# Disentangling the intricate variation patterns in the diploid-polyploid complex *Veronica* subsect. *Pentasepalae* Benth. (*Veronica* L., Plantaginaceae sensu APG IV). Relationship with the Quaternary climatic oscillations

---

# Desentrañando los intricados patrones de variación en el complejo diploid-poliploide complex *Veronica* subsect. *Pentasepalae* Benth. (*Veronica* L., Plantaginaceae sensu APG IV). Relación con las variaciones climáticas del periodo cuaternario

FACULTAD DE BIOLOGÍA

DEPARTAMENTO DE BOTÁNICA Y FISIOLOGÍA VEGETAL

ÁREA DE BOTÁNICA

TESIS DOCTORAL

# Disentangling the intricate variation patterns in the diploid-polyploid complex *Veronica* subsect. *Pentasepalae* Benth. (*Veronica* L., Plantaginaceae *sensu* APG IV). Relationship with the Quaternary climatic oscillations

---

# Desentrañando los intricados patrones de variación en el complejo diploid-poliploide complex *Veronica* subsect. *Pentasepalae* Benth. (*Veronica* L., Plantaginaceae *sensu* APG IV). Relación con las variaciones climáticas del periodo cuaternario

Memoria presentada por la Licenciada en Biología Noemí López González para optar al título de Doctor en Biología por la Universidad de Salamanca.

Salamanca, 2020

VNiVERSiDAD D SALAMANCA
CAMPUS DE EXCELENCIA INTERNACIONAL

VNiVERSiDAD D SALAMANCA
1218 ~ 2018

DEPARTAMENTO DE BOTÁNICA
Y FISIOLOGÍA VEGETAL

ÁREA DE BOTÁNICA

Las Doctoras **Dña. M. Montserrat Martínez-Ortega**, Catedrática de Botánica de la Universidad de Salamanca; y **Dña. Blanca M. Rojas-Andrés**, *Wissenschaftliche Mitarbeiterin* de la *Universität Leipzig*.

Certifican que:

La Tesis Doctoral titulada "**Disentangling the intricate variation patterns in the diploid-polyploid complex *Veronica* subsect. *Pentasepalae* Benth. (*Veronica* L., Plantaginaceae sensu APG IV). Relationship with the Quaternary climatic oscillations** (Desentrañando los intricados patrones de variación en el complejo diploid-poliploide complex *Veronica* subsect. *Pentasepalae* Benth. [*Veronica* L., Plantaginaceae sensu APG IV]. Relación con las variaciones climáticas del periodo cuaternario)" que presenta **Dña. Noemí López González** para optar al título de Doctor en Biología por la Universidad de Salamanca, ha sido realizada bajo su dirección, en el Área de Botánica de la Facultad de Biología de la Universidad de Salamanca y reúne todos los requisitos científicos y formales necesarios para su defensa.

En Salamanca, a 20 de abril de 2020.

DEPARTAMENTO DE BOTÁNICA
Y FISIOLOGÍA VEGETAL

ÁREA DE BOTÁNICA

La presente Tesis Doctoral está elaborada en el formato de compendio de artículos/publicaciones según la normativa aprobada por la Comisión de Doctorado y Posgrado de la Universidad de Salamanca el 15 de febrero de 2013, y consta de las siguientes publicaciones:

**Divide and conquer! Data-mining tools and sequential multivariate analysis to search for diagnostic morphological characters within a plant polyploid complex (*Veronica* subsect. *Pentasepalae*, Plantaginaceae).**

Noemí López-González[1]*, Santiago Andrés-Sánchez[1,2], Blanca M. Rojas-Andrés[1¤], M. Montserrat Martínez-Ortega[1]

[1]*Departamento de Botánica y Biobanco de ADN vegetal, Universidad de Salamanca, Salamanca, Spain,*

[2]*Departamento de Didáctica de las Matemáticas y Didáctica de las Ciencias Experimentales, Universidad de Salamanca, Salamanca, Spain*

[¤]*Current address: Department of Molecular Evolution and Plant Systematics & Herbarium (LZ), Institute of Biology, Leipzig University, Leipzig, Germany.*

*PLoS ONE* (2018) 13(6):e0199818.
DOI: https://doi.org/10.1371/journal.pone.0199818
IF = 2.776; within Q2 in Multidisciplinary Sciences, Position: 24/69.

**Plantaginaceae. In: Karol Marhold & Jaromír Kučera (eds.), IAPT/IOPB chromosome data 28.**

Luis Delgado[1,*], Blanca M. Rojas-Andrés[1,2,3], Noemí López-González[1,2], Nélida Padilla-García[1,2] & M. Montserrat Martínez-Ortega[1,2]

[1] *Departamento de Botánica y Fisiología Vegetal, University of Salamanca, 37007 Salamanca, Spain*

[2] *Biobanco de ADN Vegetal, University of Salamanca, Edificio Multiusos I+D+i, Calle Espejo s/n, 37007 Salamanca, Spain*

[3] Present address: *Department of Molecular Evolution and Plant Systematics & Herbarium (LZ), Institute of Biology, Leipzig University, Leipzig, Germany*

*Taxon* (2018) 67 (6): 1235-1245.
DOI: https://doi.org/10.12705/676.39
(IF = 3.823; within Q1 in Plant Sciences; Position: 27/228).

**Characterization of 12 polymorphic SSR markers in *Veronica* subsect. *Pentasepalae* (Plantaginaceae) and cross-amplification in 10 other subgenera.**

Noemí López-González[1,2], Eike Mayland-Quellhorst[3], Daniel Pinto-Carrasco[1,2], M. Montserrat Martínez-Ortega[1,2]

[1] *Departamento de Botánica, Universidad de Salamanca, E-37007 Salamanca, Spain*

[2] *Biobanco de ADN Vegetal, Banco Nacional de ADN, Edifi cio Multiusos I+D+i, Calle Espejo s/n, 37007 Salamanca, Spain*

[3] *Institut für Biologie und Umweltwissenschaften, Carl von Ossietzky Universität Oldenburg, Carl von Ossietzky-Str. 9–11, 26111 Oldenburg, Germany*

*Applications in Plant Sciences* (2015) 3 (10): 1500059.
DOI:10.3732/apps.1500059
(IF = 0.911; within Q3 in Plant Sciences; Position: 142/209).

