middle-East. The winter rainfall region, along the southwest coast of the African continent, contains the largest diversity of Pelargonium species, particularly in its fynbos and karoo biomes. The hot, dry summer and wet winters are characteristic of the Mediterranean climate that dominates the region, probably since the late Pliocene. During this era the region aridified, and it is possible the genus diversified in response to this climate shift (Bakker *et al.* 2005).

Chapter 2 - Scientific Problems and Aims of this Thesis

Genome Size

Within the phylogeny of the genus *Pelargonium* there seems to be a basal subdivision based on chromosome size. The small chromosome lineage containing clades A and B and the large chromosome lineage containing clade C as described by Bakker *et al*. (2004) diverged roughly 30 million years ago (Bakker *et al.* 2005). While it is known that the chromosome sizes of clades A and B do not exceed 1.5µm, and the chromosome size of clade C ranges from 1.5 to 3 µm, no study has ever directly measured the C(x)-values for the different species comprising these clades. In this thesis I estimated the C(x)-values for 46 *Pelargonium* species using flow cytometry and optimized genome size as a character over the established phylogenetic tree. I inferred the evolution of genome size in *Pelargonium* and sought explanations for the variation in DNA content between species for the genus. Differences in genome size of clades C, B, and A were correlated to ploidy levels and stomata size.

The main purpose of this thesis is to identify a possible evolutionary pattern in genome size variation in *Pelargonium,* and to correlate this variation to ploidy level and stomata size. As a secondary objective, the genome size estimates for the various *Pelargonium* species will be useful for future endeavours to sequence complete *Pelargonium* genomes. In a broader context, the findings of this study may support existing theories on genome size evolution in plants. This information is essential for research on genome evolution and structure. In addition, it will provide insight on the variety in genome size of a genus notorious for its morphological diversity and high mutation rate.

Hypotheses

Based on the literature research and the current *Pelargonium* phylogeny, the following hypotheses were tested:

- **Average genome size of each of the small chromosome clades will be significantly smaller than that of the large chromosome clade.** For each of the clades within the genus, average genome sizes were calculated. Significant differences in genome size between these groups were identified. Should the mean genome size of each of the groups in clades A and B be significantly smaller than the mean genome size of each of the groups in clade C, the division between the small and large chromosome clades will be strongly correlated to genome size. Should no significant difference in genome size be found, this raises new questions. For instance, how has the variety in basic chromosome number and ploidy level differences in *Pelargonium* occurred without altering genome size? Would there perhaps be epigenetic differences between clades despite the similar genome size?
- **Genome size variation in *Pelargonium* has a significant phylogenetic signal.** Considering the basal division in the genus based on chromosome size, and our expectation that chromosome and genome size in *Pelargonium* are linked, we expect genome size variation to be dependent on the phylogeny. By applying Pagel's transformations I tested whether genome size in *Pelargonium* shows a significant phylogenetic signal.
- **Changes in genome size are correlated with life form (annual/perennial), and guard cell length.** Annual species, such as those found in clade B, are expected to

have a significantly smaller genome size than perennial species. I expect the length of guard cells to be positively correlated to C-values. Finally, Cx-values may be negatively correlated with ploidy level. The Connecticut Project (See www.http://darwin.eeb.uconn.edu/wiki/) aims to study the genetic variation of *Pelargonium* in relation to environmental conditions such as drought. Their findings may be of use in this thesis to correlate genome size variation to environmental parameters.

- **There is a correlation between genome size variation and polyploidy in *Pelargonium*.** Cx-values may be negatively correlated with ploidy level. Multiple ploidization events seem to have occurred relatively recently in *Pelargonium*'s evolutionary history (Fig.3). Perhaps these events have coincided with genome enlargement, or conversely with genome downsizing. Correlations between genome variation and polyploidy may support a hypothesis that polyploidy is responsible for variable C-values across taxa.

Chapter 3 - Phylogenetic Tree

The *Pelargonium* phylogeny has been intensively studied and carefully reconstructed, with Bakker et al. (2004) publishing the first phylogeny (Figure 2) including all sections of *Pelargonium* recognised at the time. The phylogenetic tree was constructed based on sequences of nuclear (*ITS*, 55 taxa), chloroplast (*trnL-F*, 152 taxa) and mitochondrial (*nad1* b/c exons, 51 taxa) DNA. From this tree the five main clades, A1, A2, B, C1, and C2, were derived. The (AB)C ordering of the main monophyletic clades was further supported by Weng et al. (2012), who reconstructed the phylogeny of 58 *Pelargonium* species based on *rbcl, matK, ndhF, rpoC1,* and *trnL-F* sequences.

Ringelberg et al. (in prep.) reconstructed a phylogeny of *Pelargonium* based on a compilation of *trnL-F, ITS, nad1,* and 30 plastid indels (data from Bakker *et al.* 2004, Jones *et al.* 2009., Van Proosdij *et al.* (in review)). In total, 232 accesions covering 186 taxa were included in this phylogeny (Figure 3). Two independent Bayesian MCMC runs (see Box 2 for details) produced two consensus topologies, one showing the expected (AB)C topology but the other showing a A(BC) topology. This was an unexpected result, as clade A and B have been established as sister clades in previous phylogenetic studies in *Pelargonium,* first tentatively in Bakker *et al.* (2004), and later more firmly in Jones *et al.* (2009) and Weng *et al*. (2012). However, the posterior probability values for the BC node in the A(BC) topology tree was only 0.78, implying only limited support for this phylogenetic reconstruction. For this thesis, the phylogeny with the (AB)C topology will be presumed correct. Although the Weng study includes full character sets for each taxon, they only sampled 58 taxa, whereas the Bakker and Jones papers sampled 153 and 154 respectively. Wiens and Tiu (2012) showed empirically that adding incomplete taxa can improve a phylogeny substantially more compared to adding more characters.

Box 2 – Bayesian model specifications (adapted from Ringelberg *et al.* (in prep.)

The Bayesian analysis was conducted using MrBayes version 3.2.1 (Huelsenbeck & Ronquist 2001, Ronquist *et al.* 2011, http://mrbayes.sourceforge.net/download.php). No substitution model was selected a priori, rather the Bayesian analysis was itself used to sample the General Time Reversible (GTR) model space as detailed by Huelsenbeck *et al.* 2004. The following priors were set to vary across the partitions: nucleotide frequencies, substitution rates of the GTR matrix, the shape parameter of the rate variation's gamma distribution, the proportion of invariable sites and the overall rate. A relaxed Thorne-Kishino 2002 clock model with underlying strict clock model was used to create an ultrametric, rooted phylogenetic tree.

Both runs lasted 500 million generations, with a relative burn-in of 25%. The MC chain was sampled every 1000th generation. Four separate chains were used for each run, of which three were heated with a temperature of 0.5. The topology and branch length priors' starting values were derived from an ultrametric tree that was provided at the start of the analysis. The outgroup was *Pelargonium antidysentericum*.

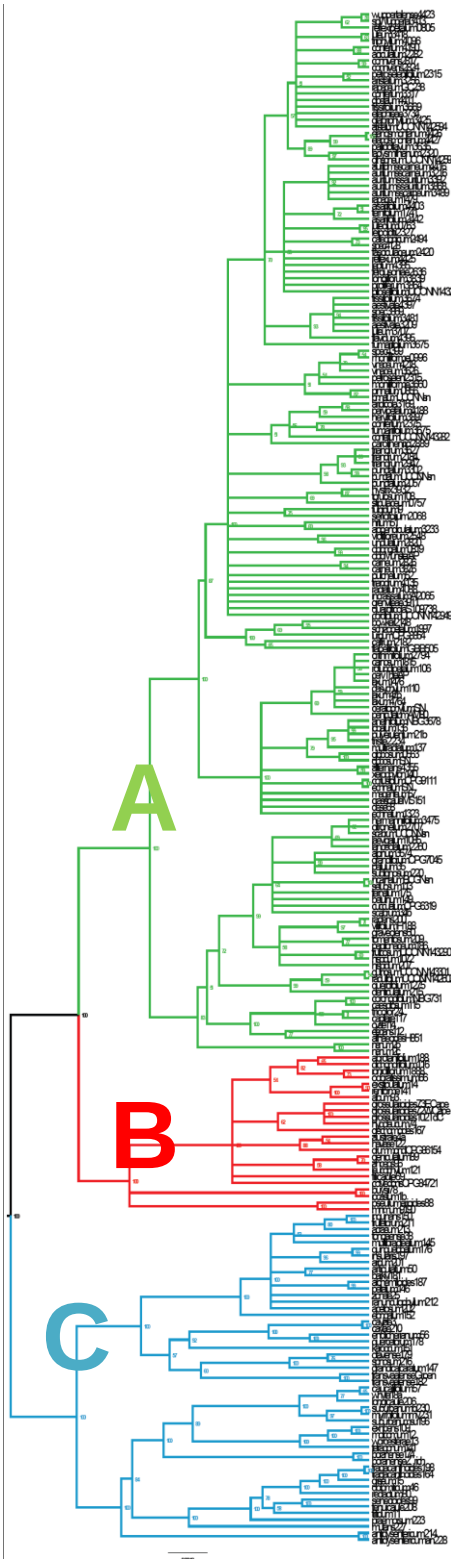


Figure 3 – Bayesian 50% majority rule consensus tree for Pelargonium *from Ringelberg et al. (in prep.) based on 500 million generations under a relaxed Thorne-Kishino 2002 clock model with underlying strict clock model. Posterior probabilities are shown in percentages for each node.*

In light of this updated phylogeny, the distribution of ploidy levels and basic chromosome numbers can be reassessed (Figure 4).

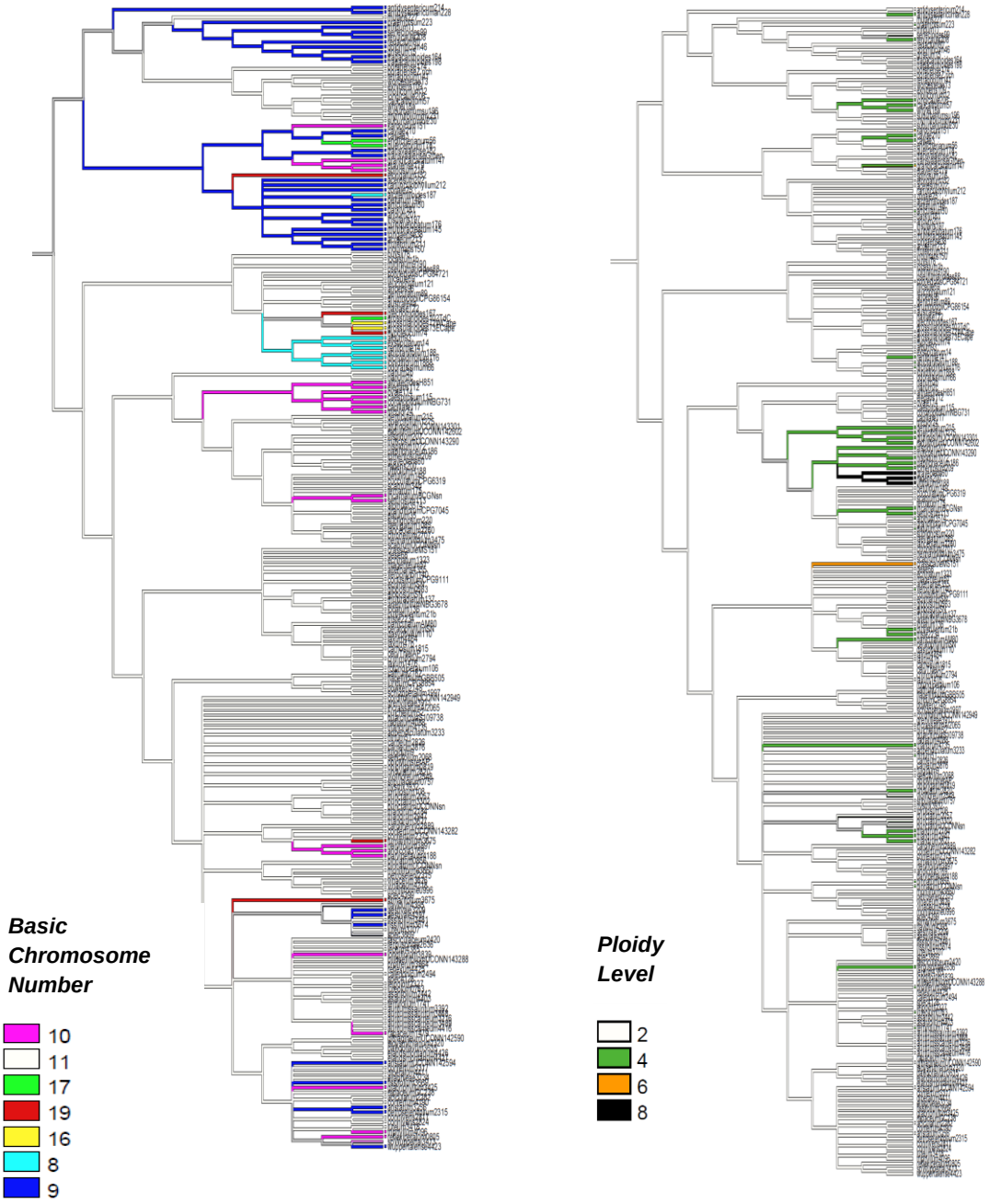


Figure 4 – Two Bayesian trees showing character traces of basic chromosome number and ploidy level across the Pelargonium *phylogeny. Left: Basic chromosome numbers. In the C-clade in particular there seems to be a general trend towards lower chromosome numbers. Species with sixteen or more chromosomes may have been the product of interspecies hybridization. Right: Ploidy level. Independent polyploidization appears to have occurred at least twenty times, thirteen of which happened in clade A. According to this tree, approximately one-fifth of the species in clade A are polyploid. Trees info on page 16.*

Chapter 4 - Karyotyping

Technique

Karyotyping is a common method of chromosome counting or inspection through fluorescence microscopy. Cell tissue from developing root or shoot tips or emerging flower buds is fixated to arrest the chromosomes in position during mitotic or meiotic division respectively. After staining with DAPI the chromosomes can be viewed under a fluorescence microscope (Figure 5).

Karyotyping has been used frequently in *Pelargonium* research to assess ploidy levels and basic chromosome numbers. In particular, papers by Gibby et al. (1990, 1996), Albers & Van der Walt (1984, 1992), and Yu & Horn (1988) contain chromosome counts based on karyograms for many *Pelargonium* species. Some of these papers, notably Gibby et al. (1996), mention the difficulty of acquiring mitotic karyograms of good quality, citing better results from fixated flower buds.

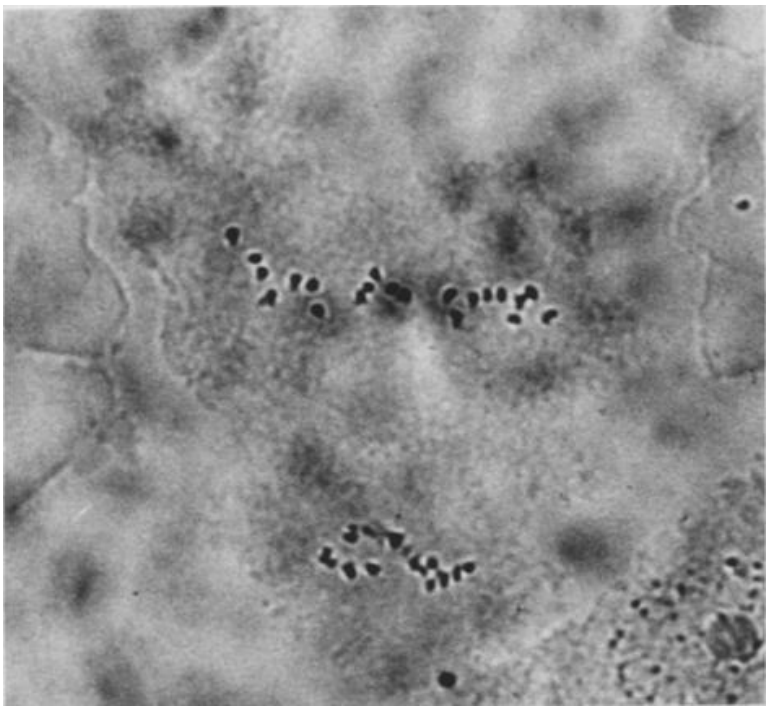


Figure 5 – Example of a karyogram of a hybrid between Pelargonium incarnatum *x* P. patulum*, n=21. The cell's chromosomes were fixated during anaphase I (upper group). From Albers & Van der Walt (1992)*

Materials and methods

To control for (endo)polyploidy/aneuploidy in accessions I used karyotyping to count the chromosomes for individual plants. For many *Pelargonium* species basic chromosome numbers and ploidy levels have been published in literature, and in some species several different ploidy levels have been observed.

In order to create a karyogram cells needed to be isolated from developing tissue, commonly root tips for mitotic and young flower buds for meiotic stages of cell division. The cellular material was fixed and stained with DAPI, so the chromosomes could be counted with a fluorescence microscope. Because the cellular material needed to be as fresh as possible, samples had to be collected on the same day as the treatment, and therefore the plants needed to be kept in a greenhouse in Wageningen. This severely limited the number of

species from which samples could be taken. The species for which material was available in the greenhouse were: *P. acetosum, P. acraeum, P. crithmifolium, P. magenteum, P. myrrhifolium,* and *P. tetragonum.* These plants were provided by Ben Groen (Kerkrade Botanical Gardens). The identity of the plants was double-checked using DNA barcoding (See Chapter 5).

Previous karyotyping studies in *Pelargonium* reported the difficulty of obtaining good mitotic divisions so meiotic divisions were preferable. However, obtaining meiotic divisions is dependent of availability of flower buds. As long as these were unavailable from the seedlings, I attempted to obtain mitotic divisions using standard WUR karyotyping protocols (ref. Hans de Jong).

Root tips were collected in the morning (9-10am) and fixated in 8-hydroxyquinoline for four hours. They were then transferred to a 3:1 ethanol-acetic acid mixture for a day. They were subsequently stored on ethanol in the fridge. Flower buds were collected in the afternoon and immediately fixated in a 3:1 ethanol-acetic acid mixture.

To make the karyograms, both root tips and flower buds were washed with MQ water to remove residual traces of ethanol, and put in cytric buffer. To dissolve the cell walls and free the chromosomes, the samples were subjected to an enzyme treatment consisting of 200 µl of a 1% solution of cellulose RS, pectolyase and cytoheliase. To this solution I added 400 µl cytric buffer, after which the samples were kept in a moist, 37ºC environment to allow the enzyme to digest the cellular material optimally.

The material was subsequently carefully squashed on a microscope slide, fixated with a 60% acetic acid solution, and finally washed with a 3:1 ethanol-acetic acid mixture. Slides were stained with DAPI and examined under a fluorescence microscope.

Results

Unfortunately neither the root tips nor the flower buds yielded slides of sufficient quality, and no metaphase or prometaphase cells were detected. Chromosome data was therefore taken from literature.

The best results were obtained by accident when I used 95% ethanol instead of 3:1 ethanol-acetic acid to fixate root tips of a *Pelargonium myrrhifolium* specimen on my first attempt (See Figure 6). These images show tightly packed chromosomes, clustered together making their exact number hard to determine. No metaphase stage cells were detected over the course of these and subsequent attempts.

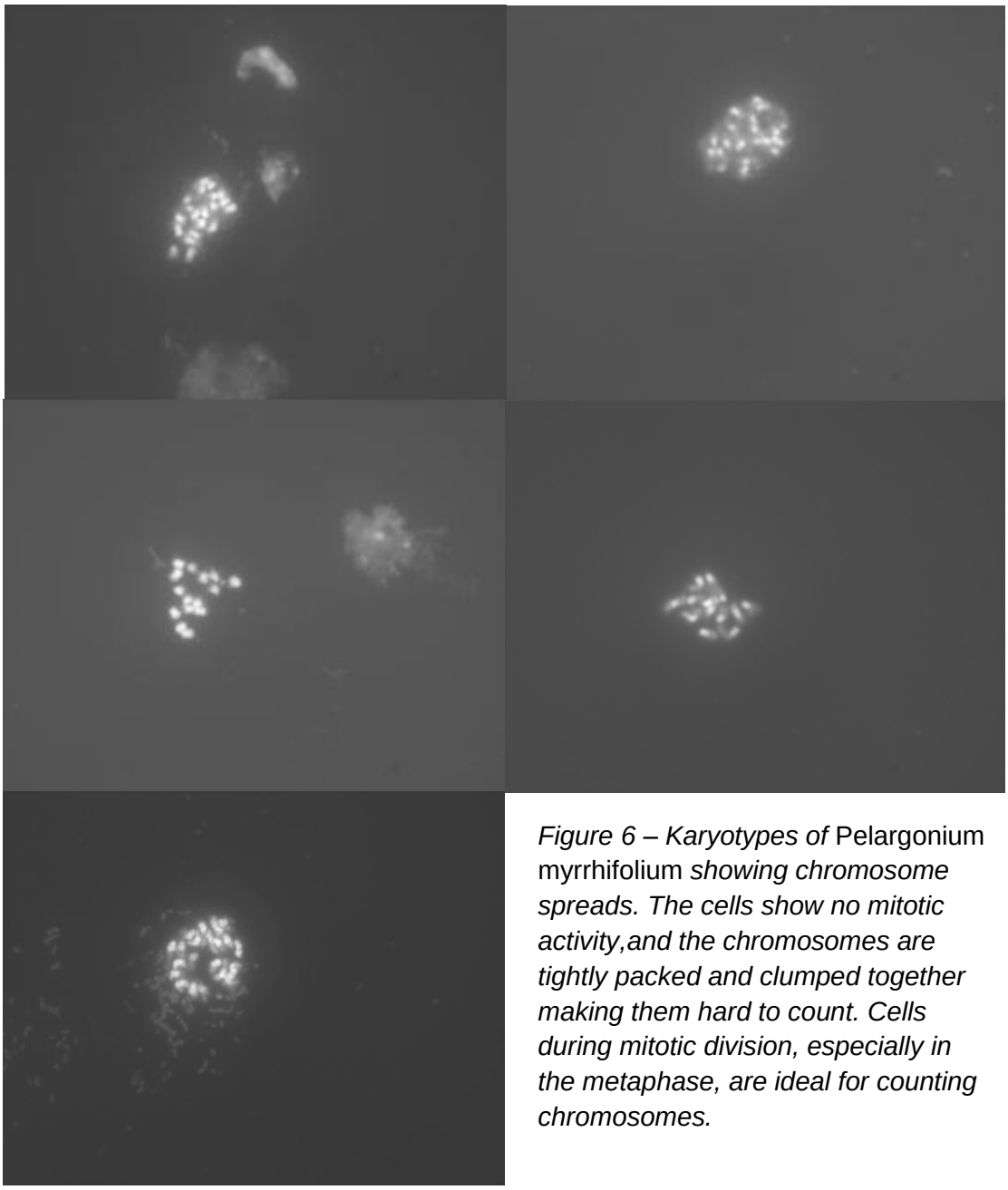


Figure 6 – Karyotypes of Pelargonium myrrhifolium *showing chromosome spreads. The cells show no mitotic activity,and the chromosomes are tightly packed and clumped together making them hard to count. Cells during mitotic division, especially in the metaphase, are ideal for counting chromosomes.*

Discussion

The main problem of the karyotyping experiment was the lack of material available, with only sixteen plants of different species to obtain root tips from, and with only two species flowering, the number of flower buds that could be obtained was even more restricted. Although chromosomes were observed in the karyograms, no cells were found to have been fixated during the (pro)metaphase, rendering chromosome counts practically impossible. To improve the chances of fixating cell tissue during (pro)metaphase, root tips should be collected earlier in the morning (before 08:00 am). Hans de Jong has suggested trying out some deviations from the standard protocol, mainly regarding the pretreatment. Ma *et al.* (1996) published results of three different pretreatment methods in roses: a 0.01mM

colchicine solution, 0.002mM 8-hydroxyquinoline, and 0.001mM 8-hydroxyquinoline and 0.01mM colchicine mix. The mixture of colchicine and 8-hydroxyquinoline yielded the best results after storage at 25ºC for four hours.

Chapter 5 - DNA Barcoding

DNA Barcoding

DNA barcoding is a method of species identification by sequencing a standardized region of the genome (e.g. www.ibol.org, www.barcodeoflife.org). By comparing the sample sequence to a pre-existing database containing sequences of the region for candidate species, a closest match can be found. DNA barcoding has many advantages over other identification techniques (e.g. morphology), but some disadvantages as well. While barcoding allows identification of organisms based solely on their genetic material, it may not be suitable for identification above the species level. Although morphologically indistinct organisms can be distinguished, it requires a comprehensive database for all known species, as well as a genetic region shared by most or all taxa of interest which is well-conserved and varies between but not within species.

To match samples to known species tree- and similarity-based methods, as well as statistical and diagnostic methods can be applied. Van Velzen *et al.* (2012) compared these methods for a simulated dataset of recently diverged species and found tree-based methods inferior to both similarity-based and diagnostic methods.

The Consortium for the Barcode of Life (CBOL) is an organisation championing the use of DNA barcoding as the standard method of species identification, striving to identify suitable genes for use as DNA barcodes and assembling a public sequence database for these genes, among other initiatives. Several genes have been proposed as standard barcode gene for plants, including ITS, rbcL and MatK (e.g. Hollingsworth 2011)

To ascertain the identity of the *Pelargonium* samples from the Hortus VU and those from Kerkrade, I used DNA barcoding following a CTAB protocol. TrnL-F was used as the target gene rather than the 'standard' barcoding genes rbcL and matK, because for *Pelargonium* these genes are less suitable for identification purposes (F.T. Bakker, pers.comm.). In addition the trnL-F sequence database published in Bakker *et al.* (2004), can be seen as a 'local' DNA barcode solution/reference library. Considering this is not an 'official' barcoding project and the aim is to swiftly check the identity of samples, I personally consider this deviation from protocol to be justifiable.

CTAB extraction

The DNA was extracted from leaf samples following standard CTAB protocol as employed in the Biosystematics Group DNA lab. In short, leaves were first dried on silica for at least an hour. Because some leaves proved difficult to dehydrate using just silica powder, I freeze-dried them at -80ºC, and in some cases used liquid nitrogen to instantly freeze and crush the leaf material. The extracted DNA samples were checked for quality and quantity using gel electrophoresis.

