# MASTERARBEIT | MASTER'S THESIS

Titel | Title

Evolution of tandem repeats and pungency of *Capsicum* (Solanaceae) species of the Pubescens clade

verfasst von | submitted by

Laura Pflügl BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2025

| | |
|---|---|
| Studienkennzahl lt. Studienblatt \| Degree programme code as it appears on the student record sheet: | UA 066 832 |
| Studienrichtung lt. Studienblatt \| Degree programme as it appears on the student record sheet: | Masterstudium Botany |
| Betreut von \| Supervisor: | Univ.-Prof. Mag. Dr. Hanna Schneeweiss Privatdoz. |

# Acknowledgements

First and foremost, I would like to thank my supervisor, Assoz. Prof. Mag. Dr. Hanna Schneeweiss, Privatdoz., for giving me the opportunity to be part of the Plant Evolutionary Cytogenetics research group and of the *Capsicum* project. I am deeply grateful for your guidance, your support during all the struggles I encountered in the lab and when writing the thesis, and for always having an open ear for me.

My sincere thanks also go to Rosi, Rebecca, and Vildana, for the great time working together in the lab and office, and for always being there with advice and support in many different situations.

I would also like to thank Mag. Dr. Johann Schinnerl for allowing me to conduct my analyses in the Phytochemistry Lab and for your valuable guidance throughout this project.

Finally, I would like to thank my family and friends for their constant support and love during this thesis and throughout my studies. I am truly grateful to have you all by my side.

# Contents

# 1. Introduction

## 1.1 Genus *Capsicum*

The dicot genus *Capsicum* belongs to the nightshade family (Solanaceae). This family encompasses genera that are economically important with cultivated species like *Solanum lycopersicum* L. (tomato) or *Solanum tuberosum* L. (potato), medicinally important taxa (*Atropa bella-donna* L., *Nicotiana tabacum* L.), many of which are used as model organisms in research (Gebhardt, 2016; Motti, 2021). The genus *Capsicum* currently comprises 43 accepted species (Figure 1), five of which are grown world-wide: *Capsicum annuum* L., *Capsicum baccatum* L. with three varieties, *Capsicum chinense* Jacq., *Capsicum frutescens* L. and *Capsicum pubescens* Ruiz & Pav.

These species are cultivated for products such as chilies, jalapeños, and bell peppers (Barboza et al., 2022). Their characteristic pungency is due to the organic compound capsaicin, which is valued not only as a component of food and spices, but also for its medicinal properties. Capsaicin has been associated with pain relief, thermoregulation, and various biological activities, including antioxidant and antimicrobial effects. (Sun et al., 2007; Adaszek et al., 2019; Hernández-Pérez & Gómez-García, 2020).

The natural distribution range of *Capsicum* spans an area from the southern United States and Mexico to Brazil and Argentina, with its highest diversity in the Andean region. It is hypothesized that *Capsicum annuum* was originally domesticated in the highlands of Mexico, while *Capsicum chinense* and *Capsicum frutescens* were domesticated in the lowland tropics of South America (W. H. Eshbaugh, 1993). These three species form a complex group that is cultivated worldwide and includes both hot and sweeter, less pungent chilies, the latter mainly consumed in Europe and Asia. The other two domesticated species, *Capsicum baccatum* and *Capsicum pubescens*, remain mainly restricted to Latin America. *Capsicum baccatum* can be found throughout the lowlands of Central and South America, whereas *Capsicum pubescens* grows naturally in the highlands and is primarily found at intermediate elevations in the Andes (Figure 2) (W. H. Eshbaugh, 1993; Pickersgill, 1997; W. Eshbaugh, 2012; Barboza et al., 2022). Archaeological evidence suggests that humans began gathering and consuming *Capsicum* fruits around 10,000 years ago (10,285–10,578 B.P.), with the earliest evidence from Peru indicating that these fruits probably originated from the Andean region (Dillehay et al., 2017).

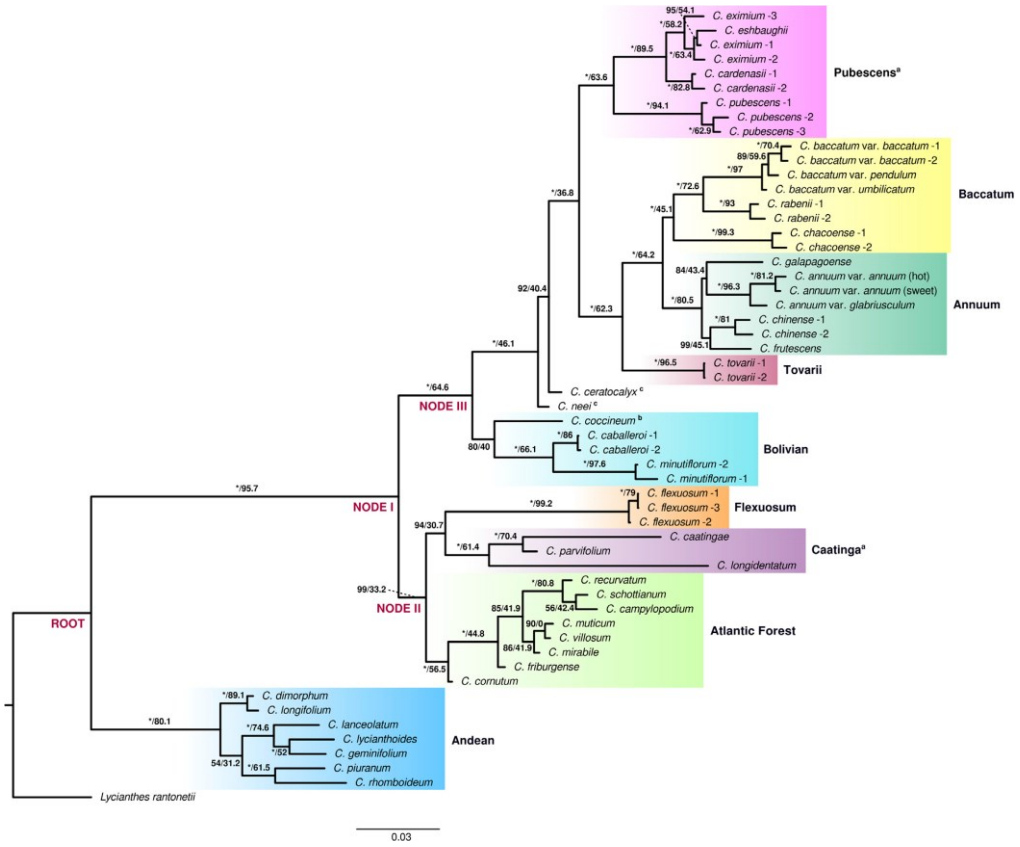



**Figure 1.** Maximum likelihood phylogenetic tree for *Capsicum*. Clades identified are shaded with different colors and labeled; a clades newly circumscribed, b provisional placement, c species not assigned to any clade. source: Carrizo García et al., 2022.

Residue analysis of pottery from the Middle to Late Preclassic periods (400 BCE to 300 CE), revealed that *Capsicum* was already used as a spice in Mesoamerican and South American cultures. Although *Capsicum* had been grown in America for thousands of years, chili peppers were introduced to Spain and Portugal much later by Christopher Columbus following his voyages in the late 15th century, and from there, they gradually spread throughout the Old World (Powis et al., 2013; Barboza et al., 2022).

### 1.1.1 Morphology and Ecology

Most *Capsicum* species grow as perennial shrubs, although exceptions exist within the *Capsicum annuum* complex, with some biennial herbaceous taxa. The leaves are typically simple in morphology, green in color, and ovate to elliptical in shape. The flowers are actinomorphic and solitary, commonly arising from the axils of the leaves. The corolla is generally stellate or rotate in form, with petals that are most often white or purple (Figure 8). While corolla colour was once used to distinguish between wild and cultivated species in a "white-" and a "purple-flowered groups", this method is no longer considered reliable for taxonomic classification (Hunziker, 2001; Moscone et al., 2006; Barboza et al., 2022). The vast

majority of species in the genus *Capsicum* are capable of self-pollination, with exceptions in the Pubescens clade, particularly in *Capsicum cardenasii* and certain accessions of *Capsicum pubescens*. Bees have been identified as effective pollinators of *Capsicum*, enabling also cross-pollination (Onus & Pickersgill, 2004; Palma et al., 2008; Barboza et al., 2022)

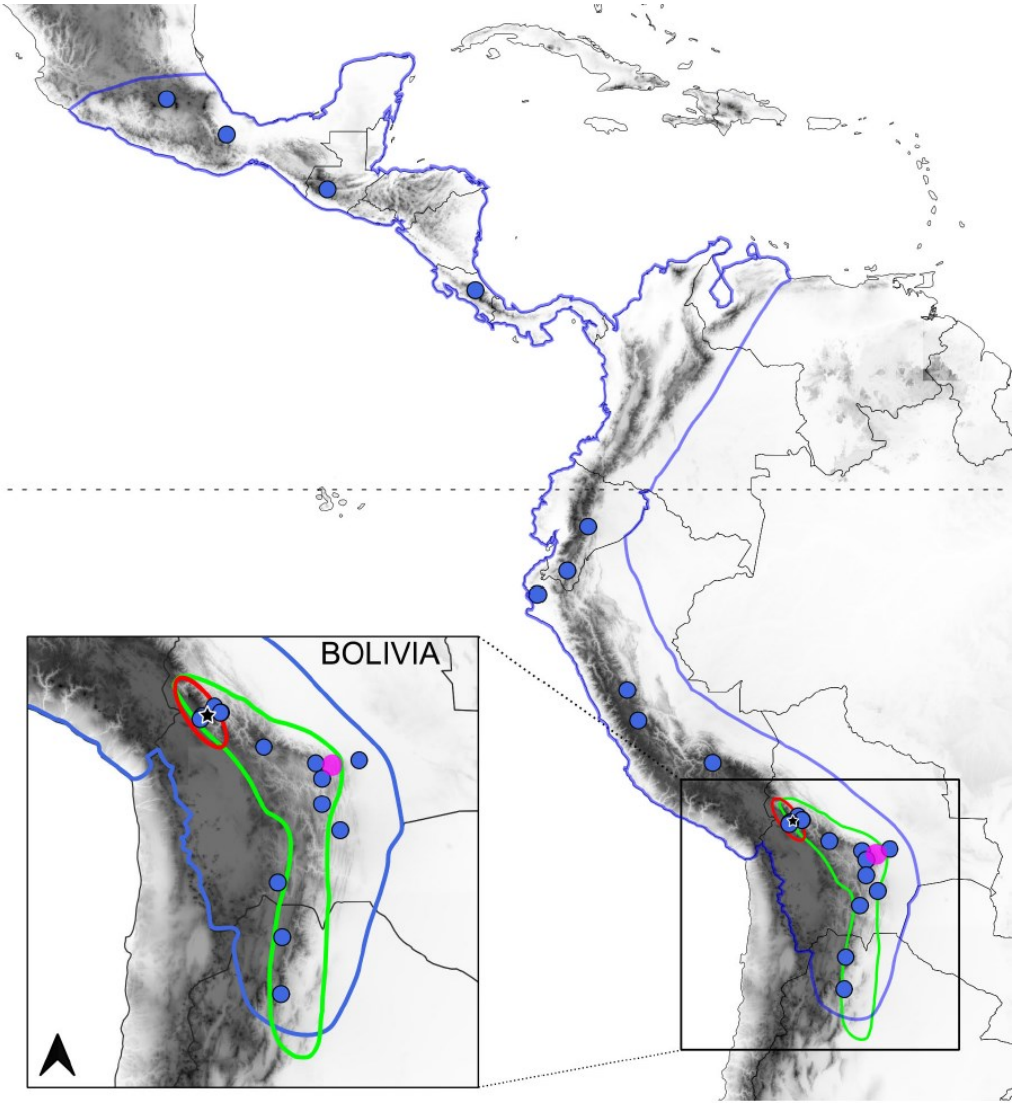



**Figure 2.** Geographic distribution of *Capsicum* species. Area of *C. pubescens* cultivation in Central-South America (blue) with the details of the native ranges of distribution of *C. eximium* (green), *C. cardenasii* (red), and *C. eshbaughii* (fuchsia). Adapted from Palombo et al., 2024.

The fruits of wild species are usually round berries, although they can also be ovoid or elliptical. Their colours range from red to yellowish-green, and occasionally orange. In cultivated species, the shapes and colours are highly variable, and the fruit is usually much bigger, elongated, cylindrical or rounded (Figure 9). Due to human-assisted selection, additional colour variants of fruits have been created, such as purple, black, white to pale yellow.

### 1.1.2 Cultivation and fruit pungency

The characteristic fruit pungency of *Capsicum* is a derived character trait, absent in species from the Andean clade, the oldest lineage within the genus. It is caused by secondary metabolites known as capsaicinoids, and primarily controlled by a single gene, *Pun*1*,* which codes for a putative acyltransferase enzyme (also called CS = capsaicin synthase), involved in the biosynthesis of capsaicinoids (Stewart Jr et al., 2005; Chen et al., 2024). A mutation, specifically a large deletion of 2.5 kB in the promoter and first exon region of this gene locus results in secondary loss of pungency, for example in fruits of *Capsicum annuum* (Stewart et al., 2007). The most common capsaicinoid compounds are capsaicin (8-methyl-*N*-vanillyl-6-nonenamide), dihydrocapsaicin (8-methyl-*N*-vanillylnonanamide) and nordihydrocapsaicin (*N*-vanillyl-7-methyloctanamide) (Figure 3) produced in the placental epidermis of the fruits, where their accumulation often results in blister-like structures (Tripodi & Kumar, 2019).

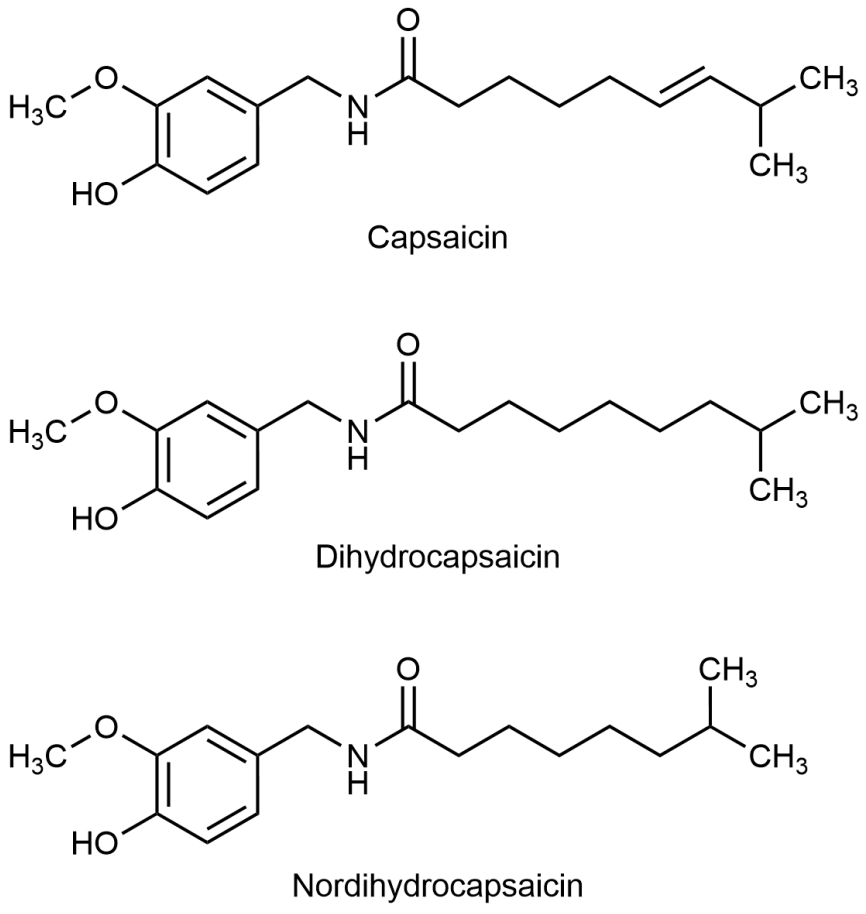



**Figure 3.** Structural formulae of capsaicin (8-methyl-*N*-vanillyl-6-nonenamide), dihydrocapsaicin (8-methyl-*N*-vanillylnonanamide) and nordihydrocapsaicin (*N*-vanillyl-7-methyloctanamide).

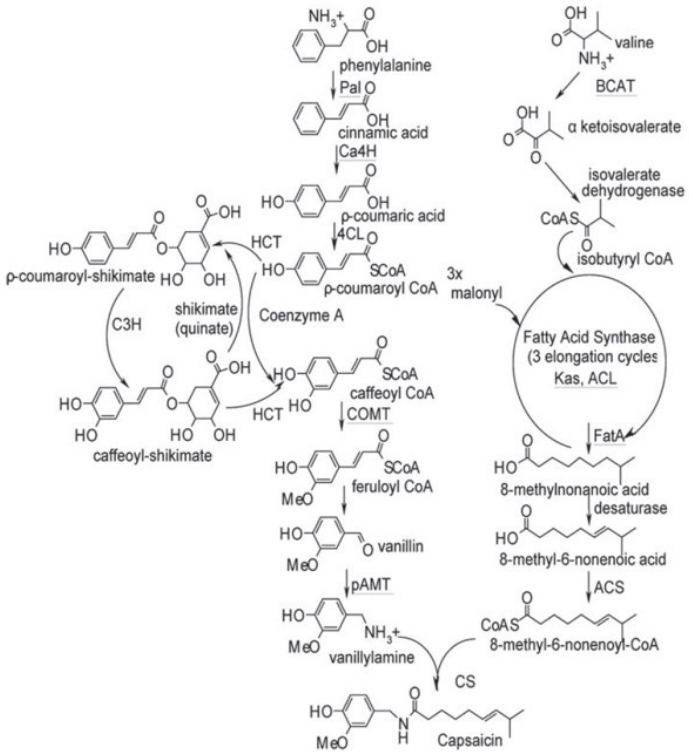



**Figure 4.** Capsaicinoid biosynthetic pathway. Adapted from Stewart et al., 2007.

The capsaicinoids of *Capsicum* fruits cause a strong sensory reaction in mammals due to the presence of pain receptors, whereas birds are insensitive to the pain these compounds cause during digestion. This differential response is ecologically significant: capsaicinoids function as a chemical deterrent to mammalian seed predators, as mammals often damage or destroy seed viability during digestion. In contrast, birds transport seeds over longer distances without harming them and reducing their viability and are therefore important for seed dispersal (Tewksbury & Nabhan, 2001; Jordt & Julius, 2002). All wild *Capsicum* species typically produce small, red fruits with a soft stem at maturity, facilitating effortless removal by birds, strongly indicating avian dispersal as their primary, natural seed dispersal mechanism (DeWitt & Bosland, 2009). In addition to deterring mammalian herbivores, capsaicinoids also serve as a chemical defense mechanism against pathogens such as insects and fungi. Capsaicin has been shown to reduce seed infections caused by the well-known agricultural pathogen *Fusiarum* (Tewksbury et al., 2008). Pungency is traditionally measured using the Scoville Heat Unit (SHU) scale, a subjective organoleptic dilution-based taste test, although high-performance liquid chromatography (HPLC) is now more commonly used for more precise capsaicin quantification. Nevertheless, the results are often afterwards converted into SHU, since this scale is widely recognized and remains the familiar way of expressing pungency levels to both

consumers and breeders (Scoville, 1912; Bosland & Baral, 2007). Capsaicinoid levels can vary, even throughout plant development. Capsaicin synthesis typically begins a few days after pollination, increases during fruit maturation and may decline abruptly at later stages (Figure 5). This expression pattern can vary among cultivars and is highly influenced by environmental factors such as water availability, light intensity, temperature and nutrient levels (Park & Kim, 1975; Gurung et al., 2011; Phimchan et al., 2012; Zhang et al., 2016; Naves et al., 2019; Tripodi & Kumar, 2019).

The cultivation and breeding of *Capsicum* species focusses on increasing the fruit yield and developing larger fruits with fewer seeds. Selective breeding results in traits characteristic of domestication syndrome, such as fruits that easily detach, specific colours and fruit shapes and a wide range of pungency levels, from completely non-pungent (sweet peppers) to highly pungent. A notable example is “Trinidad Moruga Scorpion”, a variety of *Capsicum chinense*, one of the most pungent landraces identified to date (Bosland et al., 2012). As a result of domestication, natural seed dispersal mechanisms in *Capsicum* have been suppressed, alongside the reduction of chemical defenses such as capsaicinoid production, which are essential for survival of wild populations (Luna-Ruiz et al., 2018; Barboza et al., 2022). From an agronomic perspective, breeding programmes also focus on enhancing resistances of the varieties to abiotic and biotic stress. This includes tolerance to drought and heat, as well as resistance to fungal pathogens and viral infections (Aloni et al., 2001; Barchenger et al., 2019; Tripodi & Kumar, 2019).

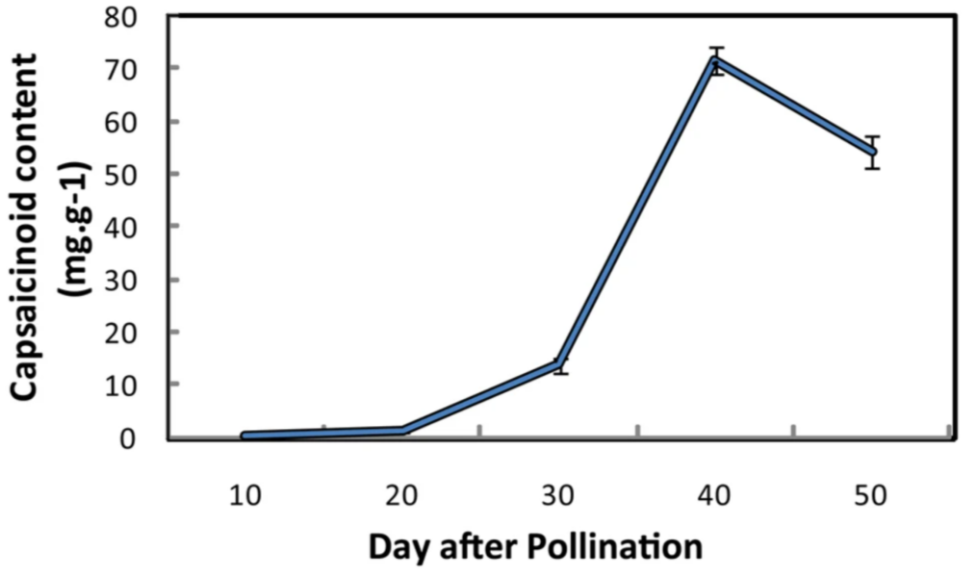



**Figure 5.** Capsaicinoid content of fruit placenta at different development stages from *Capsicum pubescens* (‘Guijiangwang’). Adapted from (Zhang et al., 2016).

### 1.1.3 Evolution and Cytogenetics

All *Capsicum* species are exclusively diploid. Two base chromosome numbers are known, either *x* = 12 or *x* = 13 (Moscone et al., 2006). The currently accepted hypothesis, supported by phylogenetic data, is that the *x* = 12 has been an ancestral chromosome number and *x* = 13 represents ascending dysploidy and has evolved independently in two different lineages (Carrizo García et al., 2022). The chromosome number *x* = 13 is found only in Andean and Atlantic forest clades, which consist of exclusively wild species. In contrast, all domesticated species, including *Capsicum pubescens*, possess a base chromosome number of *x* = 12. Furthermore, the chile pepper species vary significantly (nearly 5-fold) in their genome sizes due to the high content of repetitive DNA, up to 81 % of the genome (Qin et al., 2014). This genome fraction encompasses dispersed and tandem repeats, the latter notably including satellite DNAs. All the repeats contribute to a different extent to chromosomal differentiation. Variation in karyotype structure, genetic and genomic diversification, and restructuring point to a relatively rapid diversification that took place in the ancestral distribution area of the genus in South America. The diversification followed a broad clockwise migration across South and Central America, spanning from the Andes to Brazil and back before reaching Mesoamerica. During this period nine clades evolved: Andean, Atlantic Forest, Caatinga, Flexuosum, Bolivian, Tovarii, Annuum, Baccatum, Pubescens (Carrizo García et al., 2016, 2022).

Cytogenetic studies in *Capsicum* have provided insights into genome organization, size variation, heterochromatin abundance and distribution and repetitive DNA content. Genome size measurements (1C-value) using flow cytometry, show high variation within the genus, but also some interspecific variation ranging from 1.6 pg in *Capsicum lanceolatum* to 7.8 pg in *Capsicum flexuosum* (V. Kapetanowic et al., unpubl.). Heterochromatin content also varies among the genomes from 1.72 % in *Capsicum chacoense* to 16.82 % in the largest genome of *Capsicum flexuosum* (Scaldaferro et al., 2013). Most *Capsicum* species have GC-rich heterochromatin, although exceptions with AT-rich heterochromatin have been found, notably in *Capsicum pubescens* (Moscone et al., 2006; Scaldaferro et al., 2013; Scaldaferro & Barboza, 2023). Ribosomal DNA loci, including both 35S (“Nucleolus organizing regions” = NOR) and 5S rRNA genes, have been mapped in several *Capsicum* species, including wild and cultivated ones, using fluorescent *in situ* hybridization (FISH) and their activity has been evaluated via AgNOR banding (Moscone et al., 2006; Scaldaferro et al., 2016). Typically, the

amount of mapped 45S rDNA loci ranged between one and up to 30 pairs, in the diploid set of chromosomes, nevertheless just one locus for 5S rDNA was detected in all analyzed *Capsicum* species in that study. In all analyzed *Capsicum* species, the number of NORs was consistently two, with the exception of *Capsicum annuum* and *Capsicum rhomboideum*, where each displayed only one NOR (Moscone et al., 2006; Scaldaferro et al., 2016). Furthermore, phylogenetic analyses using RAD-sequencing have been used for the reconstruction of phylogenetic relationships within the genus (Carrizo García et al., 2016, 2022; Palombo et al., 2024). A total of 43 species were identified in the genus *Capsicum* through RAD-sequencing, which were divided into nine distinct clades. Also, *Capsicum pubescens* was placed in one clade with closely related *Capsicum cardenasii, Capsicum eshbaughii* and *Capsicum eximium*.