Adicionalmente se incluyen dos artículos, correspondientes a los capítulos V y VI. El segundo se encuentra en fase de preparación mientras que el primero ha sido enviado para su publicación en una revista indexada en Journal Citation Reports:

**Genetic similarities versus morphological resemblance: Formation of polyploids in a Mediterranean biodiversity hotspot.**

Noemí López-González[a,b], Javier Bobo-Pinilla[a,b], Nélida Padilla-García[a,b], J. Loureiro[d], S. Castro[d], Blanca M. Rojas-Andrés[c], and M. Montserrat Martínez-Ortega[a,b]

[a] *Departamento de Botánica y Fisiología vegetal, University of Salamanca, E-37007 Salamanca, Spain*

[b] *Biobanco de ADN Vegetal, University of Salamanca, Edificio Multiusos I+D+i, Calle Espejo s/n, 37007 Salamanca, Spain*

[c] *Department of Molecular Evolution and Plant Systematics & Herbarium (LZ), Institute of Biology, Leipzig University, Johannisallee 21-23, 04103 Leipzig, Germany.*

[d] *Centre for Functional Ecology, Department of Life Sciences, University of Coimbra, Calçada Martim de Freitas, 3000-456 Coimbra, Portugal.*

Manuscrito enviado a la revista *Molecular Phylogenetics and Evolution.*
Manuscript Number: MPE-D-20-00083

**From a Mediterranean hotspot to the mesic grassland areas of Europe: Exploring the genetic, morphological and ecological diversity of the Austrian speedwell.**

Este manuscrito se encuentra actualemente en fase de preparación.

VNiVERSiDAD D SALAMANCA
CAMPUS DE EXCELENCIA INTERNACIONAL

VNiVERSiDAD D SALAMANCA
1218 ~ 2018

DEPARTAMENTO DE BOTÁNICA
Y FISIOLOGÍA VEGETAL

ÁREA DE BOTÁNICA

Las Doctoras **M. Montserrat Martínez-Ortega**, Catedrática de Botánica de la Universidad de Salamanca; y **Dña. Blanca M. Rojas-Andrés**, *Wissenschaftliche Mitarbeiterin* de la *Universität Leipzig*.

Autorizan:

Que la Tesis Doctoral titulada "**Disentangling the intricate variation patterns in the diploid-polyploid complex *Veronica* subsect. *Pentasepalae* Benth. (*Veronica* L., Plantaginaceae sensu APG IV). Relationship with the Quaternary climatic oscillations** (Desentrañando los intricados patrones de variación en el complejo diploid-poliploide complex *Veronica* subsect. *Pentasepalae* Benth. [*Veronica* L., Plantaginaceae sensu APG IV]. Relación con las variaciones climáticas del periodo cuaternario)" sea presentada en la modalidad de compendio de artículos/publicaciones (Comisión de Doctorado y Posgrado, 15 de febrero de 2013).

Y para que así conste expiden y firman el presente certificado en Salamanca, a 20 de abril de 2020.

La Comisión Académica del Doctorado en Biología y Conservación de la Biodiversidad

Autoriza:

La presentación de la Tesis Doctoral titulada "**Disentangling the intricate variation patterns in the diploid-polyploid complex *Veronica* subsect. *Pentasepalae* Benth. (*Veronica* L., Plantaginaceae sensu APG IV). Relationship with the Quaternary climatic oscillations.** (Desentrañando los intrincados patrones de variación en el complejo diploide-poliploide de *Veronica* subsect. *Pentasepalae* Benth. (*Veronica* L., Plantaginaceae sensu APG IV). Relación con las variaciones climáticas del periodo cuaternario.)" para su tramitación en la modalidad de compendio de artículos/publicaciones (Comisión de Doctorado y Posgrado, 15 de febrero de 2013).

En Salamanca, a 13 de marzo de 2020

# PREAMBLE

º

# PREÁMBULO

This PhD thesis has been developed at the University of Salamanca (Spain) within the framework of the doctoral program *"Biología y Conservación de la Biodiversidad"*. This PhD thesis has been funded by the Spanish *Ministerio de Economía y Competitividad* through the projects *"Mecanismos evolutivos, filogenia, taxonomía y patrones filogeográficos en* Veronica *subsect.* Pentasepalae *Benth. (*Veronica *L., Plantaginaceae sensu APG II)"* (reference CGL2009-07555), and *"Desentrañando los intricados patrones de variación en el complejo diploide-poliploide de* Veronica *subsect.* Pentasepalae *Benth. (*Veronica *L., Plantaginaceae sensu APG III). Relación con el ultimo máximo glacial (UMG) y la recuperación climática del Holoceno Medio"* (reference CGL2012-32574) and the Spanish *Ministerio de Educación* through a PhD scholarship within the program "*Subprograma de Formación del Profesorado Universitario*" (reference AP2010-2968).

The research project has been conducted at the University of Salamanca in the Department of Botany and Plant Physiology and the Plant DNA Biobank. Further experiments and data analyses related with this PhD thesis project have also been carried out during research stays performed at the Institute of Biology and Environmental Sciences (Carl von Ossietzky University of Oldenburg, Oldenburg, Germany), the *Centre for Functional Ecology* (University of Coimbra, Coimbra, Portugal), the Institute of Botany of the Czech Academy of Sciences (Průhonice, Czech Republic), and the Florida Museum of Natural History (University of Florida, Gainesville, USA).