Results and discussion

After the PCRs were completed, samples of the amplified products were brought to the Greenomics lab (WUR), where they were sequenced single-stranded using high-throughput DNA analysis. The raw data were then converted to estimates of the DNA sequences for each accession with Codon Code Aligner (http://www.codoncode.com/aligner/). This resulted in 14 sequences of roughly 400 sites each.

To match the obtained accession sequences to a species name, I aligned them to the known TrnL-F sequences for *Pelargonium* in Mesquite using MAFFT. This resulted in a matrix of 1009 characters. I then produced a RAxML tree to see where the accessions fitted in the phylogeny (Figure 7).

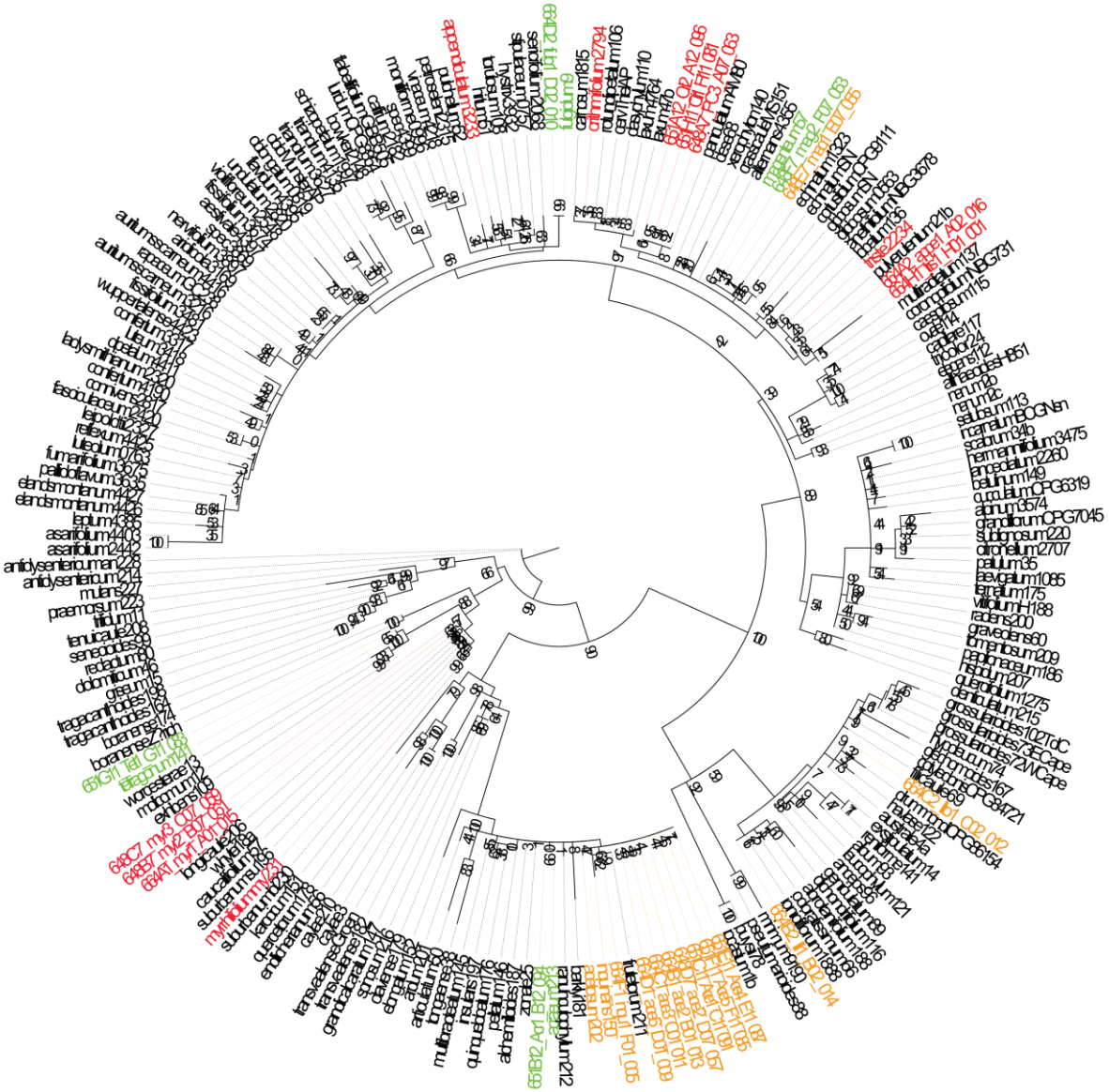


Figure 7 – RAxML bipartitions tree based on the TrnL-F *gene database aligned with MAFFT for* Pelargonium *with sampled accessions included. Accession names in green match the identity of their proposed species, while names in red match a different species. Names in orange are included in polytomies with their proposed species or match a suspected synonym of their proposed species in the case of* P. crithmifolium/paniculatum*.*

From the RAxML tree it appears that five accessions match the species they were supposed to be according to their labels (Figure 8). *P. acetosum* is included in polytomy with the accession labelled as *acetosum* and the accession labelled as *P. crithmifolium* is matched with *P. paniculatum,* a suspected synonym for *crithmifolium* (Knuth 1912, in Van der Walt 1977), which may explain a mislabelling. These two accessions are therefore very close to their proposed species. Finally, the accessions labelled *P. myrrhifolium* and *P. appendiculatum* seem to have been misidentified. As these accessions are placed in polytomies with other species, however, their true identity is difficult to ascertain.

*Figure 8 – Table showing the closest match for each accession according to the DNA barcode sequences. The Ace, Acr, Ful, Mag and Tet accession appear to match their proposed species well (*P. acetosum, P. acraeum, P. fulgidum, P. magenteum *and* P. tetragonum *respectively). Inq matches both the proposed species,* P. inquinans*, as well as* P. frutetorum*. The Cri accessions do not match* P. crithmifolium*, but are grouped together with* P. paniculatum. *Since* P. crithmifolium *and* P. paniculatum *are morphologically similar, the Cri accessions may have been misidentified and should be labelled as* P. paniculatum*. The Myr accessions also do not match their proposed species,* P. myrrhifolium*. However, they do not perfectly match any other species nor with themselves, so establishing their identity still necessitates morphological analysis. The same goes for the Appe accession. Finally, the Tri accession is grouped together with* P. multiradiatum *in the barcode tree, but with a long branch indicating significant morphological dissimilarity.*

Accession	**Proposed Species**	**Closest Match**	**Relationship**
Ace1	*Acetosum*	*Acetosum/Ace2-6*	Polytomy
Ace2	*Acetosum*	*Acetosum/Ace1/Ace3-6*	Polytomy
Ace3	*Acetosum*	*Acetosum/Ace1-2/Ace4-6*	Polytomy
Ace4	*Acetosum*	*Acetosum/Ace1-3/Ace5-6*	Polytomy
Ace5	*Acetosum*	*Acetosum/Ace1-4/Ace6*	Polytomy
Ace6	*Acetosum*	*Acetosum/Ace1-5*	Polytomy
Acr1	*Acraeum*	*Acraeum*	Sister taxa
Appe1	*Appendiculatum*	*Lobatum/Triste/Pulverulentum/Anethifolium*	Monophyly
Cri1	*Crithmifolium*	*Paniculatum/Cri2/Cri3*	Polytomy
Cri2	*Crithmifolium*	*Paniculatum/Cri1/Cri3*	Polytomy
Cri3	*Crithmifolium*	*Paniculatum/Cri2/Cri3*	Polytomy
Ful1	*Fulgidum*	*Fulgidum*	Sister taxa
Inq1	*Inquinans*	*Inquinans/Frutetorum*	Polytomy
Mag1	*Magenteum*	*Magenteum/Mag2*	Polytomy
Mag2	*Magenteum*	*Magenteum*	Sister taxa
Myr1	*Myrrhifolium*	*Longicaule*	Sister taxa
Myr2	*Myrrhifolium*	*Whytei/Caucalifolium/Myr3*	Polytomy
Myr3	*Myrrhifolium*	*Whytei/Caucalifolium/Myr2*	Polytomy
Tet1	*Tetragonum*	*Tetragonum*	Sister taxa
Tri1	*Triste*	*Multiradiatum*	Sister taxa

Conclusions

Certain suspected identities of plants based on morphology were confirmed by the DNA barcoding results. Accessions labelled as *P. acraeum, P. fulgidum, P. magenteum* and *P.*

tetragonum matched their proposed species in the Rx-ML tree. Accessions labelled *P. acetosum* and *P. inquinans* matched their supposed species as well, but were also grouped in polytomies with different species. In the case of *P. acetosum*, the accessions shared a polytomy with their proposed species as well as the branch containing the polytomy shared by *P. inquinans, P. frutetorum* and the accession Inq1. Based on the flower morphology I would confirm the identity of the Ace accessions as *P. acetosum.* The Cri accessions are matched with *P. paniculatum* rather than their proposed species, *P. crithmifolium,* which is plausible because these species are morphologically similar.

The Myr accessions are grouped with *P. longicaule, P. whytei,* and *P. caucalifolium.* However, based on flower morphology *P. caucalifolium* and *P. whytei* do not seem likely identities. The Tri1 accession is grouped together with *P. multiradiatum*, although it still show a lot of unique substitutions. Because of the distinct leaf and flower morphology however, it is unlikely that a specimen of *P. triste* would be mistaken for *P. multiradiatum*.

RAxML is a quick method for creating Maximum Likelihood trees, which can mean the tree produced is not optimal. In addition, papers like Van Velzen et al. (2012) highlight the poor performance of tree-based methods for barcode matching. For instance, the evolution of the genes on which the tree is based may not be identical to the evolution of the taxa they represent. Unfortunately, the barcode sequences were of insufficient quality to apply a diagnostic methods such as BLOG to improve the sequence matching.

Erroneous matches may also be due to the poor quality of some of the sequences obtained. The poor quality of the sequences in turn can be attributed to the difficulty of crushing the desiccated or frozen leaf material for some species.

Chapter 6 - Flow Cytometry

Technique

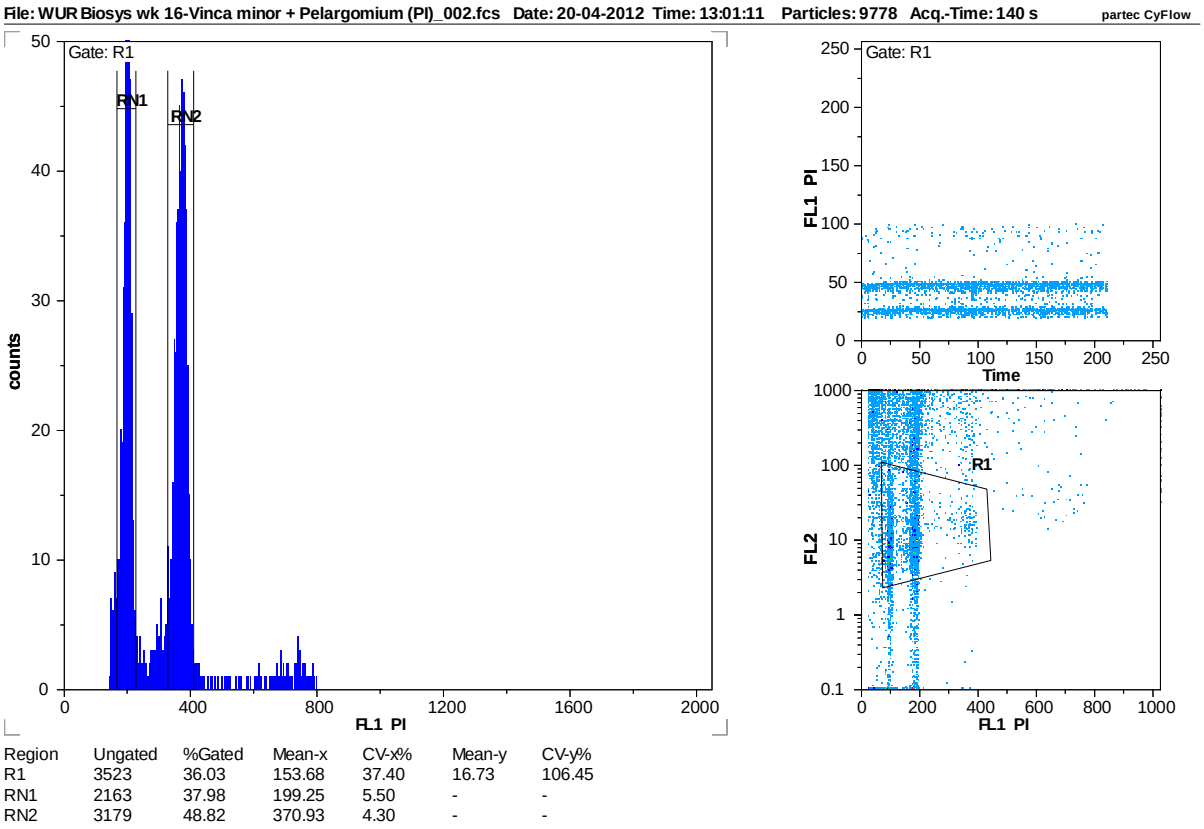
Flow cytometry is currently the most widely used method to estimate nuclear DNA content for either genome size measurements or ploidy level estimations. It is a fast and effective measuring technique, with most flow cytometer allowing 100 to 1,000 nuclear measurements per second. It requires only a small amount of plant material, usually a few square millimetres of leaf, and is therefore non-destructive.

Before flow cytometry became the most popular choice, Feulgen densitometry was used to measure nuclear DNA content. While Feulgen densitometry shares many of the advantages flow cytometry offers, it is a more time-consuming method and requires more extensive sample preparation.

The basic premise of flow cytometry is the detection of cellular particles and properties by scattering of light. Cells are suspended in a medium and pass individually through a laser beam. Forward and sideward scattering of light is measured by detectors in line with and perpendicular to the beam. In addition, fluorescence detectors measure the number of attached fluorescent particles, and the corresponding fluorescence peak can be used to estimate genome size (See for an example Figure 9).

Nuclei can be stained with DAPI or propidium iodide (PI), but in a paper by Dolezel et al. (1998) it was shown that genome size estimates obtained by DAPI staining do not agree well with Feulgen densitometry measurements, whereas PI staining did produce comparative results. This is due to the preferential binding of DAPI to AT-rich regions in the genome. In one sample, the difference between the DAPI estimate and the Feulgen densitometry estimate was as high as 58%.

Grotkopp et al. (2004) used flow cytometry to correlate genome size variation to life history traits in *Pinus,* using seed samples rather than leaf material. The use of seeds instead of leaves allows genome estimations of high quality in species with high concentrations of secondary metabolites in the leaves, which would otherwise inhibit flow cytometry measurements (e.g. Jedrzejczyk and Sliwinska 2010).



Region	Ungated	%Gated	Mean-x	CV-x%	Mean-y	CV-y%
R1	3523	36.03	153.68	37.40	16.73	106.45
RN1	2163	37.98	199.25	5.50	-	-
RN2	3179	48.82	370.93	4.30	-	-

Figure 9 – Example of a DNA histogram showing two peaks corresponding to nuclear DNA content of the control (RN1) and a P. bowkerii *sample (RN2). CV denotes the coefficient of the variation, a measure for the quality of the measurements that corresponds to the width of the peaks. If the width is very narrow it indicates the accuracy of the genome size estimate is very high, and the CV-value will be very low (0-5%).*

Since flow cytometry requires a control measurement against which other measurements can be compared, a plant of known ploidy level and genome size must be chosen. The plant database of the botanical gardens of Kew, England, only holds one record for a *Pelargonium* species' genome size: *P. radens,* with a C-value of 8.1 pg (Greilhuber 1988, in Kew Royal Botanical Gardens 2012). *P. radens* is an octaploid species, which made it less suited as a control candidate as diploid species are preferred.

Other studies have used a variety of different control plants, with the garden pea (*Pisum sativum*) being particularly popular. *P. sativum* is a diploid species with a C-value of approximately 9 pg. This is quite close to the value found for *P. radens*, but because *P. radens* is an octaploid species, C-values for diploid *Pelargonium* species are likely to measure roughly a quarter the size. Therefore a control plant with a genome size between 1 and 3 pg was the most desirable.

The actual flow cytometry procedure was carried out by Plant Cytometry Services, a company specialized in flow cytometry in Schijndel, the Netherlands. Plant Cytometry Services indicated that *Pelargonium* samples are difficult to accurately measure, possibly due to the build-up of secondary metabolites in the cells (Walker *et al.* 2006). They

recommended using young leaves for the procedure to minimize interference of these metabolites. The planned sample size was three individuals per species. However, due to limitations in species and accessions availability, for some species only one individual was analyzed while for others two to four individuals were sampled.

To study the intraspecific variation of genome size in species (Šmarda and Bureš 2006). Koopman *et al.* (2002, unpublished) studied inter- and intraspecific differences in DNA content in *Lactuca* (Asteraceae) and found that of the 23 species used in the experiment, six showed significant intraspecific variation for genome size. To determine the intraspecific variation of genome size in a species a sample size of at least ten accessions per species is recommended (Koopman *et al.* 2002, Jakob *et al.* 2004). For the scope of this thesis that is impractical, and so intraspecific variation was estimated for some species for which three or more accessions had been sampled.

Sampling

Pelargonium samples were collected in the botanical gardens of the Vrije Universiteit in Amsterdam and the Westfahlische Universitat in Munster in 2012. Collections in Amsterdam were made on the 17th of April, and collections in Munster were made on two separate occasions, once on the 8th of June and once on the 9th of August. In both the VU and Munster collections some sampled were contaminated with white fly (*Bemisia*), both adults and eggs. This may have resulted in some small peaks showing up in the flow cytometry histograms around the 1.04 and 2.06 pg values (the genome size of male and female whiteflies respectively (Brown *et al.* 2005)).

Plant material was collected from the botanical gardens of the Westfählische Wilhelms-Universität in Münster, Germany. The choice of species that were included in the analysis therefore depended on the availability of material of those species in the collection in Münster. Chase *et al.* (2005) laid down the following sampling guidelines for genome size studies across phylogenies:

- Species that represent morphological variation in a clade should be sampled adequately, for a proper representation of taxonomic diversity.
- Within the major clades there should be an even sampling of species, without focussing on basal or terminal clades.

Following these two guidelines would reduce sampling bias. Unfortunately due to lack of material for certain clades (notably *Hoarea*) sampling was more uneven than planned. The species that were sampled are shown in Appendix 3. I took care to sample young, fresh leaves to minimize the presence of interfering metabolic compounds. Leaf samples were stored in plastic zip bags with moisturised tissue paper added to prevent dehydration. Some samples from both the Munster and VU collections were contaminated with white fly (*Bemisia*).

Results

The flow cytometry measurements themselves were made at Plant Cytometry Services in Schijndel following standard protocol (see Appendix 4). Due to financial constraints (PI staining is three times as expensive as DAPI staining), in this thesis I opted to estimate the majority of the samples with DAPI staining.

DAPI binds preferentially to AT-rich regions and therefore estimates of genome size based on this method assume an equal proportion of AT to CG base pairs in the genome, which is rarely the case. Propidium Iodide is a more expensive measurement method, but does stain both AT and CG base pairs, therefore yielding more accurate estimations of genome size. To improve the accuracy of the DAPI measurements the proportion of AT to CG base pairs was calculated by measuring both DAPI and PI stained material from 12 species (see Figure 10). From the difference between these paired measurements the ratio could be calculated and each DAPI measurement was corrected using this value.

Figure 10 – Estimates of genome sizes for 12 Pelargonium *species using both DAPI and PI staining. An average conversion factor based on the difference between each pair of measurements is used to correct the DAPI measurements for differences in AT-CG base pair proportions. Two accessions were sampled for* P. triste*, and for* P. fulgidum *a leaf sample and a petal sample (19fulgAA) were sampled. The petal sample was taken to estimate the interference of secondary metabolites in the leaf cells.*

Accession	**Species**	**DAPI (pg)**	**PI (pg)**	**Factor (DAPI/PI)**
9 cauc 02	*P.caucalifolium*	*2.3103*	*2.91*	*0.793918*
4 exst 02	*P.exstipulatum*	*1.2231*	*1.50*	*0.8154*
8 fulg 02	*P.fulgidum*	*1.1778*	*1.50*	*0.7852*
18 grav 02	*P.graveolens*	*3.5334*	*4.18*	*0.845311*
10 inqu 02	*P.inquinans*	*1.7365*	*2.37*	*0.7327*
1 tris 02	*P.triste*	*3.4126*	*4.18*	*0.816411*
2 tris 03	*P.triste*	*2.5368*	*3.05*	*0.831738*
7 acet 01	*P.acetosum*	*1.8875*	*2.36*	*0.744658*
20 carn 01	*P.carnosum*	*1.1778*	*1.47*	*0.885172*
11 cote 01	*P.cotyledonis*	*0.7852*	*0.98*	*0.855667*
19 fulg AA	*P.fulgidum*	*1.2231*	*1.53*	*0.767177*
13 gibb 01	*P.gibbosum*	*1.0721*	*1.34*	*0.831738*
5 ioni 02	*P.ionidiflorum*	*1.1929*	*1.49*	*0.772835*
14 karo 01	*P.karooicum*	*3.3522*	*4.20*	*0.702326*

For *P. fulgidum* I also sampled the petals of one accession (see Figure 11). The leaves of *Pelargonium* species build up a store of secondary metabolites which are known to interfere with flow cytometry readings (e.g. Doležel and Bartoš, 2005). Since the petals do not contain these compounds, measuring them alongside a regular leaf sample should how the impact of any secondary metabolites on the genome size estimates. Apart from this control measurement I always ensured to sample young, fresh leaves which should contain very few metabolic compounds compared to older leaves.

Vinca minor was used as a control species because its genome size (2C-value 1.51 pg) was expected to be close to that of most *Pelargonium* species, an assumption based on the genome size of the octaploid *P. radens* (2C-value 8.1pg).

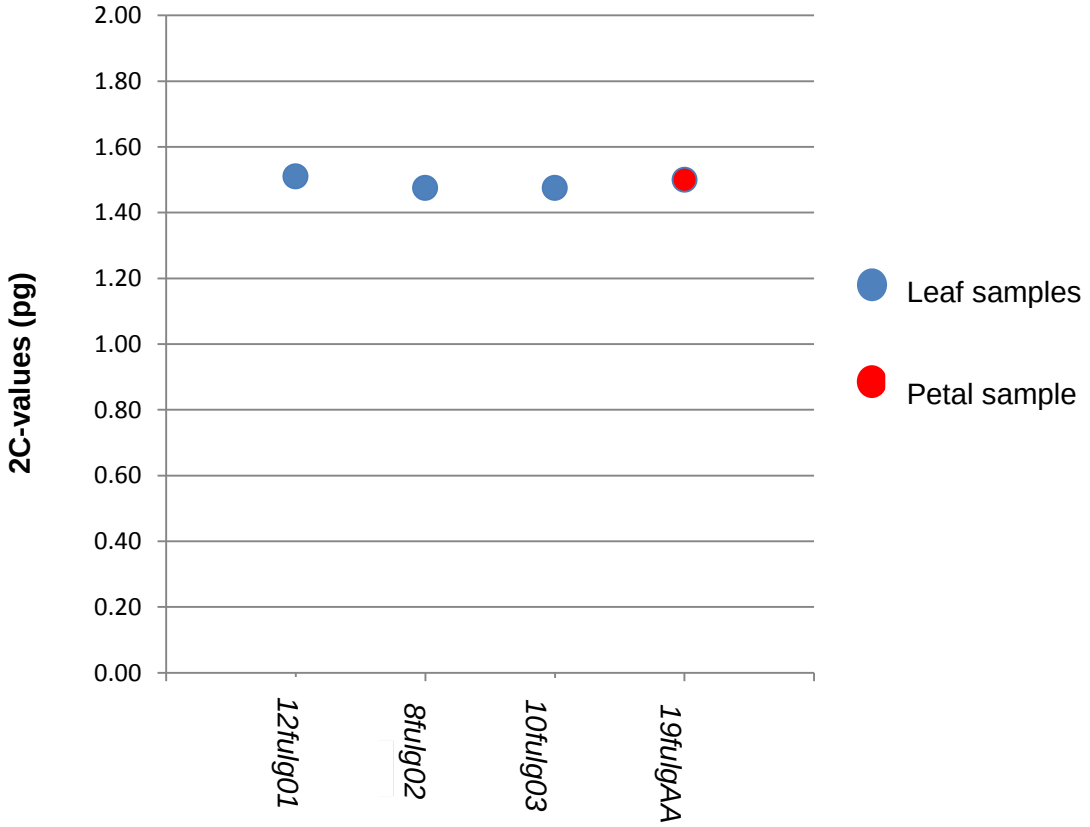


Figure 11 – Comparison of genome size measurements based on leaf and petal samples. All leaf samples were taken from different individuals of P. fulgidum, *while the petal sample was taken from the same accession as sample 10fulg03. The genome size estimate from the petal sample does not differ significantly from any of the leaf samples. This indicates that interference of secondary metabolites may be minimal in young, fresh leaves of* Pelargonium. *However, more petal measurements should be conducted to firmly support this possibility.*

Weng *et al.* data integration

During the experimental phase of my thesis a paper was published by Weng et al., which featured genome size estimates for 28 species of *Pelargonium.* They based their estimates for each species on two separate flow cytometry measurements, using two different controls: *Arabidopsis thaliana* (2C-value: 0.04pg) and trout erythrocytes (2C-value: 1.51 pg). They took the average for each of these paired measurements as their genome size estimate. However, upon inspection of their raw data I noticed the measurements made using *A. thaliana* as a control yielded 2C-value estimates roughly half as high as those made with trout erythrocytes as the control. For my own measurements I sampled ten species that were also included in the Weng et al. paper: *P.cotyledonis, P. cucullatum*, *P. exstipulatum, P. fulgidum, P. dichondrifolium, P. alchemilloides, P. tetragonum, P. grossularioides, P. tongaense, P. transvaalense, P. reniforme.*