### 1.1.4 Pubescens clade

*Capsicum pubescens* derives its name from the distinctive hairy (pubescent) leaves that set it apart from other pepper species. It was first formally described by Ruiz and Pavon, who encountered it in cultivation in Peru (DeWitt & Bosland, 2009). The domestication of *Capsicum pubescens* is estimated to have occurred around 6,000 years ago in the Andean region of Bolivia (DeWitt & Bosland, 2009). The species is a cultigen known from cultivation and with no confirmed wild populations. It typically grows as a scrambling subshrub or shrub, up to four meters tall and can develop woody stems. Its flowers are characteristically purple and its pungent fruits range in colour from yellow to orange to red with brown-blackish seeds. It is primarily grown in the highlands of Peru and Bolivia, with its current distribution assumed to be shaped largely by human activity. It is commonly known as rocoto in Peru, locoto in Bolivia and manzano in Mexico (Barboza et al., 2022; Palombo & Carrizo García, 2022; Palombo et al., 2024).

*Capsicum pubescens* forms a clade with three other wild species: *Capsicum eshbaughii*, *Capsicum eximium* and *Capsicum cardenasii*, collectively known as the Pubescens clade. This group belongs to the *Capsicum* crown group together with the sister clades Annuum, Baccatum and Tovarii (Carrizo García et al., 2022). However, recent phylogenomic analyses suggest that Pubescens clade encompasses two distinct lineages or subclades: one consisting of *Capsicum pubescens* and the other of *Capsicum cardenasii* and very closely related *Capsicum eshbaughii* and *Capsicum eximium*. The Pubescens clade is genetically distinct and largely reproductively isolated from other *Capsicum* groups. Successful hybridization is mostly

limited to within this clade with reciprocal crosses between *Capsicum eshbaughii* and both *Capsicum eximium* and *Capsicum cardenasii* reported. Natural hybrids have also been observed between *Capsicum pubescens* and its wild relatives *Capsicum eximium* and *Capsicum cardenasii* (Pickersgill, 1997; Scaldaferro, 2020; Palombo et al., 2024). Hybridization involving the cultigen *Capsicum pubescens* and the three related wild taxa suggests some degree of genetic compatibility within the clade despite strong barriers to gene flow with species from other lineages (Palombo et al., 2024).

## 1.2 Repetitive DNA in plant genomes

Plant genomes vary considerably, not only in their sizes and chromosome numbers, but also in chromosomal structure and organization. Among the angiosperms, genome sizes vary by more than 800 times between one of the smallest known genome in flowering plants of *Genlisea aurea* with roughly 2.1 Mbp and to the massive genome of *Paris japonica* with approximately 149 Gbp (Greilhuber et al., 2006; Pellicer et al., 2010; Weiss-Schneeweiss & Schneeweiss, 2013). Despite these differences, gene content remains relatively conserved across species, which supports the hypothesis that, in addition to polyploidy, genome size expansion is driven primarily by non-coding and repetitive elements rather than an increase in functional, protein-coding genes (= C-value paradox). The two main contributors to genome size variation and increase are whole genome duplications and the accumulation of repetitive DNA sequences, including both major groups of repetitive DNAs: dispersed retrotransposons and tandemly arranged satellite DNAs. These repeats not only play a role in changes of genome size in the evolution, but also influence genome structure in general, with repetitive DNAs being highly enriched in the centromeric and telomeric regions of chromosomes (Weiss-Schneeweiss et al., 2013; Pellicer et al., 2018).

### 1.2.1 Dispersed repeats

Dispersed repeats are DNA sequences with multiple copies scattered throughout the genome, rather than arranged in tandem arrays. The most abundant type of dispersed repeats in plant genomes are transposable elements (= TEs) encompassing DNA transposons, retrotransposons and helitrons (Almojil et al., 2021). These sequences have the ability to move within the genome via different mechanisms generally referred to as transposition. TEs, specifically DNA transposons, were first discovered by Barbara McClintock in *Zea mays* (maize) and were initially viewed as "controlling elements" due to their influence on gene expression

and chromosomal behavior (Ayarpadikannan & Kim, 2014). Because of their mobility they play a central role in shaping genome size, structure, and evolution in plants (McClintock, 1948; Pray & Zhaurova, 2008; Wells & Feschotte, 2020; Hayward & Gilbert, 2022). Transposable elements are classified into two major groups based on their transposition mechanism. Class I TEs or retrotransposons, transpose via a copy-and-paste mechanism. These elements are first transcribed into mRNA, which is then reverse transcribed back into DNAs and these new copies are inserted at new locations in the genome (Wicker et al., 2007). As a result, each transposition event massively increases the copy number of the element, contributing significantly to genome size expansion (Neumann et al., 2006; Weiss-Schneeweiss & Schneeweiss, 2013). Retrotransposons are further divided into non-LTR retrotransposons with LINEs (= Long Interspersed Nuclear Elements) and SINEs (= Short Interspersed Nuclear Elements), which lack long terminal repeats, and LTR retrotransposons, which have identical long terminal repeats (= LTRs) flanking their coding regions (Han, 2010; Wells & Feschotte, 2020).

LTR retrotransposons are especially abundant in plant genomes and are represented by two superfamilies: Ty1-*copia* and Ty3-*gypsy* (Kumar & Bennetzen, 1999). Both have similar structure with an LTR at each end, and a gag gene, which encodes a virus-like particle and a pol gene, the most conserved region of the element. The pol gene contains several genes for key enzymes responsible for transposition: protease (= PR), reverse transcriptase (= RT), RNase H (= RH), and integrase (= INT). The arrangement of these domains differs between the two superfamilies. In Ty3-*gypsy*, the typical order is PR–RT–RH–INT, whereas in Ty1-*copia* elements, the integrase domain appears earlier in the sequence, with the order PR–INT–RT–RH (Havecker et al., 2004; Benachenhou et al., 2013; Maxwell, 2020; Wells & Feschotte, 2020) (Figure 6). In contrast, non-LTR retrotransposons transpose via a similar RNA intermediate but lack long terminal repeats. These elements encompass autonomous LINEs that can range in size from a few hundred base pairs to up to roughly 9,000 base pairs. Despite lacking LTRs, they are capable of self-replication and movement because they carry genes for the essential enzymatic machinery for retrotransposition. Most LINEs contain an endonuclease and a reverse transcriptase, which enable them to cut the target DNA and reverse-transcribe their RNA into a new insertion site at a different genomic location (Han, 2010). Their ability to function independently makes LINEs important contributors to genome size expansion and structural variation, although they are generally less abundant in plant genomes than LTR

retrotransposons (Almojil et al., 2021). SINEs are short sequences, typically ranging from 100 to 400 base pairs, and are non-autonomous. They do not encode any functional proteins, including those required for their own transposition. Instead, SINEs need the enzymatic machinery of active LINEs to mobilize within the genome (Schmitz, 2012).

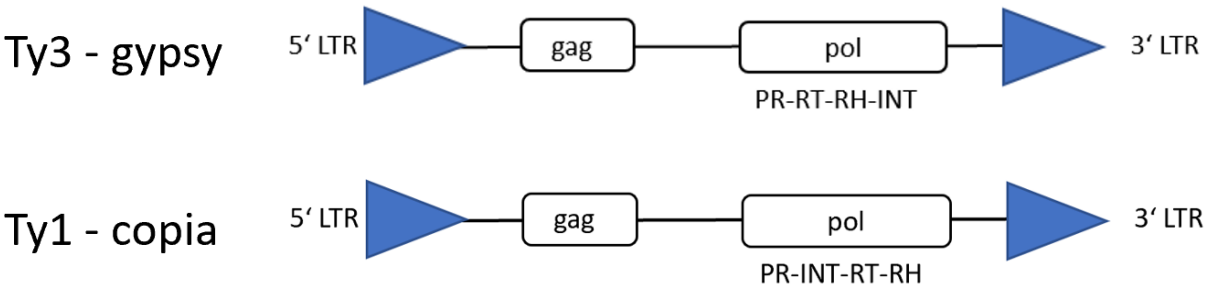



**Figure 6.** Schematic representation of the most common LTR-Retrotransposons, Ty3-gypsy and Ty1-copia. LTR = long terminal repeats, gag gene = group specific antigen, pol gene with PR = protease, RT = reverse transcriptase, RH = RNase H, INT = integrase.

Class II transposable elements, also known as DNA transposons, move via a general “cut-and-paste” mechanism. Unlike retrotransposons, their transposition does not involve an RNA intermediate. Instead, the element is excised from its original location and reinserted at a new chromosomal site, typically without increasing its copy number (Muñoz-López & García-Pérez, 2010). Despite this, DNA transposons can be found in high abundance in some plant genomes (Almojil et al., 2021). The most common type of DNA transposons in plants are those with terminal inverted repeats (= TIRs) and a gene encoding a transposase, the enzyme responsible for recognizing the TIRs and mediating the excision and reintegration process (Figure 7). These elements are structurally simple but functionally impactful, as their movement can lead to insertions, deletions, or chromosomal breaks. Another group within Class II transposons are the helitrons, which employ a “peel-and-paste” mechanism based on rolling-circle replication (Xiong et al., 2016). Rather than being excised as a double-stranded DNA fragment, helitrons transpose through a single-stranded DNA intermediate, initiated at a 3’ end that often includes a small hairpin structure (Figure 7). Helitrons can increase their copy number during transposition (Lal & Hannah, 2005). They typically encode for a helicase and a replication protein, and their activity involves Replication Protein A (= RPA), which binds to single-stranded DNA in eukaryotic cells. Notably, helitrons lack TIRs, distinguishing them structurally from many other DNA transposons.

### 1.2.2 Tandemly arranged repeats

Tandemly arranged repeats are very common in plant genomes, typically localized in specific chromosomal regions such as centromeres, pericentromeres, telomeres and interstitial heterochromatic blocks. Unlike dispersed repeats, which are scattered throughout the genome, tandem repeats are organized as adjacent copies of short DNA motifs. Those arrays are highly variable in size and sequence composition between species. There are two main types of tandem repeats: coding tandem repeats, such as ribosomal DNA (= rDNA) clusters and non-coding satellite DNAs (Jo et al., 2009).

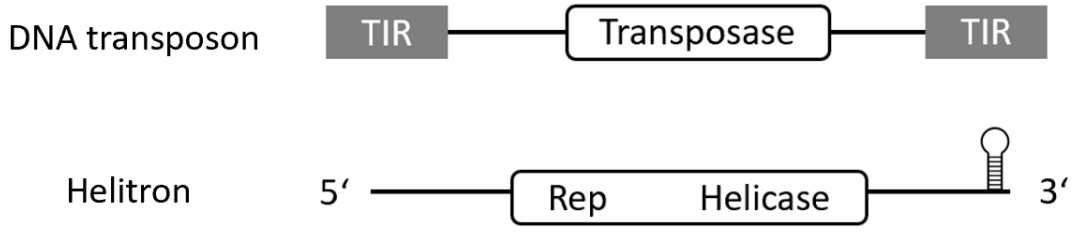



**Figure 7.** Schematic representation of the most common Class II transposable elements, DNA transposons and helitrons. TIR = Terminal inverted repeats. Rep = Replication protein.

rDNA consists of two highly conserved types of tandemly repeated genes: 5S rDNA and 35S rDNA (in some literatures also described as 45S rDNA). The 35S rDNA unit comprises the 18S, 5.8S and 25S rRNA genes separated by two different internal transcribed spacers (= ITS1 and 2) and surrounded by external transcribed spacer (= ETS) and intergenic spacer (= IGS) at both ends and is transcribed by RNA polymerase I as a single precursor (Hemleben et al., 2004). This pre-rRNA is later further processed and is spliced into the mature 18S, 5.8S and 25S rRNAs. Chromosomal regions carrying the 35S rDNA are also called “Nucleolus organizing regions” (= NORs) as their transcription is associated with the nucleolus. In contrast, 5S rDNA is transcribed independently by RNA polymerase III and mostly not linked to the position of 35S rDNA on chromosomes. These two rDNA types are often in different regions of the chromosome or on different chromosomes (Weiss-Schneeweiss & Schneeweiss, 2013) and their loci numbers also evolve independently. 5S rRNA genes are separated by non-transcribed spacers (= NTS) (Garcia et al., 2017; Rosselló et al., 2022). The 5S and 35S rDNA encode for structural RNA components of the small and large subunits of ribosomes, which are responsible for the translation of mRNAs into proteins in all living cells. Therefore, rDNA

repeats are found in all eukaryotic genomes with at least one locus present in every cell type ("house-keeping genes") (Garcia et al., 2012; Rosselló et al., 2022).

Due to the high copy number, tandem arrangement and highly conserved sequences, 35S and 5S rDNA loci are used as reliable cytogenetic markers in plants. These loci can be visualized in chromosomes using fluorescence *in situ* hybridization (FISH), enabling the identification of specific chromosomal regions or specific chromosomes and thus allowing for studies on karyotype evolution, chromosomal rearrangements and genome organization. Since the localization and abundance of the 35S and 5S rDNA are mostly genus- or species-specific, those markers are often used in comparative cytogenetic analyses, providing insight into processes of genome diversification in diploid and polyploid taxa (Weiss-Schneeweiss & Schneeweiss, 2013). Moreover, the ITS region of the 35S rDNA is a widely used molecular marker for phylogenetic analyses, due to a high evolution rate and sequence variability even at interspecific levels (Baldwin et al., 1995).

### 1.2.3 Satellite DNAs

Satellite DNAs (= satDNAs) are non-coding, tandemly repeated DNA motifs, typically located in the heterochromatic regions of chromosomes, such as centromere, pericentromere or telomere, but also interstitially (Garrido-Ramos, 2017). These repeats are in general longer than 100 bp, although also shorter variants exist: minisatellites (9 to 100 bp) and microsatellites (2 to 6 bp), which are both mainly associated with euchromatic regions. Even though satellite DNAs are not coding for proteins, they are functionally important in, for example, structural chromosome organization. Their association with key chromosomal regions, as the centromere, contributes to processes like chromosome pairing and segregation during cell division and structural rearrangements (Csink & Henikoff, 1998; Schueler & Sullivan, 2006; Garrido-Ramos, 2015; Puppo et al., 2020). The monomer sequence composition, arrays' length, copy numbers and chromosomal localizations of satellite DNAs vary substantially across genera and species due to their high evolution rate (Macas et al., 2006; Weiss-Schneeweiss & Schneeweiss, 2013).

Due to such high levels of variability, satellite DNAs are used as cytogenetic markers for chromosome identification and karyotype analysis. They are mainly used in comparative cytogenetics to study genome organization, but also species differentiation and chromosomal evolution. The evolution of satellite DNAs is defined by species-specific amplifications and/or

loss of satellite DNA families in diverging lineages. Closely related species inherit a similar or common set on satellite repeat families, but differential and independent amplification and copy number reduction via unequal cross-over result in distinct satellite DNA profiles. Concerted evolution of satellite DNAs via homogenization allows for rapid divergence of these repeats in related species (Csink & Henikoff, 1998; Biscotti et al., 2015; Thakur et al., 2021). As a result, those evolutionary processes lead to genus- or species-specific satellite repeat profiles, which are used for evolutionary or cytogenetic inferences (Weiss-Schneeweiss & Schneeweiss, 2013). Similar to the rDNA loci, presence, loci and copy numbers of satellite DNAs can also be analyzed using fluorescence *in situ* hybridization (FISH) that allows for physical mapping of known DNA motifs in chromosomes. Bioinformatic tools, like the RepeatExplorer, enable identification and quantification of the repetitive DNA elements from next-generation DNA sequencing data (Novák et al., 2013, 2020).

## 1.3 Aims of the study

The aim of this study is to investigate both the genomic and chemical diversity within the Pubescens clade of *Capsicum*, with a focus on the evolution and dynamics of satellite DNA sequences and the variation in capsaicinoid content and composition across species. The four recognized species of the Pubescens clade, *Capsicum cardenasii*, *Capsicum eshbaughii*, *Capsicum eximium* and *Capsicum pubescens*, are analyzed. Due to the lack of healthy mature plant material, *Capsicum eximium* could not be included in the cytogenetic analyses.

A primary objective is to characterize satellitome profiles of the three species of the Pubescens clade and to examine the abundances and chromosomal localization of shared satellite DNAs using FISH. Selected satellite DNAs (satellites CapSat1, CapSat4, CapSat5, CapSat6, CapSat7, CapSat8 and CapSat17) are physically mapped on the chromosomes of the three species to establish species-specific distribution patterns that may serve as cytogenetic markers for individual chromosome identification. In addition, 35S rDNA loci are mapped to assess their position in relation to satellite DNAs. All the data are used to infer patterns of evolution of individual satellite DNA families and assess the relationships of the three genomes based on tandem repeat dynamics. In addition, the genome sizes of all species belonging to the Pubescens clade are measured using flow cytometry for more insights on the genomic diversity.

In parallel, a broad comparative analysis of secondary metabolites is conducted using fruit material, not only from the Pubescens clade, but also from selected, closely related, wild relatives and all remaining cultivated species of *Capsicum*. Methanol extraction followed by High Performance Liquid Chromatography (HPLC) is used to quantify the content and obtain profiles of composition of capsaicinoids. This enables the comparative analyses of chemical diversity between domesticated and wild taxa, but also among different *Capsicum pubescens* varieties. Furthermore, the *Pun*1 gene, which encodes an enzyme that is involved in the capsaicin biosynthesis, is amplified using Polymerase Chain Reaction (PCR) in several *Capsicum* species to confirm its presence.

Together, the combined cytogenetic and chemical analyses aim to provide more insight into the evolutionary dynamics of satellite DNAs and secondary metabolite profiles in the genus *Capsicum*, with a special focus on the Pubescens clade.

# 2. Material and methods

## 2.1 Material

Four closely related species belonging to the Pubescens clade were analyzed: *Capsicum cardenasii*, *Capsicum eshbaughii*, *Capsicum eximium* and *Capsicum pubescens*. These plants were grown from seeds collected in South America and were cultivated for several years in the greenhouse of the Botanical Garden of the University of Vienna. Additionally, fruits (either grown locally or commercially available) of all five world-wide cultivated species of *Capsicum* (*Capsicum annuum, Capsicum baccatum, Capsicum chinense, Capsicum frutescens, Capsicum pubescens*), close wild relatives of those and *Capsicum rhomboideum,* representing most basal and non-pungent taxon of Andean clade, were used for analyses of capsaicinoid content.

**Table 1**. List of plants used in this study for Fluorescence *in situ* hybridization and High Performance Liquid Chromatography, including their ID and geographical origin (*CCG stands for Carolina Carrizo García).

| Taxon | Provenance | ID/Voucher | Method |
|---|---|---|---|
| *C. cardenasii* | Bolivia: LA Paz, Sorata | ca_3656 | FISH |
| *C. eshbaughii* | Bolivia: Santa Cruz, Samaipata | CCG_LF67* | FISH |
| *C. pubescens* | Market | CCG* | FISH |
| *C. rhomboideum* | | LP7 | HPLC |
| *C. eximium* | Totora, Bolivia | LP1 | HPLC |
| *C. eximium* | Tomina, Bolivia | LP2 | HPLC |
| *C. chacoense* | | LP3 | HPLC |
| *C. cardenasii* | | LP25 | HPLC |
| *C. pubescens* | Villa serrano, Bolivia | LP4 | HPLC |
| *C. pubescens* | Cusco, Peru | LP5 | HPLC |
| *C. pubescens* | Yungas area, La paz market, Bolivia | LP6 | HPLC |
| *C. pubescens* | Trijullo, Peru | LP16 | HPLC |
| *C. pubescens* | Huancane, Bolivia | LP17 | HPLC |
| *C. pubescens* | La Merced Market, Mexico | LP18 | HPLC |
| *C. pubescens* cv. CAP363 | Guatemala | LP19 | HPLC |
| *C. pubescens* | Coroico, Bolivia | LP20 | HPLC |
| *C. pubescens* | San Isidro, Bolivia | LP21 | HPLC |
| *C. pubescens* cv. "Turbo Pube" | Seeds bought online | LP23 | HPLC |
| *C. pubescens* | Moro Moro, Bolivia | LP24 | HPLC |
| C. anuum var. *annuum* cv. "Lippenstift" | | LP10 | HPLC |
| *C. anuum* var. *annuum* cv. "Royal Black" | | LP11 | HPLC |

| | | | |
|---|---|---|---|
| *C. baccatum* var. *pendulum* | | LP8 | HPLC |
| *C. baccatum* var. *umbilicatum* | | LP9 | HPLC |
| *C. chinense* cv. Trinidad Moruga (yellow) | | LP13 | HPLC |
| *C. chinense* cv. "Gänseschnabel" | | LP14 | HPLC |
| *C. chinense* cv. Jelly Bean | | LP15 | HPLC |
| *C. chinense* #2040 | | LP22 | HPLC |

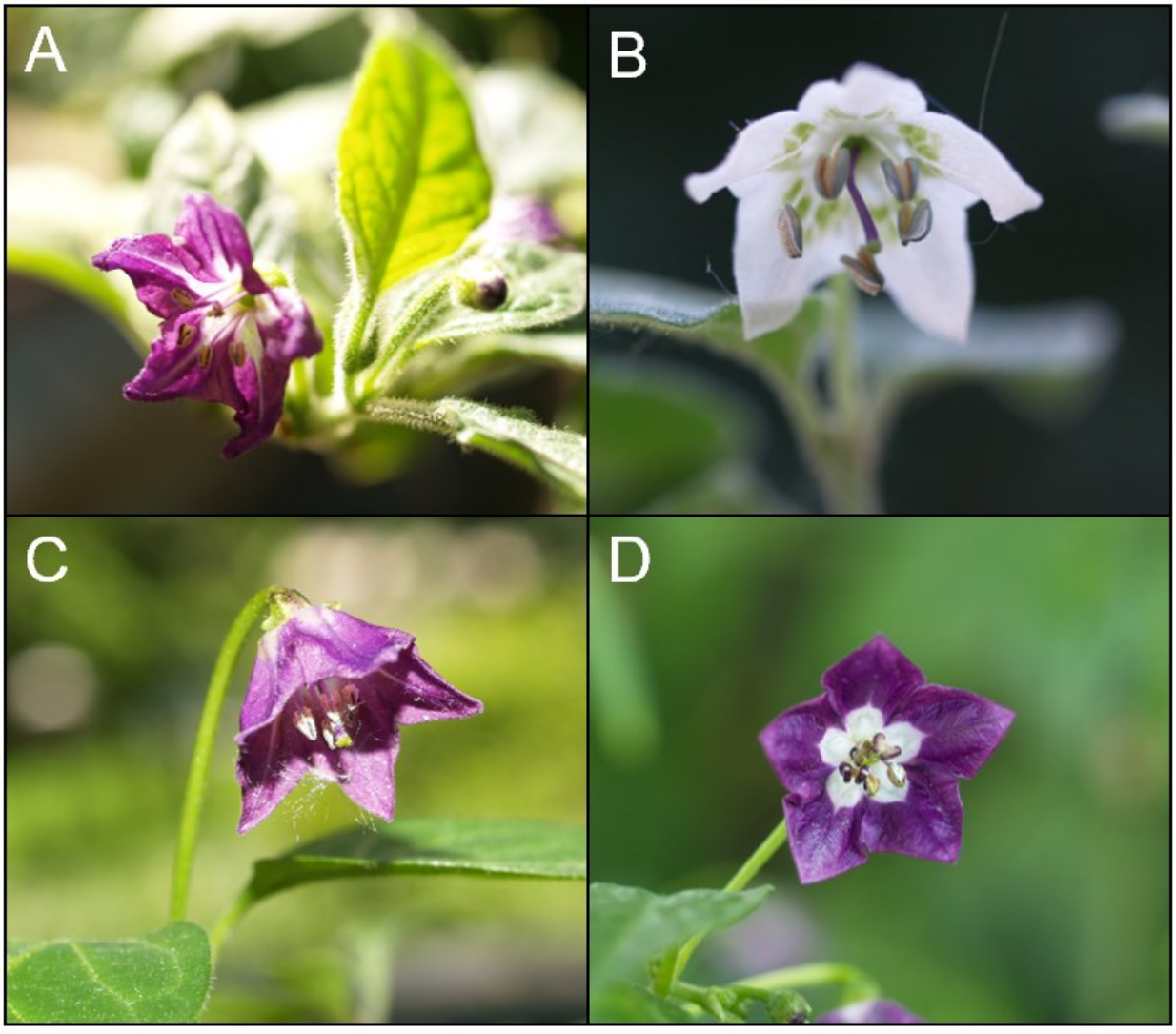



**Figure 8**. Flowers of selected *Capsicum* species showing characteristic differences in size, color, and morphology. A, *C. cardenasii*. B, *C. eshbaughii*. C, *C. pubescens* (Comarapa rojo). D, *C. pubescens* (GEB 3858). Photo credits: 8A-C, Laura Pflügl. 8D, Vildana Kapetanowic.