Additionally, field surveys related to the PhD thesis project have been carried out all across Europe including the following countries: Albania, Austria, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, France, Germany, Greece, Hungary, Italy, Montenegro, Republic of Kosovo, Republic of North Macedonia, Romania, Serbia, Slovakia, Slovenia, Spain, and Switzerland.

This manuscript is composed of VII chapters:

In Chapter I, an overview of the main topics covered in this PhD thesis is provided. The study group, *Veronica* subsect. *Pentasepalae* is here introduced with a specific focus on the diploid-polyploid '*V. austriaca – V. orbiculata* complex' and the '*V. austriaca complex*'. Additionally, the main goals of the thesis are presented.

In Chapter II, a thorough morphometric study of *V.* subsect. *Pentasepalae*—including for the first time all taxa currently recognized within the subsection for which sufficient material was available—is presented. A classical multivariate analysis (Discriminant Analysis) and two well-known data-mining techniques (Decision Trees and Artificial Neural Networks) were used in order to search for discriminant morphological characters to allow accurate taxon identification in a taxonomically intricate species group. This chapter has been published as a research article in the journal *Plos One*.

In Chapter III, vouchered chromosome counts and ploidy level estimates for several species from *Veronica* subsect. *Pentasepalae* are provided. Some of them represent the first chromosome count available for particular species. This chapter has been published as a Chromosome Data Series in the journal *Taxon*.

In Chapter IV, a new set of polymorphic microsatellite markers for *Veronica* subsect. *Pentasepalae* is reported. Twelve pairs of primers were identified and optimized using a microsatellite-enriched library method and 454 GS-FLX technique. Amplification success for these markers in the cross-transferability tests extends their potential

usefulness to other subgenera. This chapter has been published as a Primer Note in the journal *Applications in Plant Sciences*.

In Chapter V, a study that explores the genetic structure and relationships of the species belonging to the "*V. austriaca – V. orbiculata* diploid-polyploid complex" is presented. This study is focused on the Western Balkans, where active hybridization and species diversification processes have been and are notably active. Nuclear microsatellites markers, plastid DNA regions and ploidy level estimations were employed to assess the genetic structure and evolutionary dynamics of this polyploidy complex. Approximate Bayesian computation analyses are combined with species paleodistribution models to evaluate hypothesis regarding evolutionary history. The manuscript related to this chapter has been submitted to the journal *Molecular Phylogenetics & Evolution.*

In Chapter VI, a study about the perennial herbs included in the '*V. austriaca* complex' is presented. A combined approach using data from nuclear microsatellite markers, plastid DNA and ploidy levels for investigating patterns of intraspecific biodiversity is applied. Morphological and ecological variation is also evaluated though leaf and fruit measurements and present potential distributions respectively. The manuscript related to this chapter is currently in preparation for submission.

In Chapter VII, the main conclusions of this PhD thesis are displayed.

Lastly, the Appendix comprises published papers included in other PhD Dissertations in which I have participated as co-author.

AGRADECIMIENTOS

Este es el *behind the scenes* de la tesis, la parte en la que aparece toda esa gente sin la que no hubiera podido llegar hasta aquí. Y es mucha gente, lo siento si no aparecéis todos.

Los primeros tenéis que ser vosotros: mamá, papá, David, esto no habría sido posible sin vosotros. Sois las mejores personas que conozco. Gracias.

A mis lindas directoras de tesis, Montse y Blanca, gracias por acompañarme todos estos años: desde aquel viaje por los alrededores de Paris que acabó en Frankfurt, pasando por los meses que vivimos juntas, las campañas de muestreo (con sus anécdotas asociadas, que darían para más páginas que la tesis), el día a día en el departamento, hasta estas últimas semanas. Gracias por el apoyo y por enseñarme tanto.

Al resto del *Veronica freak group*, Nel y Xime, gracias por esas carreras para ser las primeras en ver la *Veronica* (aunque siempre gane Xime) y gracias por las aventuras que hemos compartido (¡aun tengo el llavero del hotel de Italia!). Gracias a Xime y a Quique por hacer que viera las plantas de otra manera y gracias al resto de gente de Botánica. En especial a Luis que siempre está ahí para cualquier duda, a Sergio por su serenidad y a Juan Carlos por salvarme más de una vez (inciso: gracias a toda la gente que alguna vez me ha ayudado con el papeleo, desde la señora rubia a María José, porque es un alivio que alguien te diga "tranquila que está todo").

Gracias a los pollitos, a todos los que habéis pasado por aquí, porque sois el alma del departamento. Gracias a Teresa, la pipeta más rápida del salvaje oeste. Gracias Víctor por el apoyo en los últimos meses, (alerta *spoiler*) he añadido a Askell Löve en la introducción acordándome muchísimo de ti. Manu, mi favorito para los asuntos de ordenador (y para arrancar *Verbascum*…). Muchísimas gracias a mis compañeros del primer Iter: David, Pinto, Santi, (Bobo, tu no fuiste, pero estabas, como los chinos) y los que os habéis incorporados en segundas y terceras ediciones: Paula y Walias. De mayor quiero ser como vosotros. Gracias a todos porque habéis hecho que la tesis, el día a día, sea un poquito más fácil. Repetiremos los viajes. Si tenemos dinero. Cuando nos dejen salir. Bueno y cuando nos dejen viajar. Igual ponen límite al número de personas que viajan juntas. Mira, nos tomamos una cerveza y luego ya vamos viendo.