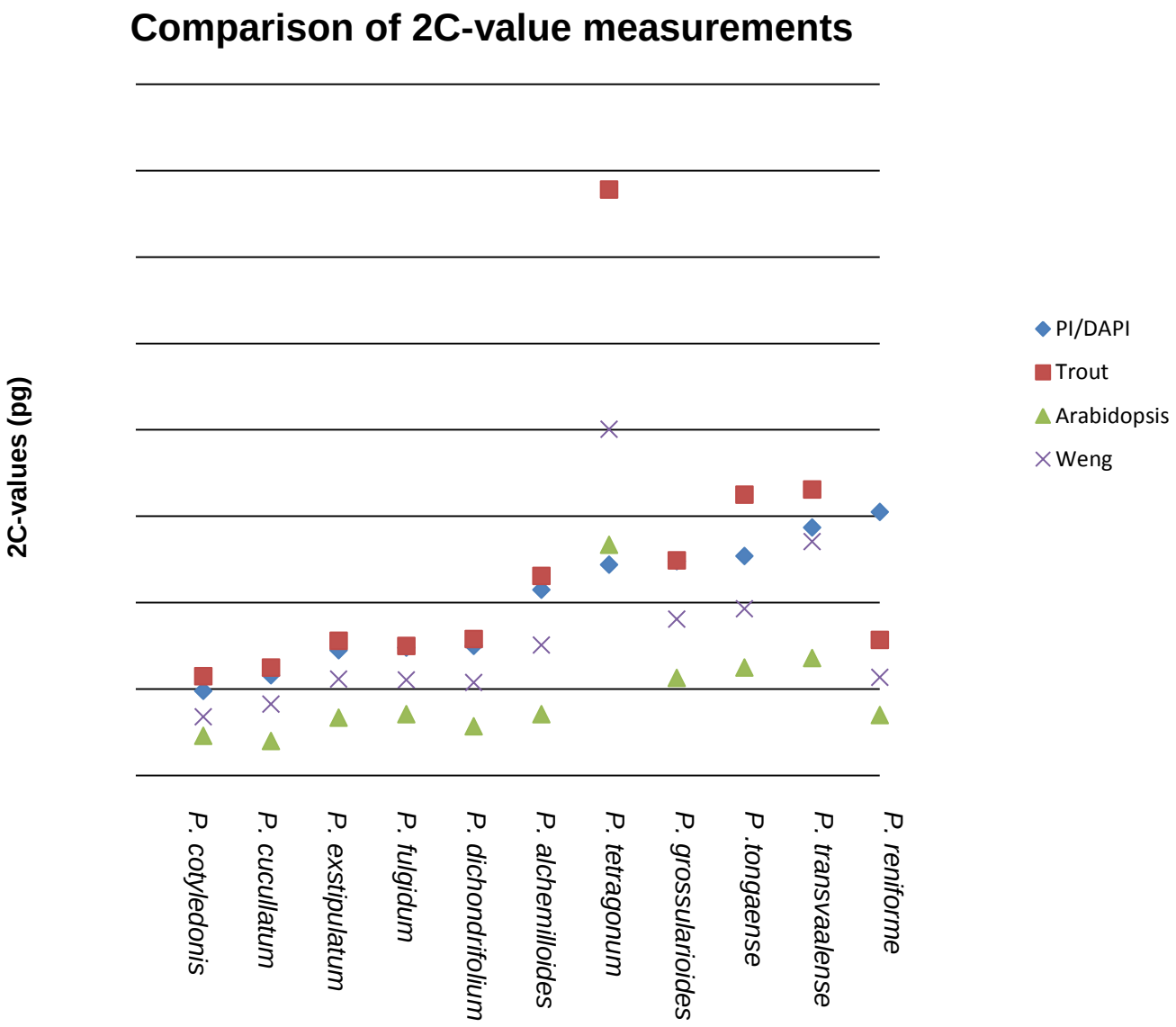


Figure 12 – Comparison of genome size measurements both this and the Weng study. The individual trout and Arabidopsis *controlled measurements are shown alongside the combined average values (as published in Weng et al. (2012)) and my own estimates based on either PI or DAPI stained samples. Apart from* P. tetragonum *and* P. reniforme*, all species show a low genome size estimate for the* Arabidopsis *controlled measurements compared to the trout and* Vinca minor *controlled measurements. While the underlying cause for this is as of yet unclear, I conclude from these results that the trout measurements are the most accurate estimations for genome size from the Weng* et al. *paper, compared to the averaged or individual* Arabidopsis *controlled measurements.*

I found that eight of the ten species yielded genome size estimates similar to the trout controlled measurements from Weng *et al.* (see Figure 12). The *Arabidopsis* controlled measurements for these species were consistently lower (about half the size in fact) of both my own and the trout controlled estimates. I consider this an indication that the *Arabidopsis*

controlled estimates are inaccurate, and that taking the average of these and the trout controlled measurements as a genome size estimate results in an underestimation of genome sizes in *Pelargonium*. I therefore took the trout controlled measurements as accurate estimations of genome size for the *Pelargonium* species analysed in the Weng *et al.* study. The two species for which the trout controlled measurements and my own measurement were significantly different (*P. tetragonum* and *P.* reniforme) may have been accessions of a different ploidy level.

Results and discussion

The individual flow cytometry measurements are shown in a table in Appendix 3. The genome size of *Pelargonium* is quite small with the smallest measured 1C-value being 0.475 pg for *P. appendiculatum* and the highest at 3.48 pg for *P. schizopetalum* (although Weng *et al.* (2012) reported 3.59 pg for *P. luridum*). Small genome sizes are common in the angiosperms (see Figure 13). The average 1C-value was around 1.20 pg (standard deviation 0.585 pg), which is more than a hundredfold smaller than the largest known angiosperm C-value belonging to *Paris japonica* at 152.23 pg. Note that this average value is calculated from all measured samples, and is not corrected for phylogenetic relationships. For the phylogenetic mean, see the continuous optimizations in Chapter 8.2.

1C-value distribution in angiosperm species

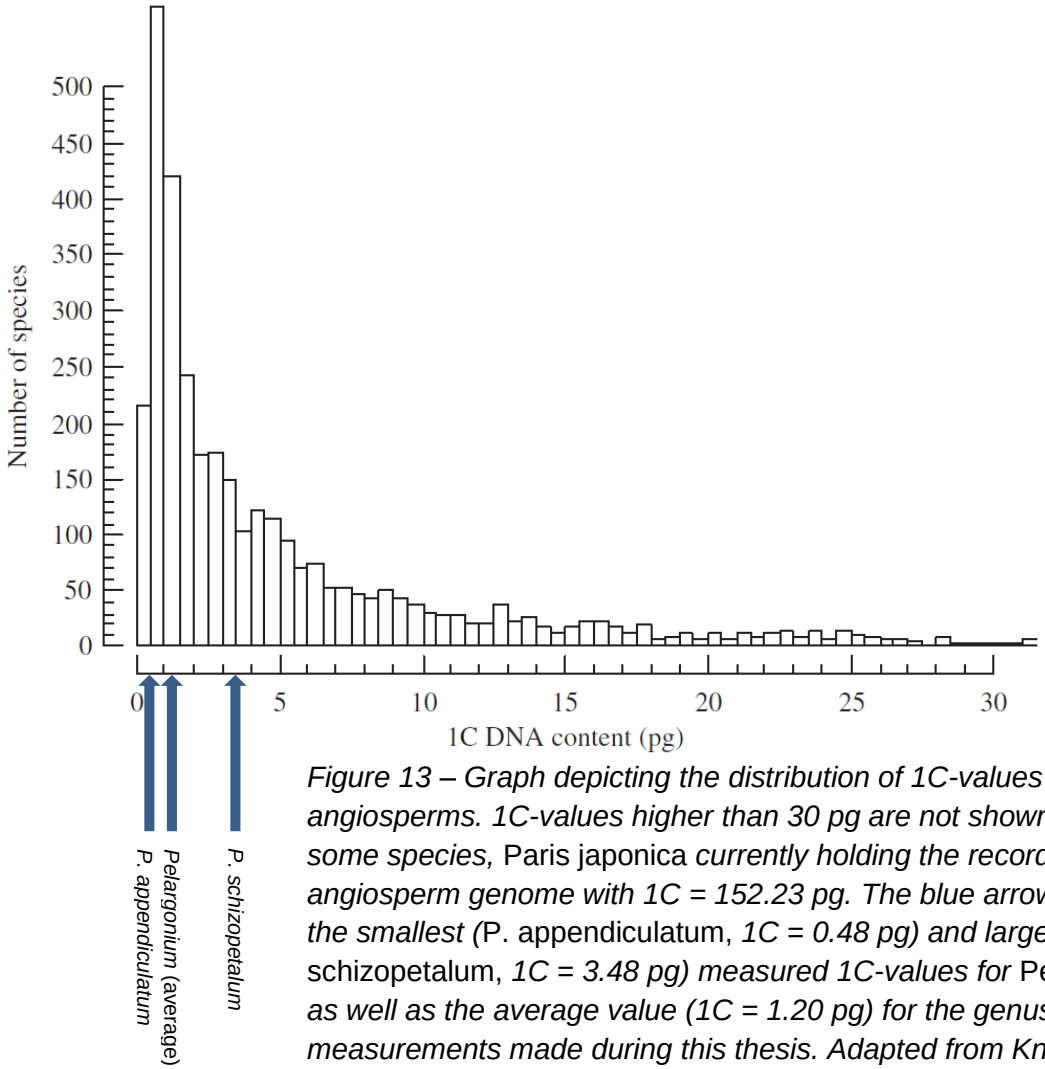


Figure 13 – Graph depicting the distribution of 1C-values across the angiosperms. 1C-values higher than 30 pg are not shown but do occur in some species, Paris japonica *currently holding the record for largest angiosperm genome with 1C = 152.23 pg. The blue arrows show where the smallest (*P. appendiculatum, *1C = 0.48 pg) and largest (*P. schizopetalum, *1C = 3.48 pg) measured 1C-values for* Pelargonium *are, as well as the average value (1C = 1.20 pg) for the genus based on the measurements made during this thesis. Adapted from Knight* et al. (2005).

When we arrange the measured 2C-values by clade, we can compare the range and relative distribution of the genome size estimates (see Figure 14). It seems clades A2 and B have 2C-values ranging between 1 and 3 pg, with most 2C-values within clade A1 also occupying this range, although A2 has some outlying species between 4-5 and 7-8 pg. Clades C1 and C2 occupy a slightly higher range, between 2-3 pg, with some outlying species as well.

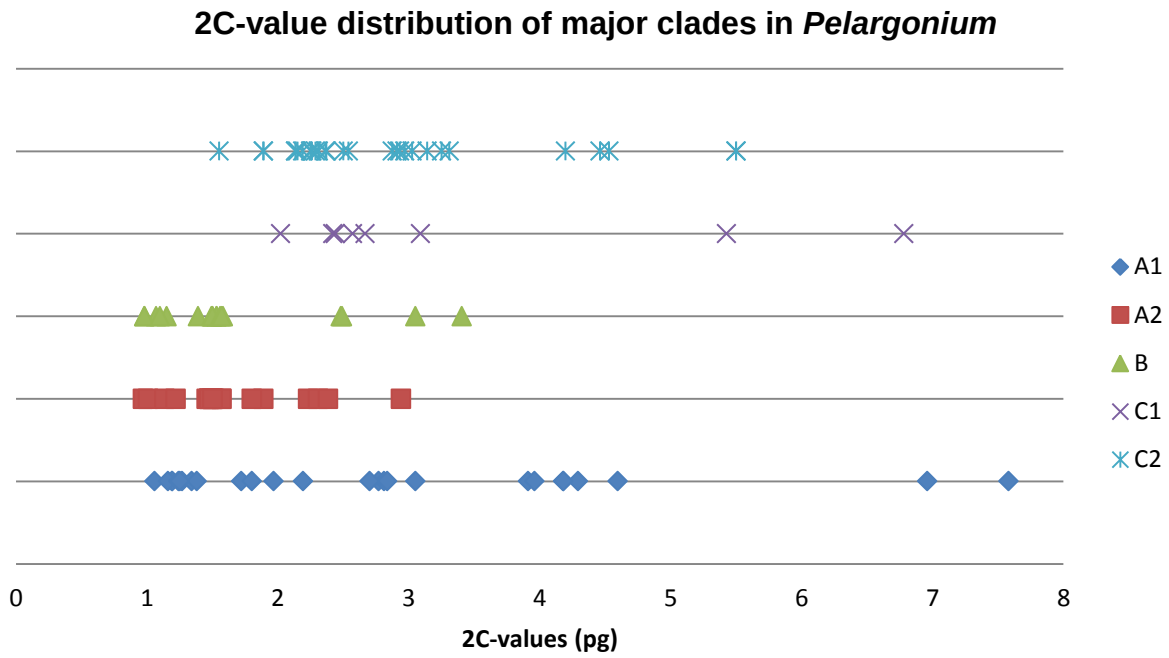


Figure 14 – Scatter plot showing the 2C-values for various Pelargonium *species organised by clade (A1, A2, B, C1, and C2). The C1 and C2 subclades show a lower limit of about 1.5 pg, with most measurements falling between 2 -3 pg, and a few outliers at higher values. Clade B and subclades A1 and A2 have a lower limit of 1 pg, with most measurements ranging between 1 and 3 pg, with a few outliers in A1.*

The Cx-values of some polyploid species can be calculated using chromosome data from literature, while for diploid species the Cx-value is the same as 1C-value. A list of species for which Cx-values could be determined for this thesis can be found in Appendix 3. The smallest estimated 1Cx-value belongs to *P. reniforme* at 0.29 pg, about 1.5 times as small as the smallest 1C-value of *P. appendiculatum,* 1C = 0.475 pg. The largest estimated 1Cx-value is 2.75 pg and belongs to *P. endlicherianum*, a diploid species with a high basic chromosome number (seventeen). The average Cx-value is 0.88 pg, with a standard deviation of 0.38 pg.

Although the genome size of *Pelargonium* species is quite small, there is quite some variation in both C- and Cx-values within the genus, as shown by the high standard deviations for both. This variation will be further explored in Chapter 8.

Chapter 7 - Stomata measurements

A correlation between guard cell length and plant C-values has been established in a wide variety of angiosperms (Beaulieu *et al.* 2008). This relationship is thought to be based on the enlargement of cells with increasing genome size. In addition, stomata density commonly increases with lower C-values. However, the significance of these correlations has been shown to be growth form dependent, with trees and shrubs for example not showing a positive correlation between guard cell length and C-values at all.

To test for a possible correlation between C-values and guard cell length, stomata images were obtained for 37 accessions of *Pelargonium* species in the botanical gardens of the Westfahlische Universitat Munster. Leaf samples were stored in a moist environment before being transferred to a buffer solution consisting of 10mM MES-KOH, 10mM KCl, 50mM $CaCl_2$, and 10µM ABA, for approximately two hours in order to close all stomata. Samples were placed in the dark to facilitate the closure of stomata.

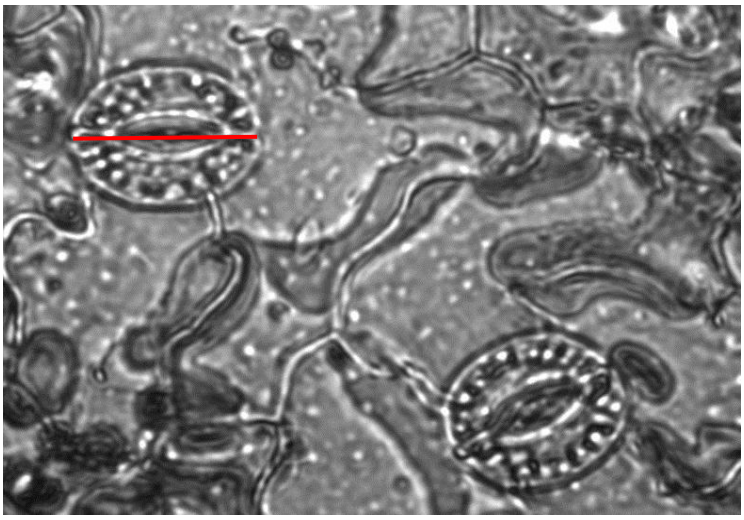


Figure 15 – Microscopic image of stomata of Pelargonium caucalifolium*. Size of the guard cells was measured using the program Openlab. The red line indicates the length measured to estimate guard cell lengths.*

In order to measure the length of the guard cells the images were taken at a fixed magnification of 400x and were loaded in the program Openlab (see Figure 15 for an example). A line was drawn through the centre of the stomata and the length of the line in micrometres was taken as the length of the guard cells. Depending on the quality of the images, 5 to 20 measurements were made per slide. To estimate intraspecific variety four individuals of *P. grandiflorum* were compared.

Results and discussion

The average guard cell length for each species sampled is shown in Figure 16.

Figure 16 – Table summarizing average guard cell lengths per species sampled. Accessions are also shown; additional letters in brackets indicate different individuals from the same accession.

Species	Accession	Mean Guard Cell Length (µm)
grandiflorum	STEU1018	30.15158
grandiflorum	STEU1018(a)	24.54961
grandiflorum	STEU1763	21.64357
grandiflorum	STEU1763(a)	23.51574
grandiflorum	STEU808	24.53122
grandiflorum	STEU808(a)	25.28617
grandiflorum	STEU808(b)	25.24333
desertorum	STEU2857	27.9628
cucullatum	STEU1120	34.30674
scabrum	STEU1000	31.17319
gibbosum	STEU721	23.55825
ceratophyllum	Ac4167	31.75713
fulgidum	STEU482	26.98836
dichondrifolium	872371933	26.31741
ionidiflorum	STEU1874	24.78853
karooicum	2967	29.12623
stipulaceum	757	26.03864
multiradiatum	STEU3191	35.5686
suburbanum	STEU785	27.98853
hispidum	STEU1006	28.3985
otaviense	943	29.60921
alchemilloides	STEU2751	31.26924
quercifolium	STEU1275	31.85817
mutans	4120	30.78772
acraeum	1975	35.15794
xerophyton	STEU910	34.45206
xerophyton	STEU1324	35.09036
tetragonum	Ac2472	30.86908
grossularioides	STEU429	28.4465
tongaense	10710	29.00177
antidysentericum	STEU2972	28.02208
pulverulentum	1236	36.23565
transvaalense	STEU779	27.23663
caucalifolium	STEU1682	30.60466
caylae	STEU2198	33.21843
reniforme	STEU1698	27.63233
sidoides	88331057650	27.22605
vitifolium	STEU731	45.32309
praemorsum	STEU1575	34.16375
crassicaule	412	25.9332
mollicomum	STEU4251	28.68453

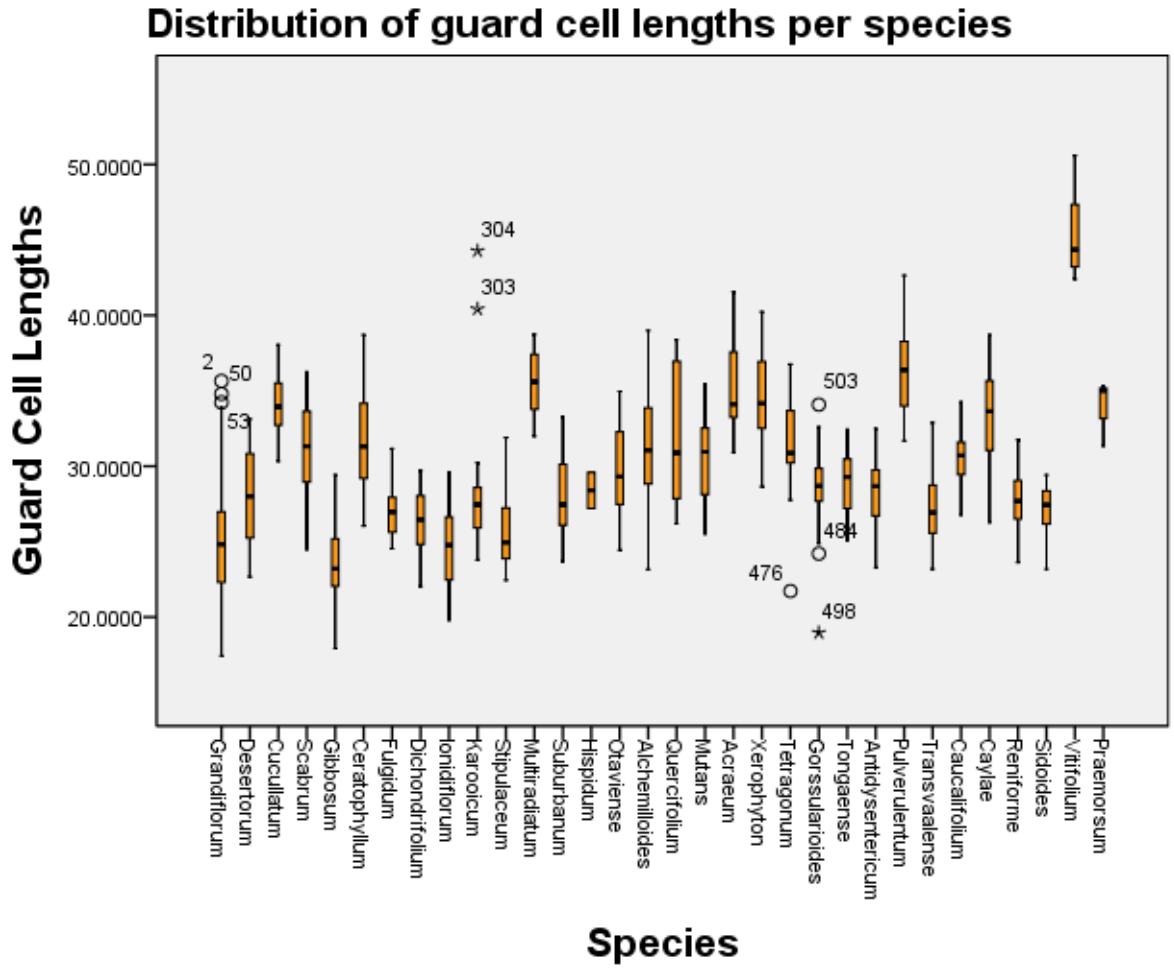


Figure 17 – Boxplots showing the distribution of Pelargonium *guard cell lengths per species used in this study. Some species show a very high variation in guard cell length compared to others (e.g.* P. grandiflorum *and* P. alchemilloides*). However, these differences in intraspecific variation are at least partly due to differences in sample size for each species (see Appendix 8.2).*

The variation in guard cell length within species was generally quite considerable (see Figure 17). In most cases the variation was normally distributed (Shapiro-wilk), although some outlying values were present. Species that did not show a normal distribution for guard cell length were *P. grandiflorum* (p=0.015)*, P. karooicum* (p=0.001) and *P. praemorsum* (p=0.012). These significant variations in guard cell length within species may be explained by cases of endopolyploidy or aneuploidy. A Kruskal-Wallis test showed differences in guard cell length among species are significant (p=0.000).

Conclusions

Pelargonium seems to have a significant interspecific variation in guard cell length based on these data. In subsequent analyses these differences can be correlated with C-values. For some species, intraspecific variation in guard cell length was also present. These may be examples of endopolyploidy or aneuploidy.

Chapter 8 - Evolutionary inferences

In order to reconstruct the evolutionary history of genome size in *Pelargonium* several analyses were conducted. To see whether the variation in genome size showed a correlation to the phylogeny of *Pelargonium* I performed Pagel's transformations, testing the various models to see which one fitted the data best. To map the evolutionary history of genome size in *Pelargonium*, I optimized the data over the phylogeny and reconstructed the ancestral values of the nodes. Because genome size varies on a continuous scale, I used R's APE package to optimize the data without discretizing it using the function ACE. R was chosen for its simple interface and customizable functions, compared to other optimization programs such as TNT. I also optimized the data in a discretized fashion in Mesquite to compare the two methods.

8.1 - Pagel's transformations

To test whether the genome size variation in *Pelargonium* has a significant phylogenetic signal, Pagel's transformations were performed on both the MultiTree and SingleTree in R. The following models were tested by fitting them to the tree using the function fitContinuous: Brownian motion, Lambda, Ornstein Uhlenbeck, Delta, Kappa, and White. The Brownian motion model assumes no phylogenetic signal for trait evolution, assuming any variance is due to random divergence. The Ornstein-Uhlenbeck model is similar to brownian motion, but with a centralizing tendency. The Lambda model multiplies the branch lengths by λ and sees for which value the model fits the data best. The value of λ can only be between 0 and 1, where a value around 0 indicated the phylogenetic signal is very weak (the tree can essentially be seen as a polytomy for the purposes of trait variance), whereas a λ approaching 1 shows a very strong phylogenetic signal. The Delta model tests whether the evolution of a trait increases or decreases over time, while the Kappa model tests for punctuated evolution.

For the majority of trees in the MultiTree set, as well as the SingleTree, the Lambda model fitted the Cx-values best (Figure 18). The average value for λ for the MultiTree set was 0.81, which corresponds to a strong phylogenetic signal. The λ value for the SingleTree was lower at 0.5, which is due to the polytomies that are present in the consensus tree.

Figure 18 – Tables showing best model fits for the SingleTree and MultiTree after performing Pagel's transformations in R. Fit estimations are based on the Akaike Information Criterion (AIC), with low values representing a better fit.

Trait	Tree Set	BM	λ	δ	κ	OU	EB	White
Cx-values	MultiTree	29.8334	0.766695	20.98688	7.285	4.619935	31.833	12.8815
	SingleTree	38	0	22	4.3	12	40	9.8
C-values	MultiTree	25.42	2.6945	14.88715	8.0985	0.72597	27.42	0.72187
	SingleTree	37	0	21	1.4	2.2	39	0.15

The C-values yielded similar results to the Cx-values for the SingleTree analysis; the Lambda model fitted the majority of topologies the best as well. However, the White Noise model also had a low AIC value at 0.15, considerably lower than the 9.8 for the Cx-values. The MultiTree analysis showed an equal fit for the Ornstein-Uhlenbeck and White Noise model, each averaging an AIC value of 0.7. The Lambda model had a mean AIC value of 2.7, which is high compared to the SingleTree and Cx-values Lambda results.

An explanation for this might be that taxa with abnormal C-values (for example, species with a high ploidy level), are grouped in clades with a low posterior probablity, or even in polytomies in the consensus tree. Within the various topologies of the MultiTree dataset the phylogenetic signal of the C-values is possibly lost as these taxa are rearranged quite frequently, and the pattern of C-value variation in the genus might resemble a more random process (Ornstein-Uhlenbeck) or appear to follow a normal distribution (White Noise).