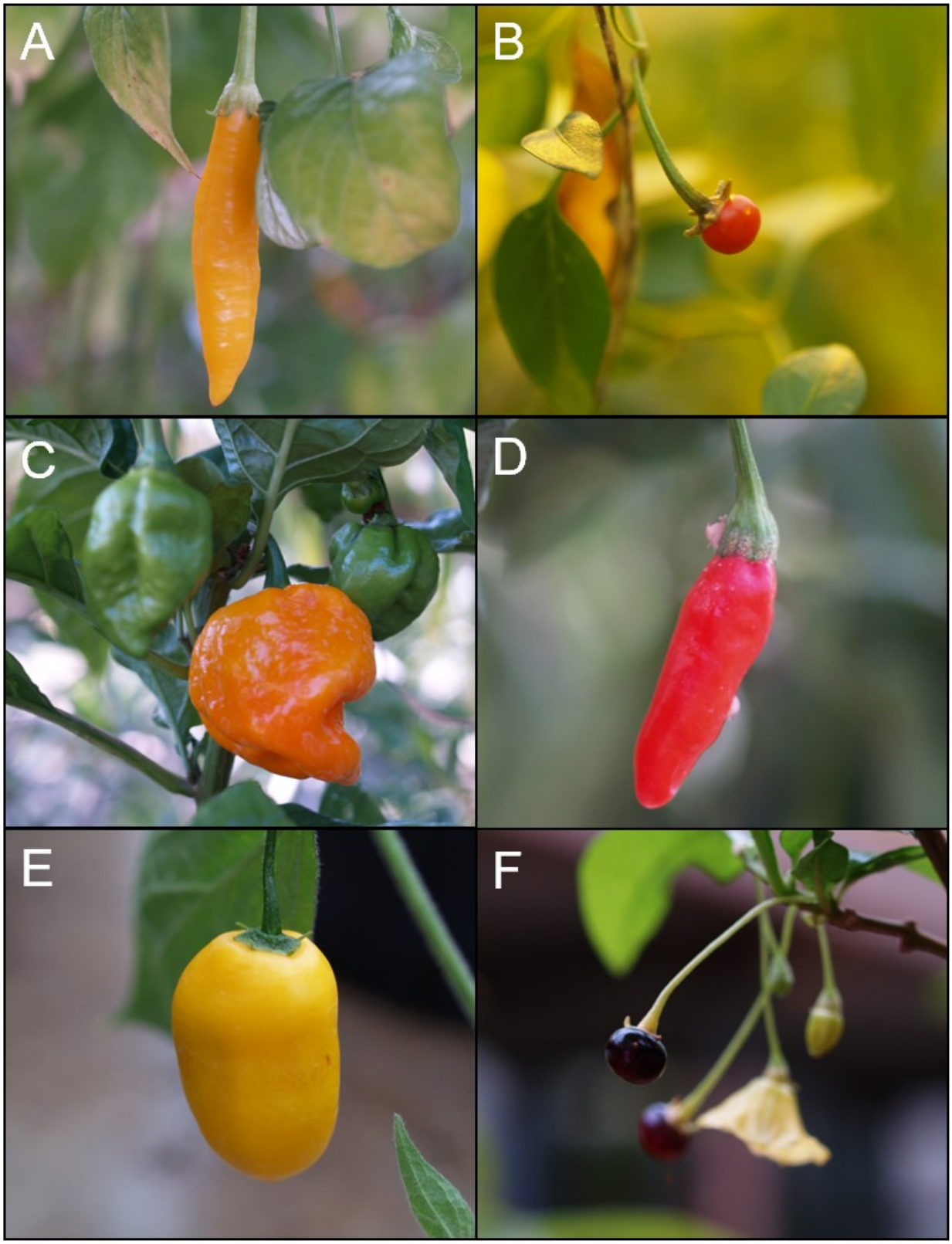



**Figure 9.** Fruits of selected *Capsicum* species showing characteristic differences in size, color, and morphology. A, *C. baccatum* var. pendulum. B, *C. cardenasii*. C, *C. chinense* var. Trinidad Moruga. D, *C. frutescens*. E, *C. pubescens*. F, *C. rhomboideum*. Photo credits: Rosemarie Benthen and Vildana Kapetanowic.

## 2.2 Methods

### 2.2.1 Flow cytometry

Genome sizes of all four species belonging to the *Pubescens* clade: *C. cardenasii*, *C. eshbaughii, C. eximium* and *C. pubescens*, were measured. *Solanum pseudocapsicum* (1C = 1.2946; Temsch et al., 2010) was used as the internal standard for the measurements. The sample preparation was followed the available protocol (Baranyi & Greilhuber, 1996) with some modifications.

Approximately 25 mg of fresh leaf material of the analyzed individuals and the standard organism were co-chopped in 1.1 mL of cold isolation buffer (0.1 M citric acid 1-hydrate and 0.5 % Triton X-100; pH = 1.5). Additionally, one sample of only *Solanum pseudocapsicum* was prepared for the calibration of the instrument and to identify the peaks of the object versus the standard. The liquid fraction containing nuclei was filtered into tubes using a 30 µm nylon mesh filter to remove any leftover leaf tissue and debris. Two thirds of the sample were stained with propidium iodide (PI), and one third was kept as backup. The samples were incubated with the propidium iodide solution (pH = 9.5) at 4 °C in darkness overnight. The RNase treatment suggested in the protocol was omitted, because it increased the CVs.

The measurements were conducted using the CyFlow Space flow cytometer (Sysmex Austria GmbH, Vienna, Austria), equipped with a 100 mW green laser operating at a wavelength of 532 nm. The fluorescence emitted by the stained nuclei was measured for 3,333 particles with the gain set to 360. Data acquisition and visualization were performed using the software CyFlow WA (Sysmex Austria GmbH, Vienna, Austria). In each histogram at least two distinct peaks were observed on the x-axis, corresponding to the G1 phase of the nuclei of the standard and the G1 phase of the measured organism, showing the relative fluorescence intensity. Each individual was measured three times and the 1C-value was calculated as the average value of the three measurements.

Calculation of the genome size with the following formula:

$$1C\ value\ object = \frac{mean\ G1\ peak\ position\ object}{mean\ G1\ peak\ position\ standard} \times 1C\ value\ standard$$

### 2.2.2 Material pre-treatment and fixation for chromosome analysis

Actively growing healthy, white, and intact root meristems were carefully collected with forceps and cleaned in water to get rid of any soil particles and debris. Subsequently the roots

were placed in a 0.002 M solution of 8-hydroxyquinoline (Sigma-Aldrich, Vienna, Austria) for two and a half hours at room temperature followed by another two and a half hours incubation at 4°C. For the species *Capsicum eshbaughii* and *Capsicum pubescens* a longer incubation time of three hours each was needed. 8-hydroxyquinoline induces chromosome condensation and inhibits the spindle formation, leading to better chromosome separation. Afterwards the roots were rinsed with tap water to get rid of the remaining 8-hydroxyquinoline solution and immediately fixed in a 3 : 1 mixture of ethanol (EtOH) and acetic acid (Sigma Aldrich, Vienna, Austria). The fixations were incubated for three hours at room temperature and then stored at -20 °C until use.

### 2.2.3 Chromosome preparation – enzymatic digestion

The roots were removed from the fixative and chromosomal spreads were prepared following the protocol of Jang & Weiss-Schneeweiss (2015) with minor modifications. Briefly, the root tips were washed twice in citrate buffer (pH = 4.8, 0.1 M citric acid monohydrate and 0.1 M trisodium citrate dehydrate; Sigma Aldrich, Vienna, Austria), 10 minutes each. Afterwards the roots were incubated in a prewarmed enzyme-mix (0.5 % pectolyase, 0.5 % cytohelicase, 1 % cellulase in citrate buffer, pH = 4.8; Sigma Aldrich, Vienna, Austria) to digest the cell walls at 37 °C in an incubator for approximately 15 to 20 minutes. The enzyme mix was removed, material washed with fresh citrate buffer and the material was transferred with the pipette onto a pre-cleaned slide into a drop of 60 % acetic acid. The meristems were dissected under a stereo microscope using entomological needles, the coverslip was carefully placed over the material and the sample was gently squashed to evenly spread the chromosomes. The quality of the spreads was examined using phase contrast (DIC Ⅲ) on an Axio Imager M2 microscope (Carl Zeiss, Vienna Austria). Slides with well-spread, complete and condensed chromosome plates were frozen at -80 °C for approximately 30 minutes, the coverslips were removed and the slides were air-dried. The slides were stored at 4 °C until further use.

### 2.2.4 Fluorescent *in situ* hybridization (FISH)

The prepared slides were re-fixed in the ethanol : acetic acid fixative (3 : 1 solution) for 15 minutes at room temperature, dehydrated in 96 % ethanol for 15 minutes, and air-dried for 45 minutes at room temperature.

The slides were incubated for 5 minutes in 2x SSC at room temperature. Subsequently, they were further placed in a pre-warmed Coplin-jar containing pepsin (10 µg/mL, pH = 2; Sigma

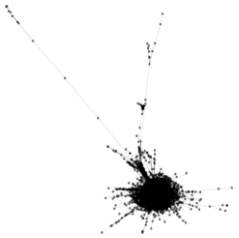
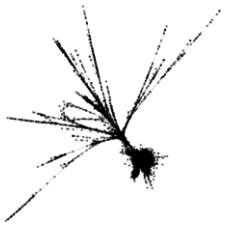
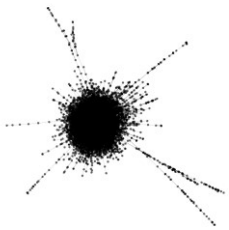
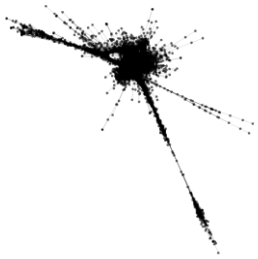
Aldrich, Vienna, Austria) and incubated at 37 °C for 20 minutes. Pepsin removes nuclear and cytoplasmic proteins, which might interfere with the probe binding. The slides were then washed in three changes of 2xSSC, for 5 minutes each at room temperature, before they were incubated in 4 % paraformaldehyde aqueous solution (VWR chemicals, Vienna, Austria) to re-fix the histones. Following the incubation, the slides were washed in three changes of 2xSSC, for 5 minutes each at room temperature and dehydrated in cold ethanol series (70 % and 96 % ethanol) for three minutes each. Subsequently, the slides were air-dried in an upright position at room temperature for 30 to 40 minutes.

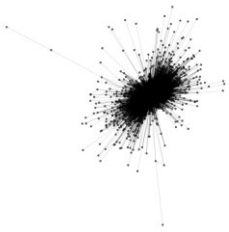
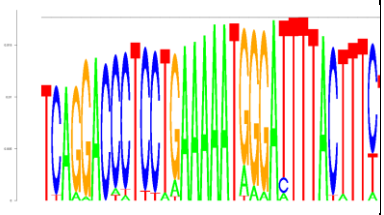
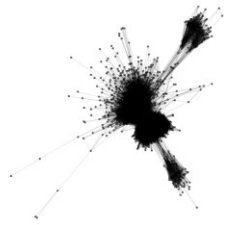
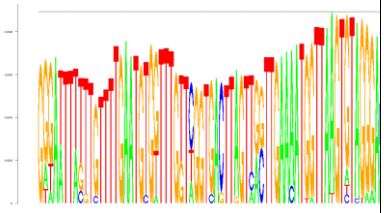
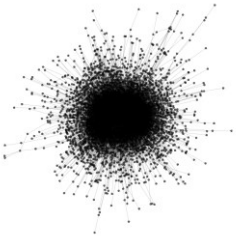
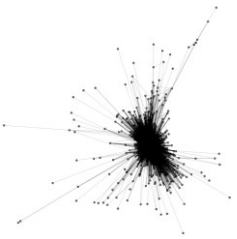
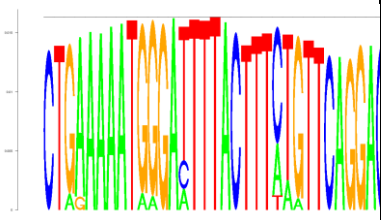
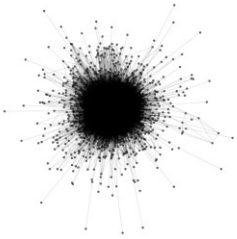
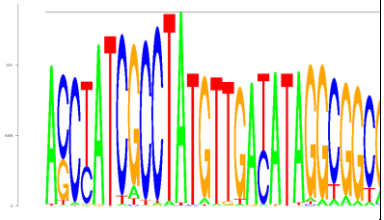
Hybridization Master-Mix (MM) of 90 µL was prepared containing 10 % dextran sulphate, 0.02xSSC and 100 ng/µL of blocking DNA (Salmon Sperm DNA; Sigma Aldrich, Vienna, Austria) in sterile water.

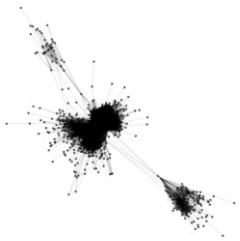
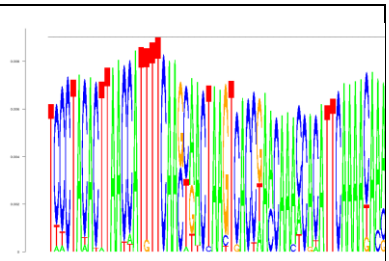
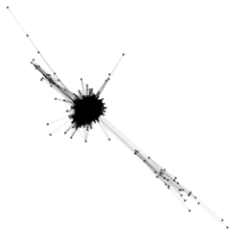
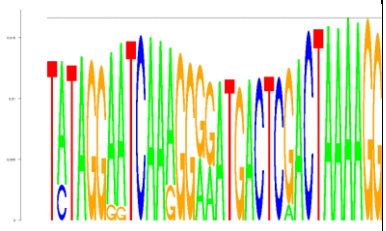
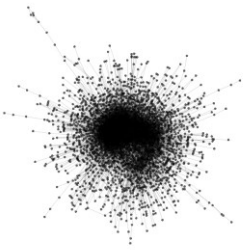
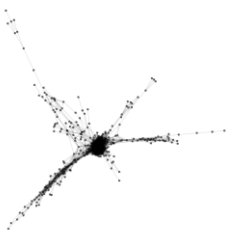
Several satellite DNAs identified in *Capsicum* genomes (H. Weiss-Schneeweiss and R. Benthen, unpubl.) were used as probes (Table 2). Briefly, 2.5 µL of each of two probes (two probes were always hybridized simultaneously) were added to 45 µL of the hybridization Master-Mix. The solution was then denatured at 98 °C in a thermoblock (Eppendorf, Vienna, Austria) for three minutes and immediately placed on ice for five minutes to stabilize the single-stranded DNA. 8 µL of the denatured hybridization mix was pipetted directly onto the sample on the slide, a coverslip was added and sealed with Fixogum. The slides were denatured on a PCR *in situ* block (PeqLab, Vienna, Austria) using the following program: 4 minutes at 72°C, 1 minute at 65°C, 1 minute at 55 °C and 1 minute at 45 °C. Finally, the slides were transferred into a prewarmed chamber and incubated at 37 °C overnight.

After overnight hybridization, coverslips were removed washed in 2x SSC at room temperature. The slides were then incubated at 39 °C in 2xSSC for 5 minutes, in 0.2xSSC for 5 minutes and in 2xSSC with Tween20 for 5 minutes.

**Table 2.** Satellite DNAs including the monomer sequence, graph and sequence logo from RepeatExplorer, used for localization in the chromosomes via FISH. (H. Weiss-Schneeweiss and R. Benthen, unpubl.)

| **Satellite name (incl. species)** | **Proportion [%] (reads)** | **Cluster and monomer length** |
|---|---|---|
| Sequence monomer | Graph | Sequence logo |
| Sat1Capsicum_eshbaughii | 0.21 (18 821 reads) | CL88_37bp |
| TGTTCAGGACCCTCCTGAAAAATGGGATTTTACTTTC |  | Contig |
| Sat5Capsicum_eshbaughii | 0.19 (16 592 reads) | CL104_76bp |
| TTTCAAGTGAATAATTCACCTAACAAACCATATTCATAACACAAAAATCACCTTTAAAGTCCCTATACTTAACCAT |  | Contig |
| Sat8Capsicum_eshbaughii | 0.31 (26 974 reads) | CL54_183bp |
| AACGAGGCATTTTCGGGCCAAATCGGTGTGCTATAGCCCACGGTTTCGGGGGTGGTCCCGGGGTCCGGGCGTGGTGTGGGCCGAAAATCGACGGGGGCATGCCAGGGAGGTTGGGGATCATCCCAGTGTGGTCCGTTTAATTTTTCGTGGCCATTTGGGTGGACAAACGGGCGAAACGGC |  | Contig |
| Sat17Capsicum_eshbaughii | 0.1 (8 925 reads) | CL174_112bp |
| ACTATTGCCTTAGAGGCAAGACGGCTCAAGGACTTTCAAACTACAAATTATGCCCCACGGGCAAGAGTTGACTCTACCACATTATGTTTTCATTTATGAGATGTTCTACAA |  | Contig |
| Sat1Capsicum_cardenasii | 1.5 (18 021 reads) | CL3_37bp |

| | | |
|---|---|---|
| TCAGGACCCTCCTGAAAA<br>ATGGGATTTTACTTTCTG<br>T |  |  |
| Sat5Capsicum_cardenasii | 0.19 (2 193 reads) | CL80_76bp |
| GGGATTTTTTTGTTTTGA<br>ATGTGGTTTGTTCGGTGA<br>CTTAGTTGGTTGAAAATG<br>GTTAAGTGTAGGGACTTT<br>AAAG |  |  |
| Sat8Capsicum_cardenasii | 0.4 (4 704 reads) | CL28_183bp |
| CCACACCACGCCCGGACC<br>CCGGGACCACCCCCAAAA<br>CCGTGGGCTATAGCACAC<br>CGATTTGGCCCGAAAATG<br>CCCCGTTCGCGCCGTTTC<br>GCCCGTTTGTCCACCCAA<br>ATGGCCACGAAAAATTAA<br>ACGGACCACACTGGGATG<br>ATCCCCAACCTCCCTGGC<br>ATGCCCCGGTCGATTTTT |  | Contig |
| Sat1Capsicum_pubescens | 0.61 (4 474 reads) | CL23_37bp |
| CTGAAAAATGGGATTTTA<br>CTTTCTGTTCAGGACCCT<br>C |  |  |
| Sat4Capsicum_pubescens | 0.39 (2 854 reads) | CL40_33bp |
| ACCTATCGCCTATGTTGA<br>TATAGGCGGCGCCTA |  |  |
| Sat5Capsicum_pubescens | 0.21 (1 542 reads) | CL67_76bp |

| | | |
|---|---|---|
| TCCCTACACTTAACCATTTTCAAGCAACTAAGTCACCGAACAAACCACATTCAAAACAAAAAATTCCCCTTTAAAG |  |  |
| Sat6Capsicum_pubescens | 0.29 (2 091 reads) | CL60_41bp |
| TATAGGAATCAAAGGGGATGACTCGACTAAAAGGTACGTCT |  |  |
| Sat8Capsicum_pubescens | 0.4 (2 909 reads) | CL39_183bp |
| GGCATTTTCGGGCCAAATCGGTGTGCTATAGCCCACGGTTTTGGGGGTGGGCCCGGGGTCCGGGCGTGCTGTGGGCCAAAAATTGATCGGGGCATGCCAGGGAGGTTGGGGATCATCCCAGTGTGGTCCGTTTAATTTTTTGTGGCCATTTGGGTGGACAAACGGGCGAAACGGCGCGAA |  | Contig |
| Sat17Capsicum_pubescens | 0.058 (421 reads) | CL102_112bp |
| TCAACTCTTGCCCGTGGGGCATAATTTGTAGTTTGAAAGTCCTTGAGCTAAGTCTTGCCTCTAAGGCAATAGTTGTAGAACATCTCATAAATGAAAACATAACGTGGTAGA |  | Contig |

The probes were commercially synthesized and labelled either with biotin or with digoxygenin. For the biotin and digoxygenin detection, 50 µL of blocking solution (= BS) containing 1 % BSA (Bovine serum albumin; (Sigma Aldrich, Vienna, Austria) in 2x SSC with 0.2 % Tween20 (Sigma Aldrich, Vienna, Austria), was added onto the sample. The slides were covered with a thin foil to prevent drying out and incubated for 20 minutes at 37 °C in a wet box. Afterwards, the slides were incubated in 50 µL of detection solution containing Streptavidin-Cy3 (red fluorescence, 1:250 in BS; Sigma Aldrich, Vienna, Austria) and anti-digoxygenin-FITC (green fluorescence, 1:100 in BS; Sigma Aldrich, Vienna, Austria) for 60 minutes at 37 °C in the wet

box, in darkness. The incubation was followed by three final washing steps at 39 °C, twice in 2xSSC at 39 °C for 7 minutes each, and once in 2xSSC with 0.2 % Tween20 for 7 minutes.

The access of liquid was then removed and slides were mounted in 10 µL of antifade buffer Vectashield (Vector Laboratories, Szabo-Scandig, Vienna, Austria) containing 2 µg/µL DAPI (4',6-Diamidin-2-phenylindol; Sigma Aldrich, Vienna, Austria). The chromosomes were counterstained using DAPI, which makes them appear blue in the fluorescence microscope. A coverslip was added and sealed with Fixogum. The slides were stored at 4 °C overnight to stabilize the fluorescence. Chromosomal localization of the probes was analyzed using the epifluorescence microscope AxioImager M2 (Carl Zeiss, Vienna, Austria) with filters for Cy3, FITC and DAPI. Complete sets of mitotic metaphase chromosomes were photographed using CCD camera and ZEN imaging software ZEN, version 3.8.2 (Carl Zeiss, Vienna, Austria). Images were contrasted in Adobe Photoshop using only the functions that applied equally to the whole image. The taken pictures were used to digitally cut out the chromosomes and arrange them in pairs and by their sizes, the karyotypes were assembled in Corel Photo Paint (version 25.2).

### 2.2.5 Reprobing of the slides for FISH

Slides with well spread and complete plates of chromosomes were reprobed with second set of the probes. First, originally used satellite DNA probes were washed off and then the slides were hybridized with new probe combinations. First the immersion oil was cleaned off the slides and the Fixogum was carefully removed. The slides were incubated in 2xSSC at 37 °C for 15 minutes to remove the coverslips. Afterwards, the slides were incubated in three changes of 4xSSC with 0.2 % Tween20 for 20 minutes each at room temperature, and in 2xSSC for 10 minutes at room temperature. The preparations were then dehydrated for three minutes in cold 70 % ethanol, and for three minutes in 96 % EtOH and finally air-dried in an upright position for approximately 40 minutes.

The following denaturation, hybridization and detection steps were performed as described before. The only difference was that the stringent washes were carried out at 37 °C for 4 minutes.

### 2.2.6 High Performance Liquid Chromatography (HPLC) analyses

Selected *Capsicum* fruits (Table 1) were prepared for HPLC analyses that allow to determine the capsaicinoid content. The material was first completely dried and stored at -18 °C in silica

gel. For the extraction of the capsaicinoids, whole fruits, including seeds, were ground with a mortar to a fine powder under the fume hood. The material was then weighed, and 30 mg of ground material was transferred into an Eppendorf tube. To each tube 300 µL of methanol (MeOH) was added, before placing the samples in an ultrasonic bath (ultrasonic cleaner, VWR, Vienna, Austria) for 15 minutes. The solutions were then centrifuged for 15 minutes at 13,900 rpm at room temperature, and the supernatant was transferred into weighed**,** 2 mL Eppendorf tubes. The steps of adding 300 µL of MeOH, incubation in an ultrasonic bath, centrifugation and collection of the supernatant were repeated three times for each sample. The tubes containing the supernatant were left open overnight to dry under the fume hood.

To ascertain that the samples were completely dried, they were vacuum centrifuged (Thermo Savant ISS100 SpeedVac Concentrator System) at medium drying rate for two hours. The material was then weighed and mixed with methanol and acetonitrile (1:1) to a final concentration of 6 mg/mL. For most samples a stock solution was prepared first, and then the dilution.

For each sample, 400 µL of extract (6 mg/mL) was transferred into HPLC vials prior to analysis. The measurements were performed using an Agilent 1100 series HPLC system equipped with a UV diode-array detector (Agilent Technologies, Vienna, Austria). Separation was carried out on a Hypersil BDS-C18 column (3 µm, 100 x 4 mm) using a mobile phase consisting of an aqueous buffer consisting of 1 mL tetrabutyl ammonium hydroxide and 1 mL *ortho*-phosphoric acid and methanol.

This method produced both chromatograms and UV spectra, with the peak area in the chromatogram correlating directly with the capsaicinoid concentration in the sample. For quantification, a calibration curve was established using standard solutions of capsaicin (from *Capsicum* sp., ≥95 %, suitable for HPLC, Sigma Aldrich, Vienna, Austria) at the following concentrations: 500 µg/mL, 250 µg/mL, 125 µg/mL, 62.5 µg/mL, 31.25 µg/mL and 6.25 µg/mL. Each standard was measured in triplicate and the mean peak areas were used to generate the calibration curve (x-axis: capsaicin concentration, y-axis: peak area). This calibration curve was then further used for the quantification of capsaicinoids in the unknown samples.

Furthermore, to evaluate the potential loss of capsaicinoids during extraction and drying, a recovery test was performed. For this, known amounts of the capsaicin standard solutions were added to the samples prior to extraction. *Capsicum annuum* and *Capsicum rhomboideum*

were used, as those do not naturally contain capsaicin. The following solutions were used: 50 µL with the concentration of 2 mg/mL, 50 µL with the concentration of 1 mg/mL and 50 µL with the concentration of 31,25 mg/mL. Each measurement was repeated twice. After standard processing, the extraction using the ultrasonic bath, centrifuge and vacuum drying, the capsaicin content was quantified via HPLC. The measured concentration was then compared to the theoretical amount added before. This allowed for the calculation of the recovery rate, indicating the extent of capsaicin loss during sample handling and preparation.

### 2.2.7 PCR amplification of the *Pun*1 gene

Polymerase Chain Reaction (PCR) was used to amplify a fragment of *Pun*1 gene (*Pun*1 exon and part of the intron) from genomic DNAs. Already available DNA extracts from *Capsicum annuum*, *Capsicum baccatum* var. *pendulum*, *Capsicum cardenasii*, *Capsicum chinense*, *Capsicum eximium*, *Capsicum frutescens*, *Capsicum pubescens* and *Capsicum rhomboideum* were used. *Capsicum rhomboideum* served as a negative control for the trail runs, as this species belongs to the Andean clade that has diverged before the pungency evolved, thus this taxon lacks the *Pun*1 gene entirely. The preliminary trial PCR runs were performed with *Capsicum annuum*, *Capsicum pubescens* and *Capsicum rhomboideum* to optimize reaction conditions. These tests included varying the extension time and comparing different DNA polymerase master mixes. The genomic DNA was diluted 1:50 prior to the PCR.

Each reaction included 1x DreamTagGreen MM, 0.25 µM of each forward and revers primer (C. Carrizo Garcia, unpubl.), DMSO and genomic DNA template in a total volume of 20 µL. The PCR amplification was carried out in the thermal cycler PeqStar (PeqLab, Vienna, Austria) with the following conditions: initial denaturation at 95 °C for 4 minutes followed by 30 cycles of denaturation at 95 °C for 45 seconds, annealing at 60 °C for 1 minute and extension at 72 °C for 70 seconds, and final extension at 72 °C for 5 minutes.

PCR products were checked on 1.5 % agarose gel by gel electrophoresis, dyed with GelRed (Biotium, Fremont, USA). The electrophoresis was performed at 100 V for 30 minutes and the PCR products were visualized with the Gel Doc 2000 (Bio-Rad, Hercules, USA).

# 3. Results

## 3.1 Genome size measurements

To estimate the genome sizes of species within the Pubescens clade, flow cytometry measurements were conducted for all four species: *Capsicum cardenasii*, *Capsicum eshbaughii*, *Capsicum eximium* and *Capsicum pubescens*. For each species, three independent measurements were performed and in every run *Solanum pseudocapsicum* (1C = 1.2946 pg) was included as an internal reference standard.

The resulting fluorescence histograms (Figure 10) consistently showed two distinct peaks, the first representing nuclei of *Solanum pseudocapsicum* and the second corresponding to nuclei from the respective *Capsicum* species. These peaks represent the detected nuclei in the G1 phase of the cell cycle, where the DNA content is unreplicated. Mean genome sizes for each species were calculated based on the three replicates.