Muchísimas gracias a toda la gente maravillosa que me he ido cruzando en las estancias y en los viajes en general. Gracias a los que me acogisteis: Dirk, Zuzanna, Pam & Doug y, sobre todo, gracias a Silvia y a Joao, da gusto conoceros, es un placer seguir trabajando con vosotros. Gracias a la gente que conocí por el camino: Leslie, Manu, por los martes (y jueves) de *language exchange*; gracias Victoria por las *parade* y las *homecoming* y por todas las americanadas que he vivido, sin ti no habría sido posible; a mi indio astrónomo particular,

Sankalp; a Pau por las excursiones a ver mariposas, armadillos, caimanes, serpientes… un montón de bichos (y gracias a Mercé por acogernos en su casa cuando fuimos a verte); gracias Matt por tu inestimable ayuda con los Mac; gracias Evgeny por llevarme de un lado a otro y por las copas de vino que compartimos; gracias Simon por aparecer cuando necesitaba ayuda con la bici, o en el laboratorio, o con cualquier cosa; gracias Ana Juan por la visita a Oldenburg (y por más cosas, pero me acuerdo mucho de esa visita); gracias Heidi por tu alegría y aquella cena maravillosa en tu casa; gracias Mariana porque conviertes el trabajo de citómetría en un arte; gracias Magda porque me encantó viajar y hablar y ver pelis y todo contigo; gracias Inés por la ayuda desde el primer momento cuando fuimos de visita al jardín y gracias Bruno porque (más o menos) desde aquel día has sido un pollito más, por eso y por el *megahit* "híbrido del prado morado". Espérame en tu casona.

Y, por supuesto, gracias a las de siempre. A Bea, Aleja, Ire (sister), Anita, Glo, Marta, Ana, Merino, Amaya, Estefa, Natalia (la que cumple años el 17 de oct.), Sari, Alberto, Bober, Tere, Mance, Lara, Lau, Casas, Anuar, Juanillo. Podría agradeceros tanto a cada uno de vosotros que se me haría esto muy largo. Y además soy bastante rancia. Así que va en general: gracias a todos por ser tan buenas personas, por las risas con chistes de no tan buenas personas, por las quedadas para cenar (y las apuestas de dichas cenas), por las "cervezas de emergencia" y las de "ya era hora que nos juntásemos, ¿no?", por las llamadas de ciento en viento que duran horas. Por darme ánimo y por aguantarme. Gracias porque con vosotros la vida es mejor.

Y luego estás tú ("Y luego está la…" ¿la qué? ¿quién empieza un capítulo así?). Pura, irreductible, tú. Muchísimas, muchísimas… (ejem). Muchísimas gracias.

Gracias a todos. Haberme encontrado con vosotros, vuestras visitas y el haberos conocido me hace reconciliarme con la humanidad (luego pongo la tele y se me pasa).

PD- Invito a una cerveza a todo el que me diga la referencia a J. L. Perales.

# INDEX
•
# ÍNDICE

## CHAPTER 1 • CAPÍTULO 1

### General introduction



### Introducción general

## CHAPTER 2 • CAPÍTULO 2

---

**Divide and conquer! Data-mining tools and sequential multivariate analysis to search for diagnostic morphological characters within a plant polyploid complex (*Veronica* subsect. *Pentasepalae*, Plantaginaceae)**

***¡Divide y vencerás! Herramientas de minería de datos y análisis multivariante secuencial para la búsqueda de caracteres morfológicos diagnósticos en un complejo poliploide de plantas* (Veronica *subsect.* Pentasepalae*, Plantaginaceae)***

## CHAPTER 3 • CAPÍTULO 3

---

**Plantaginaceae. In: Karol Marhold & Jaromír Kučera (eds.), IAPT/IOPB chromosome data 28.**

***Plantaginaceae. Editores: Karol Marhold & Jaromír Kučera. Asociación Internacional para la Taxonomía Vegetal. Datos cromosómicos 28***

## CHAPTER 4 • CAPÍTULO 4

**Characterization of 12 polymorphic SSR markers in *Veronica* subsect. *Pentasepalae* (Plantaginaceae) and cross-amplification in 10 other subgenera.**

***Characterización de 12 marcadores SSR polimórficos en* Veronica *subsect.* Pentasepalae *(Plantaginaceae) y amplificación cruzada en otros 10 subgéneros.***



## CHAPTER 5• CAPÍTULO 5

Genetic similarities versus morphological resemblance: Formation of polyploids in a Mediterranean biodiversity hotspot.

*Similaridad genética versus semejanza morfológica: formación de poliploides en un punto caliente de biodiversidad en el Mediterráneo.*

## CHAPTER 6 • CAPÍTULO 6

---

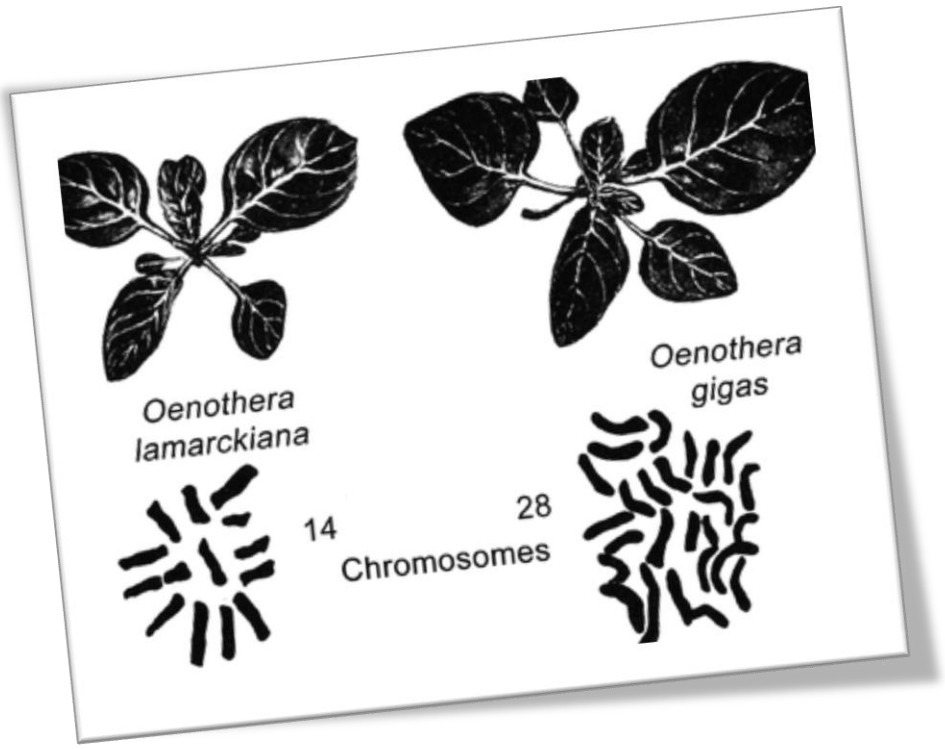
From a Mediterranean hotspot to the mesic grassland areas of Europe: Exploring the genetic, morphological and ecological diversity of the Austrian speedwell