The values for the parameter lambda are shown in Figure X, with the exception of the SingleTree C-value parameter as the lambda model was not the best fitting model there. A lambda value represents the strength of the phylogenetic component in the evolution of a character. A lambda value of 1 indicates a strong phylogenetic signal, while a value of 0 indicates no phylogenetic component is present at all. The average value of lambda for both the C- and Cx-values of the MultiTree set are 0.81, which indicates a strong phylogenetic signal. The lambda value for the Cx-value SingleTree is intermediate at 0.5, which is lower than that of the MultiTree set but that is expected as the polytomies in the consensus tree dampen the phylogenetic signal.

Figure 19 – Lambda values for the MultiTree sets and Cx-value SingleTree set. Standard deviations for the MultiTree sets are also shown.

Trait	**Tree Set**	**Mean λ**	**Std. Dev.λ**
Cx-values	MultiTree	0.81	0.01
	SingleTree	0.5	N.A.
C-values	MultiTree	0.81	0.09

8.2 - Character Optimizations

Technique

Optimizing characters over phylogenies can be done using a continuous or a discrete approach. Continuous optimization is preferable for genome size as discretization forces certain similar values apart while joining more dissimilar ones depending on the threshold values chosen for each discrete compartment. Fortunately, it is now possible to optimize continuous characters without discretizing them. MacClade (Maddison and Maddison 1992) and Mesquite (Maddison and Maddison 2005) already offered the possibility of optimizing continuous characters, but require single value input, rather than value ranges. In addition, they are unable to compute optimal trees. Goloboff *et al.* (2006) applied a continuous character optimization method in TNT and found the addition of continuous characters to phylogenetic studies highly informative.

To optimize genome size as a continuous character over the *Pelargonium* phylogeny however, I chose for the Ape package in the programming software R. The function 'ace' can be applied to both the MultiTree and a fully resolved version of the SingleTree set to obtain ancestral state estimates. R was chosen over other methods because of the freedom it offers in editing and combining functions and its graphic qualities.

A discrete approach was also taken to compare results from both methods. The program Mesquite was used to perform the discrete optimization using parsimony.

Discrete

I performed a discrete parsimony optimization analysis in Mesquite (See Figure 20) in order to reconstruct the ancestral states for both Cx and C-values for *Pelargonium.* Blue and cyan branches denote relatively small genome sizes whereas green, yellow and black indicate respectively medium sized and larger ones. Grey branches indicate that the ancestral state for that node is equivocal.

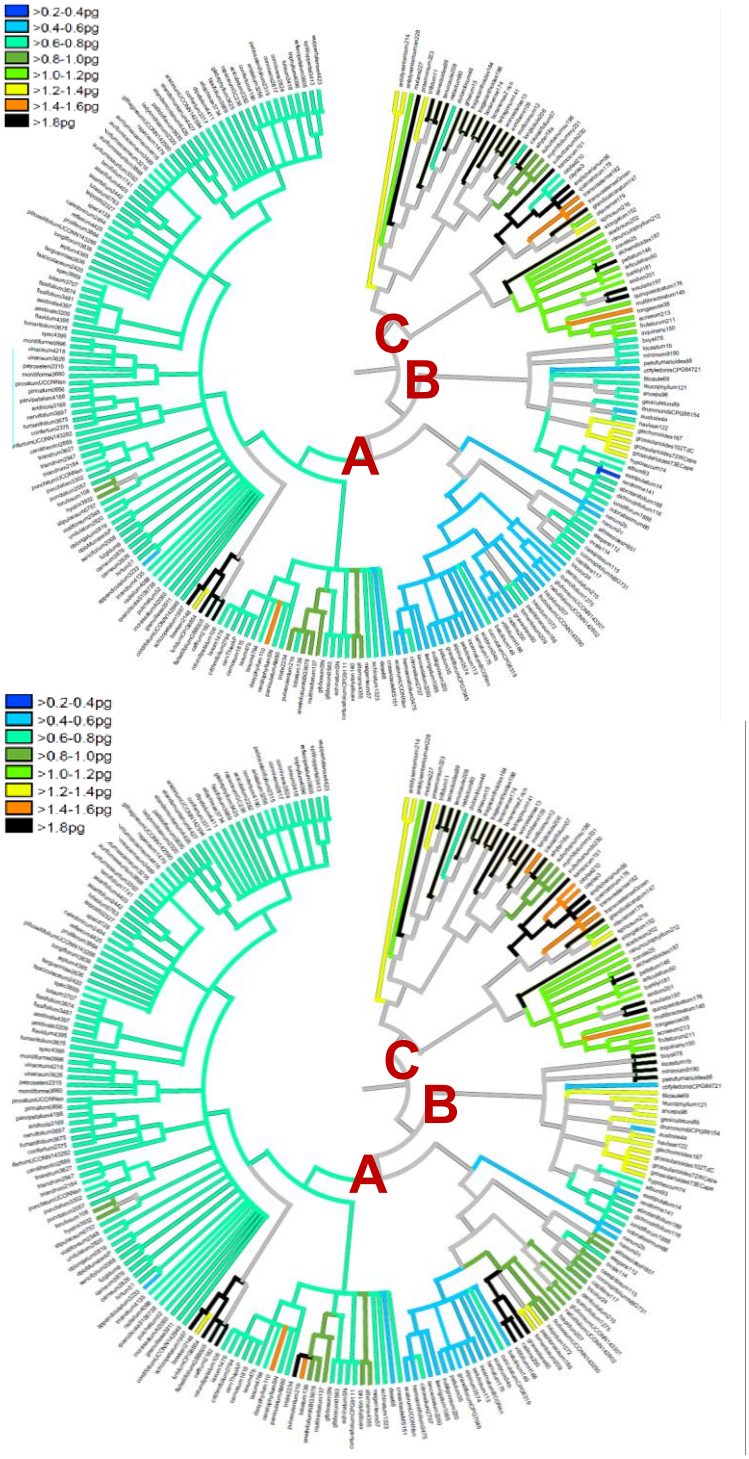


Figure 20 – Discrete parsimony optimization of Cx-values (a) and C-values (b) in Pelargonium. *The tree shown is the SingleTree. Grey branches denote equivocal ancestral states. The ancestral node of clade C for Cx-values is equivocal with a range 0.4->1.8 pg whereas those for clades A and B are equivocal with a range of 0.4-0.8 pg. For C-values clades C and B have an ancestral state with a range of 0.4->1.8 pg, while clade A maintains its smaller range of 0.4-0.8 pg. It is important to note that the C-values used for tree a are 1C-values, which means the the C-values for diploids are identical to their respective Cx-values.*

For the MultiTree set I used the function ‘Summarize state changes over trees’ in Mesquite to get an overview of the switches in C- and Cx-values across the different topologies. From the summary (shown in Figure 21), it seems both C- and Cx-values in *Pelargonium* have more frequently increased than decreased in *Pelargonium*.

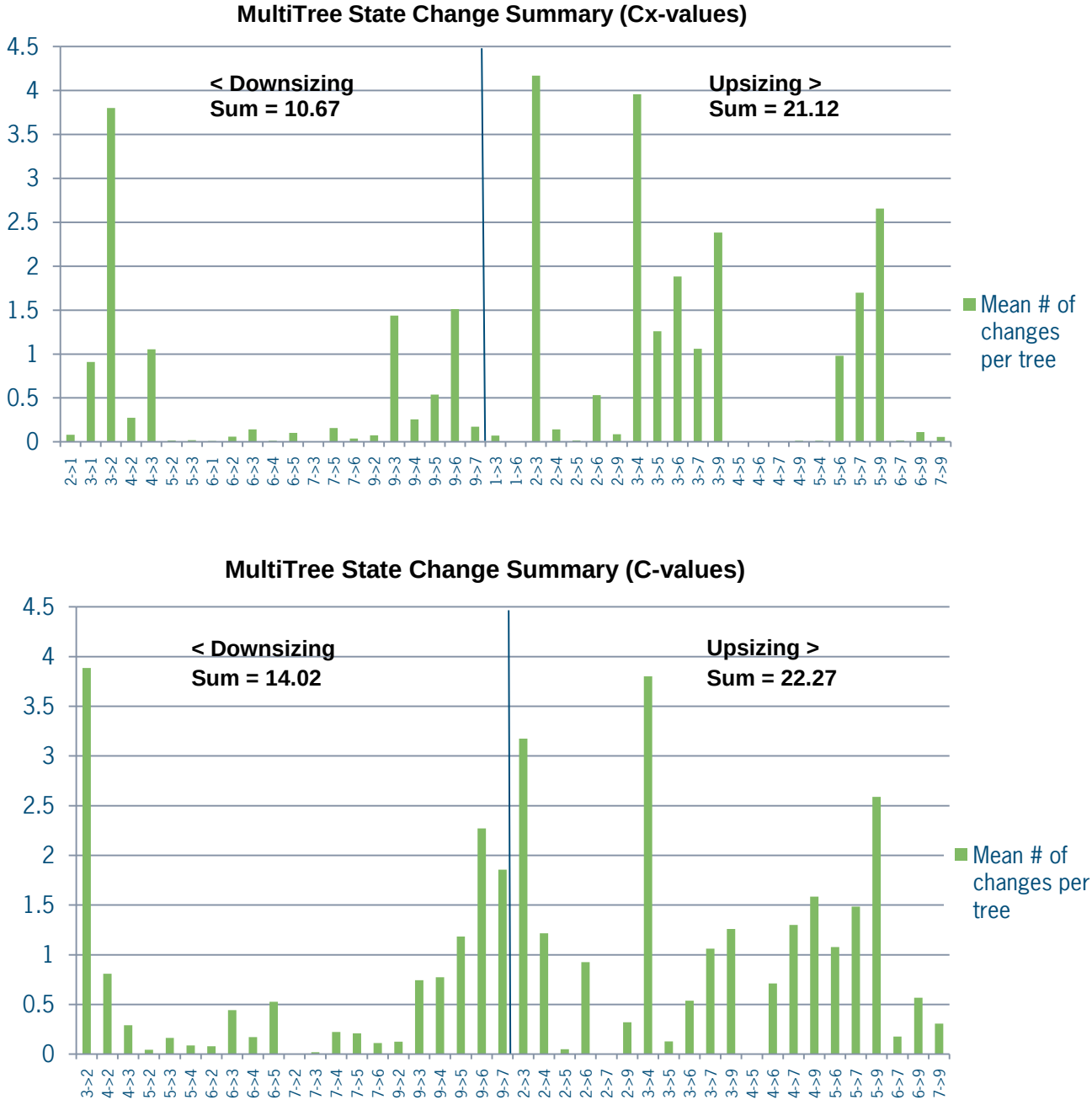


Figure 21 – State change summaries of the MultiTree discrete optimizations of Cx- (top) and C-values (bottom). State changes that did not occur in any tree are not shown. The mean frequencies of occurence of a given state change per tree are given on the y-axis. State changes resulting in a genome size decrease are shown on the left, genome size increases are shown on the right, divided by a blue line. By summing up the mean frequencies of genome size increase and decrease, it shows genome upsizing occurs more frequently across the genus’ evolutionary history than genome downsizing.

To see how this compared to the evolution of genome size in the main clades, I summarized the state changes for the clades A, B and C in separate graphs (Figure 22). These show that the instances of Cx-value increase or genome ‘upsizing’ mainly occur in clades A and C, whereas clade B actually shows more frequent instances of genome downsizing.

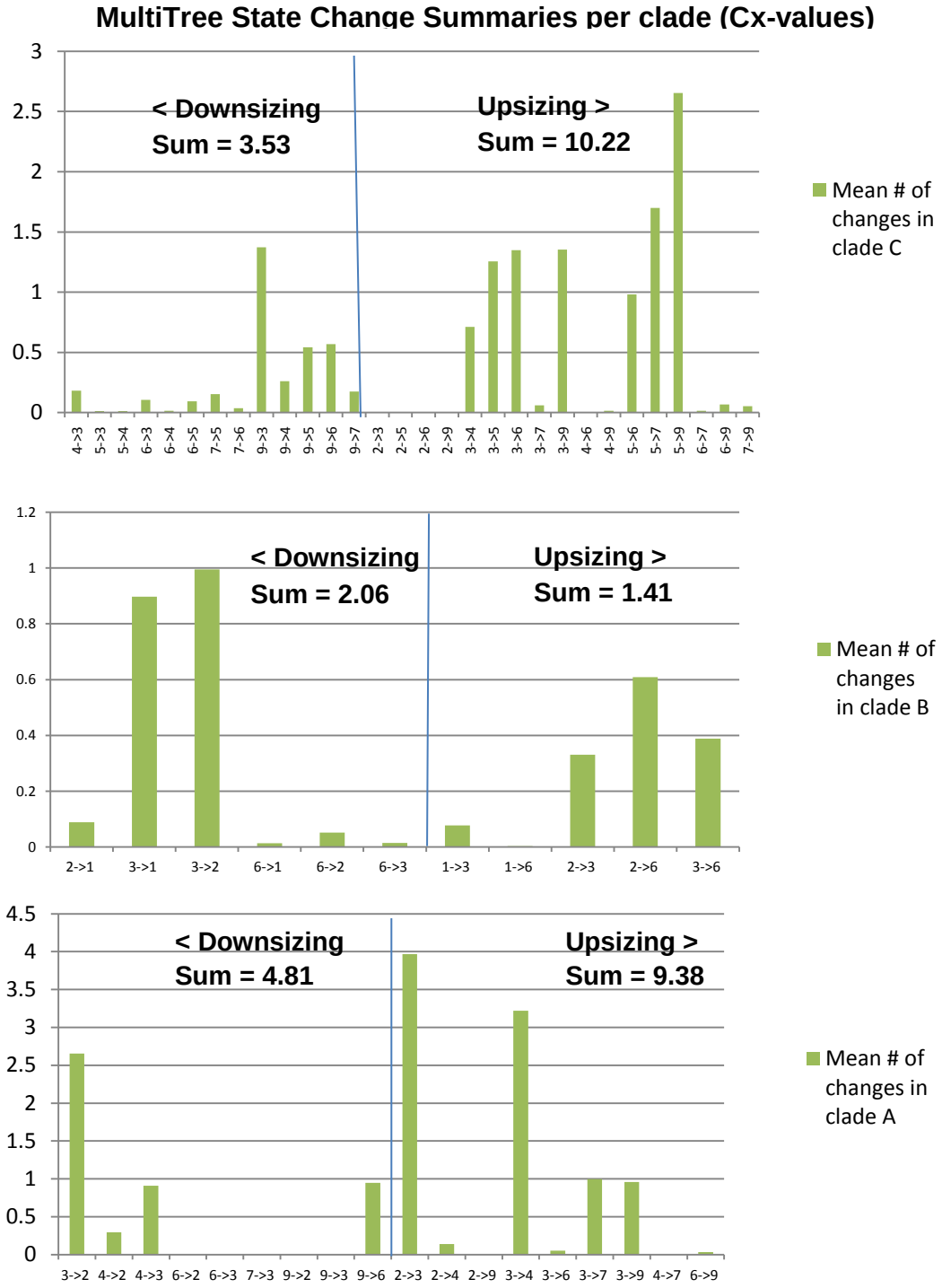


Figure 22 – MultiTree state change summaries for the major clades in Pelargonium *(C, B and A), showing mean frequency of state changes in Cx-values per tree. Summing up the mean frequency of Cx-value change shows clades C and A show more frequent instances of Cx-value increase than decrease, whereas clade B shows more frequent instances of Cx-value decrease.*

Separating the state change summary for the C-values (Figure 23) shows that clade A also has seen more frequent events of C-value increase than decrease. Clade B also features twice as many occurences of C-value increase than instances of reduction.

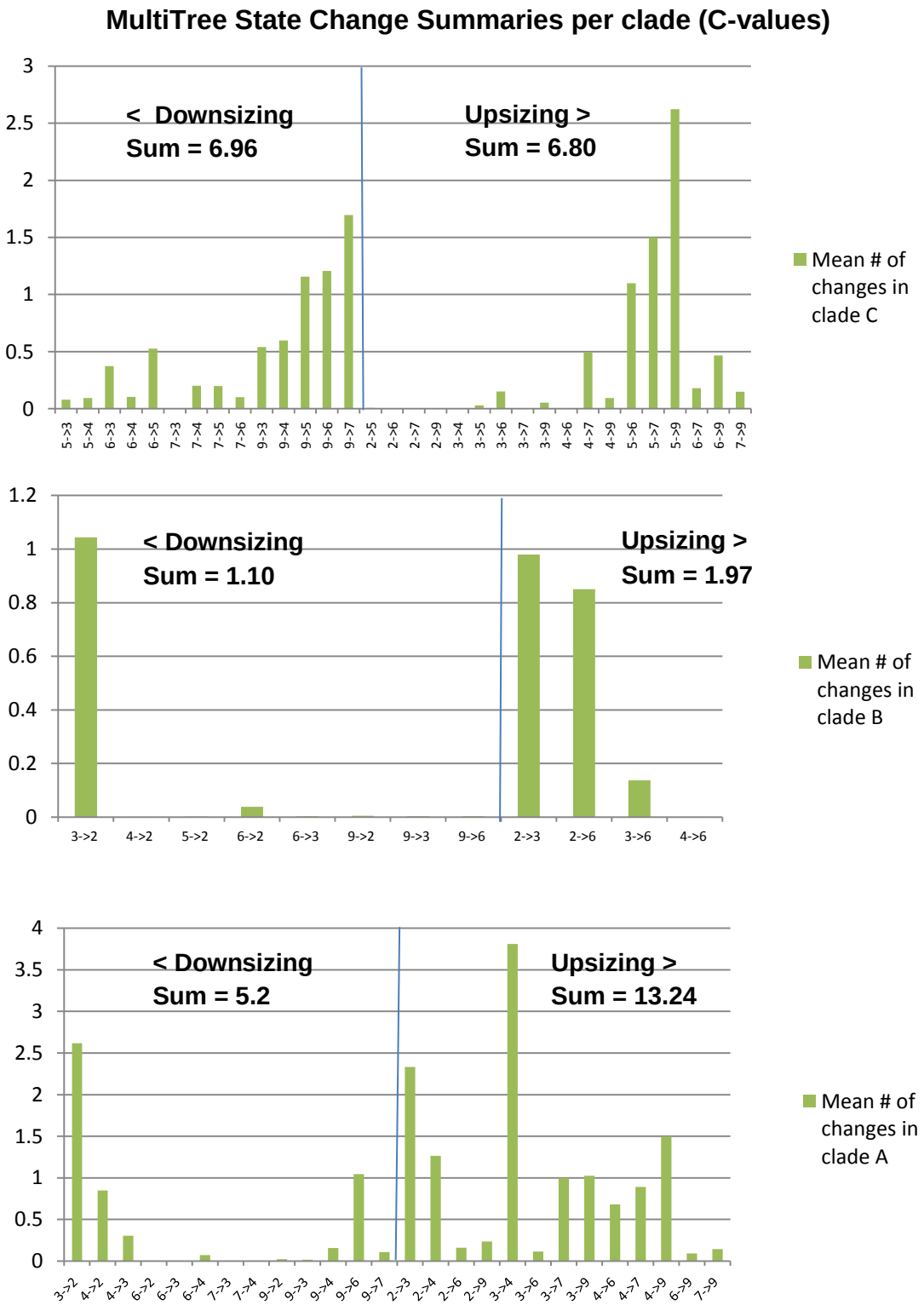
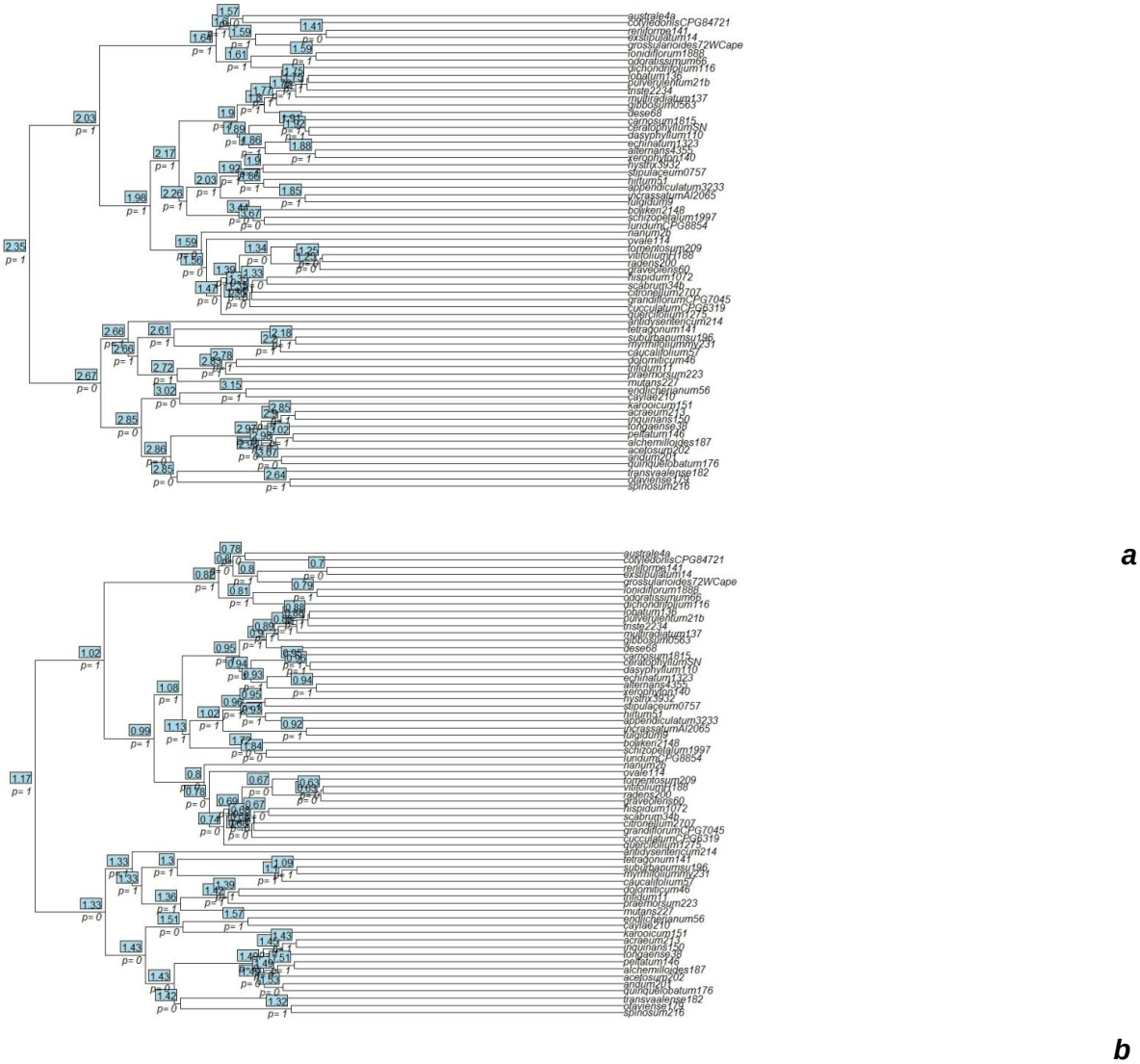


Figure 23 – MultiTree state change summaries for the major clades in Pelargonium *(C, B and A), showing mean frequency of state changes in C-values per tree. Summing up the mean frequency of C-value change shows clades B and A show more frequent instances of C-value increase than decrease.*

Continuous

I used the package ‘Ape’ in R to estimate ancestral states using continuous character optimization with the function ‘ace’. Ace computes ancestral states for continuous characters by fitting a Brownian motion model and using either maximum likelihood or least squares methods. I applied a restricted maximum likelihood (REML) method, which first estimates the ancestral state for the basal node, after which it calculates the variance for the Brownian motion using the residual log-likelihood. The REML method is preferred over the standard ML method because the latter has a downward bias towards the variance of the Brownian motion model whereas REML is an unbiased estimate. Figure 24 shows the SingleTree with ancestral states as estimated by ace.



a

b

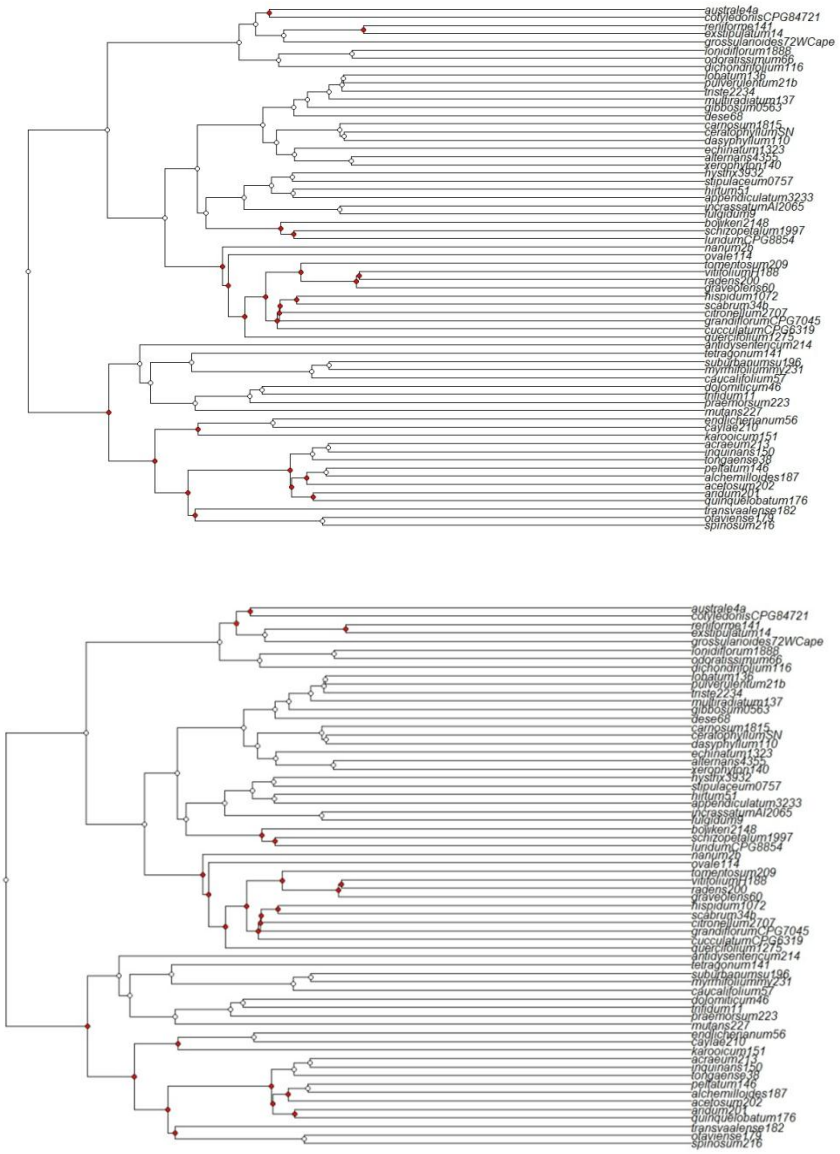
Figure 24 – Continuous optimization of C-values (a) and Cx-values (b) for the SingleTree (unsampled taxa were pruned).C-values are shown in blue boxes, with the significance values based on the randomized control shown below (nodes with p=0 were significantly different from the randomized values). Branch lengths are the result of multiplication by factor λ following Pagel’s transformations. This did not influence the ancestral state estimates.