**Table 3.** Results of DNA content measurements for the Pubescens clade using flow cytometry, including the 1C value [pg] with standard deviation (= SD), chromosome number and genome size in MBp.

| Taxon | 1C value [pg] ± SD | Chromosome number | Genome size [MBp] |
|---|---|---|---|
| *Capsicum cardenasii* | 4.5991 ± 0.0150 | 2*n* = 24 | 4498 |
| *Capsicum eshbaughii* | 4.1155 ± 0.0061 | 2*n* = 24 | 4025 |
| *Capsicum eximium* | 4.1555 ± 0.0121 | 2*n* = 24 | 4064 |
| *Capsicum pubescens* | 4.5910 ± 0.0128 | 2*n* = 24 | 4490 |

Flow cytometry measurements showed well defined differences in genome size among the species within the Pubescens clade. *Capsicum eshbaughii* exhibited the smallest genome size (4.12 pg/ 4025 MBp), while *Capsicum eximium* had a slightly higher nuclear DNA content (4.16 pg/ 4064 MBp). In contrast, *Capsicum cardenasii* and *Capsicum pubescens* possessed larger genome sizes (4.60 pg/ 4498 MBp and 5.59 pg/ 4490 MBp, respectively). The results were consistent across the three replicates for each species and the CVs were below 5 %.

## 3.2 Localization of satellites DNAs in the Pubescens clade

Fluorescence *in situ* hybridizations was applied to localize and map the satellite DNA sequences of CapSat1, CapSat4, CapSat5 and CapSat8 on metaphase chromosomes of the three selected species from the Pubescens complex. Furthermore, preliminary analyses on the localization of CapSat6, CapSat7, CapSat17 and 35S rDNA were conducted in *Capsicum eshbaughii* and *Capsicum pubescens*.

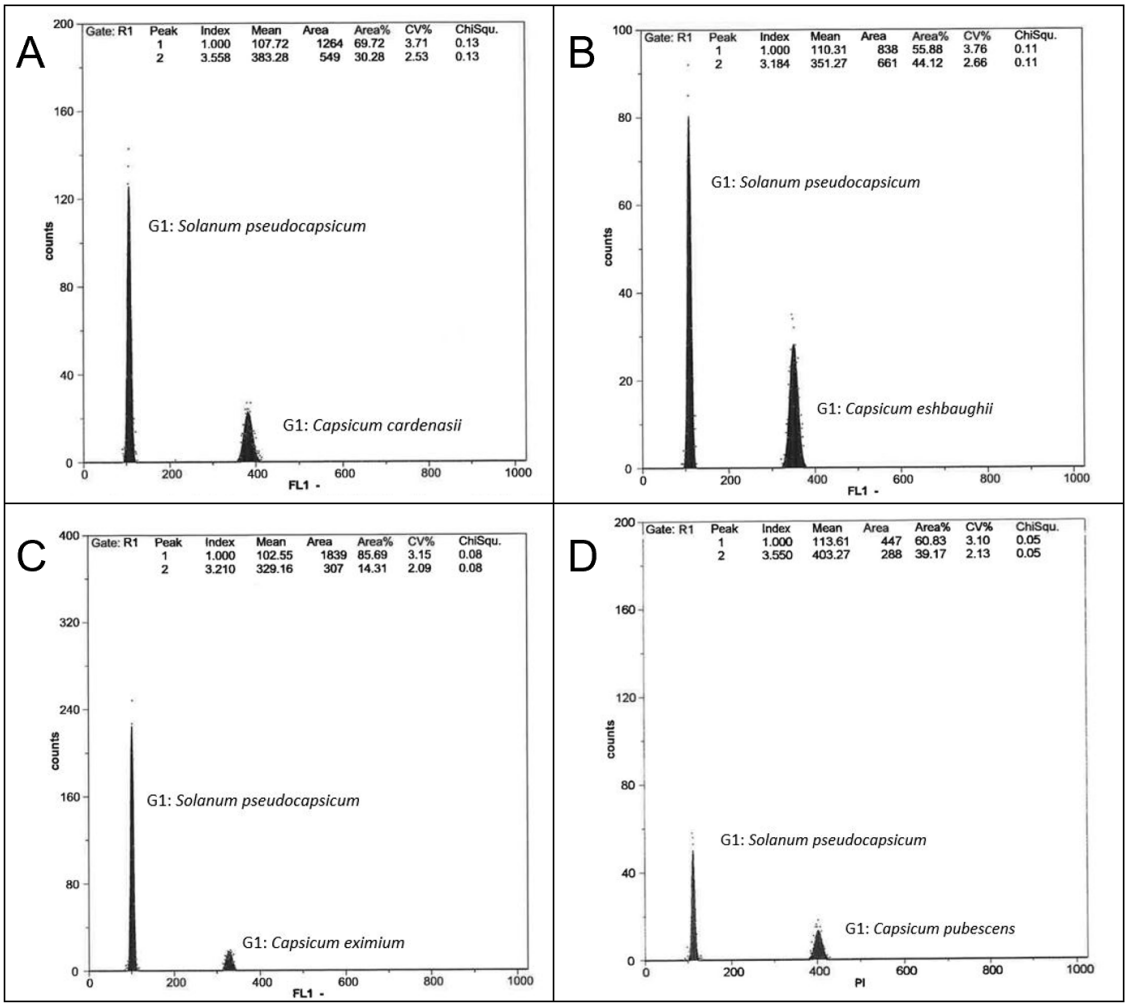



**Figure 10:** Flow cytometry histograms of one individual per each species of the Pubescens clade. A*, C. cardenasii*; B, *C. eshbaughii*; C, *C. eximium*; D, *C. pubescens*.

The satellite DNA sequences were mapped in two probe combinations: CapSat1 and CapSat8 and CapSat4 together with CapSat5. The probes were labeled with biotin (red signals) and digoxygenin (green signals), biotin was used always for CapSat8 and CapSat5 and digoxygenin for CapSat1 and CapSat4. Due to high similarity in chromosome size and morphology it was not possible to identify individual chromosomes, as they lack distinguishable chromosomal landmarks like NORs or secondary constrictions and centromeres were not always identifiable. Therefore, to facilitate the visualization and comparison of the satellite DNA signal loci observed in the three of four species of Pubescens clade, idiograms were created. Those represent a haploid set of chromosomes with all chromosomes of equal size without the position of centromeres indicated to focus on the relative position of hybridization signals rather than physical differences. *Capsicum eximium* was not included in this study due to a lack of plant material.

### 3.2.1 *Capsicum cardenasii*

The FISH analysis in *Capsicum cardenasii* revealed multiple hybridizations signals of CapSat11 with loci distributed across six chromosomes. The signals were primarily located in the subtelomeric of the chromosomes. The signals varied in appearance, some were band-like while others appeared dot-like. The strength of all signals was comparable, with the exception of the signal of CapSat1 on chromosome 7, which was weaker compared to the others (Figure 11).

CapSat8 showed a much more restricted distribution, with bigger loci observed only on two chromosomes. The subterminal locus on chromosome 3 appeared almost band-like, while on chromosome 5 it was a dot-like locus near the pericentric region. No additional loci of CapSat 8 were detected on other metaphase plates of *Capsicum cardenasii*, supporting the limited presence of this sequence in this species (Figure 11).

The CapSat4 hybridized to five chromosomes, with loci predominantly in subterminal positions of the chromosomes. The intensity of the signals was generally weaker than that of CapSat1, but they were consistently detected in all plates. The loci ranged in size from dot-like to broader band-like. Chromosome 6 carried two loci, one on each of the arms of the chromosome. On the short arm, the signal appeared a bit smaller, in contrast to the more diffuse band on the long arm (Figure 11).

The loci of CapSat5 were detected on three chromosomes, either in the subterminal region or near the pericentric region of the chromosome. Chromosome 8 carried both CapSat4 and CapSat5 loci. The signal intensities were slightly weaker as those of CapSat8 (Figure 11).

### 3.2.2 *Capsicum eshbaughii*

In *Capsicum eshbaughii*, strong signals for CapSat1 were detected across six chromosomes. The loci were only found in the subterminal regions of the chromosomes. Chromosome 1 carried a duplication of the dot-like signal, resulting in four distinct dots in one locus. In *Capsicum eshbaughii*, no band-like loci observed. The strength of signals was comparable across all analyzed plates (Figure 12).

CapSat8 was detected on four chromosomes, including a dot-like duplication on chromosome 2, similar to CapSat1, but these were close enough to be treated as one locus. The dot-like signals had a compact appearance, in contrast to the band-like ones, which appeared more

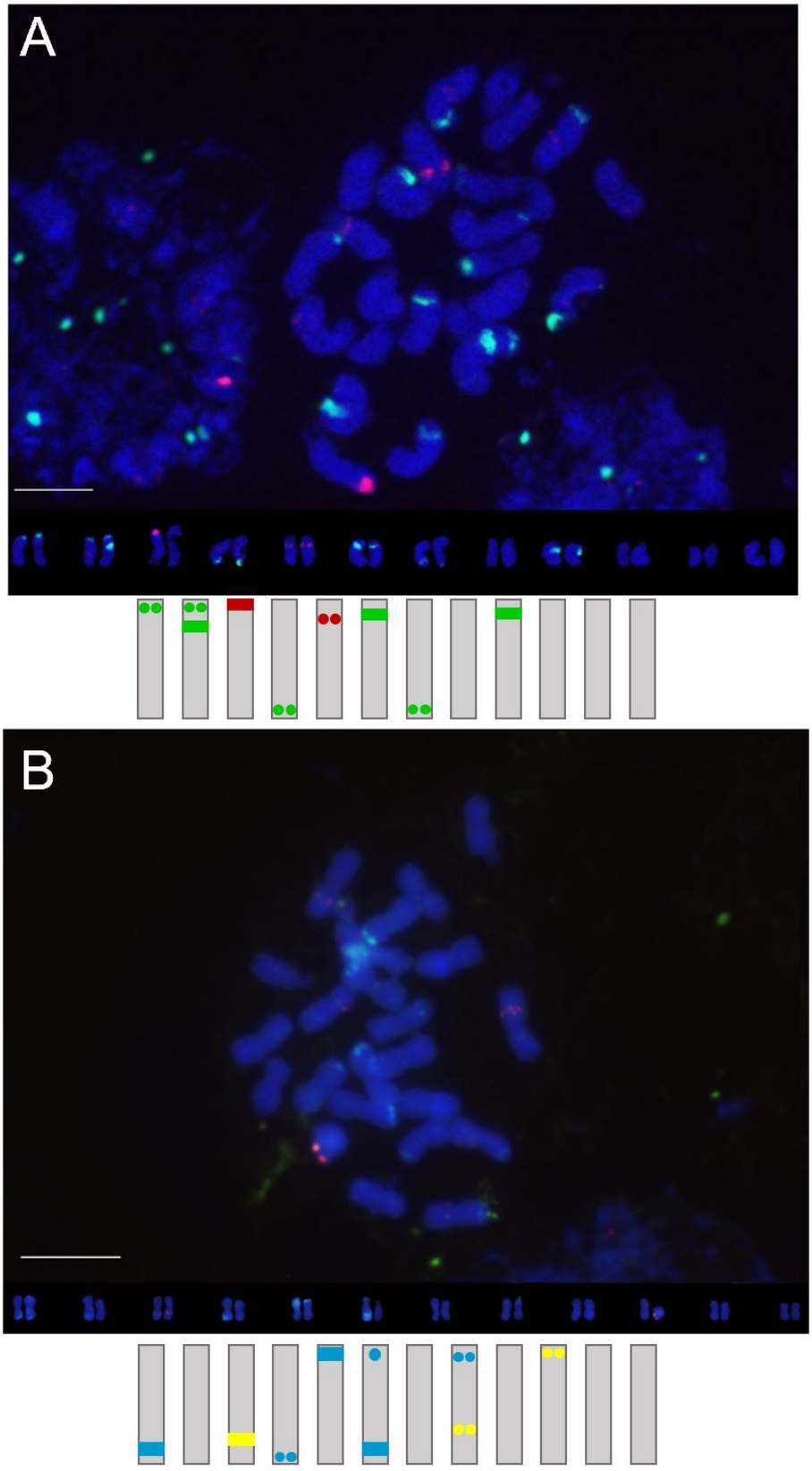



**Figure 11.** Localization of satellite DNAs CapSat1, CapSat4, CapSat5 and CapSat8 in mitotic metaphase chromosomes of *C. cardenasii* (2*n* = 24) accompanied by cut-out karyotypes and idiograms. A, localization of CapSat1 (green signals, green in idiogram) and CapSat8 (red signals, red in idiogram). B, localization of CapSat4 (green signal, blue in idiogram) and CapSat5 (red signal, yellow in idiogram). Scalebar = 5 µm

diffuse. Despite that, they were still clearly visible and strong. CapSat8 was mainly found in the interstitial region of the chromosomes (Figure 12).

The hybridization with CapSat4 resulted in clear and strong loci distributed across five chromosomes, all restricted to the subterminal regions. All signals were dot-like and no band-like loci were detected (Figure 12).

CapSat5 loci were distributed across four chromosomes, all localized in the subterminal regions. All signals had a dot-like appearance and were strong. On chromosome 1 a duplication of the signal was observed, resulting in four distinct dot-like signals, similar to these of CapSat1 and CapSat8. The remaining signals were also dot-like and no band-like loci were observed (Figure 12).

### 3.2.3 *Capsicum pubescens*

In *Capsicum pubescens* CapSat1 was detected at 13 loci across ten chromosomes, missing only from two chromosomes. The signals were both dot-like and band-like. Chromosomes 1 and 6 both carried duplicated signals, resulting in four distinct dots in one locus of each. Chromosomes 3, 6 and 7 had loci on both the short and long arms. The vast majority of loci were positioned in the subterminal regions, only one locus was slightly more interstitial. *Capsicum pubescens* possessed the highest number of CapSat1 loci among the species analyzed (Figure 13).

17 loci of CapSat8 were detected distributed across ten chromosomes, also the highest number of loci of this satellite DNA observed in this study. Similar to CapSat1, signals were strong and detected in all cells analyzed. All loci were positioned in the subterminal regions, except for chromosome 2, which not only carried loci on both the short and long arm, but additionally also one locus in the pericentric region. Dot-like duplications, resulting in four distinct dots in the same locus were detected on chromosomes 4, 7, 9 and 10. On chromosomes 3 and 10, CapSat1 and CapSat8 were detected in close proximity, appearing almost as overlapping (Figure 13).

CapSat4 was detected at five loci, each on a different chromosome. The signals were both dot-like and band-like and where all restricted to the subterminal regions of the chromosomes. Signals were strong and detected in all cells analyzed (Figure 13).

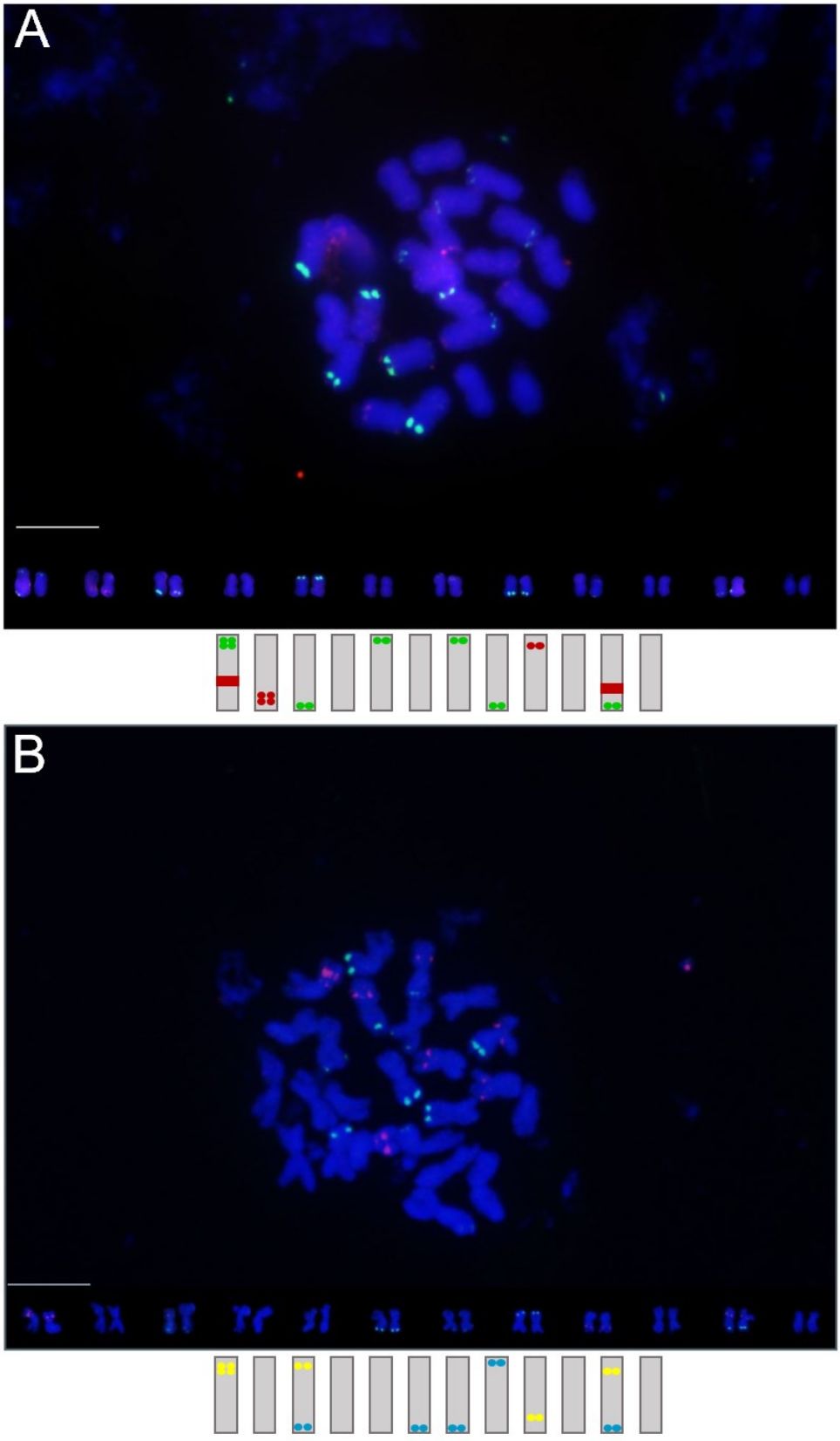



**Figure 12**. Localization of satellite DNAs CapSat1, CapSat4, CapSat5 and CapSat8 in mitotic metaphase chromosomes of *C. eshbaughii* (2*n* = 24) accompanied by cut-out karyotypes and idiograms. A, localization of CapSat1 (green signals, green in idiogram) and CapSat8 (red signals, red in idiogram). B, localization of CapSat4 (green signal, blue in idiogram) and CapSat5 (red signal, yellow in idiogram). Scalebar = 5 µm.

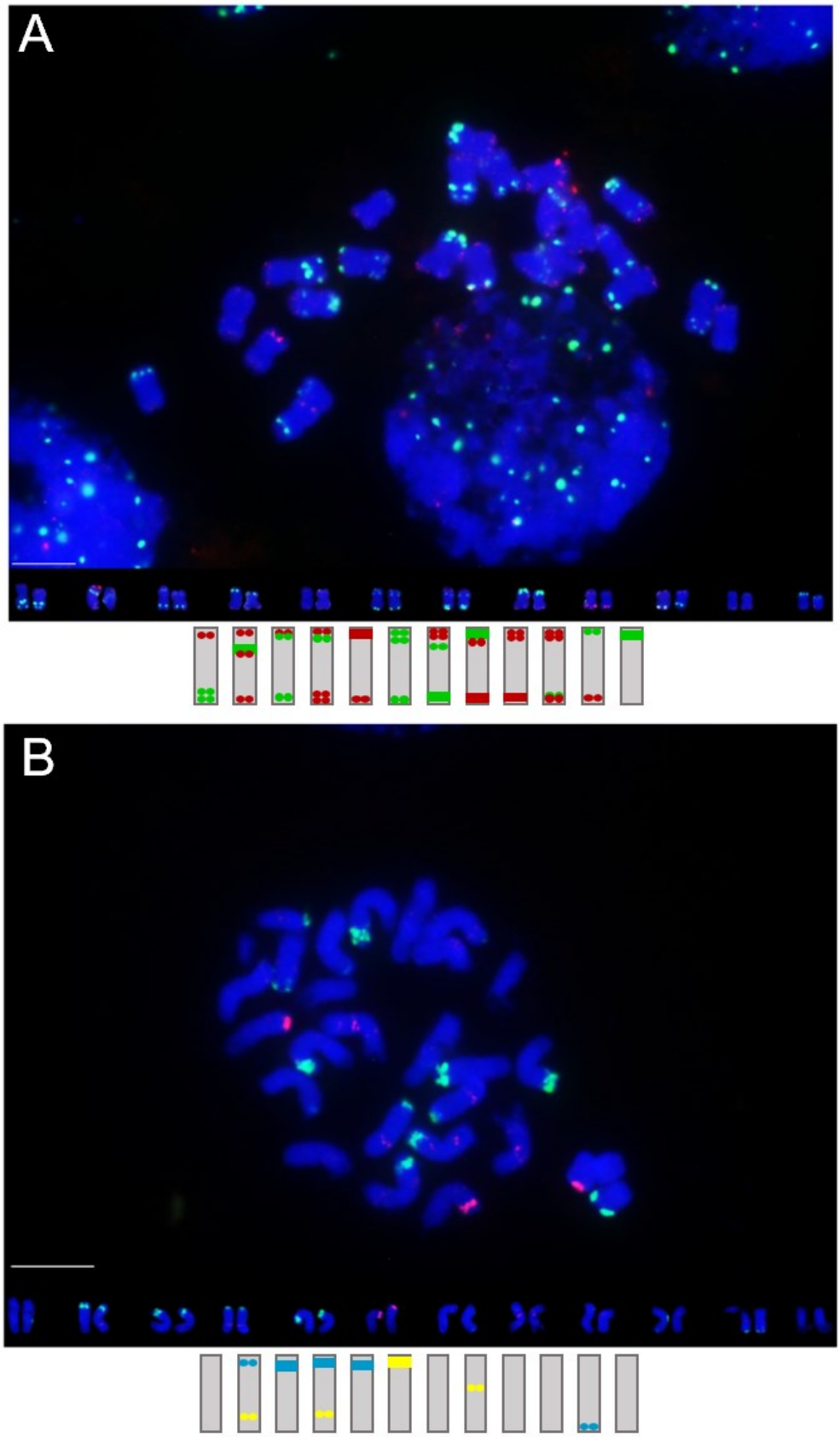



**Figure 13.** Localization of satellite DNAs CapSat1, CapSat4, CapSat5 and CapSat8 in mitotic metaphase chromosomes of *C. pubescens* (2*n* = 24) accompanied by cut-out karyotypes and idiograms. A, localization of CapSat1 (green signals, green in idiogram) and CapSat8 (red signals, red in idiogram). B, localization of CapSat4 (green signal, blue in idiogram) and CapSat5 (red signal, yellow in idiogram). Scalebar = 5 µm.

CapSat5 was found at four loci distributed across four different chromosomes. All appeared dot-like, except for the locus on chromosome 6, which was band-like. The loci positions varied, with one subterminal and three pericentric. All loci were strong and detected in all cells analyzed (Figure 13).

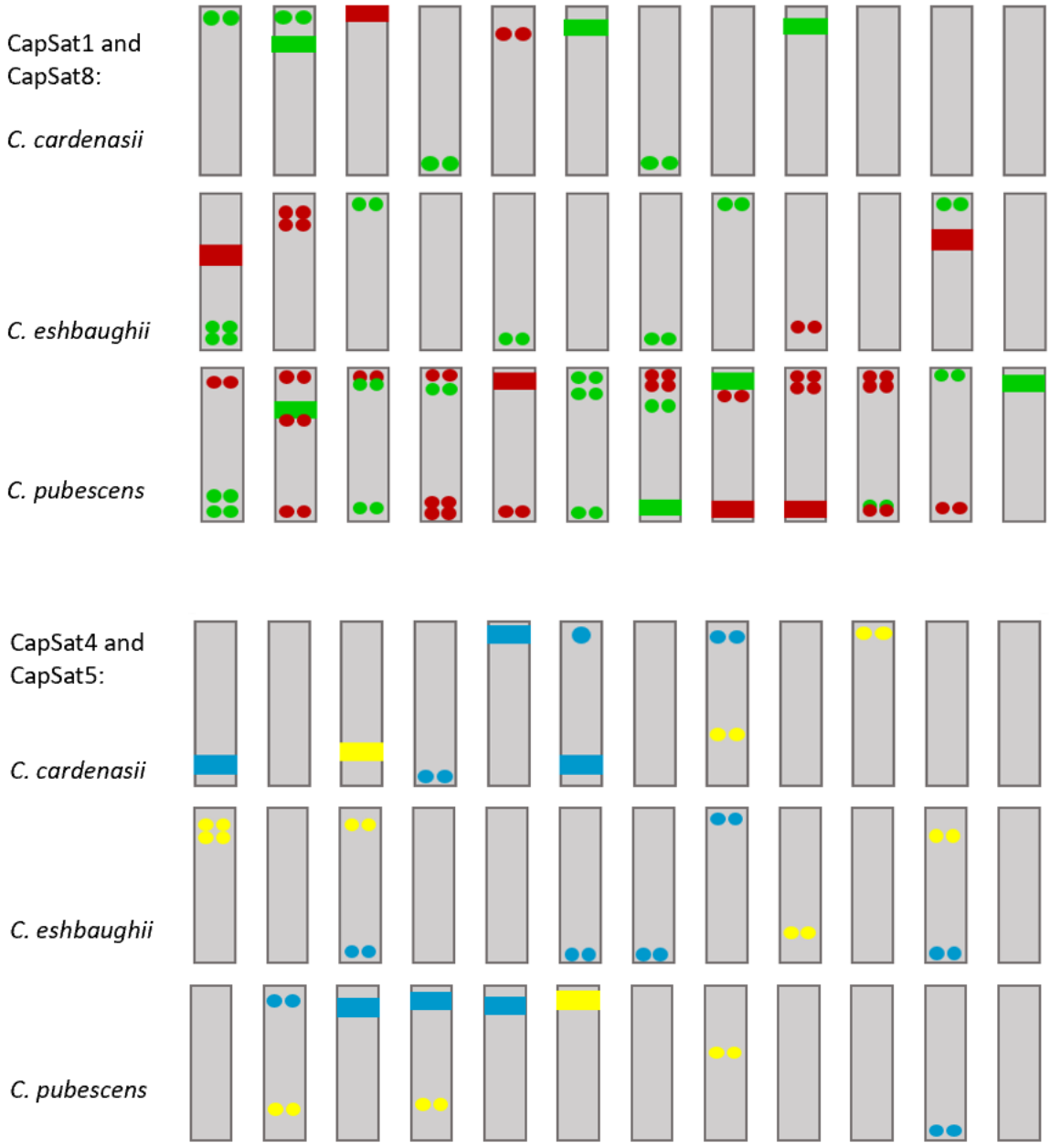



**Figure 14.** Summary of loci numbers and distribution of CapSat1 (red), CapSat4 (blue), CapSat5 (yellow) and CapSat8 (red) in *C. cardenasii*, *C. eshbaughii* and *C. pubescens*.