*Desde un punto caliente de biodiversidad mediterráneo hasta las grandes praderas mésicas de Europa: Explorando la diversidad genética, morfológica y ecológica de* Veronica austriaca

## CHAPTER 7 • CAPÍTULO 7



## APPENDIX •APÉNDICE




# CHAPTER 1

## General introduction

## 1. POLYPLOIDY AND HYBRIDIZATION

Polyploids are defined as organisms that have more than two sets of chromosomes and the process that mediates this acquisition is generally called polyploidy or whole-genome duplication (WGD).

Two types of polyploids are generally recognized according to the origin of parental genomes (Fig. 1). Autopolyploidy is defined as genome duplication within one parental genome, which results in homologous chromosome sets in the cell (diploid AA doubles to become autotetraploid AAAA). Allopolyploidy is defined as WGD associated with the merger of two more or less divergent genomes following interspecific hybridization, resulting in homoeologous chromosome sets in the cell (AA × BB ⇨ AABB). Thus polyploidy includes intra- or interspecific hybridization; however, hybridization may also occur without genome duplication giving rise to hybrids with the same number of chromosomes as the parental species (homoploid hybridization in this case).

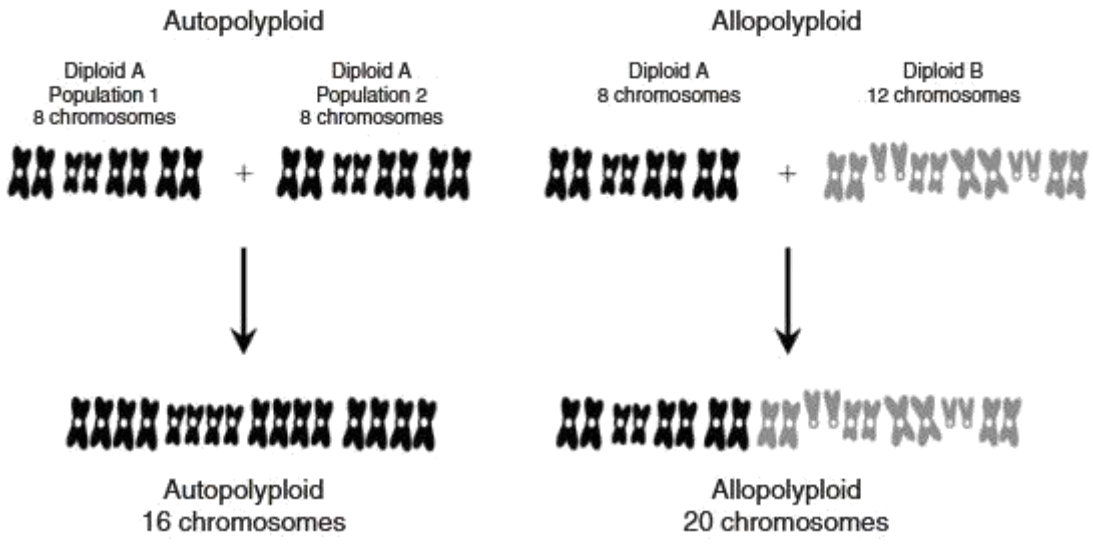



**Fig. 1.** Autopolyploid versus allopolyploid formation. Autopolyploids are formed by intraspecific hybridization, whereas allopolyploids are formed by interspecific hybridization (Tate et al., 2005).

Polyploidy was documented for the first time in plants at the beginning of the 20th century. Based on the early genetic work by Hugo de Vries (1905), Luntz (1907) and Gates (1909) discovered independently that *Oenothera lamarckiana* mutant Gigas was a tetraploid. At the same time, Kuwada (1911) proposed that an ancient chromosome duplication would have occurred in maize. A few years later, Winge (1917) introduced the term "polyploidy" and provided the most influential explanation for the origin of these plants with sets of doubled chromosomes. He proposed that these polyploid plants were fertile because genome doubling restored chromosome pairing in sterile hybrids. Further support to this hypothesis came from Müntzing (1932) who produced "experimental polyploids" (i.e. synthetic polyploids) copying natural polyploid species. He observed that hybrids between *Galeopsis pubescens* Besser and *G. speciosa* Mill. were mostly sterile. However, crossing one of these hybrids (a triploid individual) and the parental *G. speciosa* resulted in a fertile tetraploid plant. He concluded that natural polyploids may have followed a similar mechanism to avoid postzygotic reproductive isolation. This study (among others, e.g. Clausen and Goodspeed, 1925) demonstrated one of the first mechanisms of how a new species may arise from drastic genomic changes of existing species.