From the continuous optimization we can inspect the estimated ancestral state values for *Pelargonium* as a whole and its major clades (Figure 25). The large chromosome clade, C, is also estimated to have a relatively high ancestral C- and Cx-value (2.67 and 1.33 pg respectively). The smaller chromosome clades by comparison have lower ancestral estimations. For clade B, the ancestral C-value is given as 1.64 pg, and the Cx-value 0.82. For clade A, the ancestral C-value is estimated at 1.98 pg, while the Cx-value is 0.99 pg. The ancestral value for *Pelargonium* is equal to the phylonegetic mean, which for C-values is 2.35 pg and for Cx-values is 1.17 pg.

Figure 25 – Table showing the ancestral stae values for the Pelargonium *and its major clades C, B and A. The small chromosome clades are assigned a smaller ancestral C- and Cx-value than the large chromosome clade. The ancestral state for* Pelargonium *is equal to the phylogenetic mean and has an intermediate value for both the C- and Cx-values.*

Trait	**Node**	**Ancestral state (pg)**
Cx-values	*Pelargonium*	1.17
	C	1.33
	B	0.82
	A	0.99
C-values	*Pelargonium*	2.35
	C	2.67
	B	1.64
	A	1.98

Because REML uses the phylogenetic mean as its basal ancestral state and calculates subsequent ancestral states from there, deeper nodes appear to stay close to the mean trait value. These node values could be regarded as artefacts as randomizing trait values across taxa will give average values for such nodes as well, nullifying the phylogenetic component of character evolution. I therefore randomized the genome size values across the taxa 1,000 times and estimated ancestral states for each randomization. The distribution of values for each node was then taken as a normal distribution against which the original node value was tested for significance. If the original node value was either higher or lower than more than 97.5% of the randomized node values, the genome size was taken to be significantly different from the average value for that node. The results of these randomized optimizations are shown in Figure 26.



a

b

Figure 26 - The continuous optimizations of C-values (a) and Cx-values (b) for the SingleTree shown again, but this time with nodes coloured according to significance. Red nodes differ significantly from the randomized control values (p=0.000 in Figure X.). Of the ancestral nodes, only that of clade C significantly differs from the randomized data.

Discussion

The continuous and discrete optimizations yielded similar results. The exception being the inherently different values assigned to deeper ancestral nodes in the phylogeny. The discrete approach proposes a range of equivocally plausible ancestral genome sizes for clades A, B and C as well as for the genus as a whole. By definition, however, the ace function computes the phylogenetic mean for genome size as the ancestral state for the genus, and absolute values for the nodes. Considering the sampling of clades is not representative of the phylopgeny (for example, undersampling in *Hoarea*) the continuous optimization proposes an average ancestral value for genome size and subsequent genome downsizing in clades A

and B and a genome size increase in clade C. However, this pattern is more likely an artefact resulting from the optimization process rather than a most probable evolutionary history of genome size in *Pelargonium.*

The significance tests for the ace reconstruction show that a majority of nodes in *Pelargonium* do not differ significantly from a randomized analysis. An exception is clade A1, which shows a significant deviation from randomized C- and Cx-values for all nodes. Surprisingly, the octaploids in clade A1 (*P. graveolens, P. radens* and others) are assigned a lower-than-average ancestral C-value although the discrete analysis assigns them a high (>1.8pg) ancestral C-value. This may be a result of the 'downstream' approach of the ace analysis, which starts by computing the node value at the base of the tree and then works its way to the superficial nodes.

The deeper nodes in clade C are also significantly different, though most of the more superficial ones are not. It seems like the C clade shows a basal direction towards higher C- and Cx-values, but certain species have reverted back to smaller sizes later on, while others continued the upwards trend.

Some interesting shifts between C- and Cx-values can be seen in the Mesquite trees. Whereas in clade C the C-values and Cx-values are in most cases the same, for clades A1 and B we see drastic changes for polyploid species like *P. graveolens* or *P. radens*. It seems from these data that increased genome size in clade C is not caused by polyploidy events, whereas in clades A and B it is. The exception in the case of clade B is *P. grossularioides*, which maintains a high Cx-value along with its high C-value. This is probably due to the exceptionally high basic chromosome number observed in *grossularioides* (16 and 17), which gives it a similar total number of chromosomes to some tetraploids despite being diploid itself.

The most important facet missing from the current analysis is an outgroup. *Pelargonium*'s nearest living relative is the miniature genus *Hypsiocharis*, of which no samples were available from the botanical gardens I visited. An estimation for *Hypsiocharis*'s genome size would provide a context to any estimation of *Pelargonium's* ancestral genome size from which a direction of evolution could be postulated.

The discrete state change summaries of the MultiTree dataset shown that genome size changes resulting in an increase of C- and Cx-values both occur more frequently than occurences of genome size decrease. However, when we look at the major clades separately, we see that in clade B this pattern is inverted for Cx-values, and that the majority of genome upsizing has occurred in clades A and C. Clade B features a lot of annual species which are known for their low Cx-values (Knight *et al.* 2005, Herben *et al.* 2012).

Clade A shows more frequent occurences of increase in both C- and Cx-values. Considering the low level of sampling of the clade , especially in subclade A2, interpreting the state change summary of the MultiTree dataset is difficult. The results indicate that the overall trend towards lower genome sizes apparent from the flow cytometry measurements is the result of a single, ancestral change towards a smaller C- and Cx-value. This is supported by the optimization of both traits over the SingleTree, where clade A has an ancestral low genome size and subsequent increases of both C-and Cx-values in some taxa, including polyploids. More flow cytometry measurements, particularly in clade A2, are needed to support this hypothesis.

8.3 - Trait correlations

Both C-values and Cx-values were shown to correlate with different life characteristics, such as polyploidy or stomata size, in various angiosperm groups (e.g. Leitch and Bennet 2004, Beaulieu et al. 2008). These correlations may be indirect consequences of the direct influence of genome size on cell cycle rates and cell size (Gregory 2001). To test whether the known correlations between genome size and polyploidy, as well as guard cell length, are also present in *Pelargonium* I conducted several correlation tests using SPSS.

Genome size and polyploidy

Only the species for which both genome size measurements and ploidy data were available were included to test the correlation between Cx-values and polyploidy. As no karyotypes could be made for the accessions sampled for this experiment, only species that did not have multiple known ploidy levels were taken into account. Fifty species fit the criteria, of which 37 were diploid, 10 were tetraploid and 3 were octaploid. Since the data was not normally distributed, means were compared with a non-parametric Mann-Whitney U test.

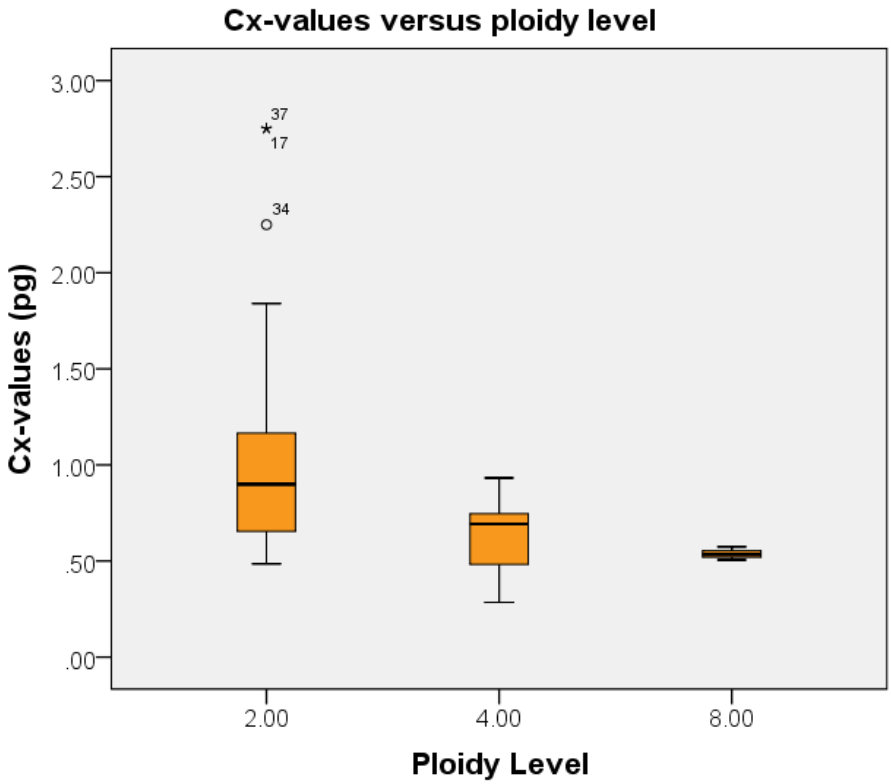


Figure 27 – Cx-values plotted against ploidy level for 50 Pelargonium *species. Diploids have significantly higher Cx-values than either tetra- or octaploids, although tetraploids and octaploids do not differ significantly from each other.*

A Mann-Whitney U test showed significant differences between average diploid and polyploid Cx-values (diploid v. tetraploid p=0.023, diploid v. octaploid p=0.016). However, between tetraploids and octaploids no significant difference between average Cx-values was found (p=0.54).

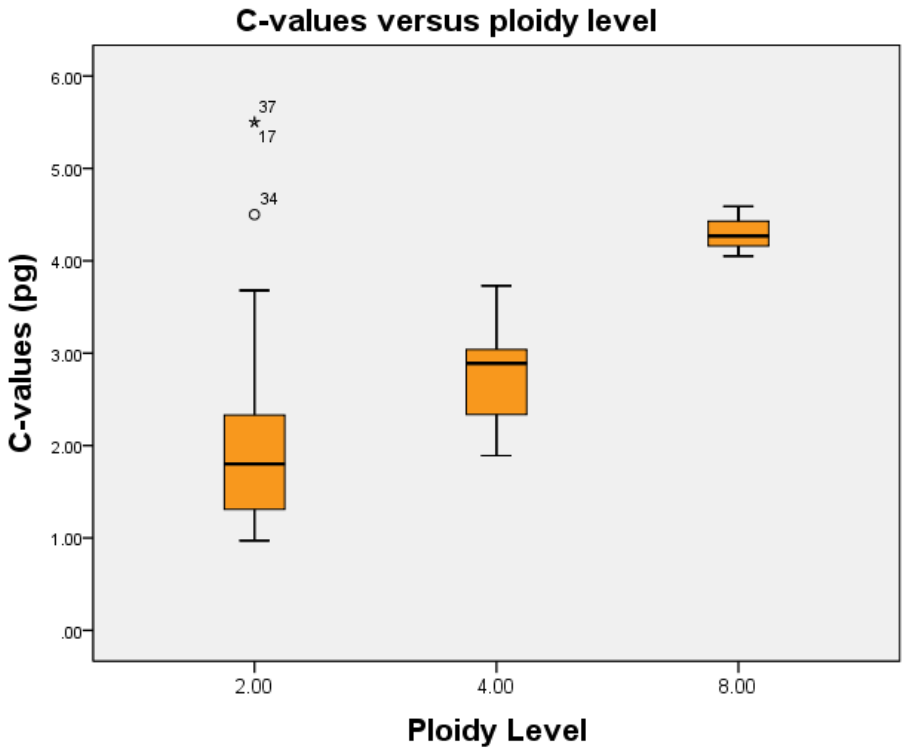


Figure 28 – C-values plotted against ploidy level. There seems to be a positive correlation between ploidy level and C-values, as expected. All three groups differ significantly from each other in average C-values (p=0.014 in all three cases).

The Mann-Whitney U test showed a significant difference between average C-values for octaploids, diploids, and tetraploids (p=0.014 for all three comparisons). Three outliers were identified among the diploid species. These were *P. peltatum P quinquelobatum. P.endlicherianum*, which have unusually high C-values and Cx-values compared to other diploids.

Looking at the total number of chromosomes, there was a strong positive correlation between the number of chromosomes and the C-value (p=0.000), while a negative correlation between Cx-values and chromosome numbers was indicated but not significant when outliers were present, although without outliers the correlation was significant (p=0.063 with outliers, p=0.004 without).

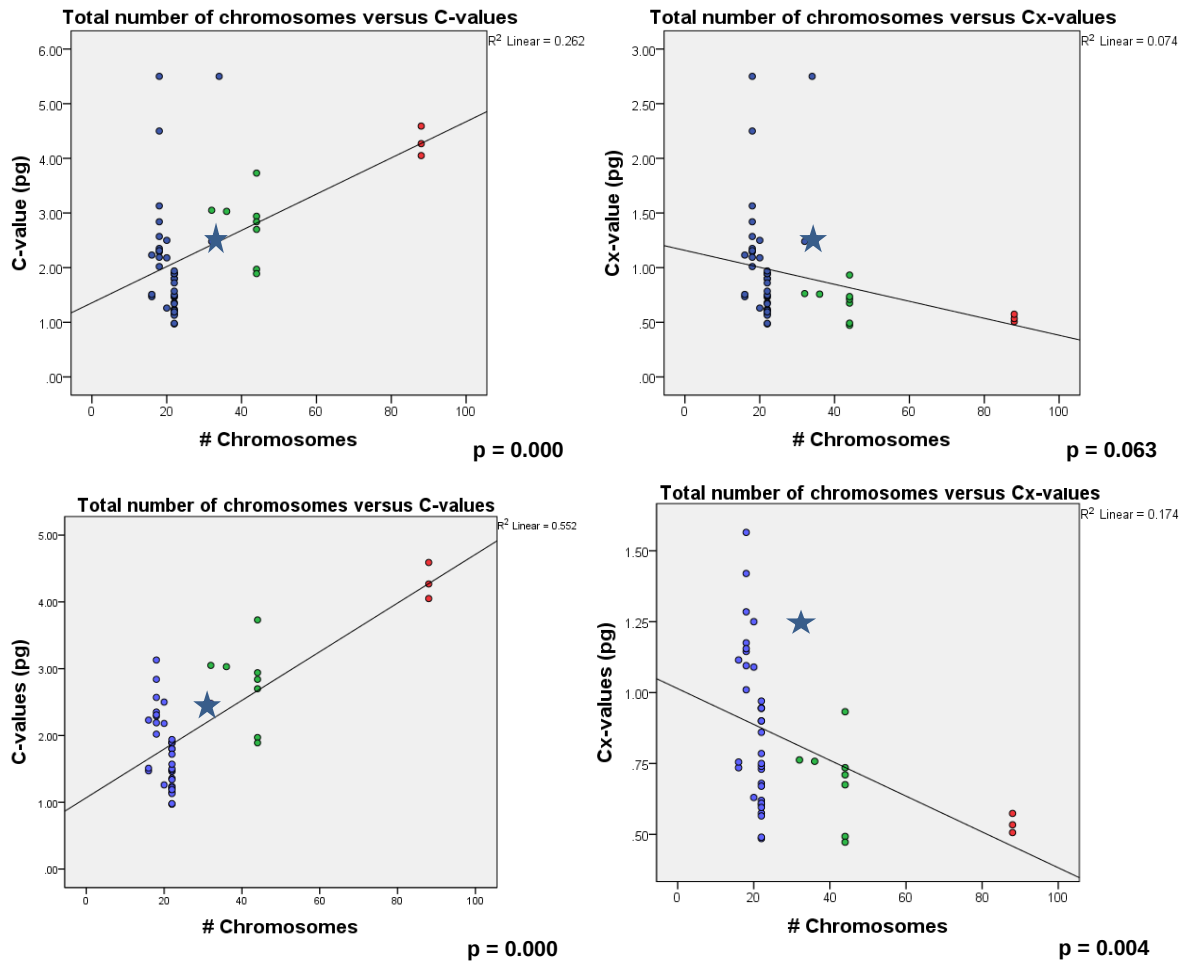


Figure 29 – Scatter plots showing correlations between absolute number of chromosomes and C-values (left) and Cx-values (right) for datasets including (top) and excluding outliers (bottom). The positive correlation between C-values and total chromosome number is significant (p=0.000), whereas the negative correlation between Cx-values and total chromosome number is not by a small margin with outliers included (p=0.063), although it is significant when outliers are excluded. Outlying species are P. peltatum, P. quinquelobatum *and* P. endlicherianum. *Blue dots are diploid species, green ones are tetraploid, and red dots are octaploid. The star indicates* P. grossularioides*, a diploid species with an unusually high basic chromosome number.*

Genome size and guard cell length

To test for a possible correlation between genome size and guard cell length was present in *Pelargonium*, the guard cell measurements made in Munster were compared to the mean genome size estimates for the species sampled there.

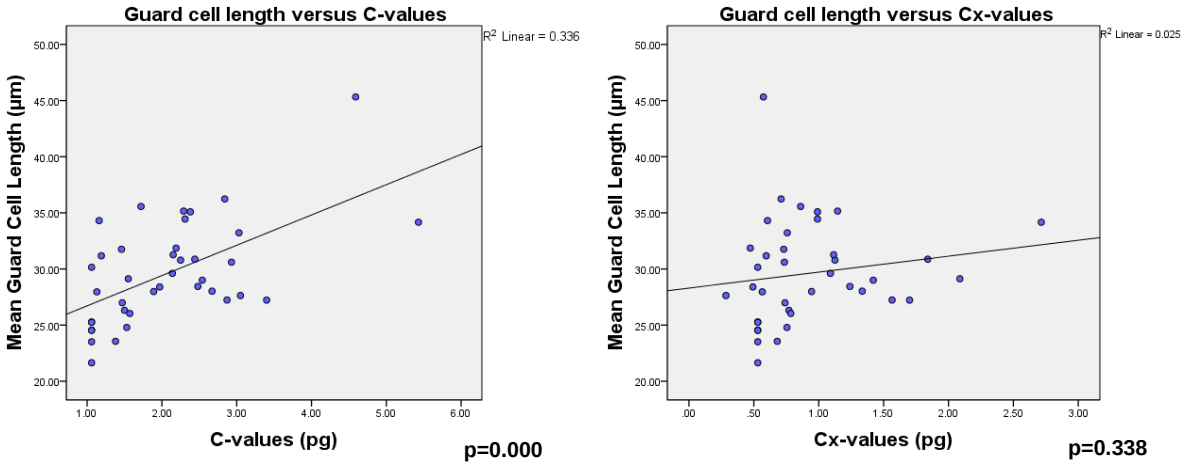


Figure 30 – Regression analyses for average guard cell length per species and mean C- and Cx-values. Guard cell length and C-values show a positive correlation (p=0.000).

The regression analysis showed that average guard cell length correlates positively with C-values (p=0.000) but not with Cx-values (p=0.338). However, as can be seen in Figure X in the C-values plot two outliers (measurements for species *P. vitifolium* and *P. praemorsum*) seem to contribute to the positive correlation. The analysis was therefore repeated without those outliers to see if the correlation was still significant (Figure 31). Repeating the regression test, the positive correlation between C-values and guard cell length still held without the outliers (p=0.009).

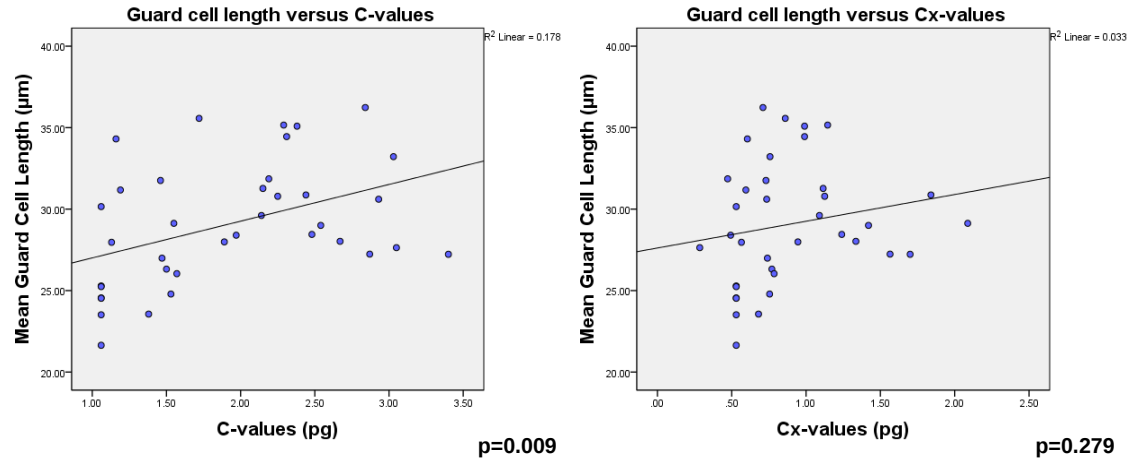


*Figure 31 – Regression analysis showing positive correlation (p=0.009) between average C-values and guard cell length without outliers (*P. vitifolium *and* P. praemorsum*). Cx-values remain insignificantly correlated to mean guard cell length (p=0.279).*

8.4 – Discussion

The strong relationship between C-values and polyploidy is well established in previous studies (e.g. Gregory 2001, Bennetzen et al. 2005) and seems to be supported for *Pelargonium.* It is sensible to assume that an increase in the number of chromosomes will inevitably lead to an increase in absolute DNA content, although this increase is not necessarily linear (e.g. Devos et al. 2002).

Pelargonium also shows a significant decrease in Cx-value with increasing ploidy level. This supports the hypothesis that polyploidization shortens the monoploid genome despite increasing C-values. However, the difference between octaploids and tetraploids in this analysis is not significant, which might be because the reduction in Cx-values could be relative and not absolute following a ploidization event. A low sample size for octaploids may also have contributed.

Diploids display a wide range of both high and low C- and Cx-values. We see this more clearly in the plots comparing absolute chromosome numbers to C- and Cx-values, where *P. peltatum, P. endlicherianum* and *P. quinquelobatum,* which have unusually high C-values for diploids, are clear outliers. In the case of *P. peltatum*, this is probably due to a higher ploidy level than assumed, as all sampled species with several possible ploidy levels were assumed to be diploid.

We see that species that are diploid but have a high basic number of chromosomes (*P. grossularioides,* indicated with a star in Figure 29*)* share the high C-value of tetraploids while maintaining a high Cx-value, whereas tetraploids show a decline in Cx-value.

The correlation between C-values and guard cell length is significant in *Pelargonium* according to these results. This supports the findings by Beaulieu et al. (2008), who proposed this correlation for all angiosperms. Although the predictive strength of the correlation is low, it is strongly significant (p=0.000 with outliers, p=0.009 without outliers). The outlying species, *P. vitifolium* and *P. praemorsum*, are expected to be outliers based on their high genome size, so sampling intermediate species to fill the gap between them and the other sampled species would be a good way to justify their inclusion.

Despite the widely different leaf morphology in the *Pelargonium* species sampled, genome size might be an important factor in determining stomata size. To determine this causative relationship, diploid species could be hybridized or otherwise subjected to polyploidy-inducing treatment (cold or N2 under pressure) to increase their C-values and subsequently measure their guard cell length. Cx-values do not appear to be correlated to guard cell length which is expected as it is the absolute DNA content that determines cell size and not the size of the monoploid genome.

The relationship between longevity and genome size in *Pelargonium* could not be tested as very few annuals had been sampled. It seems unlikely that genome size would be an important factor in determining perennial/annual life forms in *Pelargonium* as the absolute differences in genome size within the genus seem small (about 6 pg at most) whereas the literature suggests sizes of around 20 pg for obligate perennials (Knight *et al.* 2005). However, that is not to say that increasing genome size heightens the chance of becoming perennial, and to test this more annual species will need to be sampled.

Chapter 9 – Conclusions

The genus *Pelargonium* has a small genome size, as is common for angiosperms, with the average 1C-value measured for this thesis at around 1.20 pg. However, as evidenced by the respective standard deviation of 0.585 pg, the genus shows a notable interspecific variation in genome size.

The evolution of both C- and Cx-values show a moderate to strong phylogenetic signal (λ = 0.5-0.81), although C-values may also follow an Ornstein-Uhlenbeck mode of random evolution with a centralizing tendency, or even a normal distribution. Resolving the polytomies in the current *Pelargonium* phylogeny may confirm either of these possibilities.

According to the discrete character optimizations, the ancestral C-values of clades C and B lie in the range of 0.4->1.8 pg, while the range of clade A is more narrow, between 0.4-0.8 pg. For Cx-values, the range of clades A and B lies within the range of 0.4-0.8 pg, whereas the range of clade C remains equivocal at 0.4->1.8pg.

From the continuous optimizations, it seems that clade C (the large chromosome clade) had high ancestral C- and Cx-values compared to clades A and B. The ancestral C- and Cx-values for *Pelargonium* are intermediate, lying in between the ancestral values of the major clades. However, these and indeed most ancestral values calculated in the continuous optimization method do not differ significantly from a randomized control, so there may be artefact issues with the analysis.

The positive correlation between polyploidy and C-values, as well as the negative correlation between polyploidy and Cx-values, both occur in *Pelargonium*. These correlations have been found in many angiosperm groups and these findings therefore support the broader view that polyploidization may lead to genome downsizing despite increasing actual DNA content. The positive correlation between guard cell length and C-values is also significant in *Pelargonium*.