### 3.2.4 Reprobing

Chromosome reprobing has also been attempted. Initially hybridized probes were washed off and the same slides were hybridized with a new satellite DNA probe set (Figure 15, Figure 16).

*Capsicum cardenasii* had been hybridized with CapSat1 and CapSat8, images were taken and the probes were washed off. The slide was subsequently hybridized with CapSat4 and CapSat5. Although a few additional loci were detected, the results could not be used for further analysis, because the originally mapped probes were not washed away efficiently resulting in signal overlay. Additionally, some chromosomes were lost during the second hybridization step. Reprobing has also been attempted on the preparations used for whole chromosome painting. The results were better because the original oligo paint probes were completely removed (Figure 7B reprobing with CapSat4 and CapSat5).

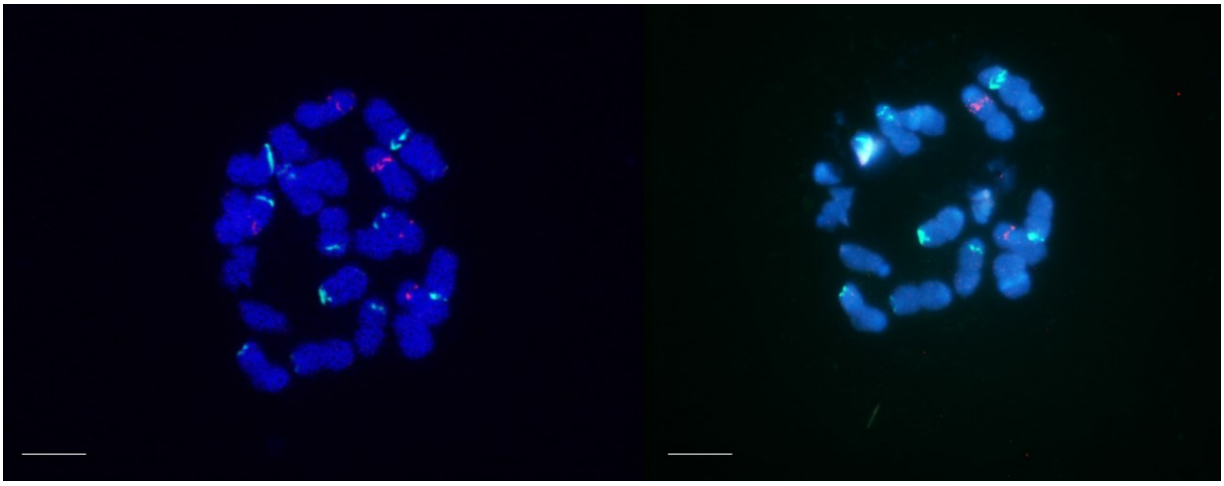


**Figure 15.** Reprobing of chromosomes of *C. cardenasii* after first hybridization round with CapSat1 and CapSat8 (left; CapSat1 with green signals and CapSat8 with red signals) with CapSat4 and CapSat5 (right; CapSat4 with green signals and CapSat5 with red signals). Scalebar = 5 µm.

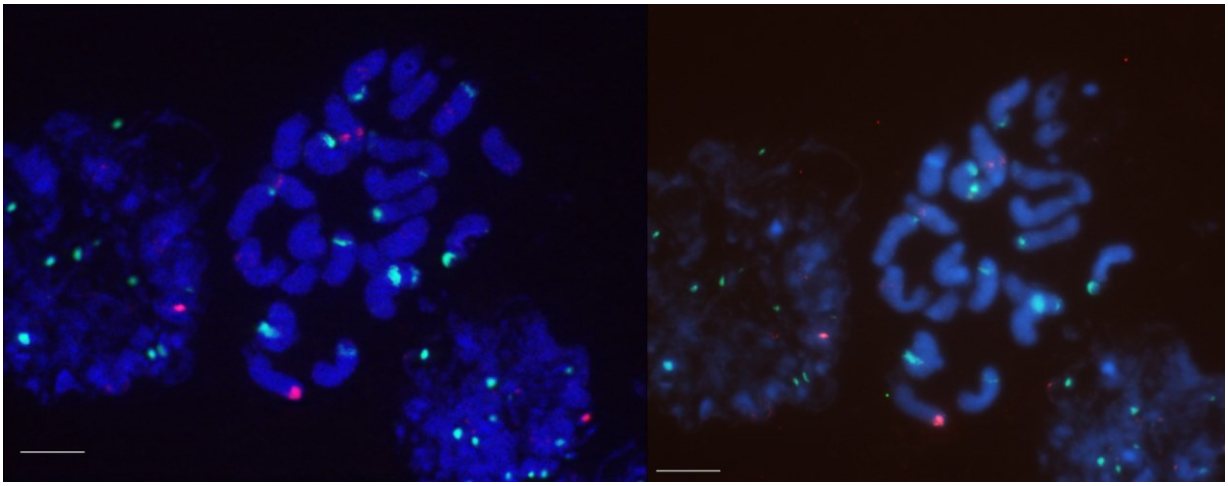


**Figure 16.** Reprobing of chromosomes of *C. cardenasii* after first hybridization round with CapSat1 and CapSat8 (left; CapSat1 with green signals and CapSat8 with red signals) with CapSat4 and CapSat5 (right; CapSat4 with green signals and CapSat5 with red signals). Scalebar = 5 µm.

### 3.2.5 Preliminary data on other satellite DNAs abundance and localization

Additional satellite DNA probes were localized in chromosomes of *Capsicum eshbaughii* and *Capsicum pubescens*: probe combination of CapSat6 and CapSat17 (Figure 17), CapSat7 and CapSat8 and CapSat7 with 35S rDNA (Figure 18).

In *Capsicum eshbaughii*, CapSat6 was detected in five loci, mainly located in the interstitial and subterminal regions of the chromosomes. Exclusively dot-like strong signals were found. CapSat17 was not detected in this species at all (Figure 17). In *Capsicum pubescens*, CapSat6 was present at three loci, with two loci positioned on the opposite arms of the same chromosome. One locus was larger, and band-like, while the other loci had a dot-like appearance. CapSat17 was detected at five loci, both dot-like and band-like (Figure 17).

In *Capsicum eshbaughii* two loci of 18S rDNA (part of 35S rDNA) were observed in the haploid chromosome set. This sample was also hybridized with CapSat7, for which no signal was detected (Figure 18).

The combination of CapSat7 and CapSat8 was tested in all three species, but the results were inconclusive. In most of the preparations, the hybridization produced either a single, overlapping signal of both probes or no detectable signal at all in the condensed chromosomes. In the interphase nuclei signals were observed more often, but also with inconsistent patterns. Only in *Capsicum pubescens* discrete signals were observed, but only in incomplete metaphase plates. The signals for CapSat8 were in general very weak and faded with longer exposure under the fluorescence microscope (Figure 18).

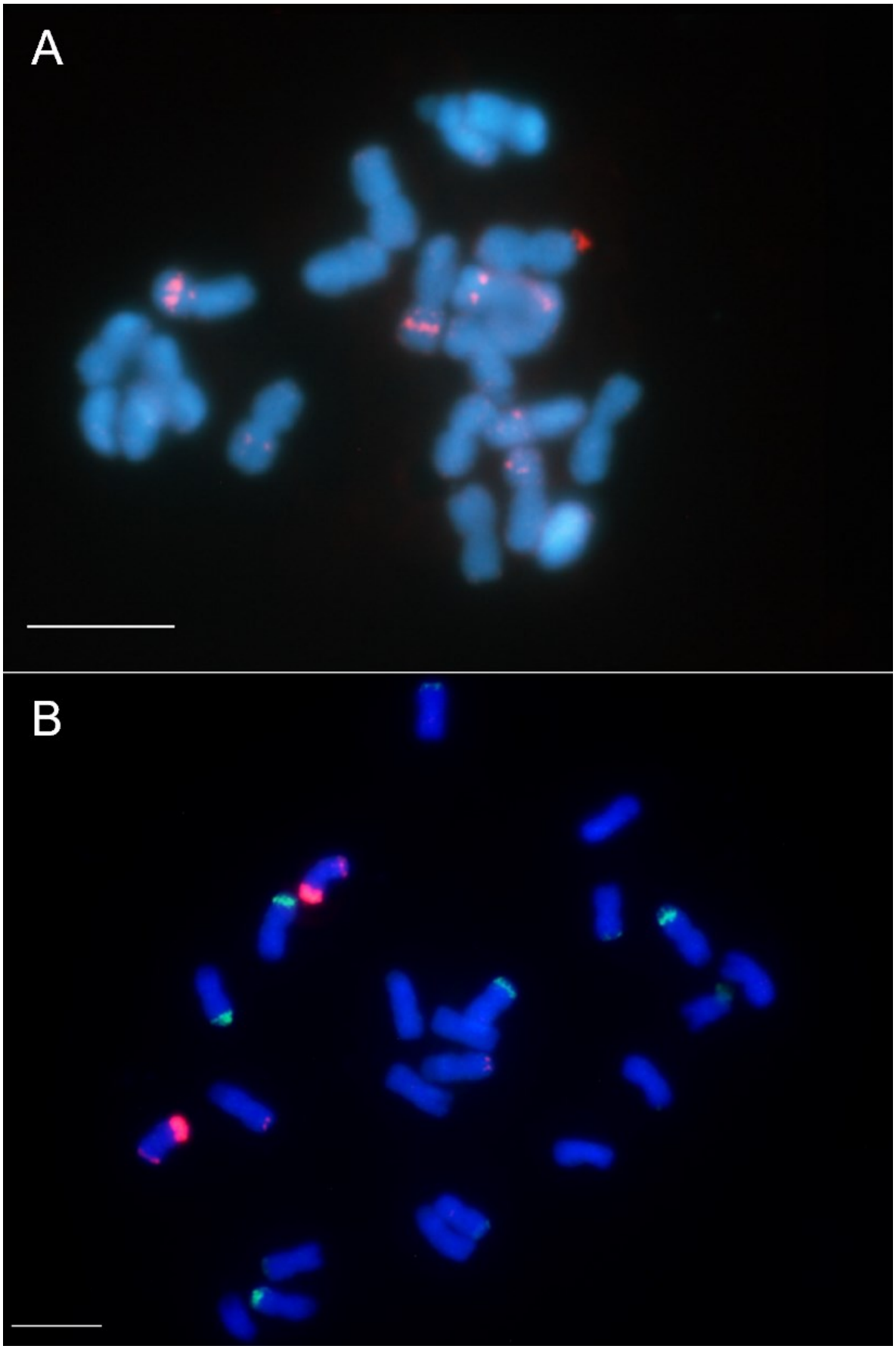



**Figure 17.** Localization of satellite DNAs CapSat6 (red signal) and CapSat17 (green signal) in mitotic metaphase chromosomes. A, *C. eshbaughii*. B, *C. pubescens*. Scalebar = 5 µm.

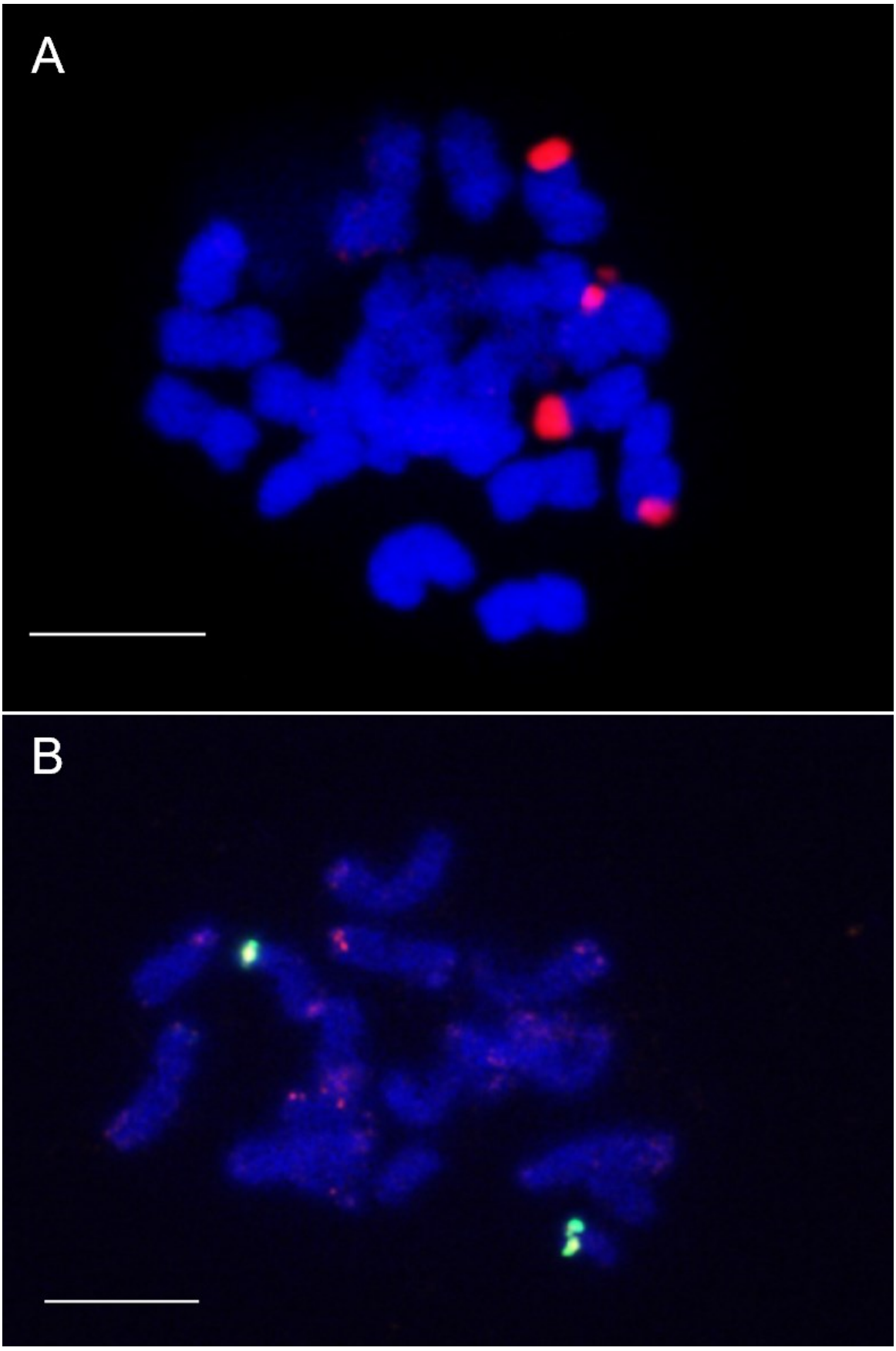



**Figure 18.** Localization of 35S rDNA, CapSat7 and CapSat8 in mitotic metaphase chromosomes. A, *C. eshbaughii*, with 35S rDNA (red signal) and no signals of CapSat7 (green signal). B, *C. pubescens*, with CapSat7 (green signal) and CapSat8 (red signal). Scalebar = 5 µm

## 3.3 Capsaicinoid quantification

Capsaicinoids were quantified using HPLC in 25 dried, whole fruits samples, representing three species of Pubescens clade (*Capsicum cardenasii*, *Capsicum eximium* and *Capsicum pubescens*), two other wild species (*Capsicum chacoense* and *Capsicum rhomboideum*) and all other domesticated species (*Capsicum annuum*, *Capsicum baccatum*, *Capsicum chinense* and *Capsicum frutescens*) (Table 1). The number of *Capsicum pubescens* samples was higher to evaluate the levels of variations in this species encompassing both pungent and non-pungent varieties.

Quantification was based on a calibration curve described by the equation $y = 10.587x – 10.926$ with an $R^2$ value of 0.9999. The recovery test resulted in a value of 96 %, indicating the extraction was highly efficient and that only minimal amounts of material were lost during this process.

The peak areas obtained from the HPLC chromatograms were used for the quantification of the capsaicinoids, at 230 nm. Individual peaks were identified based on their retention time and UV spectra (Figure 19). Among the detected compounds, the peak with the lowest retention time corresponds to nordihydrocapsaicin, followed by capsaicin and dihydrocapsaicin. As the peak area is directly proportional to the concentration of the corresponding capsaicinoid, these values were used to calculate the amount of capsaicinoid in each sample, expressed as capsaicinoids in µg/g dry material. The calculation was calibrated for the amount of plant material used for extraction. For each sample, three replicate measurements were performed and the mean value from the peak areas was used for further analysis (Table 3).

The capsaicinoid concentrations in the dried fruit samples varied substantially among the analyzed taxa (Figure 20, Table 4). The measured values ranged from undetectable amounts in several samples, including *Capsicum rhomboideum*, *Capsicum annuum* var. *annuum* cv. "Lippenstift", *Capsicum chinense* cv. "Gänseschnabel", *Capsicum pubescens* cv. CAP363, *Capsicum chinense* #2040 and *Capsicum pubescens* cv. "Turbo Pube", to exceptionally high concentrations of 11.88 mg (= 11877.71 µg) of capsaicinoids per gram of dry material in *Capsicum chinense* cv. “Trinidad Moruga” (yellow). Among the wild accessions *Capsicum eximium* from Totora and Tomina (Bolivia) reached 4.97 mg and 3.09 mg of capsaicinoids per

gram of dry material, while *Capsicum cardenasii* contained 4.46 mg/g dry material. *Capsicum chacoense* also exhibited substantial capsaicin levels with 1.73 mg/g dry material (Figure 20).

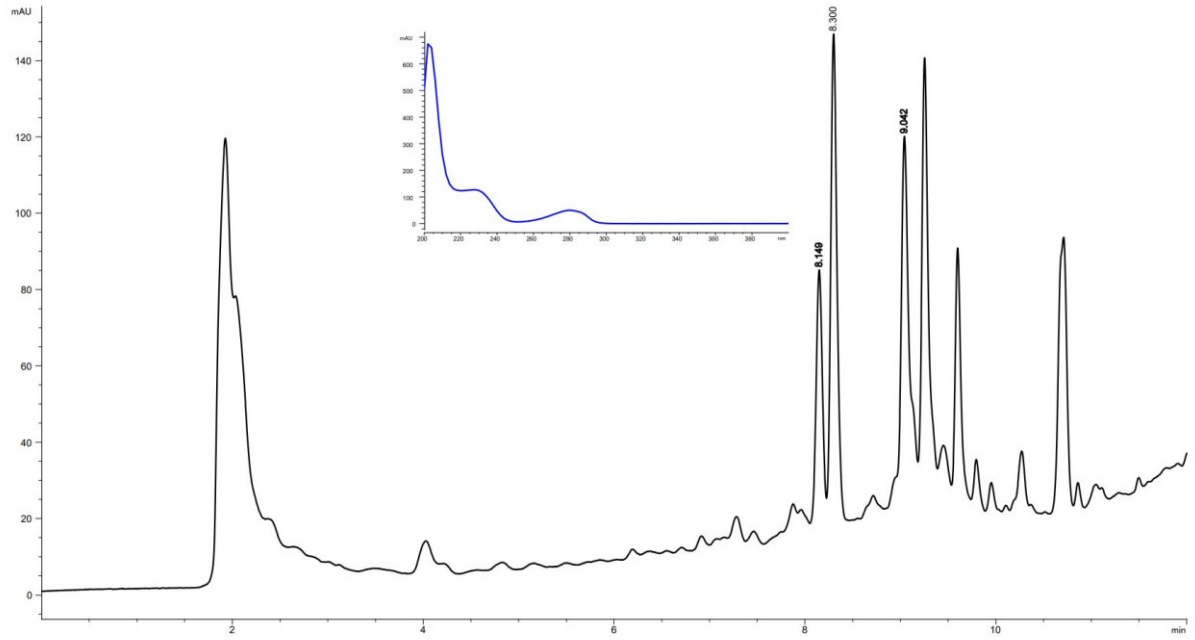



**Figure 19.** Example HPLC chromatogram of *C. eximium* with corresponding UV spectrum (blue). The major peaks represent individual capsaicinoids, identified based on their retention time and UV spectra. From the lowest to highest retention time, the peaks correspond to nordihydrocapsaicin**,** capsaicin, and dihydrocapsaicin.

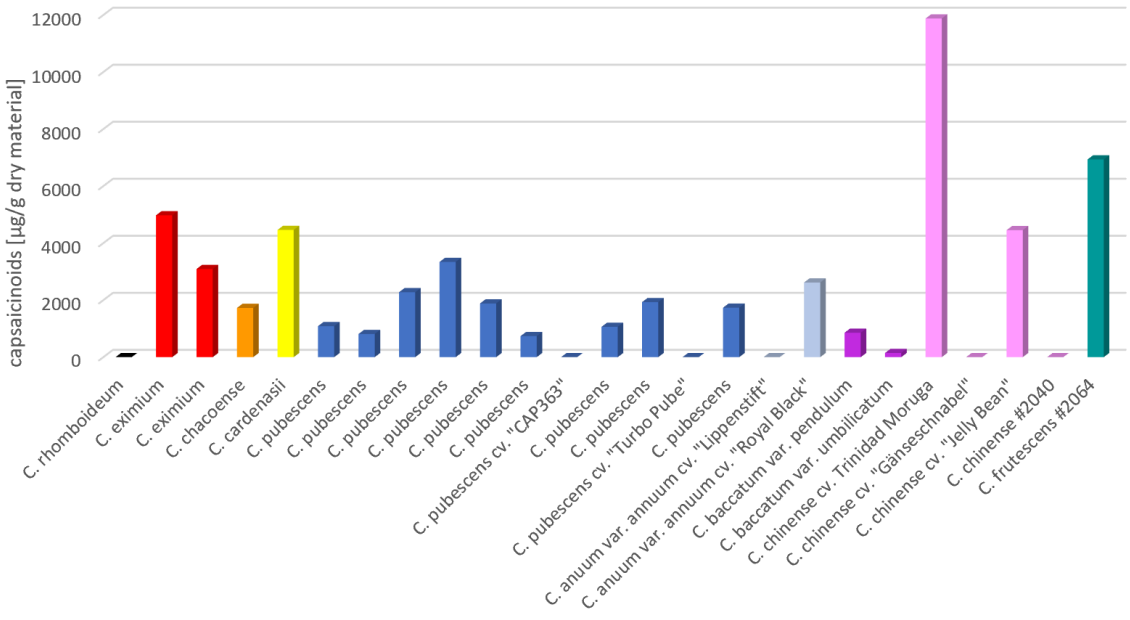



**Figure 20.** Mean total capsaicinoid amounts [µg/g dry material] across all analyzed species of *Capsicum*. Values represent the sum of all detected capsaicinoid derivatives, with means calculated from three replicate measurements per sample. Red, orange and yellow colours used for wild species, and blue, purple, pink colours for cultivated species.

Within cultivated species *Capsicum chinense* cv. “Trinidad Moruga” (yellow) exhibited by far the highest capsaicin concentration, reaching 11.88 mg/g dry material. This was followed by *Capsicum frutescens* #2064 with 6.93 mg/g dry material and *Capsicum chinense* cv. “Jelly Bean” with 4.45 mg capsaicinoids per gram dry material. While *Capsicum annuum* var. *annuum* cv. "Lippenstift" showed no capsaicinoids after extraction, *Capsicum annuum* var. *annuum* cv. "Royal Black” had a relatively high capsaicinoid value with 2.62 mg/g dry material. The *Capsicum pubescens* accessions showed moderate capsaicinoid levels, ranging mostly between 738.14 µg/g dry material and 3.34 mg/g dry material. Two of those *Capsicum pubescens* accessions, cv. "CAP363" and “Turbo Pube” had no detectable amount of capsaicinoids. The *Capsicum baccatum* var. *umbilicatum* and var. *pendulum* species exhibited comparatively low concentrations, in comparison to the other cultivated taxa, with 146.82 and 857.01 µg/g dry material (Figure 20).

Within *Capsicum pubescens*, the capsaicinoid levels varied widely among the analyzed accessions, ranging from complete absence in *Capsicum pubescens* cv. "CAP363" from Guatemala and “Turbo Pube”, which was grown from seeds bought online, to relatively high concentrations in other accessions (Figure 21). The highest value was measured in the sample from Trujillo, Peru with 3.34 mg/g dry material, followed by *Capsicum pubescens* from the Yungas area and from San Isidro, both from Bolivia with 2.28 mg and 1.93 mg/g dry material, respectively. Intermediate levels of capsaicinoids were measured in the sample #4 rojo, Huancane, Bolivia (1.88 mg/g dry material) and the sample Moro Moro from Bolivia (1.73 mg/g dry material). Lower concentrations were found in the sample from Cusco, Peru (1.09 mg /g dry material) and from Coroico, Bolivia (1.06 mg/g dry material). The lowest amounts of capsaicinoids, less than 1 mg/g dry material were detected in the samples from Villa serrano, Bolivia (812.23 µg/g dry material) and in the fruits bought at the Merced Market in Mexico (738.14 µg/g dry material).

A comparison of capsaicinoid concentrations between wild and cultivated *Capsicum* species revealed a distinct distribution pattern (Figure 22). Wild species possessed a higher median capsaicinoid content compared to the cultivated ones, along with a broader interquartile range, indicating a bigger variability in pungency in the latter group. While the maximum non-outlier values for wild samples were slightly below 5 mg/g dry material, the cultivated ones

contained two prominent outliers, reaching approximately 6.8 mg/g dry material and 12 mg/g dry material, respectively. Overall, wild *Capsicum* species had higher capsaicinoid concentrations albeit with overall lower levels of variation than the cultivated analyzed samples.