During the 30s, polyploidy was a relatively well-studied topic. It was identified as the greatest difference between plant and animal speciation in the modern synthesis (Dobzhansky, 1937). Indeed, it was the most important amendment to Darwin and Wallace's theory of the origin of species (Haldane, 1959). The possible important role of ancient polyploidy in diverse lineages was accepted (Stebbins, 1985) and the polyploid origin of many plant species was unequivocal (Klekowski and Baker, 1966; Ohno, 1970) Despite this first acknowledged importance, later in the 20th century polyploidy was relegated as a largely irrelevant topic by many researchers. The

contribution of polyploidy to evolution started to be questioned (Stebbins, 1950, 1971). Referring to polyploids and interspecific hybrids Wagner (1970) stated: "In the broad picture of evolution, may these phenomena not merely be trivial aberrations and defects in the biology of plants, produced by simple genetic transformations or chance fertilizations, which plants can tolerate, but which have little or no importance to the big picture?". For some authors, the prevalence and persistence of diploids indicated that all meaningful long-term evolutionary impact relied at the diploid level, while polyploids by themselves had little evolutionary potential (Stebbins, 1971; Grant, 1981). Hybrids and polyploids were described as "blind alleys", "evolutionary dead-ends" or simply "evolutionary noise" (Wagner 1970; Stebbins 1971). During this time, researchers as A. Mützing, C. D. Darlington and A. Löve, maintained that polyploidy and hybridization were both common and evolutionary remarkable for plants (Darlington, 1963; Löve, 1964; Müntzing, 1965), although they were proved to be less influential than Stebbins (Ramsey and Ramsey, 2014).

The idea of polyploidization as a process with little evolutionary relevance started to be refuted during the genomics era (2000 – present) (Adams and Wendel, 2005). Several studies provided proof on the evolutionary potential of polyploids as many sequenced eukaryotic genomes displayed evidence of ancestral polyploidy (e. g. Gu et al., 2002; Wong et al., 2002; Becak and Kobashi, 2004; Paterson, 2005; Van de Peer and Meyer, 2005; Yu et al., 2005) indicating that polyploidy might confer long-term evolutionary flexibility (Comai, 2005). There is now consensus about polyploidy to have occurred at least once in the evolution of seed plants (Jiao et al., 2011; Wendel, 2015). Indeed, the evolutionary success of flowering plants has been associated with genome doublings (Fawcett et al., 2009; Van de Peer et al., 2009).

Polyploidization is not important only at ancient scales; it can alter plant morphology, phenology, physiology and/or ecology within only one or a few generations (Levin, 2002). The dynamic nature of polyploid genomes —with alterations in gene content, gene number, gene arrangement, gene expression and transposon activity— represent a source for novel biodiversity (Soltis et al., 2014). While some combinations will not prosper, certain unique combinations of newly generated variation may result in enormous evolutionary potential and adaptive capabilities (Soltis, 2013). Thus, polyploidy represents a significant force in plant evolution, at temporal scales ranging from ancient to contemporary, and with deep consequences from the molecular to the ecological level (Adams and Wendel, 2005).

Nowadays research on polyploidy is as vigorous as ever (Barker et al., 2016). Important advances have been made towards better predicting the outcomes of polyploidy (e.g. Hollister and Gaut, 2009; Xiong and Pires, 2011; Wendel, 2015; Steige and Slotte, 2016). However, patterns of evolution following polyploidy are inconsistent and establishing a paradigm of polyploidy is challenging (Soltis et al. 2016).

Although many questions remain about the consequences of genome duplication, two findings from molecular studies of polyploid complexes are of particular relevance. First, auto- and allopolyploidization are both common and widespread in nature (Soltis et al., 2007; Spoelhof et al., 2017) and these two forms represent the extremes of a continuum. It is important to realize that the evolutionary origin of all natural polyploids (i.e., both auto- and allopolyploids) involves hybridization between more or less related genomes (Parisod and Senerchia, 2012; Abbott et al., 2013). Consequently, hybridization and polyploidization are designations for complex evolutionary processes occurring in nature which are profoundly interrelated (Fig. 2) (Marques et al., 2018).