9.1 - Future prospects

For future research in genome size evolution in Pelargonium priority must lie in setting up a good protocol for accurate flow cytometry measurements of Pelargonium plants. This thesis has not included the proper trials outlined by Dolezel *et al.* (2007) to check for the accuracy of the flow cytometry readings. In addition, propidium iodide staining should be used for each flow cytometer reading to enhance accuracy. Petal samples should also be examined together with standard leaf samples to assess the effect of secondary metabolites on the flow cytometry measurements.

This study has due to limited funds been unable to conduct proper intraspecific comparisons of genome size. To assess the variation within species in C- and Cx-values multiple genome size measurements should be made of at least one representative species for each clade (C1, C2, B, A1, and A2). A diploid species with basic chromosome number eleven would be suitable and can be found in each major clade.

Proper chromosome counts for each accession sampled for flow cytometry readings should be done, at least for species with known polyploid individuals. Without chromosome counts, flow cytometry measurements can only be used for estimating Cx-values, by assuming ploidy levels for species that have not been karyotyped.

When a good flow cytometry protocol for *Pelargonium* has been established, the existing genome size estimates can be checked and supplemented. In particular, sections *Hoarea* and *Ligularia* are currently lacking in genome size estimates, whilst being the most specious clades in the genus. Annual species, mostly found in clade B, should also be targeted for flow cytometry, to see if longevity correlates to genome size in *Pelargonium*. Finally, more polyploid species, in particular species with multiple ploidy levels such as *P. karooicum*, require more attention. Their genome size estimates, together with karyotyping can solidify the relationship between polyploidy and Cx-values indicated in this thesis.

The guard cell measurements can be complemented by stomata density data, as this characteristic is also known to correlate with genome size in some angiosperms. In addition, the guard cell data can be supplemented for species with currently few guard cell length estimates, and new species records can be made and added.

The character optimizations can be redone when more data is available. The discrete approach can be rerun with the current available data; the compartmentalization of the genome size range for this thesis was done without taking the distribution of C- and Cx-values across the genus into account. The program Morphocode (http://bio.kuleuven.be/sys/mc/) could be used to properly assign compartments to both traits based on gap weighting.

Now that we know *Pelargonium* shows variation in genome size between species, a logical question is where this variation comes from. Part of the variation is correlated to ploidy levels and chromosome numbers, but the variation among species with similar ploidy and total number of chromosomes may be correlated to repetitive or transposable DNA content. Perhaps (partial) genome sequencing can be combined with flow cytometry measurements to explore this.

Acknowledgements

I owe thanks to many people and institutions who helped me during this thesis, either with advice, materials, data, or moral support:

First of all my supervisor, Freek Bakker, for his advice and support during this thesis.

Hans de Jong, who supervised my work on the karyotypes and provided space in the greenhouse for several *Pelargonium* plants. Also thanks to Dorá Szinay and José van de Belt for their assistance in the lab.

Thanks also to Lars Chatrou for his advice and in particular his support during the Mainz conference. Also thanks to Ronald van den Berg and Eric Schranz for their advice, Wilma Twigt and Nynke Groendijk-Wilders for helping out with practical and financial issues, André van Proosdij for providing some *Pelargonium* samples, Pulcherie Bissiengou for her assistance during the DNA barcoding labwork, Robin van Velzen and Martijn Staats for help with the data analysis and for providing general advice, and all the other people at the WUR Biosystematics group for providing such a nice and friendly work atmosphere.

I'd also like to thank my fellow MSc thesis students and friends, Jens Ringelberg, Sander van Leeuwen, Sara van de Kerke, Rens Holmer, Jeike van de Poel, Rutger Wilschut, Mark Lammers and Erik-Jan Bijleveld for invaluable help and frequent distractions from work on the infamous Island.

Thanks to Joachim Roeschenbleck from the Westfahlische Universitat in Munster for providing access to the botanical gardens in that wonderful city and for his help during my stay there. Thanks as well to Jorge Kudla for providing access to the facilities at his workgroup. I'm also much obliged to Kathrin Schluking and Leonie Steinhorst for their assistance during the stomata measurements, and all the other people at the Arbeitsgruppe Kudla for their friendliness and helpfulness.

Thanks to Gerard Geenen from Plant Cytometry Services for expert advice on the genome size measurements and for carrying out the flow cytometry work.

I'd like to thank Ben Groen of the Kerkrade botanical gardens for providing some *Pelargonium* plants for use in the barcoding work.

Thanks as well to the people at the VU botanical gardens for their cooperation and the *Pelargonium* samples obtained there.

I would also like to thank the organizers of the 21rst international symposium "Biodiversity and Evolutionary Biology" in Mainz for allowing me to participate in the conference and present my work there. Thanks also to the Alberta Mennenga Stichting for funding my stay and conference fee.

Finally, I'd like to thank my family and friends for their cotinuous support.

References

Albers F., Van der Walt J.J.A., *Untersuchungen zur Karyologie und Mikrosporogenese von Pelargonium sect. Pelargonium (Geraniaceae)* (German); Plant Systematics and Evolution 147: 177-188 (1984)

Albers F., Gibby M., Austmann M., *A reappraisal of Pelargonium sect. Ligularia (Geraniaceae);* Plant Systematics and Evolution 179: 257-276 (1992)

Bakker F.T., Culham A., Daugherty L.C., Gibby M., *A trnL-F based phylogeny for species of Pelargonium (Geraniaceae) with small chromosomes;* Plant Systematics and Evolution 216: 309-324 (1999)

Bakker F.T., Culham A., Hettiarachi P., Touloumenidou T., Gibby M., *Phylogeny of Pelargonium (Geraniaceae) based on DNA sequences from three genomes;* Taxon 53(1): 17-28 (February 2004)

Bakker F.T., Culham A., Marais E.M., Gibby M., *Nested radiation in Cape Pelargonium;* Plant Species-level Systematics (Chapter 4), International Association for Plant Taxonomy, Institute of Botany, University of Vienna (2005)

Barringer B.C., *Polyploidy and self-fertilization in flowering plants;* American journal of Botany 94: 1527-1533 (2007)

Beaulieu J.M., Leitch I.J., Patel S., Pendharkar A., Knight C.A., *Genome size is a strong predictor of cell size and stomata density in ansiosperms;* New Phytologist 179(4): 975-986 (2008)

Bennett M.D, *Nuclear DNA Content and Minimum Generation Time in Herbaceous Plants;* Proceedings of the Royal Society B 181(1063): 109-135 (June 1972)

Bennetzen J.L., Kellogg E.A., *Do plants have a one way ticket to genetic obesity?;* The Plant Cell 9: 1509-1514 (1997)

Bennetzen J.L., Ma J., Devos K.M., *Mechanisms of Recent Genome Size Variation in Flowering Plants;* Annals of Botany 95: 127-132 (2005)

Bollback J.P., *SIMMAP: Stochastic character mapping of discrete traits on phylogenies;* BMC Bioinformatics 7(88): (2006)

Bretagnolle F., Thompson J.D., *Gametes with the somatic chromosome number: mechanisms of their formation and role in the evolution of autopolyploid plants;* New Phytologist 129(1): 1-22 (January 1995)

Brown J.K., Lambert G.M., Ghanim M., Czosnek H., Galbraith D.W., *Nuclear DNA content of the whitefly* Bemisia tabaci *(Aleyrodidae: Hemiptera) estimated by flow cytometry;* Bulletin of Entomological Research 95: 309-312 (2005)

Chase M.W., Hanson L., Albert V.A., Whitten W.M., Williams N.H., *Evolution and Genome Size in Subtribe Oncidiinae (Orchidaceae);* Annals of Botany 95: 191-199 (2005)

Chumley T.W., Palmer J.D., Mower J.P., Fourcade H.M., Calie P.J., Boore J.L., Jansen R.K., *The complete chloroplast genome sequence of Pelargonium x hortorum: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants;* Molecular Biology and Evolution 23(11): 2175-2190 (2006)

Comai L., *The advantages and disadvantages of being polyploid;* Nature Reviews Genetics 6: 836-846 (2005)

Devos K.M., Brown J.K.M., Bennetzen J.L., *Genome Size Reduction through Illegitimate Recombination Counteracts Genome Expansion in Arabidopsis;* Genome Research 12: 1075-1079 (2002)

Doležel, J., Greilhuber J., Lucretti S., Meister A., Lysák M.A., Nardi L., Obermayer R., *Plant Genome Size Estimation by Flow Cytometry: Inter-laboratory Comparison;* Annals of Botany 82(A): 17-26 (1998)

Doležel J., Bartoš J., *Plant DNA Flow Cytometry and Estimation of Nuclear Genome Size:* Annals of Botany 95: 99-110 (2005)

Doležel J., Greilhuber J., Suda J., *Estimation of nuclear DNA content in plants using flow cytometry;* Nature Protocols 2(9): 2233-2244 (2007)

Fawcett J.A., Van de Peer Y., *Angiosperm polyploids and their road to evolutionary success;* Trends in Evolutionary Biology 2(1) (2010)

Francis D., Davies M.S., Barlow P., *A Strong Nucleotypic Effect on the Cell Cycle Regardless of Ploidy Level;* Annals of Botany 101: 747-757 (2008)

Franks P.J., Freckleton R.P., Beaulieu J.M., Leitch I.J., Beerling D.J., *Megacycles of atmospheric carbon dioxide concentration correlate with fossil plant genome size;* Philosophical Transactions of the Royal Society B 367: 556-564 (2012)

Gibby M., Westfold J., *A Cytological Study of Pelargonium sect. Eumorpha (Geraniceae);* Plant Systematics and Evolution 153: 205-222 (1986)

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

Gibby M., Hinnah S., Marais, E. M., Albers, F., *Cytological variation and evolution within Pelargonium section Hoarea (Geraniaceae);* Plant Systematics and Evolution 203: 111-142 (1996)

Goloboff P.A., Mattoni C.I., Quinteros A.S., *Continuous characters analyzed as such;* Cladistics 22(6): 589-601 (December 2006)

Gregory T.R., *Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma;* Biological Reviews 76(1): 65-101 (February 2001)

Gregory T.R., *Animal Genome Size Database;* http://www.genomesize.com (2005)

Greilhuber J., *"Self-tanning" - a new and important source of stoichiometric error in cytophotometric determination of nuclear DNA content in plants;* Plant Systematics and Evolution 158: 87-96 (1988)

Grielhuber J., Doležel J., Lysák M.A., Bennett M.D., *The Origin, Evolution and Proposed Stabilization of the Terms 'Genome Size' and 'C-value' to Describe Nuclear DNA Contents;* Annals of Botany 95: 255-260 (2005)

Grotkopp E., Rejmánek M., Sanderson M.J., Rost T.L., *Evolution of genome size in pines (Pinus) and its life-history correlates: supertree analysis;* Evolution 58(8): 1705-1729 (2004)

Hardie D.C., Gregory T.R., Hebert P.D.N., *From Pixels to Picograms: A Beginners' Guide to Genome Quantification by Feulgen Image Analysis Densitometry;* Journal of Histochemistry & Cytochemistry 50(6): 735-749 (2002)

Hawkins J.S., Proulx S.R., Rapp R.A., Wendel J.F., *Rapid DNA loss as a counterbalance to genome expression through retrotransposon proliferation in plants;* PNAS 106(42): 17811-17816 (October 2009)

Hollingsworth P.M., *Refining the DNA barcode for land plants;* Proceedings of the National Academy of Sciences 108(49): 19451-19452 (2011)

Jakob S.S., Meister A., Blattner F.R., *The Considerable Genome Size Variation of Hordeum Species (Poaceae) is Linked to Phylogeny, Life Form, Ecology and Speciation Rates;* Molecular Biology and Evolution 21(5): 860-869 (2004)

Jedrzejczyk I., Sliwinska E., *Leaves and Seeds as Materials for Flow Cytometric Estimation of the Genome Size of 11 Rosaceae Woody Species Containing DNA-Staining Inhibitors;* Journal of Botany 2010, 9 pages (2010)

Kazazian jr. H.H., *et al., Mobile elements: Drivers of Genome Evolution;* Science 303: 1626-1632 (2004)

Kew Royal Botanical Gardens, *Plant C-value Database;* www.data.kew.org/cvalues (2012)

Knight C.A., Ackerley D.D., *Variation in nuclear DNA content across environmental gradients: a quantile regression analysis;* Ecology Letters 5(1): 66-76 (2002)

Knight C.A., Molinari N.A., Petrov D.A., *The Large Genome Constraint Hypothesis: Evolution, Ecology and Phenotype;* Annals of Botany 95: 177-190 (2005)

Knight C.A., Beaulieu J.M., *Genome Size Scaling through Phenotype Space;* Annals of Botany 101: 759-766 (2008)

Kolodziej H., Kayser O., Radtke O.A., Kiderlen A.F., Koch E., *Pharmacological profile of extracts of Pelargonium sidoides and their constituents;* Phytomedicine 10(4): 18-24 (2003)

Koopman W.J.M., Hadam J., Doležel J., *Evolution of DNA content and base composition in Lactuca (Asteraceae) and related genera;* Zooming in on the lettuce genome 97-124 (2002)

Kraaijeveld K., *Genome Size and Species Diversification;* Evolutionary Biology 37: 227-233 (2010)

Leitch I.J., Bennett M.D., *Polyploidy in angiosperms;* Trends in Plant Science 2(12): 470-476 (1997)

Leitch I.J., Bennett M.D., *Genome downsizing in polyploid plants;* Biological Journal of the Linnean Society 82: 651-663 (2004)

Leitch I.J., Kahandawala I., Suda J., Hanson L., Ingrouille M.J., Chase M.W., Fay M.F., *Genome size diversity in orchids: consequences and evolution;* Annals of Botany 104: 469-481 (2009)

Leitch I.J., Leitch A.R., *Genome Size Diversity and Evolution in Land Plants;* Plant Genome Diversity Vol. 2, DOI 10.1007/978-3-7091-4_19, Springer-Verlag Wien (2013)

Linder H.P., *The radiation of the Cape flora, Southern Africa;* Biological Reviews 78: 597-638 (2003)

Ma Y., Islam-Faridi M.N., Crane C.F., Stelly D.M., Price H.M., Byrne D.H., *A New Procedure to Prepare Slides of Metaphase Chromosomes of Roses;* HortScience 31(5): 855-857 (1996)

Masterson J., *Stomatal Size in Fossil Plants: Evidence for Polyploidy in Majority of Angiosperms;* Science 264(5157): 421-424 (April 1994)

Matzk F., Hammer K., Schubert I., *Coevolution of apomixis and genome size within the genus Hypericum;* Sexual Plant Reproduction 16: 51-58 (2003)

Matzke M.A., Matzke A.J.M., *Polyploidy and transposons*; Trends in Ecology and Evolution 13(241) (1998*)*

Park M., Park J., Kim S., Kwon J., Park H.M., Bae I.K., Yang T., Lee Y.H., Kang B., Choi D., *Evolution of the large genome in Capsicum annuum occurred through accumulation of single-type long terminal repeat retrotransposons and their derivatives;* The Plant Journal 1-13 (2011)

Pagel M., *Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters;* Proceedings of the Royal Society (B) 244: 37-45 (1994)

Pagel M., *Inferring evolutionary processes from phylogenies;* Zoologica Scripta 26: 331-348 (1997)

Pagel M., *Inferring the historical patterns of biological evolution;* Nature 401: 877-884 (1999)

Pellicer J., Fay M.F., Leitch I.J., *The largest eukaryotic genome of them all?* Botanical Journal of the Linnean Society 164: 10-15 (2010)

Ramsey J., Schemske D.W., *Pathways, mechanisms and rates of polyploid formation in flowering plants;* Annual Review of Ecology and Systematics 29: 467-501 (1998)

SanMiguel P., Tikhonov A., Jin Y., Motchoulskaia N., Zakharov D., Melake-Berhan A., Springer P.S., Edwards K.J., Lee M., Avramova Z., Bennetzen J.L., *Nested retrotransposons in the Intergenic Regions of the Maize Genome;* Science 274(5288): 765-768 (November 1996)

Šmarda P., Bureš P., *Intraspecific DNA content variability in Festuca pallens on different geographical scales and ploidy levels;* Annals of Botany 98: 665-678 (2006)

Soltis D.E., Soltis P.S., *Polyploidy: recurrent formation and genome evolution;* TREE 14(9): 348-352 (September 1999)

Soltis D.E., Soltis P.S., Tate J.A., *Advances in the study of polyploidy since Plant speciation;* New Phytologist 161: 173-191 (2003)

Te Beest M., Le Roux J.J., Richardson D.M., Brysting A.K., Suda J., Kubešová M., Pyšek P., *The more the better? The role of polyploidy in facilitating plant invasions;* Annals of Botany 109: 19-45 (2012)

Van der Walt, J.J.A., *Pelargoniums of Southern Africa*; Vol. 1: p.11, Purnell & Sons S.A. (Pty) Ltd., Cape Town (1977)

Van Velzen R., Weitschek E., Felici G., Bakker F.T., *DNA Barcoding of Recently Diverged Species: Relative Performance of Matching Methods;* PLoS One 7(1)): e30490. doi:10.1371/journal.pone.0030490

Veselý P., Bureš P., Šmarda P., Pavlíček T., *Genome size and DNA base composition of geophytes: the mirror of phenology and ecology?* Annals of Botany 109: 65-75 (2012)

Vinogradov A.E., *Genome in toto;* Genome 42(2): 361-362 (1999)

Vinogradov A.E., *Genome size and extinction risk in vertebrates;* Proceedings of the Royal Society London B 271: 1701-1705 (2004)

Walker D.J., Moñino I., Correal E., *Genome size in Bituminaria bituminosa (L.) C.H. Stirton (Fabaceae) populations: separation of "true" differences from environmental effects on DNA determination;* Environmental and Experimental Botany 55(3): 258-265 (2006)

Weiss-Schneeweiss H., Greilhuber J., Schneeweiss G.M., *Genome size evolution in holoparasitic Orobanche (Orobanchaceae) and related genera;* American Journal of Botany 93(1): 148-156 (2006)

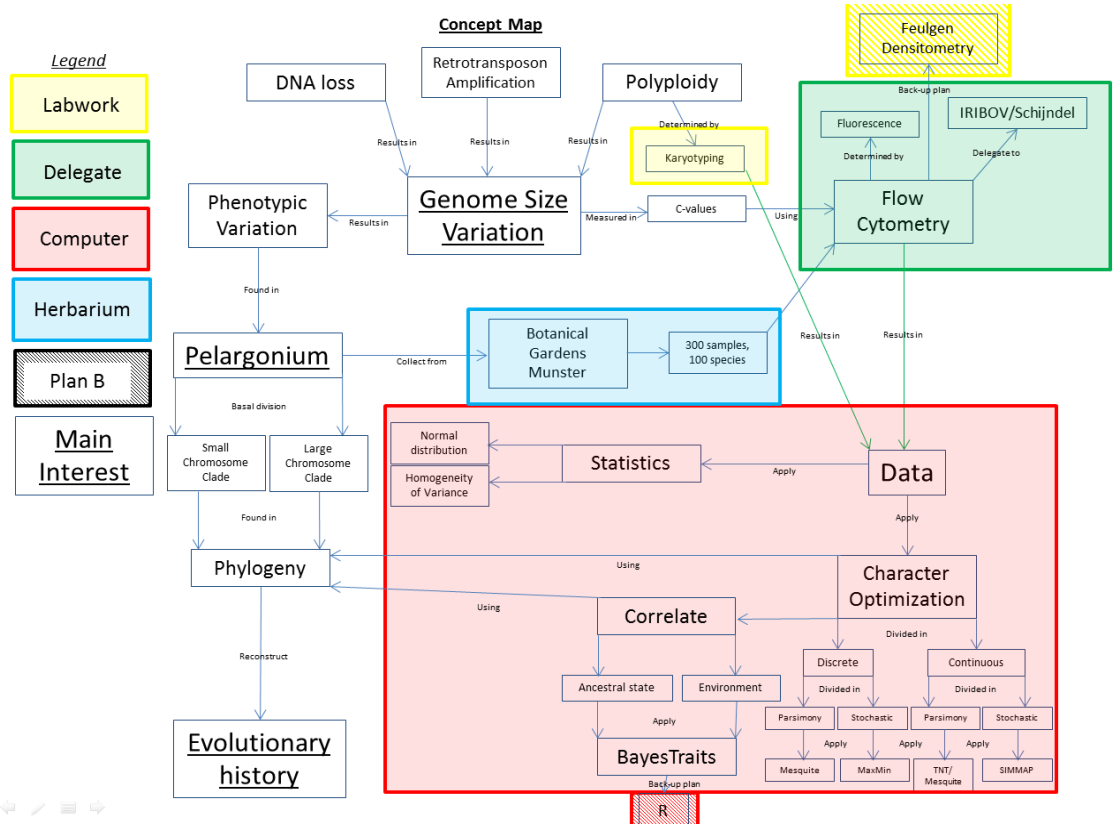
Wiens J.J., Tiu J., *Highly Incomplete Taxa Can Rescue Phylogenetic Analyses from the Negative Impacts of Limited Taxon Sampling,* PLoS One 7(8): e42925. doi:10.1371/journal.pone.0042925 (August 2012)

Wendel J.F., *Genome evolution in polyploids;* Plant Molecular Biology 42: 225-249 (2000)

Yu S., Horn W.A.H., *Additional chromosome numbers in* Pelargonium *(*Geraniaceae*);* Plant Systematics and Evolution 159: 165-171 (1988)

Appendices

Appendix 1 – Concept Map



Appendix 2 – PCR & Cycle Sequencing profiles

PCR

To amplify the trnL-F genes in the samples a PCR protocol for the particular gene was implemented. The mastermix consisted of the following:

10xbuffer Invitrogen (-MgCl)	1µl
dNTPs	0.4µl
BSA 10mg/ml	1µl
Primer trn-C 10µM	0.35µl
Primer trn-F 10µM	0.35µl
Taq polymerase Invitrogen 5U/µl	0.08µl
Water	X µl
Total	Y µl

PCR profile:

50 sec.	94ºC
50 sec.	94 ºC
50 sec.	53 ºC
150 sec.	72 ºC
7 min.	72 ºC
Infinite hold	10 ºC

PCR fragment purification

The PCR products are placed in a Qiagen PCR purification column, after which 750µl PE buffer is added. The columns are centrifuged, and the flow through is discarded. The DNA is then eluted by adding 30µl sterile MQ water and spinning the mixture down.

Cycle Sequencing

In order to determine the base order of the DNA strands, the Cycle Sequencing method was used. The master mix was composed as follows:

Ingredients	**Quantity per sample**
buffer	2.0µl
DETT	2.0µl
Primer Forward	0.5µl
MQ Water	4.5µl
Total	9.0µl

PCR profile:

Infinite hold	10 ºC

*exact composition: 180µl MQ water, 135µl 7.5M NH4Ac, 1ml cold 96% ETOH (2.5 Vol.)