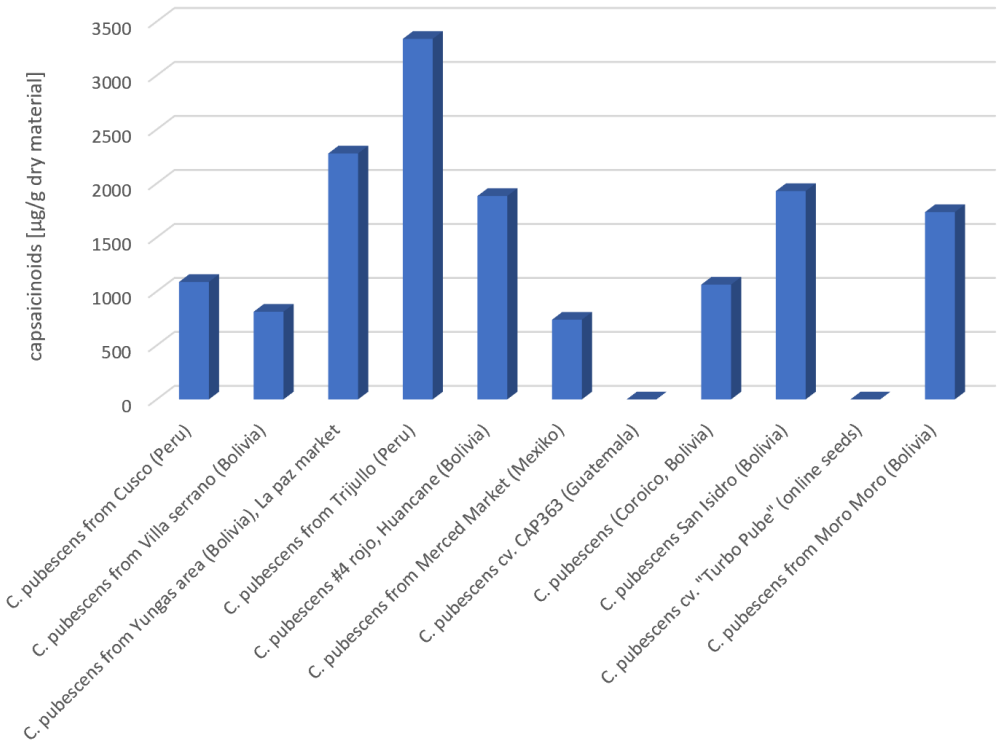



**Figure 21.** Mean total capsaicinoid amounts [µg/g dry material] in the *C. pubescens* accessions. Values represent the sum of all detected derivatives, averaged from three replicate measurements.

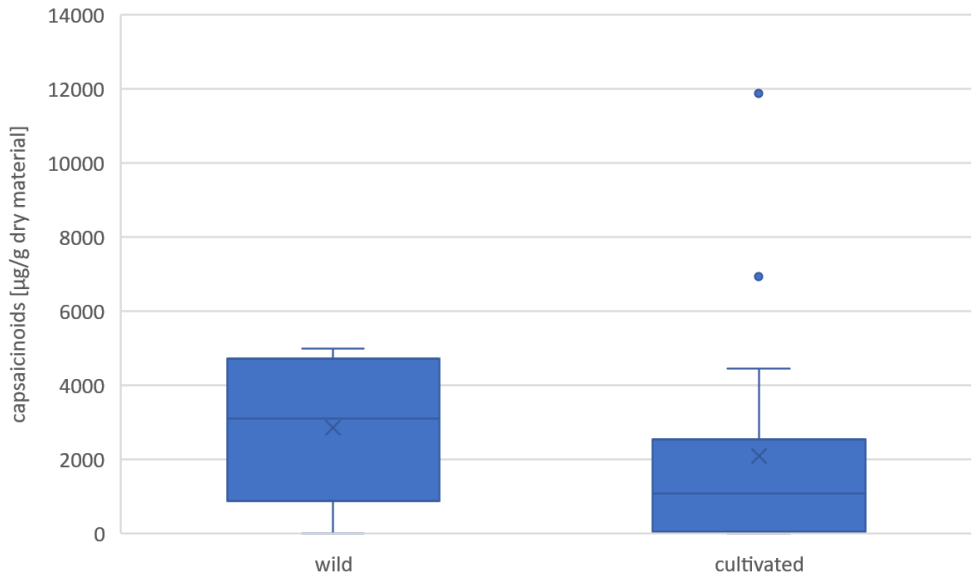



**Figure 22.** Comparison of mean total capsaicinoid amounts [µg/g dry material] between wild and cultivated *Capsicum* species presented in boxplots. Values represent the sum of all capsaicinoid derivatives, with means calculated from three replicate measurements per accession. The x-axis differentiates wild versus cultivated groups, the y-axis shows capsaicinoid content [µg/g dry material].

The data allowed also to identify and compare the individual capsaicinoid derivates based on their retention times. Altogether, three major compounds were detected, nordihydrocapsaicin (A1), capsaicin (A2) and dihydrocapsaicin (A3). In one sample, *Capsicum chinense* cv. "Trinidad Moruga", a fourth derivate was found, but only in a small proportion (1.78 %). This one could not be identified, but it is likely to be homodihydrocapsaicin.

The distribution of the three main derivates varied a lot across species (Figure 23, Table 5). Dihydrocapsaicin (A3) was especially prominent in several wild species such as *Capsicum eximium*, *Capsicum chacoense* and *Capsicum cardenasii*, where it accounted for more than 60 % up to almost 70 % of the total capsaicinoid fraction. In one *Capsicum pubescens* accession from Mexico, dihydrocapsaicin was the most abundant derivate. In contrast, another *Capsicum pubescens* sample lacked dihydrocapsaicin entirely (LP18, *Capsicum pubescens* from La Merced Market, Mexico). Apart from that, dihydrocapsaicin was present in all samples but mainly remaining below 40 %. Capsaicin (A2) was present in all analyzed samples, with relative abundances ranging from 12 % in *Capsicum eximium* up to 78 % in *Capsicum chinense* cv. "Trinidad Moruga". In several species capsaicin was the dominant compound, especially in the cultivated species such as *Capsicum annuum*, *Capsicum baccatum*, *Capsicum chinense*, *Capsicum frutescens* and also some *Capsicum pubescens* accessions. Nordihydrocapsaicin (A1) was present in nearly all accessions, with the exception of *Capsicum baccatum* var. *umbilicatum*. While A1 was generally the least abundant derivative with an average of 20 % overall, in some *Capsicum pubescens* accessions it contributed up to 45 % of the total capsaicinoid fraction.

Although capsaicin and dihydrocapsaicin are the dominant compounds across most *Capsicum* species, the balance between derivatives is highly variable. Particularly in *Capsicum pubescens*, a high degree of intraspecific variation was observed. The individual accessions differ strongly in their capsaicinoid profiles, and thus no overall pattern for *Capsicum pubescens* could be established.

### 3.4 PCR amplification of the *Pun*1 gene

To detect the presence of the *Pun*1 gene, which encodes for a putative acyltransferase involved in capsaicinoid biosynthesis, PCR amplification was performed. The same set of species used for HPLC analysis was included: *Capsicum annuum*, *Capsicum baccatum* var. *pendulum*, *Capsicum cardenasii*, *Capsicum chinense*, *Capsicum eximium*, *Capsicum frutescens*,

*Capsicum pubescens*. Due to a lack of plant material and DNA extracts, *Capsicum chacoense* and *Capsicum eshbaughii* were not analyzed.

Clear bands corresponding to the expected fragment size of *Pun*1 were observed in most pungent species, including *Capsicum annuum*, *Capsicum baccatum* var. *pendulum*, *Capsicum chinense*, *Capsicum frutescens* and *Capsicum pubescens*. Notably, no amplification was detected in *Capsicum cardenasii* and *Capsicum eximium*, despite their pungency. In *Capsicum pubescens* two bands were visible, one at the expected size and a second, fainter corresponding to longer DNA fragment (Figure 24).

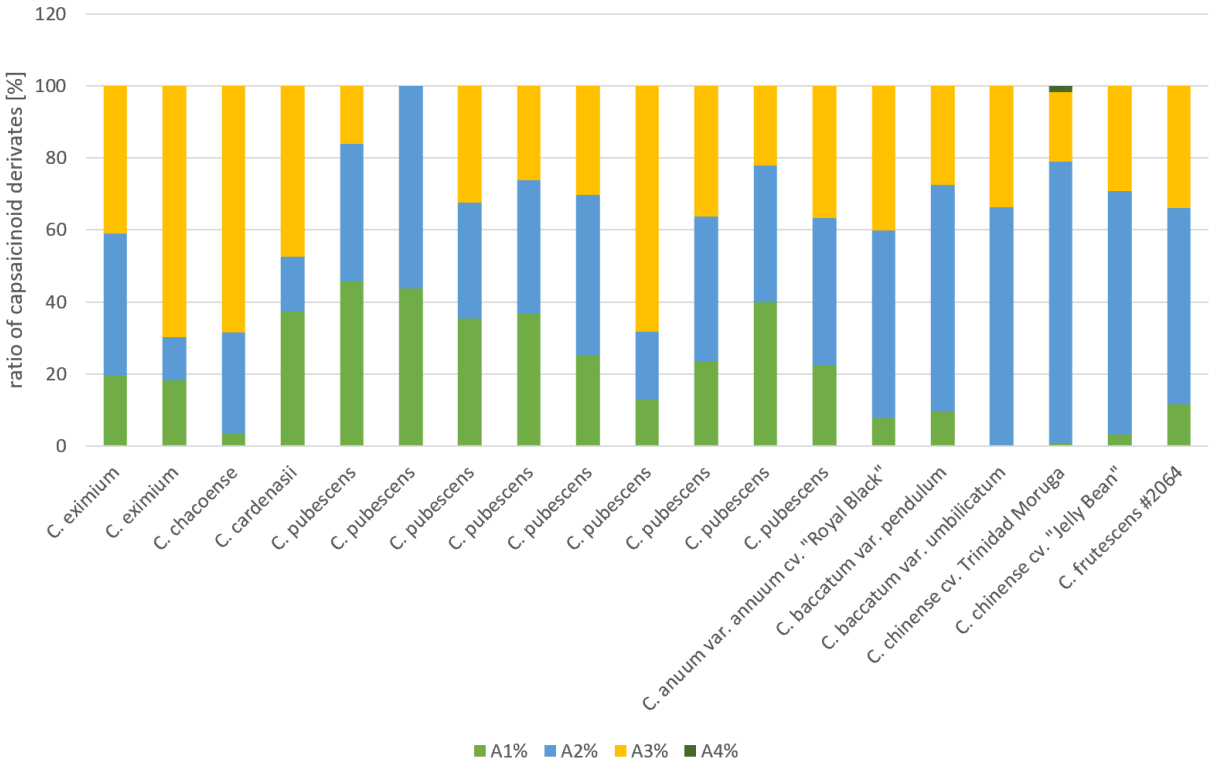



**Figure 23.** Stacked column chart showing the relative proportions of the capsaicinoid derivates. Samples where no derivate was detected were left out for this chart. A1 = Nordihydrocapsaicin, A2 = Capsaicin, A3 = dihydrocapsaicin. A4 = not identified. Ratios were calculated from mean values of three independent HPLC measurements per sample. X-axis represent the species, y-axis indicates the percentage contribution of each derivative to the total capsaicinoid content.

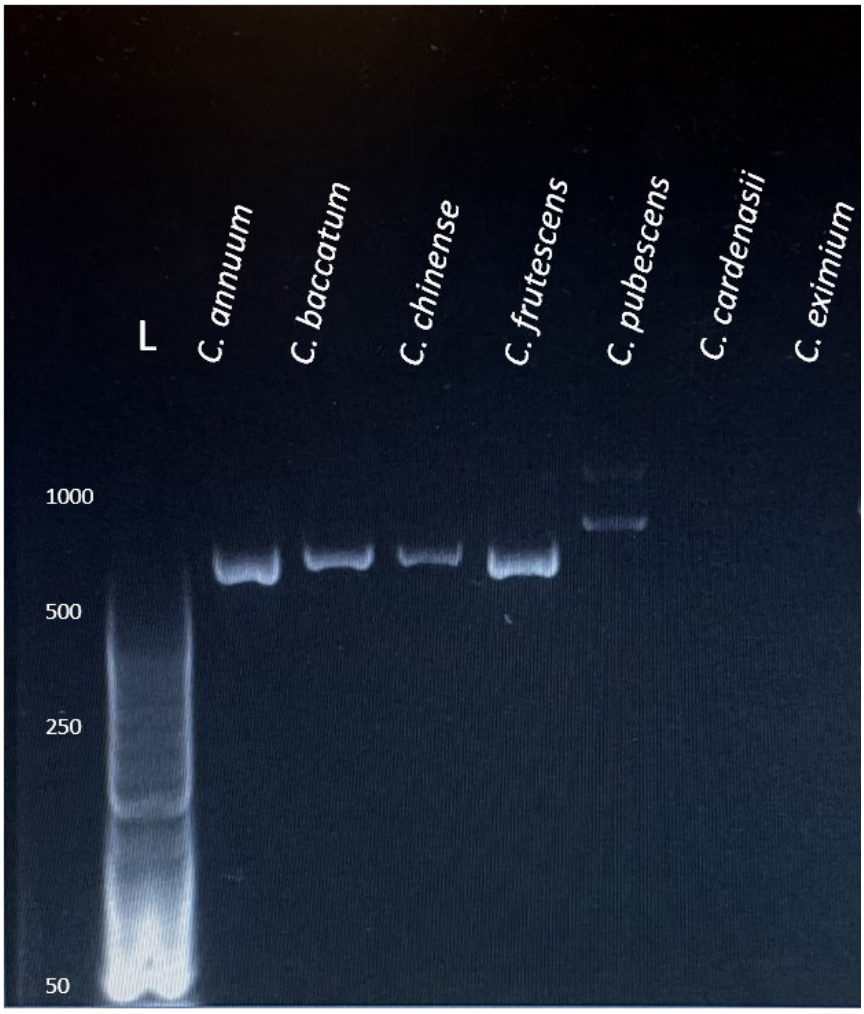



**Figure 24.** PCR amplification of *Pun*1 gene in analyzed *Capsicum* species. Species names are indicated above the lanes. L - DNA ladder (50 bp).

# 4. Discussion

This study focused on the Pubescens clade, combining cytogenetic and chemical approaches for a better understanding of its genomic architecture and metabolic diversity. A primary objective was to investigate genome size variation and to characterize the satellitome profiles of the three species belonging to this clade. In particular, the abundances and chromosomal distributions of shared satellite DNAs were examined using FISH, with the goal of establishing species- and chromosome-specific patterns. In addition, the chemical diversity of the clade was established by quantifying the capsaicinoid content, especially across multiple *Capsicum pubescens* accessions and comparing them with both wild and cultivated *Capsicum* species from other clades. PCR amplification of the *Pun*1 gene, involved in the capsaicin biosynthesis, confirmed its presence in pungent species. Together, these data provide insights into the evolutionary distinctiveness of the Pubescens clade at both the genomic and metabolic level.

## 4.1 Genome size and satellite DNA evolution in the Pubescens clade

Genome sizes of all four analyzed *Capsicum* species representing Pubescens clade were similar, with two subgroups. *Capsicum eshbaughii* and *Capsicum eximium* exhibited lower genome sizes than *Capsicum cardenasii* and *Capsicum pubescens* with higher genome sizes. In general, the genome size is mainly influenced by the presence of repetitive DNA types (Weiss-Schneeweiss et al., 2013; Pellicer et al., 2018).

Satellite DNAs represent an important type of repetitive DNA in plant genomes. Even though they are non-coding, they are abundant and contribute significantly to genome size variation and are also important for chromosomal organization and evolution due to their fast-evolving nature (Macas et al., 2006; Weiss-Schneeweiss & Schneeweiss, 2013; Garrido-Ramos, 2015). In this study the satellitome profiles of *Capsicum cardenasii*, *Capsicum eshbaughii* and *Capsicum pubescens*, all belonging to the Pubescens clade, were analyzed, with the aim of identifying their abundance, diversity and chromosomal localization for the establishment of species-specific distribution patterns. The satellites DNAs CapSat1, CapSat4, CapSat5, CapSat6, CapSat7, CapSat8 and CapSat17 were physically mapped on the chromosomes also to create cytogenetic markers for individual chromosome identification. While certain satellite DNA repeats (CapSat1, CapSat4, CapSat8) were distributed across multiple chromosomes in most species, reflecting their origin from a common ancestral library of satellite DNA repeats, others exhibited a much more restricted chromosomal distribution (CapSat5, CapSat6,

CapSat7, CapSat17). This pattern points to lineage-specific diversification and divergence, likely driven by different mechanisms including rolling-circle replication with reinsertion of the replicated repeats and unequal cross-overs (Feliciello et al., 2006; Garrido-Ramos, 2017).

The comparative FISH analysis revealed both common conserved and species-specific distribution patterns of the satellite DNAs mapped (Figure 14). CapSat1, CapSat4, CapSat5 and CapSat8 were present in all three analyzed species, *Capsicum cardenasii*, *Capsicum eshbaughii* and *Capsicum pubescens*, indicating their common origin in the ancestor of the clade. Nevertheless, the number of loci and their localization highly varied between the species, reflecting the species-specific diversification. CapSat1 was the most abundant satellite DNA across the three species analyzed, with loci distributed across multiple chromosomes. *Capsicum pubescens* displayed the highest copy number of CapSat1 loci with additional band like loci through amplification and also dot-like duplications compared to *Capsicum cardenasii* and *Capsicum eshbaughii*. CapSat1 was almost exclusively localized in the subterminal region of the chromosomes, with one exception of one pericentric locus in *Capsicum pubescens*. CapSat8 was detected in *Capsicum cardenasii* and *Capsicum eshbaughii* in only a few chromosomes, in contrast to *Capsicum pubescens*, where it was amplified across almost all chromosomes. The loci of this satellite DNA were also mainly subterminal, with only a few detected in the pericentric regions in *Capsicum eshbaughii* and *Capsicum pubescens* (Figure 14).

CapSat4 and CapSat5 were less abundant than CapSat1 and CapSat8. *Capsicum eshbaughii* and *Capsicum pubescens* had the highest number of loci of satellite DNAs CapSat4 and CapSat5, whereas only a few were detected in *Capsicum cardenasii*. These loci were also mostly subterminal, with few pericentric ones, including one locus of CapSat4 in *Capsicum pubescens*. In general, fewer duplications within the same locus were observed compared to CapSat1 and CapSat8 (Figure 14).

Overall, *Capsicum pubescens* exhibited the most diverse satellitome profile among the three analyzed species. This was not only reflected in the greater number of loci detected, but also in the morphological diversity of the hybridization patterns. *Capsicum pubescens* was also the only species in which CapSat7 and CapSat17 were detected. Additionally, intraspecific variation in the number and distribution of satellite DNAs observed *Capsicum pubescens*, also

has been reported in other plant systems, in agreement with rapid diversification of satellite DNAs, even within species (Ruiz-Ruano et al., 2016; Garrido-Ramos, 2017; Simon, 2020).

Further analyses will be necessary to expand the understanding of satellite DNA organization in *Capsicum*. Additional probes should be localized, and more extensive reprobing experiments are required. For comparative results, testing different stringency conditions, such as different temperatures during probe removal and hybridization, would be valuable. This would allow multiple satellite DNAs to be mapped and precisely assigned to the same chromosomes. So far, studies on tandem repeats and specifically satellite DNAs in *Capsicum* have largely relied on bioinformatic approaches, such as clustering analyses with RepeatExplorer, and have mainly focused on *Capsicum annuum* (Zhou et al., 2019; Rudenko & Korotkov, 2023), with only limited inclusion of other species like *Capsicum chinense* and *Capsicum baccatum* (de Assis et al., 2023).

## 4.2 Capsaicinoid quantification in *Capsicum*

The quantification of capsaicinoids in this study revealed a distinct variation both among species and within individual accessions of *Capsicum pubescens*. While several cultivated species such as *Capsicum chinense* and *Capsicum frutescens* displayed the highest levels of capsaicinoids content, *Capsicum pubescens* accessions possessed intermediate to low concentrations and, in some cases, no detectable amounts. These findings are consistent with previous reports that describe this species as having lower levels of pungency compared to other domesticated *Capsicum* species (Zewdie et al., 2001; Meckelmann et al., 2015). Importantly, no clear geographical trend was observed within the *Capsicum pubescens* accessions examined here, as samples from both Bolivia and Peru had both high and low capsaicinoid contents. This suggests that factors such as local selection, genetic background and generally, human assisted selection may play a bigger role than geographical origin in determining capsaicin levels. An exception to the overall pattern of the *Capsicum pubescens* accessions was the sample of *Capsicum pubescens* cv. “Turbo Pube” (LP23), grown from online-bought seeds. It is advertised as extremely pungent (rated 9/10 on an online scale from 0 to 10) and despite its morphological characters corresponding to *Capsicum pubescens* (e.g., presence of black seeds in the fruit), no capsaicin was detected in this accession, even after repeated measurement of three replicates. In addition, a taste test of the fruits was conducted and also no pungency was detected organoleptically. This discrepancy may be explained by

either mislabeling or incorrect seed supply, as the chemical results did not match the expected pungency. The observed variability highlights the chemical diversity within *Capsicum pubescens* and the importance of analyzing multiple accessions when characterizing pungency profiles. It also offers an excellent and reliable tool for testing the commercially available accessions.

*Capsicum rhomboideum* was confirmed to lack capsaicinoids in this study, which is consistent with previous reports that this species does not produce pungent secondary metabolites due of the lack of the pungency metabolic pathway in whole Andean clade to which this species belongs (Carrizo García et al., 2016; Luna-Ruiz et al., 2018). Comparison of cultivated and wild species, revealed the spectrum of pungency levels. Cultivated *Capsicum annuum* accessions included a non-pungent as well as a pungent type, whereas other cultivated taxa like *Capsicum frutescens* and *Capsicum chinense* cv. Trinidad Moruga (yellow) had extremely high concentrations of over 8 mg of capsaicinoids per gram of dry material, which is consistent with its classification among the world's hottest chili cultivars (Bosland et al., 2012). This demonstrates that domestication has not led to a simple increase or reduction of capsaicinoid levels, but rather enabled the development of cultivars with broadly varying pungency levels, from non-pungent to extremely hot. Even though previous studies showed that cultivation is often associated with a reduction in the diversity of secondary metabolites, targeted breeding can lead to the increase of specific compounds like capsaicinoids (Luna-Ruiz et al., 2018; Cervantes-Hernández et al., 2022).

In contrast, wild species often displayed moderate to high levels of pungency, reflecting the ecological importance of capsaicinoids for the defense against herbivores and pathogens (Tewksbury & Nabhan, 2001; Tewksbury et al., 2006, 2008; Luna-Ruiz et al., 2018)**.** This trend was also seen here, with wild taxa exhibiting on average higher concentrations (the mean value was approximately 2.8 mg of capsaicinoids per gram of dry material) than most cultivated accessions. However, these outcomes depend strongly on sampling**.** The presence of the outliers in the samples of the cultivated group analyzed here suggests the overall tendency of cultivated accessions to have lower capsaicin levels than wild ones on average, despite broader levels of variation. Even though certain taxa are typically pungent, non-pungent variants have been documented as well (Tewksbury et al., 2006; Tripodi & Kumar,

2019). In this study, *Capsicum annuum* and *Capsicum chinense* both had pungent and non-pungent forms.

The comparison of individual forms of capsaicinoids and their ratios across the analyzed species revealed a clear trend between cultivated and wild taxa. In cultivated taxa, capsaicin was consistently the most prominent compound, often making up the majority of the total capsaicinoid fraction, whereas in wild species such as *Capsicum eximium*, *Capsicum chacoense*, and *Capsicum cardenasii*, dihydrocapsaicin was a dominant form, accounting for nearly 70 % of the total content in some samples.

Previous studies on the capsaicinoids of *Capsicum pubescens* have reported relatively stable ranges of capsaicinoid proportions, with dihydrocapsaicin being the most abundant one, followed by capsaicin and nordihydrocapsaicin (Zewdie et al., 2001; Meckelmann et al., 2015). The current data shows a much higher intraspecific variation. While several accessions had comparable capsaicin content and profiles to the reported ones, others deviated strongly. For example, dihydrocapsaicin was earlier reported to be abundant in this species (34–53 %), but it was found completely lacking in some accessions in this study. Whereas the capsaicin levels in the analyzed accessions, fit rather well within the published ranges, nordihydrocapsaicin content deviated. Since earlier studies using HPLC were based on relatively few accessions, the new dataset with more varieties analyzed here provides valuable additional data, suggesting that capsaicinoid ratios in *Capsicum pubescens* might be more variable than previously shown.

Finally, it should be emphasized that pungency levels are not fixed traits, but can vary a lot with harvest stage and the environmental conditions under which the plants are cultivated. Differences in growth temperature, soil, water and nutrient availability, light or sun exposure and fruit maturity are all known to strongly influence the capsaicinoid production and accumulation (Johnson & Decoteau, 1996; Contreras-Padilla & Yahia, 1998; Aloni et al., 2001; Gurung et al., 2011; Phimchan et al., 2012; Zhang et al., 2016; Naves et al., 2019). Moreover, biotic factors like herbivores or pathogens were shown to enhance in general secondary metabolite production and here also the capsaicinoid production, which highlights their function as ecological defense compounds (Tewksbury et al., 2008; Al-Khayri et al., 2023). The genetic makeup of each accession may also plays a role, as different genotypes may have distinct capacities for capsaicinoid biosynthesis and regulation (Zewdie & Bosland, 2000;

Cisneros-Pineda et al., 2007; Gurung et al., 2011). Consequently, the values reported here reflect the combined effects of genotype, environment and harvest stage. Even within a single plant the fruits might already differ in their capsaicinoid concentrations, which even more underscores the variability of pungency in *Capsicum*. Additionally, the efficiency of recovering placental tissue during extraction may also influence the measured capsaicinoid content.

The results of the PCR amplification of the *Pun*1 gene confirmed its presence in several pungent species, consistent with their ability to produce capsaicinoids. However, the amplification success was not uniform across all samples. For example, in *Capsicum pubescens*, only faint bands were observed, and in *Capsicum cardenasii* and *Capsicum eximium* no amplification was detected, despite these taxa being known to produce capsaicinoids. This likely reflects gene sequence divergence in the *Pun*1 locus between different *Capsicum* species, which could reduce primer binding efficiency, as the primers used in this study were originally designed for *Capsicum annuum* (C. Carrizo Garcia, unpubl.). These findings highlight the need to design and test multiple primer sets, ideally targeting conserved regions, to reliably detect *Pun*1 across diverse *Capsicum* species. Nevertheless, the detection of capsaicinoids in HPLC measurements indirectly confirmed the presence of *Pun*1, even when PCR amplification was unsuccessful.