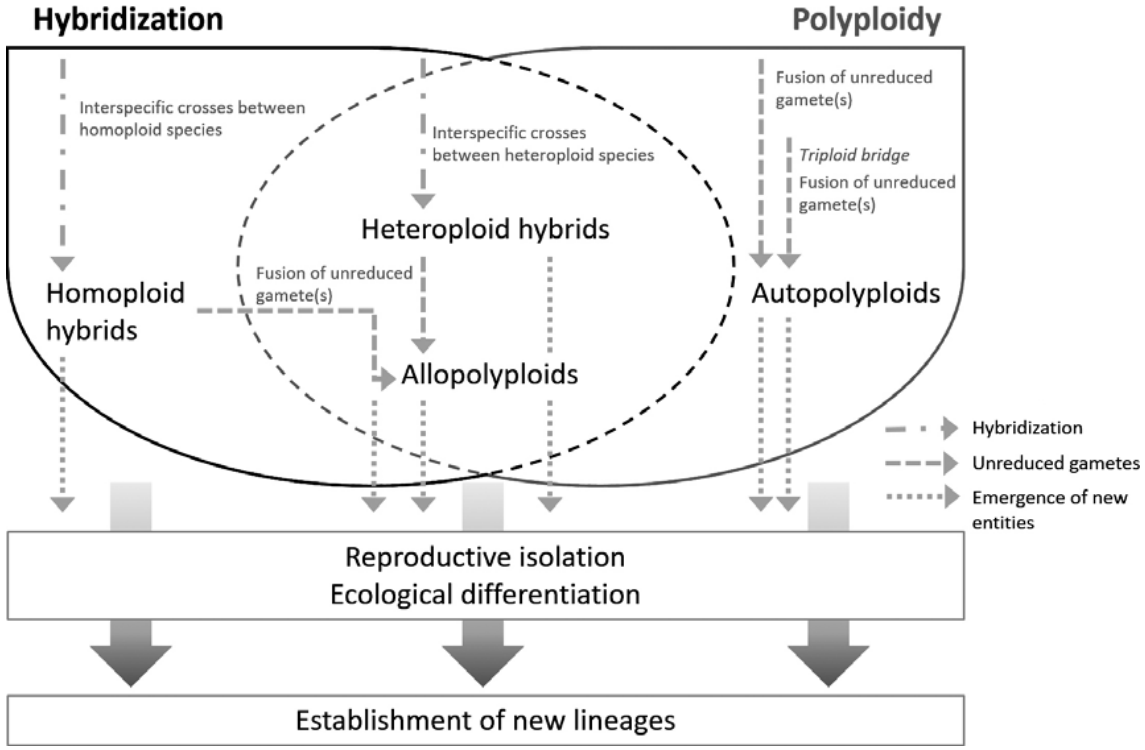



**Fig. 2.** Schematic representation of the intricate connection between hybridization and polyploidization in nature (Marques et al., 2018).

Second, molecular data have confirmed numerous independent origins of polyploids within plant genera or even single taxonomic species (Parisod and Besnard, 2007; Soltis et al., 2016); indeed, recurrent formation of polyploids seems to be the rule rather than the exception in nature (Barker et al., 2016).

Polyploidy and hybridization will be the *leitmotive* of the chapters composing the present PhD thesis, whose main objective is to contribute to the understanding of the relationships and evolutionary history of the entities within the diploid-polyploid complex *Veronica* subsection *Pentasepalae* Benth.

## 2. DEFINING AND REFINING THE STUDY GROUP: FROM VERONICA SUBSECT. PENTASEPALAE TO THE 'VERONICA AUSTRIACA-ORBICULATA COMPLEX'

### *2.1. Taxonomic framework*

The genus *Veronica* L. has been traditionally included in the family Scrophulariaceae. In the first APG classification (1998) it was transferred to the Plantaginaceae family becoming the largest genus within this family. *Veronica* includes more than 450 species mainly distributed in the Northern Hemisphere and Australasia with a few species present in Africa and South America.

The vast genus *Veronica* is currently divided into 12 subgenera, all of them monophyletic (Garnock-Jones et al., 2007). Among those represented in the Northern Hemisphere, *V.* subgenus *Pentasepalae* (Benth.) M. M. Mart. Ort., Albach & M. A. Fisch. comprises 70-75 species of perennial herbs distributed in Eurasia and North Africa. The species in this subgenus are generally characterized by a pentapartite calyx (rarely quadripartite) with the fifth sepal being significantly smaller (Albach et al., 2004), and seeds with a verrucate coat (Muñoz-Centeno et al., 2006). The base chromosome number, a phylogenetically important character within the genus, is $x = 8$. Four subsections are recognized within *V.* subgenus *Pentasepalae* according to the most recent classification (Albach et al., 2008): *V.* subsect. *Armeno-Persicae* Stroh, *V.* subsect. *Orientales* (Wulff) Stroh, *V.* subsect. *Petraea* Benth., and *V.* subsect. *Pentasepalae* Benth. The last one, *V.* subsect. *Pentasepalae*, which was proven to be monophyletic (Rojas-Andrés et al., 2015), represents the study group of the present work.

*Veronica* subsect. *Pentasepalae* was described for the first time by Bentham (1846) as the "Pentasepalae group" included within *V.* sect. *Chamaedrys* W. D. J. Koch. At the beginning, it comprised a few species, although in the following years many authors

have revisited the group proposing different taxonomic treatments with substantial variations (see Rojas-Andrés and Martínez-Ortega, 2016 for a complete review on taxonomic treatments). Different taxa have been included in each revision and many names have been applied in different senses (e. g. Watzl, 1910; Walters and Webb, 1972; Fischer, 1982; Peev, 1995; Martínez-Ortega et al., 2009; Tison and Foucault, 2014), making the taxonomy of the group highly complicated. This active taxonomic history has led to the accumulation of more than 200 names for ca. 20 taxa.

Recently, Rojas-Andrés and Martínez-Ortega have proposed a thorough nomenclatural and taxonomic treatment of *V.* subsect. *Pentasepalae* (Rojas-Andrés and Martínez-Ortega, 2016; Rojas-Andrés et al., 2016). According to this exhaustive revision, *V.* subsect. *Pentasepalae* is a diploid-polyploid complex composed of 22 perennial taxa (17 species, four subspecies and one variety). The species and infraspecific taxa are mainly distributed in Europe, with the exception of *V. krylovii* Schischk, which only occurs in Siberia and Kazakhstan; and *V. rosea* Desf., which occurs in the north of Algeria and Morocco. This taxonomic treatment has been slightly modified on the base of AFLP fingerprinting and DNA ploidy-level estimations (Padilla-García et al., 2018).

Successive studies exploring *Veronica* subsect. *Pentasepalae* have shown that it is a recently diversified complex in which genetic isolation barriers are not definitely established (Martínez-Ortega et al., 2004; Rojas-Andrés et al., 2015; Padilla-García et al., 2018). In addition, both polyploidy and hybridization have been identified to occur in the subsection and this may represent an important source of homology (Martínez-Ortega et al., 2004). Polyploidy and hybridization have been recognized as processes causing morphological alterations that make species boundaries diffuse and avoid clear-cut recognition of closely related taxa (Stace, 2000). In extreme cases morphological delimitation of species might be simply impossible (e.g. Nielsen and Olrik, 2001; Li et

al., 2017). As a consequence of polyploidy, hybridization and weak reproductive barriers, the borders between species or infraspecific taxa within *V.* subsect *Pentasepalae* are blurred. However, in the case of this subsection, we have the advantage of having at our disposal a solid and well-supported taxonomic hypothesis at a fine scale level (individuals and populations), based on genetic data, ploidy level information, overall morphology and geographical range (Rojas-Andrés and Martínez-Ortega, 2016). Multiple lines of evidence have been used to propose this taxonomic hypothesis, following an integrative taxonomic approach (Dayrat, 2005) and the general lineage species concept of De Queiroz (2005, 2007). Hence, well-established taxonomic entities can be used as a reference for an accurate selection of morphological discriminant traits.