Appendix 2 – Flow Cytometry samples

Species	Collection number	Labcode	Locality	Cv value	Cv Value	Ploidy	Section Pelargonium	Clade	Date	Clade Code
appendiculatum	VU012	13appe01	Hortus VU	0.97	0.485	2	Ligularia	A2	17-apr-12	2
bowkerii	VU001	1bowk01	Hortus VU	2.77	N.A.		Polyactium	A1	17-apr-12	1
bowkerii	VU001	2bowk01	Hortus VU	2.81	N.A.		Polyactium	A1	17-apr-12	1
caucaulifolium	VU007	8cauc01	Hortus VU	2.97	0.7425	4	Myrrhidium	C2	17-apr-12	5
dasyphyllum	VU010	11dasy01	Hortus VU	2.94	N.A.		Otidia	A2	17-apr-12	2
exstipulatum	VU002	3exst01	Hortus VU	1.39	0.695	2	Reniformia	B	17-apr-12	3
fulgidum	VU011	12fulg01	Hortus VU	1.51	0.755	2	Ligularia	A2	17-apr-12	2
graveolens	VU005	6grav01	Hortus VU	3.91	0.48875	8	Pelargonium	A1	17-apr-12	1
inquinans	VU009	10inqu01	Hortus VU	2.26	1.13	2	Ciconium	C2	17-apr-12	5
litorale	VU006	7lito01	Hortus VU	1.07	N.A.		Peristera	B	17-apr-12	3
peltatum	VU003	4pelt01	Hortus VU	4.53	2.265	2	Ciconium	C2	17-apr-12	5
peltatum	VU004	5pelt02	Hortus VU	4.46	2.23	2	Ciconium	C2	17-apr-12	5
tetragonum	VU014	15tetr01	Hortus VU	3.09	1.545	2	Chorisma	C1	17-apr-12	4
triste	VU008	9tris01	Hortus VU	3.96	0.99	4	Polyactium	A1	17-apr-12	1
xerophyton	VU013	14xero01	Hortus VU	1.01	N.A.		Cortusina	A2	17-apr-12	2
cotyledonis	SN001	11cote01	Munster	0.983234	0.491617	2	Isopetalum	B	8-jun-12	3
grandiflorum	STEU 1105	12gran01	Munster	1.058867	0.529434	2	Glaucophyllum	A1	9-aug-12	1
desertorum	STEU 2857	8dese01	Munster	1.134501	0.56725	2	Cortusina	A2	9-aug-12	2
scabrum	SL 17300	17scab01	Munster	1.191226	0.595613	2	Pelargonium	A1	8-jun-12	1
scabrum	STEU 1000	25scab02	Munster	1.191226	0.595613	2	Pelargonium	A1	9-aug-12	1
ovale	STEU 2237	6oval01	Munster	1.266859	0.633429	2	Campylia	A1	8-jun-12	1
cucullatum	STEU 490	7cucu01	Munster	1.16	0.58	2	Pelargonium	A1	9-aug-12	1
gibbosum	N.A.	13gibb01	Munster	1.342492	0.671246	2	Polyactium	A1	8-jun-12	1
gibbosum	STEU 721	11gibb02	Munster	1.380309	0.690155	2	Polyactium	A1	9-aug-12	1
ceratophyllum	Ac 4167	6cera01	Munster	1.455942	0.727971	2	Otidia	A2	9-aug-12	2
carnosum	STEU 1362	20carn01	Munster	1.474851	N.A.		Otidia	A2	8-jun-12	2
fulgidum	N.A.	8fulg02	Munster	1.474851	0.737425	2	Ligularia	A2	8-jun-12	2
fulgidum	STEU 482	10fulg03	Munster	1.474851	0.737425	2	Ligularia	A2	9-aug-12	2
ionidiflorum	STEU 1888	5ioni02	Munster	1.493759	0.74688	2	Reniformia	B	8-jun-12	3
exstipulatum	STEU 1665	4exst02	Munster	1.5	0.75	2	Reniformia	B	8-jun-12	3
fulgidum	N.A.	19fulgAA	Munster	1.5	0.75	2	Ligularia	A2	8-jun-12	2
ionidiflorum	STEU 1874	16ioni03	Munster	1.531576	0.765788	2	Reniformia	B	9-aug-12	3
odoratissimum	STEU 1190	20odor01	Munster	1.531576	N.A.		Reniformia	B	9-aug-12	3
karooicum	2967	17karo02	Munster	1.550484	N.A.		Subsucculentia	C2	9-aug-12	5
hirtum		14hirt01	Munster	1.569392	N.A.		Ligularia	A2	9-aug-12	2
stipulaceum	757	27stip01	Munster	1.569392	0.784696	2	Ligularia	A2	9-aug-12	2
dichondrifolium	872371933	9dich01	Munster	1.5	N.A.		Reniformia	B	9-aug-12	3
multiradiatum	STEU 2191	18mult01	Munster	1.720659	0.86033	2	Polyactium	A1	9-aug-12	1
alchemilloides	STEU 2731	34alch01	Munster	2.15	1.075	2	Ciconium	C2	9-aug-12	5
suburbanum	STEU 785	28subu01	Munster	1.890834	0.945417	2	Myrrhidium	C2	9-aug-12	5
hispidum	STEU 1006	15hisp01	Munster	1.966468	0.491617	4	Pelargonium	A1	9-aug-12	1
aridum	STEU 1847	3arid01	Munster	2.19	1.095	2	Ciconium	C2	9-aug-12	5
otaviense	943	21otav02	Munster	2.136643	1.068321	2	Subsucculentia	C2	9-aug-12	5
inquinans		10inqu02	Munster	2.37	1.185	2	Ciconium	C2	8-jun-12	5
quercifolium	STEU 1275	23quer01	Munster	2.193368	0.548342	4	Pelargonium	A1	9-aug-12	1
otaviense	943	16otav01	Munster	2.136643	1.068321	2	Subsucculentia	C2	8-jun-12	5
xerophyton	STEU 910	21xero02	Munster	2.231184	N.A.		Cortusina	A2	8-jun-12	2
mutans	4120	19muta01	Munster	2.250093	N.A.		Ciconium	C2	9-aug-12	5
acraeum	1975	1acra01	Munster	2.287909	1.143955	2	Ciconium	C2	9-aug-12	5
xerophyton	STEU 910	33xero04	Munster	2.306818	N.A.		Cortusina	A2	9-aug-12	2
acetosum	STEU 692	7acet01	Munster	2.363543	1.181771	2	Ciconium	C2	8-jun-12	5
grossularioides	STEU 4253	13gros01	Munster	2.48	1.24	2	Peristera	B	9-aug-12	3
xerophyton	STEU 1324	32xero03	Munster	2.382451	N.A.		Cortusina	A2	9-aug-12	2
tetragonum		15tetr02	Munster	2.420268	1.210134	2	Chorisma	C1	8-jun-12	4
tetragonum	Ac 2472	29tetr03	Munster	2.439176	1.219588	2	Chorisma	C1	9-aug-12	4
tongaense	10710	30tong01	Munster	2.54	1.27	2	Ciconium	C2	9-aug-12	5
antidysentericum		2anti01	Munster	2.666076	N.A.		Jenkinsonia	C1	9-aug-12	4
pulverulentum	1236	22pulv01	Munster	2.836251	0.709063	4	Polyactium	A1	9-aug-12	1
transvaalense	STEU 779	36tran01	Munster	2.874068	1.437034	2	Ciconium	C2	9-aug-12	5
caucalifolium		9cauc02	Munster	2.91	0.7275	4	Myrrhidium	C2	8-jun-12	5
caucalifolium	STEU 1682	4cauc03	Munster	2.930793	0.732698	4	Myrrhidium	C2	9-aug-12	5
caylae	STEU 2198	5cayl01	Munster	3.025335	0.756334	4	Ciconium	C2	9-aug-12	5
reniforme	STEU 1698	24reni01	Munster	3.05	0.7625	4	Reniformia	B	9-aug-12	3
triste	10591-1-0	2tris03	Munster	3.05	0.7625	4	Polyactium	A1	8-jun-12	1
sidoides	88331057650	26sido01	Munster	3.403502	N.A.		Reniformia	B	9-aug-12	3
karooicum	STEU 2907	14karo01	Munster	4.197652	N.A.		Subsucculentia	C2	8-jun-12	5
triste	STEU 1549	1tris02	Munster	4.18	1.045	4	Polyactium	A1	8-jun-12	1
radens	STEU 8175	12rade01	Munster	4.292194	0.536524	8	Pelargonium	A1	8-jun-12	1
graveolens	STEU 3039	18grav02	Munster	4.18	0.5225	8	Pelargonium	A1	8-jun-12	1
vitifolium	STEU 731	31viti01	Munster	4.594727	0.574341	8	Pelargonium	A1	9-aug-12	1
praemorsum	STEU 1575	35prae01	Munster	5.426694	N.A.		Jenkinsonia	C1	9-aug-12	4
schizopetalum	STEU 1532	3schi01	Munster	6.95827	N.A.		Polyactium	A1		1
alchemilloides			Weng *et a*	2.31	1.155	2	Ciconium	C2		5
alternans			Weng *et a*	1.89	0.945	2	Otidia	A2		2
australe			Weng *et a*	0.98	0.49	2	Peristera	B		3
citronellum			Weng *et a*	1.24	0.62	2	Pelargonium	A1		1
cotyledonis			Weng *et a*	1.15	0.575	2	Isopetalum	B		3
cucullatum			Weng *et a*	1.25	0.625	2	Pelargonium	A1		1
dichondrifolium			Weng *et a*	1.58	N.A.		Reniformia	B		3
dolomiticum			Weng *et a*	2.02	1.01	2	Jenkinsonia	C1		4
echinatum			Weng *et a*	1.22	0.61	2	Cortusina	A2		2
endlicherianum			Weng *et a*	5.5	2.75	2	N.A.	C2		5
exstipulatum			Weng *et a*	1.56	0.78	2	Reniformia	B		3
fulgidum			Weng *et a*	1.5	0.75	2	Ligularia	A2		2
grossularioides			Weng *et a*	2.49	1.245	2	Peristera	B		3
hortorum			Weng *et a*	3.14	N.A.		Ciconium	C2		5
hystrix			Weng *et a*	1.8	0.9	2	Ligularia	A2		2
incrassatum			Weng *et a*	1.5	0.75	2	Hoarea	A2		2
lobatum			Weng *et a*	1.8	0.9	2	Polyactium	A1		1
luridum			Weng *et a*	7.58	N.A.		Polyactium	A1		1
myrrhifolium			Weng *et a*	1.89	0.945	2	Myrrhidium	C2		5
nanum			Weng *et a*	1.1	N.A.		Peristera	B		3
quinquelobatum			Weng *et a*	5.5	2.75	2	Ciconium	C2		5
reniforme			Weng *et a*	1.57	0.3925	4	Reniformia	B		3
spinosum			Weng *et a*	2.5	1.25	2	Subsucculentia	C2		5
tetragonum			Weng *et a*	6.78	3.39	2	Chorisma	C1		4
tomentosum			Weng *et a*	2.7	0.675	4	Pelargonium	A1		1
tongaense			Weng *et a*	3.25	1.625	2	Ciconium	C2		5
transvaalense			Weng *et a*	3.31	1.655	2	Ciconium	C2		5
trifidum			Weng *et a*	2.57	1.285	2	Jenkinsonia	C1		4

Appendix 4.1 – Flow Cytometry Protocol at Plant Cytometry Services (DAPI)

Isolation of Nuclei

Fresh leaf material (≈ 1 cm^2 /20-50 mgr) is "chopped" with a sharp razor blade in 700µl ice-cold buffer, in a plastic Petri disc. (For analyses with an internal standard a piece of leaf material from this standard is chopped together with the sample leaf piece).

DNA buffer (stored at 4 ºC): CyStain UV Precise P. (Partec GmbH, Münster, Germany).

To this buffer is added:

0,1 % DTT (Dithiothreitol)

1% PVP 15 (Polyvinylpyrrolidone)

The fluorescent dye in this buffer is DAPI which selectively complexes with double-stranded DNA to give a product that fluoresce at 465 nm. DAPI has specific DNA-binding properties with preference for adenine-thymine (AT)-rich sequences.

After chopping the buffer is added up to 2 ml.. The solution, containing cell constituents and large tissue remnants, is passed through a nylon filter of 50 µm mesh size. After 15-30 min incubation time the solution with stained nuclei is sent through the flow cytometer.

Flow Cytometry

The fluorescence of the stained nuclei, passing through the focus of a light beam from the high-pressure mercury lamp, is measured by a photodiode and converted into voltage pulses.

These voltage pulses are electronically processed to yield integral and peak signals and are processed by a computer with special software into a DNA histogram. The place of the histogram peaks on the x-axis of the histogram is a measure for the relative DNA amount of the plant.

Material

Flowcytometer: CyFlow ML (Partec GmbH, Otto Hahnstrasse 32, D-4400 Münster, Germany) with a high pressure mercury lamp, OSRAM HBO 100 long life

Objective: 40 x N.A. 0,8 air (Partec)

Filter combination with DAPI:

Heat protection filter KG-1

Excitation-filters: UG-1(for UV) and BG-38.

Dichroic mirrors: TK 420 and TK 560.

Emission-filter: GG 435

Software: Flomax version 2.4 d (Partec)

Appendix 4.2 – Flow Cytometry Protocol at Plant Cytometry Services (PI)

Isolation of Nuclei

'Two-step' Method with 'Cystain PI absolute P' buffer from Partec (art. Nr.: 05-5502). Leaf material (a few cm^2 /20-50 mgr) together with leaf material of an internal standard with known DNA content is "chopped" with a sharp razor blade in 500µl Extraction Buffer (ice-cold), in a plastic petri disc. After 30 – 60 seconds of incubation 2.0 ml Staining Buffer is added. This buffer contains Propidium Iodide (PI) as fluorescent dye and RNA-se. To the buffer is also added 0,1% DTT (Dithiothreitol) and 1% Polyvinylpyrolidone. The sample, containing cell constituents and large tissue remnants of the sample and the internal standard, is passed through a nylon filter of 50 µm mesh size.

Flow Cytometry

After incubation of at least 30 minutes at room temperature, the filtered solution with stained nuclei is send through the flow cytometer. The fluorescence of the stained nuclei, passing through the focus of the light beam of a 50 mW, 532 nm green laser, is measured by a photomultiplier and converted into voltage pulses. These voltage pulses are electronically processed to yield integral and peak signals and have been processed by a computer into a DNA histogram. The DNA amount of the unknown samples will be calculated by multiplying the DNA amount of the used internal standard with the DNA ratio of the relative DNA amount of the unknown sample and the internal standard.

Material

Flow cytometer: CyFlow ML (Partec GmbH, Otto Hahnstrasse 32, D-4400 Münster, Germany) with a high pressure mercury lamp, OSRAM HBO 100 long life for UV (for use with DAPI) and a with a green diode laser 30 mW 532 nm (for use with Propidium Iodide or Ethidium Bromide)

Objective: 40 x N.A. 0,8 air (Partec)

Heat protection filter KG-1

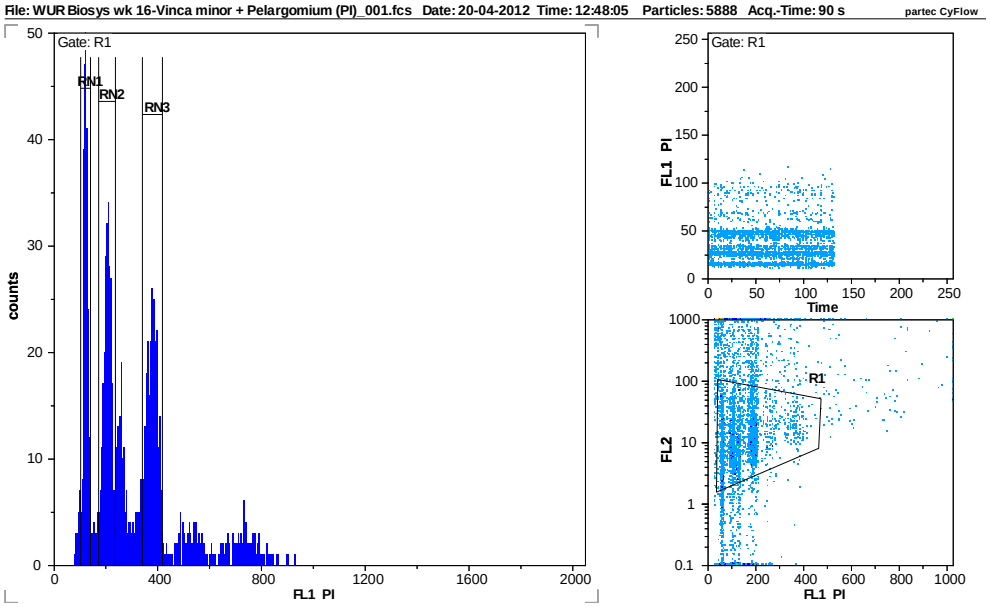
Excitation-filters: UG-1(for UV) and BG-38.

Dichroic beam splitters: TK 420 and TK 560 and 3 full mirrors.

Emission-filter: GG 435 for DAPI (UV) and RG590 for PI

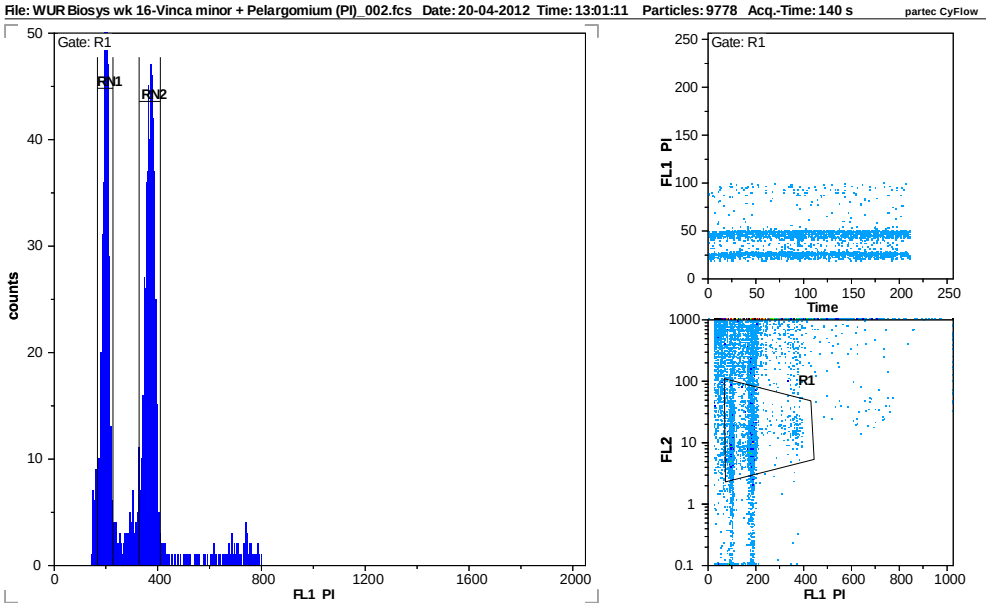
Software: Flomax version 2.4 d (Partec)

Appendix 5 – Flow Cytometry Histograms



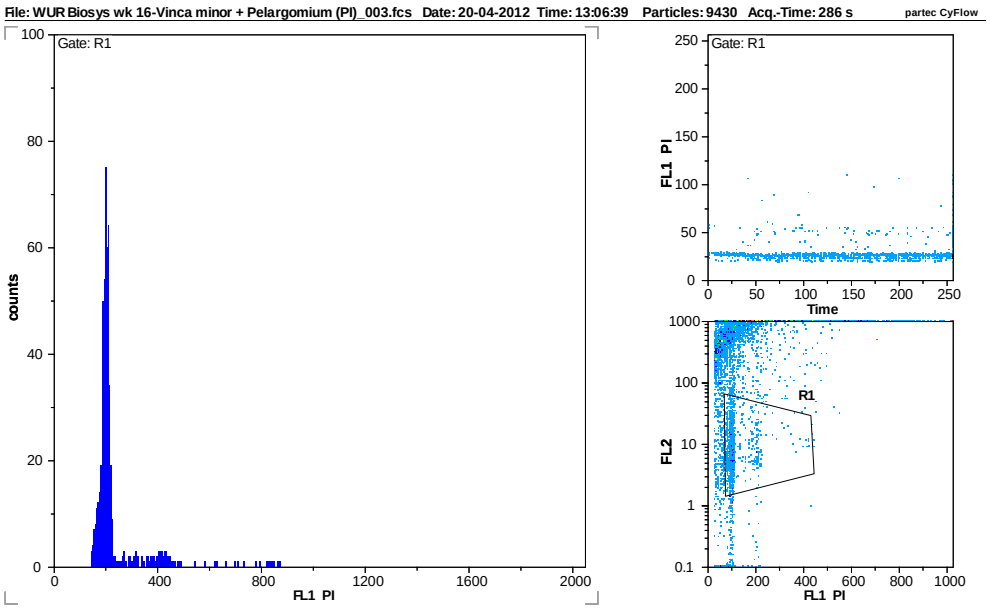
Region	Ungated	%Gated	Mean-x	CV-x%	Mean-y	CV-y%
R1	3583	60.85	139.54	53.30	16.46	101.82
RN1	1598	21.88	122.94	5.67	-	-
RN2	1128	24.64	207.24	6.49	-	-
RN3	1270	27.74	379.59	4.66	-	-

Sample: *Pelargonium bowkerii* (2C = 0.90 + 2.77 pg)



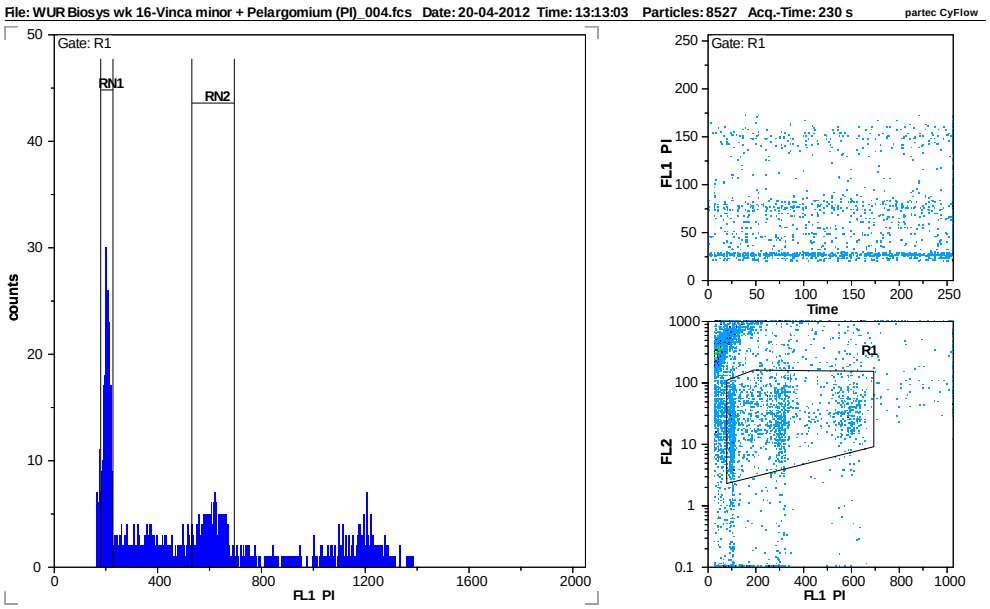
Region	Ungated	%Gated	Mean-x	CV-x%	Mean-y	CV-y%
R1	3523	36.03	153.68	37.40	16.73	106.45
RN1	2163	37.98	199.25	5.50	-	-
RN2	3179	48.82	370.93	4.30	-	-

Sample: *Pelargonium bowkerii* (2C = 2.81 pg)



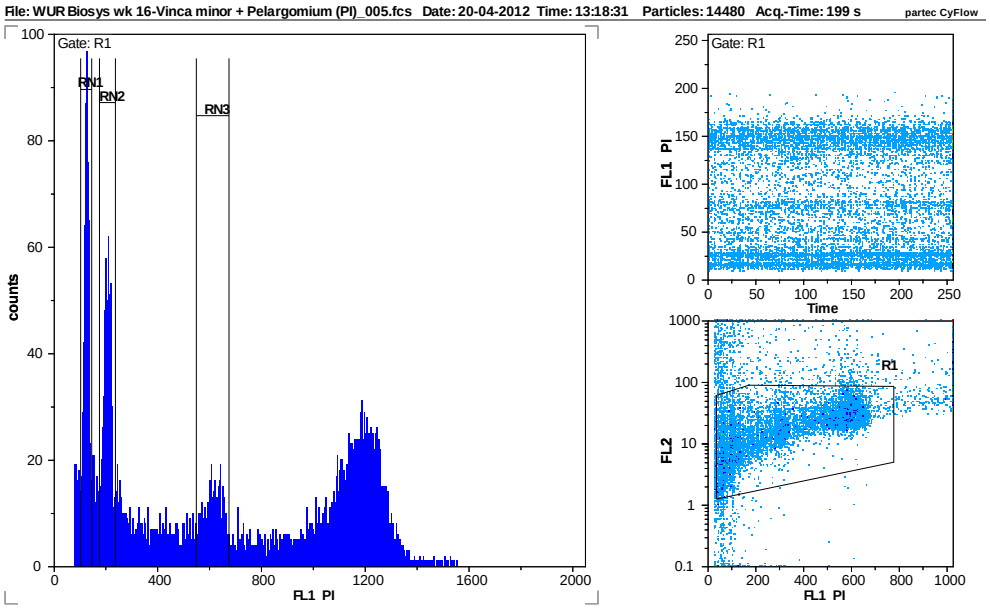
Region	Ungated	%Gated	Mean-x	CV-x%	Mean-y	CV-y%
R1	1971	20.90	107.99	34.16	10.75	90.50

Sample: *Pelargonium exstipulatum* (2C = 1.39 pg)



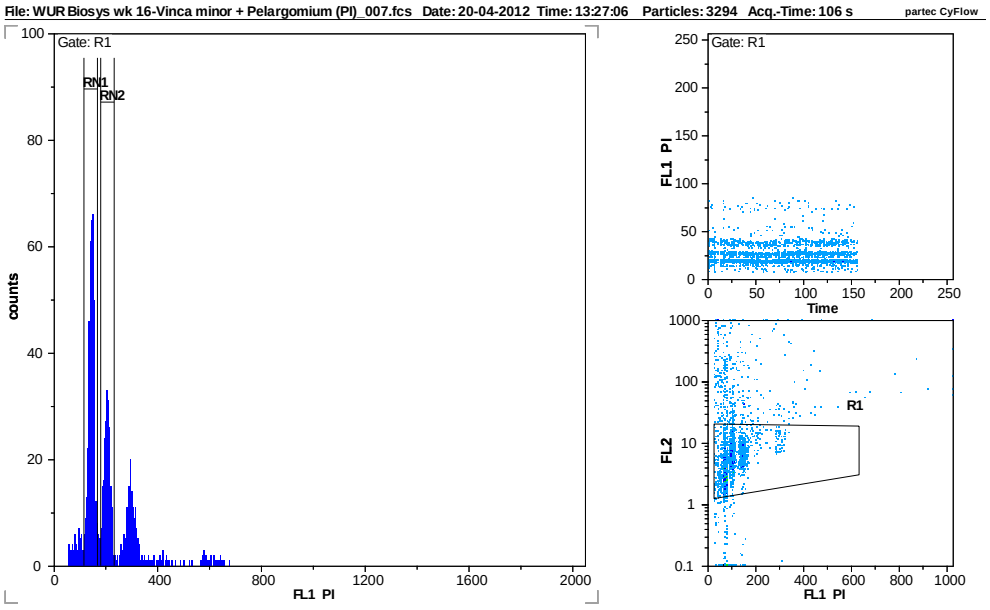
Region	Ungated	%Gated	Mean-x	CV-x%	Mean-y	CV-y%
R1	1835	21.52	247.23	69.75	33.99	86.57
RN1	1228	35.53	205.95	5.25	-	-
RN2	541	20.98	612.79	6.19	-	-

Sample: *Pelargonium peltatum* (2C = 4.53 pg)



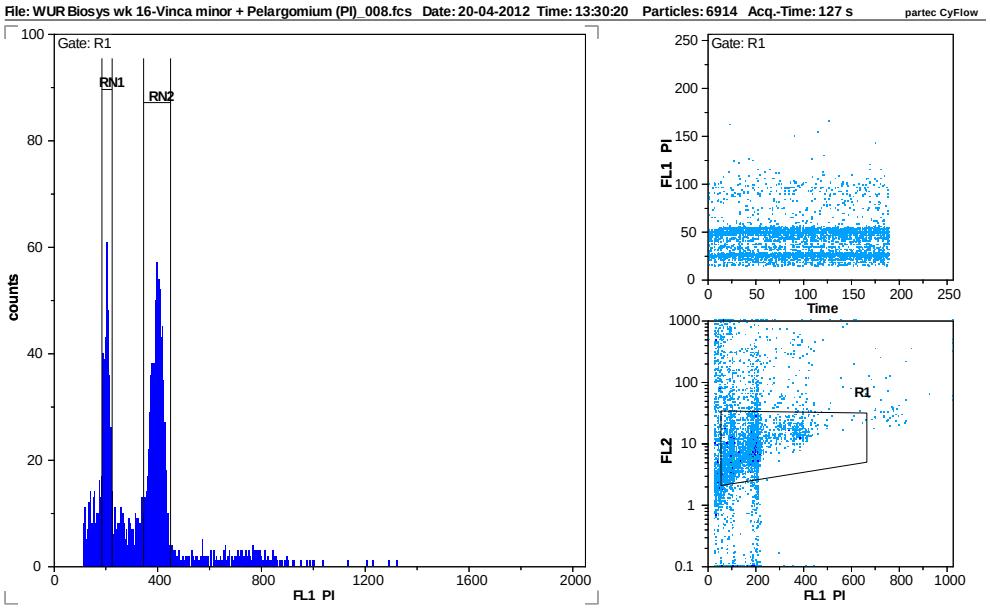
Region	Ungated	%Gated	Mean-x	CV-x%	Mean-y	CV-y%
R1	12297	84.92	318.82	70.26	20.14	73.27
RN1	2222	14.90	127.13	6.93	-	-
RN2	1892	14.14	207.31	6.46	-	-
RN3	1107	8.57	614.66	5.24	-	-