Overall, HPLC proved to be a precise analytical method for quantifying capsaicinoids and distinguishing between their individual derivatives. The results of this study expand the knowledge on *Capsicum pubescens*, by revealing not only a broad range of total capsaicinoid content, but also striking variation in derivative ratios among accessions. This highlights the considerable intraspecific diversity within the species. More generally, the comparison between wild and cultivated taxa demonstrated that all levels of pungency can be found across *Capsicum*, ranging from completely non-pungent to extremely pungent forms. While cultivated species often displayed a predominance of capsaicin, wild relatives were more frequently characterized by higher proportions of dihydrocapsaicin. Together, these findings emphasize the complex interplay of genetic, physiological, and environmental factors shaping capsaicinoid biosynthesis and provide new insights into the chemical diversity of the genus.

# 5. References


Adaszek, Ł., Gadomska, D., Mazurek, Ł., Łyp, P., Madany, J., & Winiarczyk, S. (2019). Properties of capsaicin and its utility in veterinary and human medicine. *Research in Veterinary Science*, *123*, 14–19.

Al-Khayri, J. M., Rashmi, R., Toppo, V., Chole, P. B., Banadka, A., Sudheer, W. N., Nagella, P., Shehata, W. F., Al-Mssallem, M. Q., Alessa, F. M., Almaghasla, M. I., & Rezk, A. A. (2023). Plant secondary metabolites: The weapons for biotic stress management. *Metabolites*, *13*(6).

Almojil, D., Bourgeois, Y., Falis, M., Hariyani, I., Wilcox, J., & Boissinot, S. (2021). The structural, functional and evolutionary impact of transposable elements in eukaryotes. *Genes*, *12*(6), 918.

Aloni, B., Peet, M., Pharr, M., & Karni, L. (2001). The effect of high temperature and high atmospheric CO2 on carbohydrate changes in bell pepper (*Capsicum annuum*) pollen in relation to its germination. *Physiologia Plantarum*, *112*(4), 505–512.

Ayarpadikannan, S., & Kim, H.-S. (2014). The impact of transposable elements in genome evolution and genetic instability and their implications in various diseases. *Genomics & Informatics*, *12*(3), 98.

Baldwin, B. G., Sanderson, M. J., Porter, J. M., Wojciechowski, M. F., Campbell, C. S., & Donoghue, M. J. (1995). The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden*, 247–277.

Baranyi, M., & Greilhuber, J. (1996). Flow cytometric and Feulgen densitometric analysis of genome size variation in *Pisum*. *Theoretical and Applied Genetics*, *92*, 297–307.

Barboza, G. E., Carrizo García, C., Bianchetti, L. B., Romero, M. V., & Scaldaferro, M. (2022). Monograph of wild and cultivated chili peppers (*Capsicum* L., Solanaceae). *PhytoKeys*, *200*, 1–423.

Barchenger, D. W., Naresh, P., & Kumar, S. (2019). Genetic resources of *Capsicum*. In N. Ramchiary & C. Kole (Eds.), *The Capsicum Genome* (pp. 9–23). Compendium of Plant Genomes. Springer Verlag, Cham.

Benachenhou, F., Sperber, G. O., Bongcam-Rudloff, E., Andersson, G., Boeke, J. D., & Blomberg, J. (2013). Conserved structure and inferred evolutionary history of long terminal repeats (LTRs). *Mobile DNA*, *4*(1), 5.

Biscotti, M. A., Olmo, E., & Heslop-Harrison, J. S. (2015). Repetitive DNA in eukaryotic genomes. *Chromosome Research*, *23*(3), 415–420.

Bosland, P. W., & Baral, J. B. (2007). 'Bhut Jolokia'—The world's hottest known chile pepper is a putative naturally occurring interspecific hybrid. *HortScience*, *42*(2), 222–224.

Bosland, P. W., Coon, D., & Reeves, G. (2012). 'Trinidad Moruga Scorpion' pepper is the world's hottest measured chile pepper at more than two million scoville heat units. *HortTechnology, 22*(4), 534–538.

Carrizo García, C., Barboza, G. E., Palombo, N., & Weiss-Schneeweiss, H. (2022). Diversification of chiles (*Capsicum*, Solanaceae) through time and space: New insights from genome-wide RAD-seq data. *Frontiers in Genetics*, *13*, 1030536.

Carrizo García, C., Barfuss, M. H. J., Sehr, E. M., Barboza, G. E., Samuel, R., Moscone, E. A., & Ehrendorfer, F. (2016). Phylogenetic relationships, diversification and expansion of chili peppers (*Capsicum*, Solanaceae). *Annals of Botany*, *118*(1), 35–51.

Cervantes-Hernández, F., Ochoa-Alejo, N., Martínez, O., & Ordaz-Ortiz, J. J. (2022). Metabolomic analysis identifies differences between wild and domesticated chili

pepper fruits during development (*Capsicum annuum* L.). *Frontiers in Plant Science*, *13*, 893055.

Chen, W., Wang, X., Sun, J., Wang, X., Zhu, Z., Ayhan, D. H., Yi, S., Yan, M., Zhang, L., & Meng, T. (2024). Two telomere-to-telomere gapless genomes reveal insights into *Capsicum* evolution and capsaicinoid biosynthesis. *Nature Communications*, *15*(1), 4295.

Cisneros-Pineda, O., Torres-Tapia, L. W., Gutiérrez-Pacheco, L. C., Contreras-Martín, F., González-Estrada, T., & Peraza-Sánchez, S. R. (2007). Capsaicinoids quantification in chili peppers cultivated in the state of Yucatan, Mexico. *Food Chemistry*, *104*(4), 1755–1760.

Contreras-Padilla, M., & Yahia, E. M. (1998). Changes in capsaicinoids during development, maturation, and senescence of chile peppers and relation with peroxidase activity. *Journal of Agricultural and Food Chemistry*, *46*(6), 2075–2079.

Csink, A. K., & Henikoff, S. (1998). Something from nothing: The evolution and utility of satellite repeats. *Trends in Genetics*, *14*(5), 200–204.

de Assis, R., Gonçalves, L. S. A., Guyot, R., & Vanzela, A. L. L. (2023). Abundance of distal repetitive DNA sequences in *Capsicum* L. (Solanaceae) chromosomes. *Genome*, *66*(10), 269–280.

DeWitt, D., & Bosland, P. W. (2009). *The complete chile pepper book: A gardener's guide to choosing, growing, preserving, and cooking.* Timber Press, London.

Dillehay, T. D., Goodbred, S., Pino, M., Sánchez, V. F. V., Tham, T. R., Adovasio, J., Collins, M. B., Netherly, P. J., Hastorf, C. A., Chiou, K. L., Piperno, D., Rey, I., & Velchoff, N. (2017). Simple technologies and diverse food strategies of the Late Pleistocene and Early Holocene at Huaca Prieta, Coastal Peru. *Science Advances*, *3*(5), e1602778.

Eshbaugh, W. (2012). The taxonomy of the genus *Capsicum*. In V. Russo (Eds.), *Peppers: Botany, Production and Uses* (pp. 14–28). CABI Digital Library.

Eshbaugh, W. H. (1993). Peppers: History and exploitation of a serendipitous new crop discovery. In: J. Janick and J. E. Simon (Eds.), *New crops* (pp. 132-139). Wiley, New York.

Feliciello, I., Picariello, O., & Chinali, G. (2006). Intra-specific variability and unusual organization of the repetitive units in a satellite DNA from *Rana dalmatina*: Molecular evidence of a new mechanism of DNA repair acting on satellite DNA. *Gene*, *383*, 81–92.

Garcia, S., Crhák Khaitová, L., & Kovařík, A. (2012). Expression of 5S rRNA genes linked to 35S rDNA in plants, their epigenetic modification and regulatory element divergence. *BMC Plant Biology*, *12*(1), 95.

Garcia, S., Kovařík, A., Leitch, A. R., & Garnatje, T. (2017). Cytogenetic features of rRNA genes across land plants: Analysis of the plant rDNA database. *The Plant Journal*, *89*(5), 1020–1030.

Garrido-Ramos, M. A. (2015). Satellite DNA in plants: More than just rubbish. *Cytogenetic and Genome Research*, *146*(2), 153–170.

Garrido-Ramos, M. A. (2017). Satellite DNA: An evolving topic. *Genes*, *8*(9), 230.

Gebhardt, C. (2016). The historical role of species from the Solanaceae plant family in genetic research. *Theoretical and Applied Genetics*, *129*, 2281–2294.

Greilhuber, J., Borsch, T., Müller, K., Worberg, A., Porembski, S., & Barthlott, W. (2006). Smallest angiosperm genomes found in Lentibulariaceae, with chromosomes of bacterial size. *Plant Biology*, *8*(06), 770–777.

Gurung, T., Techawongstien, S., Suriharn, B., & Techawongstien, S. (2011). Impact of environments on the accumulation of capsaicinoids in *Capsicum* spp. *HortScience*, *46*(12), 1576–1581.

Han, J. S. (2010). Non-long terminal repeat (non-LTR) retrotransposons: Mechanisms, recent developments, and unanswered questions. *Mobile DNA*, *1*(1), 15.

Havecker, E. R., Gao, X., & Voytas, D. F. (2004). The diversity of LTR retrotransposons. *Genome Biology*, *5*(6), 225.

Hayward, A., & Gilbert, C. (2022). Transposable elements. *Current Biology*, *32*(17), R904–R909.

Hemleben, V., Volkov, R., Zentgraf, U., & Medina, F. J. (2004). Molecular cell biology: Organization and molecular evolution of rDNA, nucleolar dominance, and nucleolus structure. In K. Esser, U. Lüttge, W. Beyschlag, J. Murata (Eds.), *Progress in Botany* (Vol. 65, pp. 106–146). Springer Verlag, New York.

Hernández-Pérez, T., & Gómez-García, M. del R., Valverde Me, Parades-López O., 2020, *Capsicum annuum* (hot pepper): An ancient Latin-American crop with outstanding bioactive compounds and nutraceutical potential. A review. *Comprehensive Reviews in Food Science and Food Safety*, *19*(6), 2972–2993.

Hunziker, A. T. (2001). Genera Solanacearum: The genera of Solanaceae illustrated, arranged according to a new system. A. R. G. Gantner Verlag K.-G, Ruggell, 500 pp.

Jang, T.-S., & Weiss-Schneeweiss, H. (2015). Formamide-free genomic *in situ* hybridization allows unambiguous discrimination of highly similar parental genomes in diploid hybrids and allopolyploids. *Cytogenetic and Genome Research*, *146(4), 325–331*.

Jo, S.-H., Koo, D.-H., Kim, J. F., Hur, C.-G., Lee, S., Yang, T., Kwon, S.-Y., & Choi, D. (2009). Evolution of ribosomal DNA-derived satellite repeat in tomato genome. *BMC Plant Biology*, *9*(1), 42.

Johnson, C. D., & Decoteau, D. R. (1996). Nitrogen and potassium fertility affects Jalapeño pepper plant growth, pod yield, and pungency. *HortScience*, *31*(7), 1119–1123.

Jordt, S.-E., & Julius, D. (2002). Molecular basis for species-specific sensitivity to "hot" chili peppers. *Cell*, *108*(3), 421–430.

Kumar, A., & Bennetzen, J. L. (1999). Plant retrotransposons. *Annual Review of Genetics*, *33*(1), 479–532.

Lal, S. K., & Hannah, L. C. (2005). Plant genomes: Massive changes of the maize genome are caused by Helitrons. *Heredity*, *95*(6), 421–422.

Luna-Ruiz, J. de J., Nabhan, G. P., & Aguilar-Meléndez, A. (2018). Shifts in plant chemical defenses of chile pepper (*Capsicum annuum* L.) due to domestication in Mesoamerica. *Frontiers in Ecology and Evolution*, *6*, 48.

Macas, J., Navrátilová, A., & Koblížková, A. (2006). Sequence homogenization and chromosomal localization of VicTR-B satellites differ between closely related *Vicia* species. *Chromosoma*, *115*(6), 437–447.

Maxwell, P. H. (2020). Diverse transposable element landscapes in pathogenic and nonpathogenic yeast models: The value of a comparative perspective. *Mobile DNA*, *11*(1), 16.

McClintock, B. (1948). Mutable loci in maize. *Carnegie Institution of Washington Yearbook*, *47*, 155–169.

Meckelmann, S. W., Jansen, C., Riegel, D. W., Van Zonneveld, M., Ríos, L., Peña, K., Mueller-Seitz, E., & Petz, M. (2015). Phytochemicals in native peruvian *Capsicum pubescens* (rocoto). *European Food Research and Technology*, *241*(6), 817–825.

Moscone, E. A., Scaldaferro, M. A., Grabiele, M., Cecchini, N. M., Sanchez García, Y., Jarret, R., Daviña, J. R., Ducasse, D. A., Barboza, G. E., & Ehrendorfer, F. (2006). The evolution of

chili peppers (*Capsicum*-Solanaceae): A cytogenetic perspective. Acta Hortic. 745, 137–170.

Motti, R. (2021). The Solanaceae family: Botanical features and diversity. In D. Carputo, R. Aversano, & M. R. Ercolano (Eds.), *The Wild Solanum Genomes* (pp. 1–9). Springer Verlag, Cham.

Muñoz-López, M., & García-Pérez, J. L. (2010). DNA transposons: Nature and applications in genomics. *Current Genomics*, *11*(2), 115–128.

Naves, E. R., de Ávila Silva, L., Sulpice, R., Araújo, W. L., Nunes-Nesi, A., Peres, L. E. P., & Zsögön, A. (2019). Capsaicinoids: Pungency beyond *Capsicum*. *Trends in Plant Science*, *24*(2), 109–120.

Neumann, P., Koblížková, A., Navrátilová, A., & Macas, J. (2006). Significant expansion of *Vicia pannonica* genome size mediated by amplification of a single type of giant retroelement. *Genetics*, *173*(2), 1047–1056.

Novák, P., Neumann, P., & Macas, J. (2020). Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nature Protocols*, *15*(11), 3745–3776.

Novák, P., Neumann, P., Pech, J., Steinhaisl, J., & Macas, J. (2013). RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics*, *29*(6), 792–793.

Onus, A. N., & Pickersgill, B. (2004). Unilateral incompatibility in *Capsicum* (Solanaceae): Occurrence and taxonomic distribution. *Annals of Botany*, *94*(2), 289–295.

Palma, G., Quezada-Euán, J. J. G., Meléndez-Ramirez, V., Irigoyen, J., Valdovinos-Nuñez, G. R., & Rejón, M. (2008). Comparative efficiency of *Nannotrigona perilampoides*, *Bombus*

*impatiens* (Hymenoptera: Apoidea), and mechanical vibration on fruit production of enclosed habanero pepper. *Journal of Economic Entomology*, *101*(1), 132–138.

Palombo, N. E., & Carrizo García, C. (2022). Geographical patterns of genetic variation in locoto chile (*Capsicum pubescens*) in the americas inferred by genome-wide data analysis. *Plants*, *11*(21).

Palombo, N. E., Weiss-Schneeweiss, H., & Carrizo García, C. (2024). Evolutionary relationships, hybridization and diversification under domestication of the locoto chile (*Capsicum pubescens*) and its wild relatives. *Frontiers in Plant Science*, *15*, 1353991.

Park, C.-R., & Kim, S.-D. (1975). The effect of light on the matured hot green pepper fruits during the after-ripening period*. Korean Journal of Nutrition,* 2, (91–94).

Pellicer, J., Fay, M. F., & Leitch, I. J. (2010). The largest eukaryotic genome of them all? *Botanical Journal of the Linnean Society*, *164*(1), 10–15.

Pellicer, J., Hidalgo, O., Dodsworth, S., & Leitch, I. J. (2018). Genome size diversity and its impact on the evolution of land plants. *Genes*, *9*(2).

Phimchan, P., Techawongstien, S., Chanthai, S., & Bosland, P. W. (2012). Impact of drought stress on the accumulation of capsaicinoids in *Capsicum* cultivars with different initial capsaicinoid levels. *HortScience*, *47*(9), 1204–1209.

Pickersgill, B. (1997). Genetic resources and breeding of *Capsicum* spp. *Euphytica*, *96*(1), 129–133.

Powis, T. G., Gallaga Murrieta, E., Lesure, R., Lopez Bravo, R., Grivetti, L., Kucera, H., & Gaikwad, N. W. (2013). Prehispanic use of chili peppers in Chiapas, Mexico. *PLoS One*, *8*(11), e79013.

Pray, L., & Zhaurova, K. (2008). Barbara McClintock and the discovery of jumping genes (transposons). *Nature Education*, *1*(1), 169.

Puppo, I., Saifitdinova, A., & Tonyan, Z. (2020). The role of satellite DNA in causing structural rearrangements in human karyotype. *Russian Journal of Genetics*, *56*(1), 41–47.

Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., et al. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proceedings of the National Academy of Sciences of the U S A*, *111*(14), 5135–5140.

Rosselló, J. A., Maravilla, A. J., & Rosato, M. (2022). The nuclear 35S rDNA world in plant systematics and evolution: A primer of cautions and common misconceptions in cytogenetic studies. *Frontiers in Plant Science*, *13*, 788911.

Rudenko, V., & Korotkov, E. (2023). Detection of tandem repeats in the *Capsicum annuum* genome. *DNA Research*, *30*(3), dsad007.

Ruiz-Ruano, F. J., López-León, M. D., Cabrero, J., & Camacho, J. P. M. (2016). High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Scientific Reports*, *6*(1), 28333.

Scaldaferro, M. A. (2020). Molecular cytogenetic evidence of hybridization in the "purple corolla clade of the genus *Capsicum*" (*C. eximium* × *C. cardenasii*). *Plant Biosystems*, *154*(5), 685–691.

Scaldaferro, M. A., da Cruz, M. V. R., Cecchini, N. M., & Moscone, E. A. (2016). FISH and AgNOR mapping of the 45S and 5S rRNA genes in wild and cultivated species of *Capsicum* (Solananceae). *Genome*, *59*(2), 95–113.

Scaldaferro, M. A., Grabiele, M., & Moscone, E. A. (2013). Heterochromatin type, amount and distribution in wild species of chili peppers (*Capsicum*, Solanaceae). *Genetic Resources and Crop Evolution*, *60*(2), 693–709.

Scaldaferro, M., & Barboza, G. E. (2023). Cytogenetic characterization and chromosome evolution of the striking tropical Andean species of *Capsicum* (Solanaceae). *Plant Biosystems*, *157*(3), 584–593.

Schmitz, J. (2012). SINEs as driving forces in genome evolution. *Genome dynamics*, *7*, 92–107.

Schueler, M. G., & Sullivan, B. A. (2006). Structural and functional dynamics of human centromeric chromatin. *Annual Review of Genomics Human Genetics*, *7*(1), 301–313.

Simon, A. U. (2020). Evolution of selected satellite DNAs in cultivated *Capsicum* species. MSc. thesis, University of Vienna, Vienna, Austria.

Stewart, C., Jr, Mazourek, M., Stellari, G. M., O'Connell, M., & Jahn, M. (2007). Genetic control of pungency in *C. chinense* via the *Pun*1 locus. *Journal of Experimental Botany*, *58*(5), 979–991.

Stewart Jr, C., Kang, B., Liu, K., Mazourek, M., Moore, S. L., Yoo, E. Y., Kim, B., Paran, I., & Jahn, M. M. (2005). The *Pun*1 gene for pungency in pepper encodes a putative acyltransferase. *The Plant Journal*, *42*(5), 675–688.

Sun, T., Xu, Z., Wu, C., Janes, M., Prinyawiwatkul, W., & No, H. (2007). Antioxidant activities of different colored sweet bell peppers (*Capsicum annuum* L.). *Journal of Food Science*, *72*(2), S98–S102.

Temsch, E. M., Greilhuber, J., & Krisai, R. (2010). Genome size in liverworts. *Preslia*, *82*, 63–80.

Tewksbury, J. J., Manchego, C., Haak, D. C., & Levey, D. J. (2006). Where did the chili get its spice? Biogeography of capsaicinoid production in ancestral wild chili species. *Journal of Chemical Ecology*, *32*(3), 547–564.

Tewksbury, J. J., & Nabhan, G. P. (2001). Directed deterrence by capsaicin in chillies. *Nature*, *412*(6845), 403–404.

Tewksbury, J. J., Reagan, K. M., Machnicki, N. J., Carlo, T. A., Haak, D. C., Peñaloza, A. L. C., & Levey, D. J. (2008). Evolutionary ecology of pungency in wild chilies. *Proceedings of the National Academy of Sciences of the U S A*, *105*(33), 11808–11811.

Thakur, J., Packiaraj, J., & Henikoff, S. (2021). Sequence, chromatin and evolution of satellite DNA. *International Journal of Molecular Sciences*, *22*(9), 4309.

Tripodi, P., & Kumar, S. (2019). The *Capsicum* Crop: An Introduction. In N. Ramchiary & C. Kole (Eds.), *The Capsicum Genome* (pp. 1–8). Compendium of Plant Genomes. Springer Verlag, Cham.

Weiss-Schneeweiss, H., Emadzade, K., Jang, T.-S., & Schneeweiss, G. M. (2013). Evolutionary consequences, constraints and potential of polyploidy in plants. *Cytogenetics and Genome Research*, *140 (2-4)*, 137–150.

Weiss-Schneeweiss, H., & Schneeweiss, G. M. (2013). Karyotype Diversity and Evolutionary Trends in Angiosperms. In Leitch, I. J., Greilhuber, J., Dolezel, J., Wendel, J. F. (Eds.), *Plant Genome Diversity* (Eds.) (Vol. 2) (pp. 209-230). Springer Verlag, Wien.

Wells, J. N., & Feschotte, C. (2020). A field guide to eukaryotic transposable elements. *Annual Review of Genetics*, *54*(1), 539–561.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., & Panaud, O. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, *8*(12), 973–982.

W.L. Scoville. (1912). Note on *Capsicum*. *Journal of American Pharmacists Association*, *1*, 453.

Xiong, W., Dooner, H. K., & Du, C. (2016). Rolling-circle amplification of centromeric Helitrons in plant genomes. *The Plant Journal*, *88*(6), 1038–1045.

Zewdie, Y., & Bosland, P. W. (2000). Evaluation of genotype, environment, and genotype-by-environment interaction for capsaicinoids in *Capsicum annuum* L. *Euphytica*, *111*(3), 185–190.

Zewdie, Y., Bosland, P. W., & Steiner, R. (2001). Combining ability and heterosis for capsaicinoids in *Capsicum pubescens*. *HortScience*, *36*(7), 1315–1317.

Zhang, Z.-X., Zhao, S.-N., Liu, G.-F., Huang, Z.-M., Cao, Z.-M., Cheng, S.-H., & Lin, S.-S. (2016). Discovery of putative capsaicin biosynthetic genes by RNA-Seq and digital gene expression analysis of pepper. *Scientific Reports*, *6*(1), 34121.

Zhou, H., Waminal, N., & Kim, H. (2019). *In silico* mining and FISH mapping of a chromosome-specific satellite DNA in *Capsicum annuum* L. *Genes & Genomics*, *41*(9), 1001–1006.

# 6. Supplementary data

## 6.1 Phase contrast microscopy

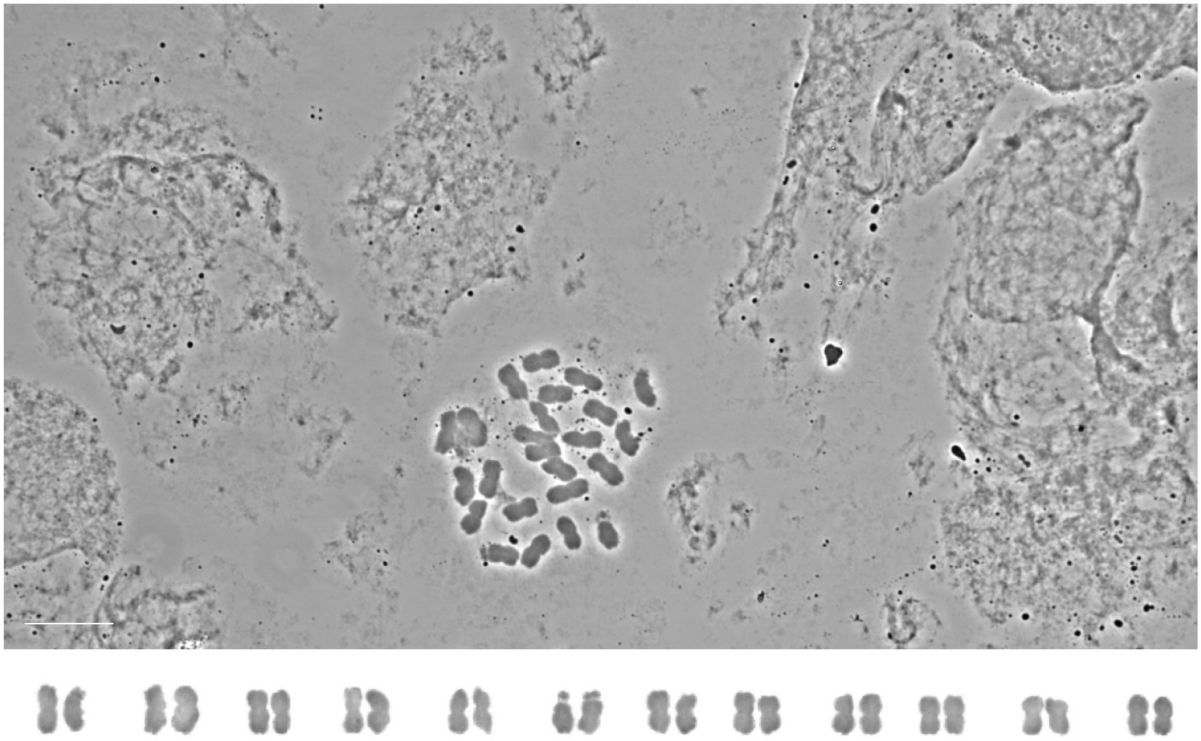


**Figure 25.** Mitotic metaphase chromosomes of *C. eshbaughii* including the karyotype below. Picture taken with phase contrast (DIC Ⅲ) on an Axio Imager M2 microscope (Carl Zeiss, Vienna Austria).