Sample: *Pelargonium graveolens* (2C = 3.91 pg)



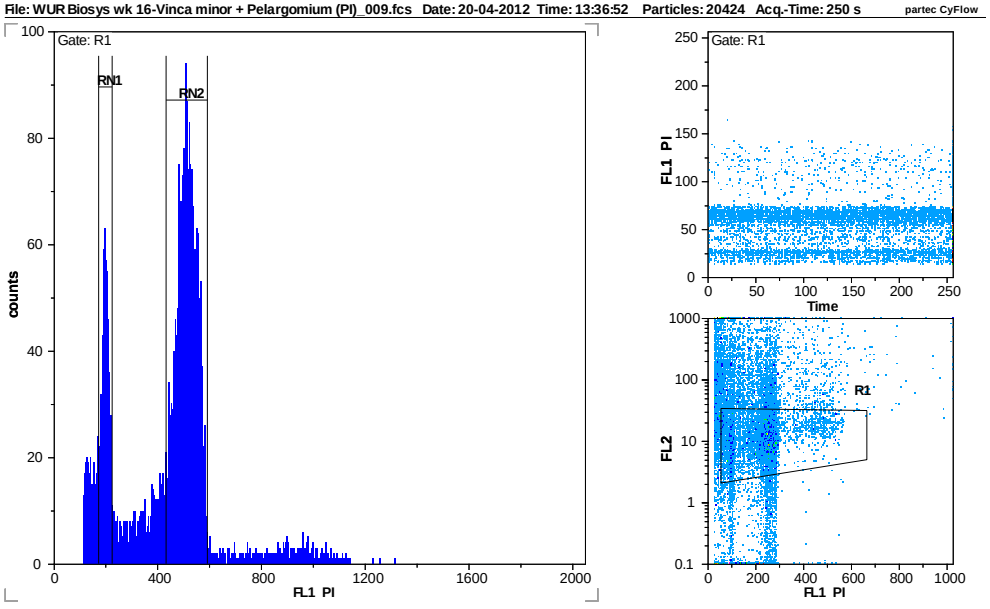
Region	Ungated	%Gated	Mean-x	CV-x%	Mean-y	CV-y%
R1	2742	83.24	98.70	44.19	5.61	63.87
RN1	1482	47.23	144.75	6.03	-	-
RN2	781	25.24	205.30	5.03	-	-

Sample: *Pelargonium litorale* (2C = 1.07 pg)



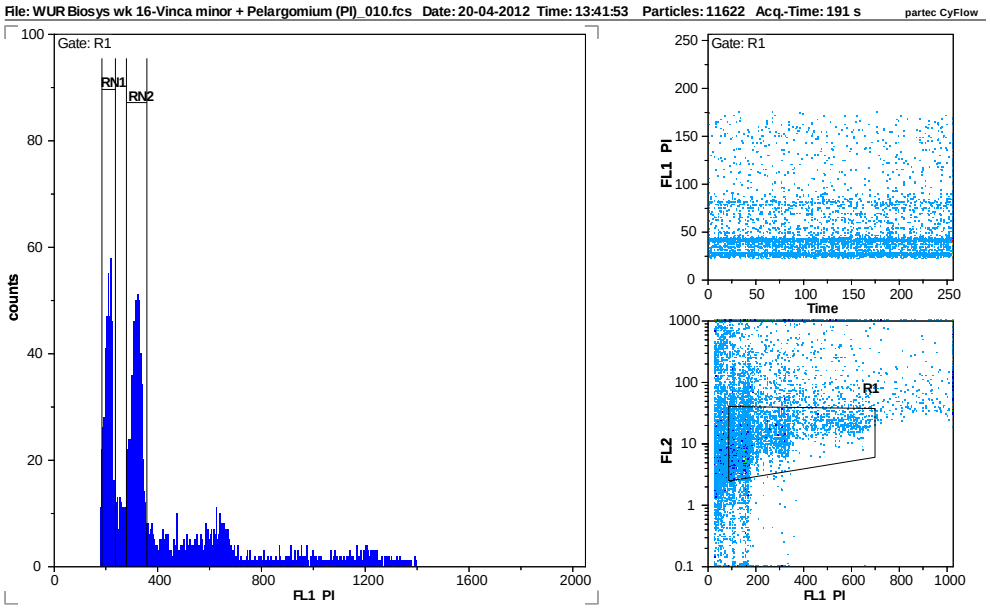
Region	Ungated	%Gated	Mean-x	CV-x%	Mean-y	CV-y%
R1	5059	73.17	168.18	42.24	8.84	59.38
RN1	1305	22.55	204.57	4.48	-	-
RN2	2778	47.95	397.00	5.32	-	-

Sample: *Pelargonium caucalifolium* (2C = 2.97 pg)



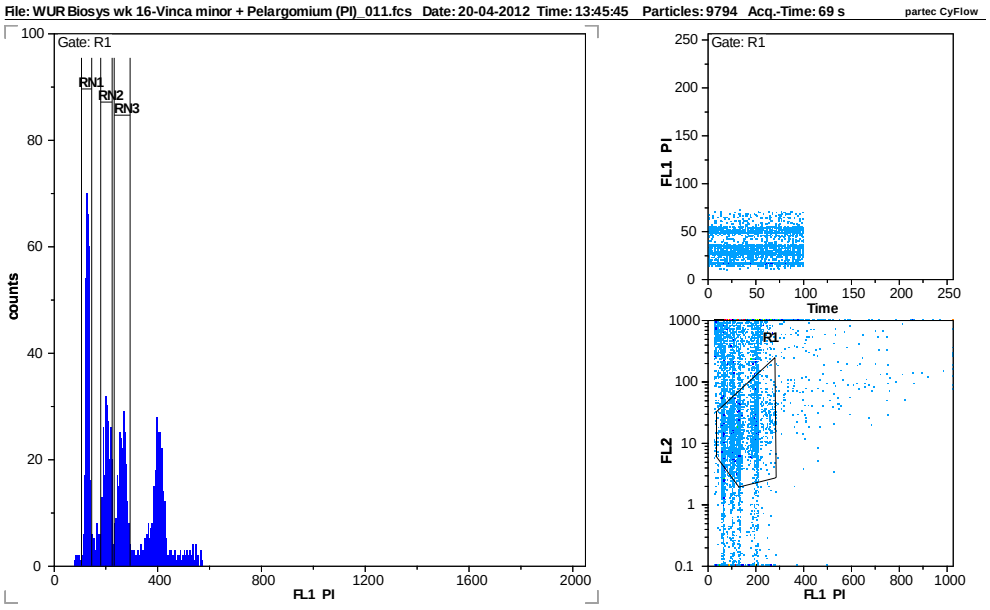
Region	Ungated	%Gated	Mean-x	CV-x%	Mean-y	CV-y%
R1	11603	56.81	219.92	39.68	12.52	57.28
RN1	2661	14.64	198.94	5.85	-	-
RN2	9713	60.92	514.49	6.87	-	-

Sample: *Pelargonium triste* (2C = 3.96 pg)



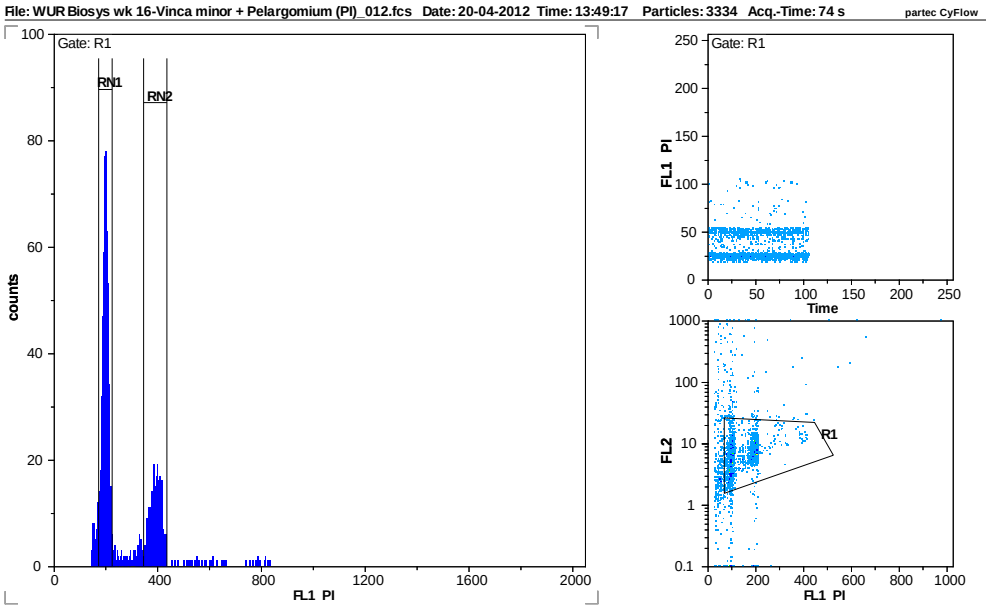
Region	Ungated	%Gated	Mean-x	CV-x%	Mean-y	CV-y%
R1	5606	48.24	194.53	63.09	11.83	69.17
RN1	2053	27.79	211.62	5.56	-	-
RN2	2649	35.52	319.45	5.45	-	-

Sample: *Pelargonium inquinans* (2C = 2.26 pg)



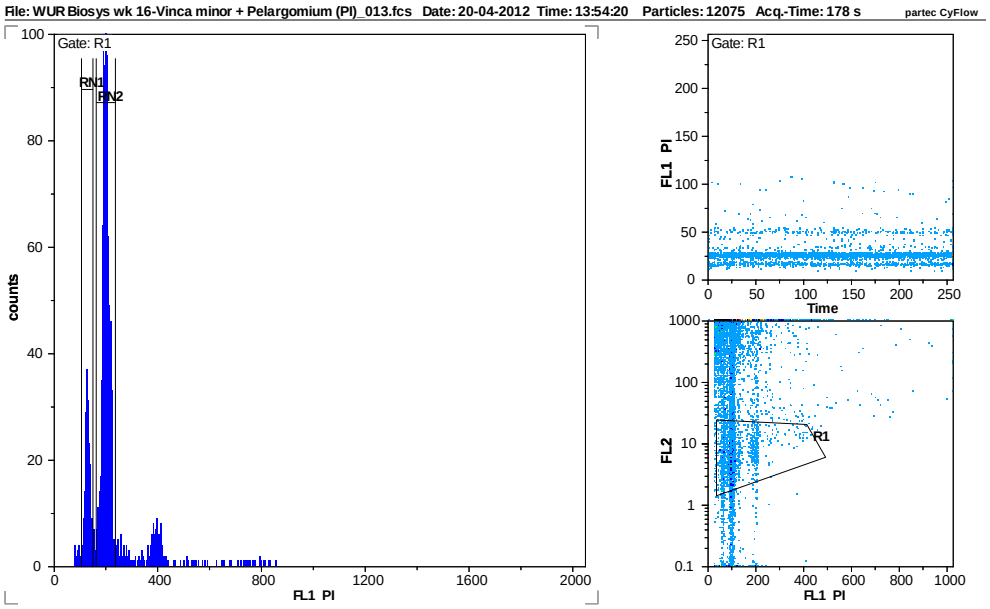
Region	Ungated	%Gated	Mean-x	CV-x%	Mean-y	CV-y%
R1	3847	39.28	128.33	44.07	19.65	94.13
RN1	2992	26.51	130.48	4.77	-	-
RN2	1493	18.90	206.14	5.25	-	-
RN3	1449	21.00	264.02	5.11	-	-

Sample: *Pelargonium dasyphyllum* (2C = 0.96 + 2.94 pg)



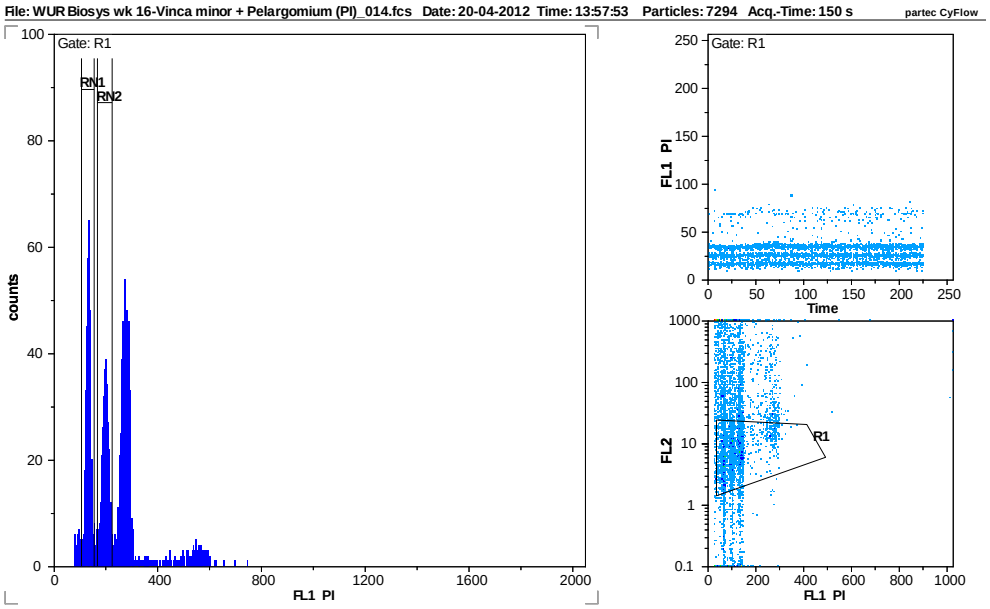
Region	Ungated	%Gated	Mean-x	CV-x%	Mean-y	CV-y%
R1	2793	83.77	131.21	40.47	6.73	55.16
RN1	1809	60.44	198.29	5.08	-	-
RN2	789	26.82	395.43	4.85	-	-

Sample: *Pelargonium fulgidum* (2C = 1.51 pg)



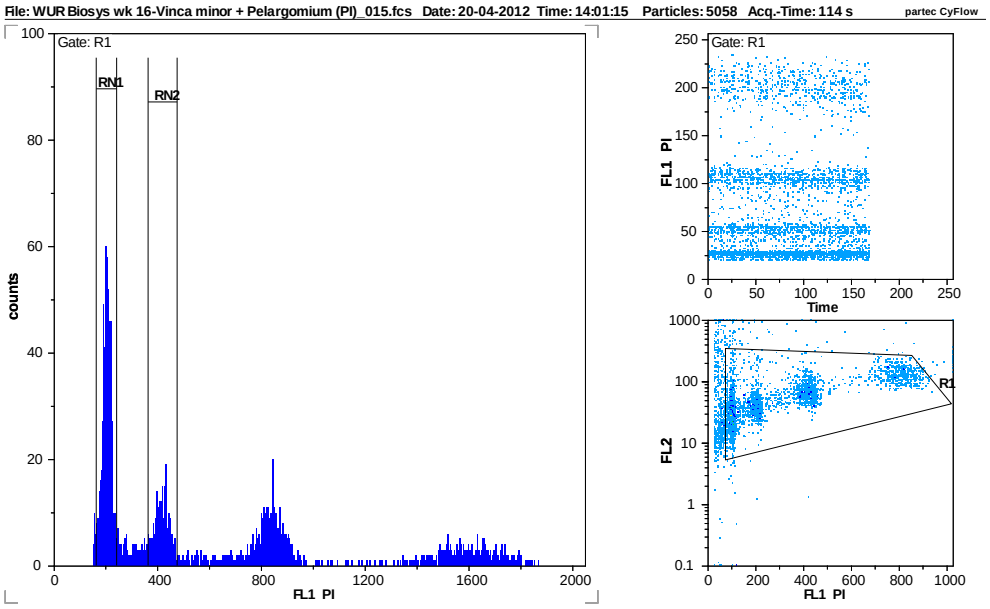
Region	Ungated	%Gated	Mean-x	CV-x%	Mean-y	CV-y%
R1	3852	31.90	104.48	38.28	7.29	67.13
RN1	2020	15.76	130.41	6.49	-	-
RN2	5637	70.79	201.43	6.17	-	-

Sample: *Pelargonium appendiculatum* (2C = 0.97 pg)



Region	Ungated	%Gated	Mean-x	CV-x%	Mean-y	CV-y%
R1	4238	58.10	113.21	40.76	8.38	60.22
RN1	2033	26.57	133.97	6.22	-	-
RN2	1460	23.24	199.76	5.73	-	-

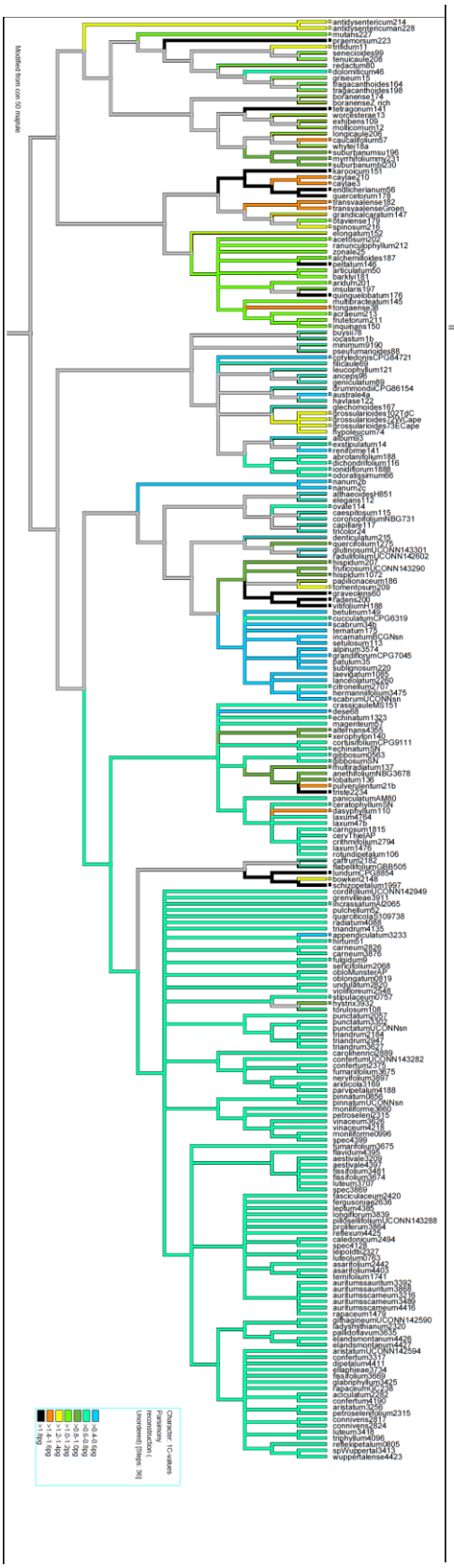
Sample: *Pelargonium xerophyton* (2C = 1.01 pg)

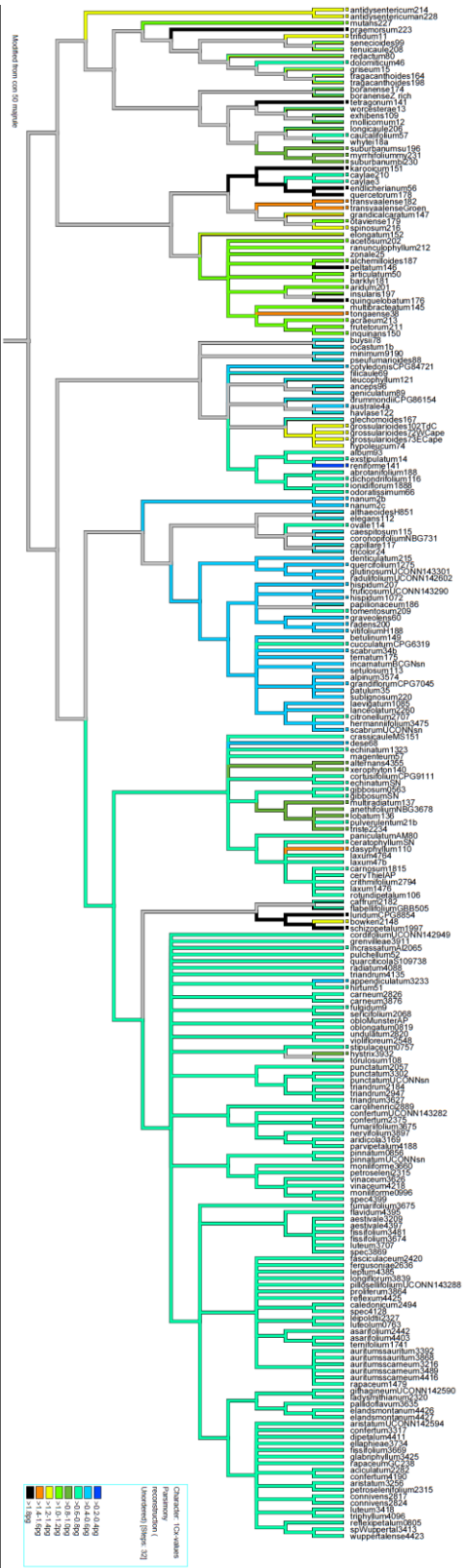


Region	Ungated	%Gated	Mean-x	CV-x%	Mean-y	CV-y%
R1	4358	86.16	279.32	83.19	58.22	77.14
RN1	1888	42.36	204.83	7.37	-	-
RN2	658	14.98	416.99	5.75	-	-

Sample: *Pelargonium tetragonum* (2C = 3.09 pg)

Appendix 6 – Discrete Character Optimizations (Square Trees)





Appendix 7 - R Scripts

Continuous optimization and randomized control of node values:

```
#Install matrixStats
install.packages("matrixStats")

#Read libraries
library(ape)
library(geiger)
library(phytools)
library(matrixStats)

setwd("N:/Common/Mathijs Nieuwenhuis/data")

#Read Trees
trees <- read.nexus("BeastMaxCladeCred86M200trees.t")

#Read Data
d <-read.csv("GenomeSize..csv")
GenomeSize <-d$DNA.Content
names(GenomeSize) <- d$Species

#Check names
name.check <- list()
all <- list()

name.check <- name.check(trees, GenomeSize)



#Delete unsampled taxa
unsampled <- name.check$Tree.not.data

treesNew <- list()

treesNew <- drop.tip(trees, unsampled)
trees <- treesNew

#Check names
name.check <- list()
all <- list()

name.check <- name.check(trees, GenomeSize)

#lambda transformations
trees <- lambdaTree(trees,0.5)

#Ancestral State Reconstruction 1
aceREML <- list()
aceRes  <- NULL

aceREML<- ace(GenomeSize, trees, type="continuous", method="REML")
aceRes <-cbind (aceRes, aceREML$ace)
```

```
#Ancestral State Reconstruction 2 (Randomizations)

  #Empty list
aceREML2 <- list()
  #Empty matrix
aceRes2 <- NULL
  #1000 repetitions
for (i in 1:1000) {
  #Randomization of data (taxon names only)
names(GenomeSize)<-sample(names(GenomeSize))
  #ace Restricted Maximum Likelihood
aceREML2[[i]] <- ace(GenomeSize, trees, type="continuous", method="REML")
  #Add nodes as column to matrix, node 65 is basal node
aceRes2<-cbind(aceRes2, aceREML2[[i]]$ace)
}

#Now we want to compare the node values of the original analysis to those
of the randomized dataset, to see if the values differ significantly from
te null-distribution.

aceP<-NULL
aceM<-NULL
for (i in 1:length(aceRes)) {
aceP<-c(aceP,sum(aceRes[i]>aceRes2[i,])/length(aceRes2[i,]))
aceM<-c(aceM,sum(aceRes[i]<aceRes2[i,])/length(aceRes2[i,]))
}
acePvals<-cbind(aceP,aceM)*2
rm(aceP,aceM)

aceMPvals<-apply(acePvals,MARGIN=1, min)

plot(trees)
a <-(round(aceREML$ace, digits=2))
p <-(paste("p=",round(aceMPvals>=0.05, digits=2)))
nodelabels(a, adj=c(1.2, -0.5))
nodelabels(p, adj=c(1.2,1), frame="n", font=3)

aceMPvals<0.05
p <- character(length(aceMPvals))
p[aceMPvals>=0.05] <- "white"
p[aceMPvals<0.05] <- "red"

plot(trees)
nodelabels(pch=21, bg=p)


#Calculate node means
NodeMeans <- rowMeans(aceRes2)
NodeSds <-rowSds(aceRes2)
```

Parsimony and Neighbour Joining barcoding tree analysis:

```
#Load Packages
library("phangorn")
library("ape")

#Read data
Pelargonium = read.phyDat("202TrnLFPelargoniumMAFFT.cgi.phy",
format="phylip", type="DNA")

#Create distance matrix
dm=dist.dna(as.DNAbin(Pelargonium))

#create UPGMA and NJ trees
treeUPGMA = upgma(dm)
treeNJ = NJ(dm)

#view trees
layout(matrix(c(1,2)), height=c(1,1.25))
par(mar=c.1,.1,.1,.1))
plot(treeUPGMA, main="UPGMA", cex=0.8)
plot(treeNJ, "unrooted", main"NJ", cex=0.5)

#Obtain parsimony score
parsimony(treeUPGMA, Pelargonium)
parsimony(treeNJ, Pelargonium)

#Find most parsimonious tree
optParsUPGMA = optim.parsimony(treeUPGMA, Pelargonium)
optParsNJ = optim.parsimony(treeNJ, Pelargonium)

#View trees
plot(optParsUPGMA, main="UPGMA", cex=0.8)
plot(optParsNJ, "unrooted", main="NJ", cex=0.5)

#export tree in newick format
```

Appendix 8 – Online Resources

8.1 Accesion photos

8.2 Large tables

8.3 Statistical data