## 6.2 Capsaicinoid quantification

**Table 3.** List of *Capsicum* fruit samples used for HPLC analyses, including the amount of dried plant material used in mg, amount of extract obtained in mg and the peak areas from the HPLC chromatograms from all four derivates. A1 = Nordihydrocapsaicin, A2 = Capsaicin, A3 = dihydrocapsaicin. A4 = not identified.

| Sample | Plant material [mg] | Extract [mg] | Peak A1 | Peak A2 | Peak A3 | Peak A4 |
|---|---|---|---|---|---|---|
| LP1a | 30,9 | 9 | 375,3839 | 747,4805 | 779,7065 | n.d. |
| LP1b | 30,2 | 9,7 | 296,0865 | 591,4268 | 622,631 | n.d. |
| LP1c | 32,7 | 10,4 | 296,669 | 586,5101 | 600,7305 | n.d. |
| LP2a | 32,1 | 8 | 185,3975 | 120,3221 | 747,0072 | n.d. |
| LP2b | 33,1 | 8 | 185,5935 | 122,4043 | 702,1759 | n.d. |
| LP2c | 31,3 | 7,6 | 201,6772 | 132,3792 | 725,2995 | n.d. |
| LP3a | 31 | 9,9 | 22,8093 | 182,888 | 436,4318 | n.d. |
| LP3b | 32,8 | 10,1 | 20,22561 | 162,2024 | 385,2621 | n.d. |
| LP3c | 31 | 10,7 | 17,68941 | 131,5777 | 340,3721 | n.d. |
| LP4a | 33,4 | 19,1 | 152,6017 | 128,6074 | 51,45502 | n.d. |
| LP4b | 31,4 | 17,7 | 166,9805 | 136,6704 | 58,72485 | n.d. |
| LP4c | 29,9 | 17 | 162,4832 | 138,0361 | 58,84112 | n.d. |
| LP5a | 30,5 | 17,2 | 109,7039 | 140,3289 | n.d. | n.d. |
| LP5b | 32 | 18,3 | 108,3121 | 141,1456 | n.d. | n.d. |

| | | | | | | |
|---|---|---|---|---|---|---|
| LP5c | 30,2 | 17 | 116,0719 | 147,9489 | n.d. | n.d. |
| LP6a | 31,8 | 15,6 | 268,4963 | 238,4839 | 235,0766 | n.d. |
| LP6b | 30,4 | 14,4 | 285,4046 | 276,9971 | 287,4693 | n.d. |
| LP6c | 33,6 | 16,2 | 245,6437 | 215,7759 | 211,562 | n.d. |
| LP7a | 31,7 | 15,3 | n.d. | n.d. | n.d. | n.d. |
| LP7b | 32,2 | 14,8 | n.d. | n.d. | n.d. | n.d. |
| LP7c | 33,3 | 15,3 | n.d. | n.d. | n.d. | n.d. |
| LP8a | 30,4 | 14,5 | 22,13393 | 161,848 | 73,04558 | n.d. |
| LP8b | 30,2 | 14,1 | 27,18609 | 165,931 | 73,40312 | n.d. |
| LP8c | 31,5 | 14,4 | 29,02638 | 175,6158 | 74,80777 | n.d. |
| LP9a | 32,8 | 19,7 | n.d. | 25,02167 | 11,81041 | n.d. |
| LP9b | 32,4 | 18,6 | n.d. | 26,05796 | 13,99113 | n.d. |
| LP9c | 30,9 | 18,4 | n.d. | 26,19058 | 13,36898 | n.d. |
| LP10a | 30,5 | 20,4 | n.d. | n.d. | n.d. | n.d. |
| LP10b | 30,4 | 20,6 | n.d. | n.d. | n.d. | n.d. |
| LP10c | 32,8 | 22,4 | n.d. | n.d. | n.d. | n.d. |
| LP11a | 30,6 | 11,6 | 62,66489 | 369,2075 | 320,7825 | n.d. |
| LP11b | 32,4 | 13,1 | 67,81176 | 481,0218 | 350,608 | n.d. |
| LP11c | 33,8 | 12,4 | 74,51357 | 541,5994 | 391,022 | n.d. |
| LP12a | 32,4 | 14 | 164,0796 | 764,6483 | 479,2979 | n.d. |
| LP12b | 30,5 | 13,4 | 157,882 | 731,4194 | 457,0377 | n.d. |
| LP12c | 11,1 | 5,5 | 170,9414 | 791,1244 | 491,6184 | n.d. |
| LP13a | 29,9 | 9,9 | 33,33117 | 2973,661 | 721,8156 | 61,61754 |
| LP13b | 31,8 | 10,8 | 39,68646 | 2921,976 | 722,9028 | 69,09886 |
| LP13c | 30,5 | 9,4 | 42,09906 | 3119,979 | 770,3831 | 74,34904 |
| LP14a | 33 | 12,7 | n.d. | n.d. | n.d. | n.d. |
| LP14b | 29,9 | 11,6 | n.d. | n.d. | n.d. | n.d. |
| LP14c | 29,6 | 10,7 | n.d. | n.d. | n.d. | n.d. |
| LP15a | 31,8 | 12,6 | 45,73418 | 931,873 | 402,2164 | n.d. |
| LP15b | 29,9 | 11,1 | 48,91113 | 980,0068 | 427,2561 | n.d. |
| LP15c | 31,9 | 12,3 | 51,56819 | 1044,671 | 445,6276 | n.d. |
| LP16a | 33,6 | 14,7 | 422,3773 | 419,8088 | 298,8915 | n.d. |
| LP16b | 30 | 13,4 | 402,4189 | 394,1335 | 276,8716 | n.d. |
| LP16c | 32 | 13,7 | 409,7401 | 416,0743 | 303,3417 | n.d. |
| LP17a | 32,9 | 18,3 | 157,4362 | 262,2512 | 182,6688 | n.d. |
| LP17b | 31,3 | 17,3 | 159,6846 | 276,4263 | 188,754 | n.d. |
| LP17c | 31,5 | 16,8 | 157,3315 | 292,265 | 197,1469 | n.d. |
| LP18a | 31,3 | 17,7 | 31,23717 | 44,68151 | 162,2057 | n.d. |
| LP18b | 30,4 | 17,9 | 27,32575 | 38,97163 | 141,5895 | n.d. |
| LP18c | 29,9 | 17,4 | 30,01001 | 45,2958 | 161,6163 | n.d. |
| LP19a | 30,9 | 15,9 | n.d. | n.d. | n.d. | n.d. |
| LP19b | 30,9 | 15,8 | n.d. | n.d. | n.d. | n.d. |
| LP19c | 30 | 15,6 | n.d. | n.d. | n.d. | n.d. |
| LP20a | 31,3 | 20,3 | 91,47665 | 143,4354 | 124,8476 | n.d. |
| LP20b | 30,1 | 18,8 | 77,0333 | 143,3167 | 132,3116 | n.d. |
| LP20c | 32,3 | 20 | 71,33731 | 122,9928 | 112,8368 | n.d. |

| | | | | | | |
|---|---|---|---|---|---|---|
| LP21a | 32,5 | 15,5 | 260,3059 | 246,3499 | 145,0279 | n.d. |
| LP21b | 32,3 | 16 | 253,5569 | 245,3799 | 144,0128 | n.d. |
| LP21c | 31,1 | 15,4 | 255,1359 | 237,9954 | 138,3948 | n.d. |
| LP22a | 30,5 | 14,4 | n.d. | n.d. | n.d. | n.d. |
| LP22b | 33,7 | 15,2 | n.d. | n.d. | n.d. | n.d. |
| LP22c | 31,8 | 14,7 | n.d. | n.d. | n.d. | n.d. |
| LP23a | 31,3 | 14,8 | n.d. | n.d. | n.d. | n.d. |
| LP23b | 30,3 | 14,6 | n.d. | n.d. | n.d. | n.d. |
| LP23c | 32,6 | 15,3 | n.d. | n.d. | n.d. | n.d. |
| LP24a | 29,7 | 16,3 | 130,7342 | 236,6637 | 212,6175 | n.d. |
| LP24b | 30,9 | 17,1 | 122,1144 | 220,904 | 198,4367 | n.d. |
| LP24c | 32,7 | 18,8 | 124,6292 | 227,0958 | 203,8689 | n.d. |
| LP25a | 32,3 | 11,6 | 562,6572 | 227,964 | 718,4536 | n.d. |
| LP25b | 31,6 | 12,9 | 516,8703 | 212,2192 | 673,4647 | n.d. |
| LP25c | 30,6 | 12,6 | 575,1591 | 233,5212 | 708,1318 | n.d. |

**Table 4.** List of fruit samples used for the HPLC analyses including the peak area sum from the HPLC chromatograms, the capsaicinoid sum calculated with the calibration curve in µg/mL, the capsaicinoid sum in µg/g dry material and mg/g dry material.

| Sample | Peak area sum | Capsaicin sum [µg/mL] | Capsaicin [µg/g dry material] | Capsaicin [mg/g dry material] |
|---|---|---|---|---|
| LP1a | 1902,570923 | 180,7402402 | 5849,198712 | 5,849199 |
| LP1b | 1510,144196 | 143,6733915 | 4757,397069 | 4,757397 |
| LP1c | 1483,909698 | 141,1953999 | 4317,902137 | 4,317902 |
| LP2a | 1052,726852 | 100,467824 | 3129,838753 | 3,129839 |
| LP2b | 1010,173737 | 96,44844966 | 2913,850443 | 2,91385 |
| LP2c | 1059,355896 | 101,0939734 | 3229,839405 | 3,229839 |
| LP3a | 642,1290817 | 61,68462092 | 1989,826481 | 1,989826 |
| LP3b | 567,6901321 | 54,65345538 | 1666,263884 | 1,666264 |
| LP3c | 489,6392117 | 47,28111945 | 1525,197402 | 1,525197 |
| LP4a | 332,6641312 | 32,45396535 | 971,6756093 | 0,971676 |
| LP4b | 362,3757858 | 35,26039349 | 1122,942468 | 1,122942 |
| LP4c | 359,360405 | 34,97557429 | 1169,751648 | 1,169752 |
| LP5a | 250,0328522 | 24,64898953 | 808,1635912 | 0,808164 |
| LP5b | 249,4577103 | 24,59466424 | 768,5832574 | 0,768583 |
| LP5c | 264,0208664 | 25,97023391 | 859,9415202 | 0,859942 |
| LP6a | 742,0568085 | 71,12334075 | 2236,583042 | 2,236583 |
| LP6b | 849,8709717 | 81,30697758 | 2674,571631 | 2,674572 |
| LP6c | 672,9815674 | 64,59880678 | 1922,583535 | 1,922584 |
| LP7a | 0 | 0 | 0 | 0 |
| LP7b | 0 | 0 | 0 | 0 |
| LP7c | 0 | 0 | 0 | 0 |
| LP8a | 257,0275345 | 25,3096755 | 832,555115 | 0,832555 |
| LP8b | 266,5202236 | 26,20631186 | 867,7586707 | 0,867759 |
| LP8c | 279,4499054 | 27,42759095 | 870,717173 | 0,870717 |
| LP9a | 36,83208275 | 4,511011878 | 137,5308499 | 0,137531 |

| | | | | |
|---|---|---|---|---|
| LP9b | 40,04909134 | 4,814875918 | 148,6072814 | 0,148607 |
| LP9c | 39,55955887 | 4,768636901 | 154,3248188 | 0,154325 |
| LP10a | 0 | 0 | 0 | 0 |
| LP10b | 0 | 0 | 0 | 0 |
| LP10c | 0 | 0 | 0 | 0 |
| LP11a | 752,6548843 | 72,12438692 | 2357,006109 | 2,357006 |
| LP11b | 899,4415207 | 85,9891868 | 2653,987247 | 2,653987 |
| LP11c | 1007,134903 | 96,16141522 | 2845,012285 | 2,845012 |
| LP12a | 1408,025772 | 134,0277484 | 4136,658901 | 4,136659 |
| LP12b | 1346,339005 | 128,2010961 | 4203,314627 | 4,203315 |
| LP12c | 1453,684189 | 138,3404353 | 12463,10228 | 12,4631 |
| LP13a | 3790,424973 | 359,0583709 | 12008,64117 | 12,00864 |
| LP13b | 3753,66423 | 355,5861179 | 11181,95339 | 11,18195 |
| LP13c | 4006,809669 | 379,4970879 | 12442,52747 | 12,44253 |
| LP14a | 0 | 0 | 0 | 0 |
| LP14b | 0 | 0 | 0 | 0 |
| LP14c | 0 | 0 | 0 | 0 |
| LP15a | 1379,823658 | 131,3639046 | 4130,940396 | 4,13094 |
| LP15b | 1456,174011 | 138,5756127 | 4634,635875 | 4,634636 |
| LP15c | 1541,866531 | 146,6697394 | 4597,797474 | 4,597797 |
| LP16a | 1141,077576 | 108,8130326 | 3238,483112 | 3,238483 |
| LP16b | 1073,42395 | 102,422778 | 3414,092598 | 3,414093 |
| LP16c | 1129,156036 | 107,686978 | 3365,218063 | 3,365218 |
| LP17a | 602,3561096 | 57,92784638 | 1760,724814 | 1,760725 |
| LP17b | 624,8648682 | 60,05392162 | 1918,655643 | 1,918656 |
| LP17c | 646,743454 | 62,1204736 | 1972,078527 | 1,972079 |
| LP18a | 238,1244202 | 23,52417306 | 751,5710242 | 0,751571 |
| LP18b | 207,8869057 | 20,66807459 | 679,8708746 | 0,679871 |
| LP18c | 236,9221001 | 23,41060735 | 782,9634567 | 0,782963 |
| LP19a | 0 | 0 | 0 | 0 |
| LP19b | 0 | 0 | 0 | 0 |
| LP19c | 0 | 0 | 0 | 0 |
| LP20a | 359,7597046 | 35,01329032 | 1118,635473 | 1,118635 |
| LP20b | 352,661644 | 34,34283971 | 1140,95813 | 1,140958 |
| LP20c | 307,1669159 | 30,04561404 | 930,2047692 | 0,930205 |
| LP21a | 651,6837311 | 62,58710976 | 1925,757224 | 1,925757 |
| LP21b | 642,9496002 | 61,76212338 | 1912,140043 | 1,91214 |
| LP21c | 631,5261993 | 60,68312075 | 1951,225747 | 1,951226 |
| LP22a | 0 | 0 | 0 | 0 |
| LP22b | 0 | 0 | 0 | 0 |
| LP22c | 0 | 0 | 0 | 0 |
| LP23a | 0 | 0 | 0 | 0 |
| LP23b | 0 | 0 | 0 | 0 |
| LP23c | 0 | 0 | 0 | 0 |
| LP24a | 580,0153656 | 55,81764103 | 1879,381853 | 1,879382 |
| LP24b | 541,4550705 | 52,17541046 | 1688,52461 | 1,688525 |

| | | | | |
|---|---|---|---|---|
| LP24c | 555,5938187 | 53,51089248 | 1636,41873 | 1,636419 |
| LP25a | 1509,074799 | 143,5723811 | 4444,965359 | 4,444965 |
| LP25b | 1402,554245 | 133,5109327 | 4225,029517 | 4,22503 |
| LP25c | 1516,812134 | 144,3032147 | 4715,791329 | 4,715791 |

**Table 5**. List of fruit samples used for the HPLC analyses including the relative distribution of the capsaicinoid derivates [%]. A1 = Nordihydrocapsaicin, A2 = Capsaicin, A3 = dihydrocapsaicin. A4 = not identified.

| Sample | A1 [%] | A2 [%] | A3 [%] | A4 [%] |
|---|---|---|---|---|
| LP1a | 19,73035 | 39,28792 | 40,98173 | 0 |
| LP1b | 19,6065 | 39,1636 | 41,2299 | 0 |
| LP1c | 19,99239 | 39,52465 | 40,48296 | 0 |
| LP2a | 17,61117 | 11,42957 | 70,95926 | 0 |
| LP2b | 18,37244 | 12,11715 | 69,51041 | 0 |
| LP2c | 19,03772 | 12,4962 | 68,46608 | 0 |
| LP3a | 3,552137 | 28,48149 | 67,96637 | 0 |
| LP3b | 3,56279 | 28,57235 | 67,86485 | 0 |
| LP3c | 3,612744 | 26,87238 | 69,51488 | 0 |
| LP4a | 45,8726 | 38,65984 | 15,46756 | 0 |
| LP4b | 46,07938 | 37,71511 | 16,20551 | 0 |
| LP4c | 45,21456 | 38,4116 | 16,37385 | 0 |
| LP5a | 43,8758 | 56,1242 | 0 | 0 |
| LP5b | 43,41902 | 56,58098 | 0 | 0 |
| LP5c | 43,96316 | 56,03684 | 0 | 0 |
| LP6a | 36,18271 | 32,13823 | 31,67905 | 0 |
| LP6b | 33,58211 | 32,59284 | 33,82505 | 0 |
| LP6c | 36,5008 | 32,06268 | 31,43652 | 0 |
| LP7a | 0 | 0 | 0 | 0 |
| LP7b | 0 | 0 | 0 | 0 |
| LP7c | 0 | 0 | 0 | 0 |
| LP8a | 8,611503 | 62,96914 | 28,41936 | 0 |
| LP8b | 10,20038 | 62,25832 | 27,5413 | 0 |
| LP8c | 10,38697 | 62,84338 | 26,76965 | 0 |
| LP9a | | 67,93445 | 32,06555 | 0 |
| LP9b | | 65,06504 | 34,93496 | 0 |
| LP9c | | 66,20543 | 33,79457 | 0 |
| LP10a | 0 | 0 | 0 | 0 |
| LP10b | 0 | 0 | 0 | 0 |
| LP10c | 0 | 0 | 0 | 0 |
| LP11a | 8,325847 | 49,05403 | 42,62013 | 0 |
| LP11b | 7,539318 | 53,48005 | 38,98063 | 0 |
| LP11c | 7,398568 | 53,77625 | 38,82518 | 0 |
| LP12a | 11,65317 | 54,30641 | 34,04042 | 0 |
| LP12b | 11,72676 | 54,32654 | 33,9467 | 0 |
| LP12c | 11,75918 | 54,42203 | 33,81879 | 0 |
| LP13a | 0,879352 | 78,45191 | 19,04313 | 1,62561 |

| | | | | |
|---|---|---|---|---|
| LP13b | 1,057273 | 77,8433 | 19,25859 | 1,840838 |
| LP13c | 1,050688 | 77,8669 | 19,22684 | 1,855567 |
| LP14a | 0 | 0 | 0 | 0 |
| LP14b | 0 | 0 | 0 | 0 |
| LP14c | 0 | 0 | 0 | 0 |
| LP15a | 3,314495 | 67,53566 | 29,14984 | 0 |
| LP15b | 3,35888 | 67,30011 | 29,34101 | 0 |
| LP15c | 3,34453 | 67,75365 | 28,90182 | 0 |
| LP16a | 37,01565 | 36,79056 | 26,19379 | 0 |
| LP16b | 37,48927 | 36,71741 | 25,79331 | 0 |
| LP16c | 36,28728 | 36,84826 | 26,86446 | 0 |
| LP17a | 26,13672 | 43,53757 | 30,32571 | 0 |
| LP17b | 25,55507 | 44,23777 | 30,20716 | 0 |
| LP17c | 24,32673 | 45,19025 | 30,48302 | 0 |
| LP18a | 13,11801 | 18,76393 | 68,11806 | 0 |
| LP18b | 13,14453 | 18,74655 | 68,10892 | 0 |
| LP18c | 12,66662 | 19,11844 | 68,21495 | 0 |
| LP19a | 0 | 0 | 0 | 0 |
| LP19b | 0 | 0 | 0 | 0 |
| LP19c | 0 | 0 | 0 | 0 |
| LP20a | 25,42715 | 39,86979 | 34,70306 | 0 |
| LP20b | 21,8434 | 40,63859 | 37,51801 | 0 |
| LP20c | 23,22428 | 40,04102 | 36,7347 | 0 |
| LP21a | 39,94359 | 37,80207 | 22,25433 | 0 |
| LP21b | 39,43651 | 38,16472 | 22,39877 | 0 |
| LP21c | 40,3999 | 37,68576 | 21,91434 | 0 |
| LP22a | 0 | 0 | 0 | 0 |
| LP22b | 0 | 0 | 0 | 0 |
| LP22c | 0 | 0 | 0 | 0 |
| LP23a | 0 | 0 | 0 | 0 |
| LP23b | 0 | 0 | 0 | 0 |
| LP23c | 0 | 0 | 0 | 0 |
| LP24a | 22,53978 | 40,803 | 36,65722 | 0 |
| LP24b | 22,553 | 40,79821 | 36,64879 | 0 |
| LP24c | 22,4317 | 40,87443 | 36,69387 | 0 |
| LP25a | 37,28491 | 15,10621 | 47,60888 | 0 |
| LP25b | 36,85207 | 15,13091 | 48,01702 | 0 |
| LP25c | 37,91894 | 15,39553 | 46,68553 | 0 |

# 7. Abstract

The genus *Capsicum* (Solanaceae) comprises 43 species distributed in South and Middle America. Five species are cultivated worldwide as economically important crops. *Capsicum* species are known for their unique pungency resulting from the presence of secondary metabolites, capsaicinoids. All pepper species are exclusively diploid with two basic chromosome numbers known, *x* = 12 and *x* = 13, and genome sizes that vary nearly 5-fold. Such genome size variation largely reflects differences in the abundances of repetitive DNA elements. This study focuses on members of the Pubescens clade (*C. cardenasii*, *C. eshbaughii*, *C. eximium*, *Capsicum pubescens*; *x* = 12) and aims to characterize the abundance and chromosomal distribution of recently identified families of tandem repeats (satellite DNAs) as well as their chemical diversity, specifically capsaicinoids content and diversity. Genome sizes were measured using flow cytometry and several families of fast evolving satellite DNAs were mapped in chromosomes of three species of Pubescens clade using fluorescence *in situ* hybridization (FISH). High performance liquid chromatography (HPLC) was applied to quantify capsaicinoids content, identify their derivatives (nordihydrocapsaicin, capsaicin, dihydrocapsaicin) and determine their ratios. These measurements were also performed for the remaining four cultivated *Capsicum* species (*C. annuum*, *C. baccatum*, *C. chinense*, *C. frutescens*) and several wild taxa from other lineages. Genomes of the four analyzed species of the Pubescens clade were all of similar sizes. Most satellite DNAs (CapSat1, CapSat4-6, CapSat8) were shared by all analyzed species, whereas two (CapSat7, CapSat17) were present only in *C. pubescens*. Most satellite DNAs displayed species-specific amplification patterns concerning numbers and localization of loci. Phytochemical analyses revealed a wide range of pungency levels, from non-pungent to extremely pungent accessions. All cultivated taxa exhibited average rather low total capsaicinoid contents, albeit with some very pungent accessions, as well as greater proportions of capsaicin. In contrast, wild species had higher overall levels of capsaicinoids, albeit dominated by dihydrocapsaicin. High levels of intraspecific variation were observed among multiple accessions of *Capsicum pubescens*, concerning both capsaicinoid content and the ratios of the derivatives, with some accessions lacking individual compounds. Together, these findings provide novel insights into the genomic organization and biochemical diversity of the Pubescens clade. The study contributes to the establishment of chromosomal markers and to a better understanding of intra- and interspecific variation of capsaicinoid composition and content in the Pubescens clade.

Die Gattung *Capsicum* (Solanaceae) umfasst 43 Arten, die in Süd- und Mittelamerika verbreitet sind. Fünf davon werden weltweit als wirtschaftlich wichtige Nutzpflanzen angebaut. *Capsicum* Arten sind für ihre einzigartige Schärfe bekannt, die auf das Vorhandensein sekundären Metaboliten, der Capsaicinoiden, zurückzuführen ist. Alle Paprika Arten sind ausschließlich diploid mit zwei bekannten Grundchromosomenzahlen, *x* = 12 und *x* = 13, und Genomgrößen, die bis um das Fünffache variieren. Diese Variation der Genomgröße spiegelt weitgehend Unterschiede in der Häufigkeit repetitiver DNA-Elemente wider. Diese Studie konzentriert sich auf Arten der Pubescens Klade (*C. cardenasii*, *C. eshbaughii*, *C. eximium*, *Capsicum pubescens*; *x* = 12) und zielt darauf ab, die Häufigkeit und chromosomale Verteilung kürzlich identifizierter Familien von Tandem-Wiederholungen (Satelliten DNAs) sowie ihre chemische Vielfalt, insbesondere den Gehalt und die Vielfalt an Capsaicinoiden, zu charakterisieren. Die Genomgrößen wurden mittels Durchflusszytometrie gemessen, und mehrere Familien schnell evolvierender Satelliten DNAs konnten mittels Fluoreszenz *in situ* Hybridisierung (FISH) in den Chromosomen von drei Arten der Pubescens Klade nachgewiesen werden. Mittels Hochleistungsflüssigkeitschromatographie (HPLC) wurde der Capsaicinoidgehalt quantifiziert, dabei wurden die Derivate (Nordihydrocapsaicin, Capsaicin, Dihydrocapsaicin) identifiziert und ihre Verhältnisse bestimmt. Diese Messungen wurden auch für die übrigen vier kultivierten *Capsicum* Arten *(C. annuum*, *C. baccatum*, *C. chinense*, *C. frutescens*) und mehrere wilde Taxa aus anderen Abstammungslinien durchgeführt. Die Genome der vier untersuchten Arten der Pubescens Klade waren alle von ähnlicher Größe. Die meisten Satelliten DNAs (CapSat1, CapSat4-6, CapSat8) kamen in allen analysierten Arten vor, während zwei (CapSat7, CapSat17) ausschließlich in *C. pubescens* vorkamen. Die meisten Satelliten DNAs zeigten artspezifische Amplifikationsmuster hinsichtlich der Anzahl und Lokalisierung der Loci. Phytochemische Analysen ergaben eine große Bandbreite an Schärfegraden, von nicht scharfen bis zu extrem scharfen Akzessionen. Alle kultivierten Taxa wiesen im Durchschnitt einen eher geringen Gesamtgehalt an Capsaicinoiden auf, wobei Capsaicin meist dominierte, allerdings wurden auch einzelne sehr scharfe Akzessionen festgestellt. Im Gegensatz dazu wiesen die wilden Arten insgesamt höhere Capsaicinoidwerte auf, wobei Dihydrocapsaicin den größten relativen Anteil hatte. Bei mehreren Akzessionen von *Capsicum pubescescens* wurde eine hohe intraspezifische Variation sowohl im Capsaicinoidgehalt als auch in den Anteilen der Derivate beobachtet, wobei in manchen Akzessionen einzelne Metabolite vollständig fehlten. Insgesamt liefern die Ergebnisse neue

Erkenntnisse über die genomische Organisation und über die biochemische Vielfalt der Pubescens Klade. Die Studie trägt zur Etablierung chromosomaler Marker sowie zu einem besseren Verständnis der intra- und interspezifischen Variation der Capsaicinoidzusammensetzung und des Capsaicinoidgehalts in der Pubescens Klade bei.