### *2.2 Investigating the patterns of variation of morphological characters in V. subsect. Pentasepalae: Data mining versus multivariate discriminant analyses and the "Divide and conquer" approach*

*Veronica* subsect. *Pentasepalae* comprises perennial herbs, suffrutex, usually woody at the base, with decumbent to erect stems that bear a vegetative apical shoot. Leaves are generally opposite, sessile to shortly petiolate, entire to bipinnatisect, and glabrous to densely covered by eglandular hairs. The apical shoot leaves are generally opposite. Inflorescences are racemose and generally axillary. Flowers are hermaphrodite, with a calyx generally smaller than the capsule, persistent, glabrous, subglabrous or pilose. The four-lobed corolla is slightly zygomorphic, larger than the calyx, violet, pale violet or dark blue, rarely whitish or pink, usually with darker veins. Fruit in capsule, bilocular, dehiscent, in general laterally compressed.

The most discriminant morphological traits are those related to leaf morphology, but other sometimes helpful diagnostic characters are plant size, growth form, number and hairiness of sepals, and fruit morphology (Martínez-Ortega et al., 2004). Specifically, the aforementioned taxonomic treatments available for *V.* subsect. *Pentasepalae* consider and use leaves as a primary source for species identification, mainly because floral features show little variation in this subsection of *Veronica* and also because they are quite ephemeral.

In the above described situation, to improve the morphological knowledge of *Veronica* subsect. *Pentasepalae*, it seems necessary to think of and apply an innovative integrative approach in terms of methodology. On the one hand, modern tools such as data-mining techniques have proven to be successful in a wide variety of fields such as marketing, chemistry or social studies (e. g. Delen et al., 2005; Fischer et al., 2006; Paliwal and Kumar, 2009; Baker, 2010; Michel et al., 2011; Strelcov et al., 2014). Data-mining is the core step in the Knowledge Discovery in Databases, and data-mining tools are thought to find and describe structural patterns in data (Maimon and Rokach, 2010). On the other hand, classical statistical analyses through multivariate discriminant analyses (i.e. multivariate morphometrics) represent a robust tool for evaluating variation patterns and have continuously been applied with successful outcomes (e.g. Henderson, 2006; Marhold, 2011; Lorenz et al., 2014).

The combination of the aforementioned techniques is applied under one of the most recurrent problem-solving methods: the “divide and conquer” strategy. In this approach, a particular problem is divided into smaller parts, which are resolved more easily, and then these solutions are used to build a complex solution for the initial problem (Fig. 3).

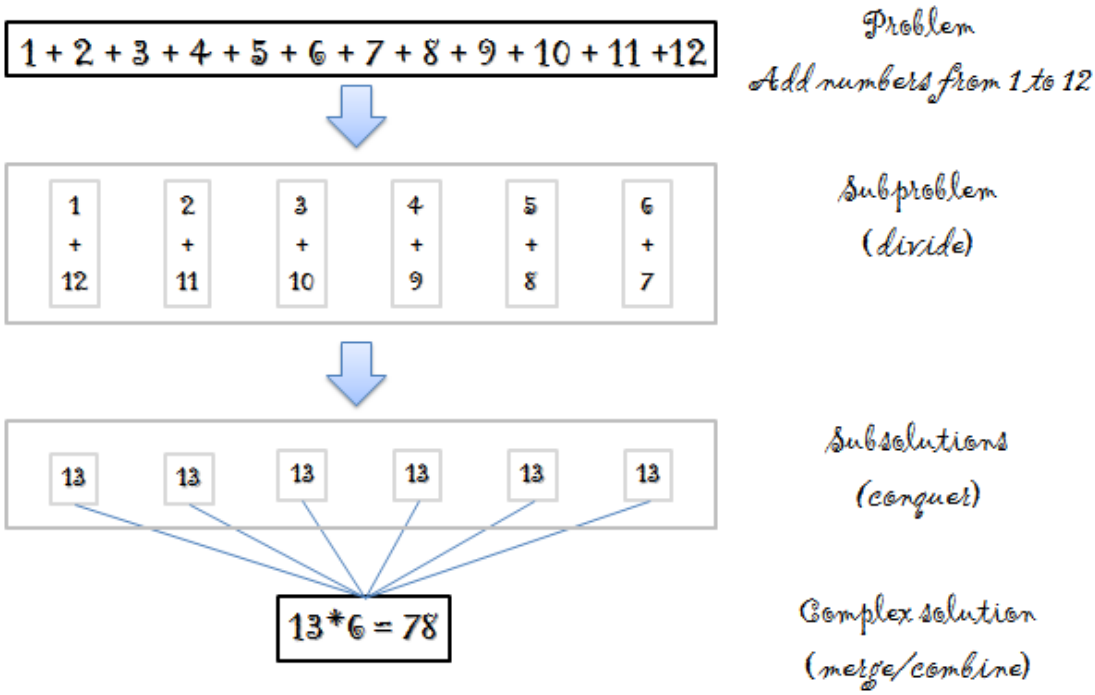



**Fig. 3.** Example of the "Divide and conquer" strategy. Modified from Magnus and Berner (2000).

### *2.3. Previous phylogenetic and evolutionary studies on V. subsect. Pentasepalae*

Several studies have previously tackled the evolutionary mechanisms and have investigated the underlying phylogenetic relationships among taxa within *V.* subsect. *Pentasepalae*. The first approach combining morphometric data and molecular markers was initiated with the PhD Thesis presented at the University of Salamanca by Martinez-Ortega (1999) and those studies that followed it, all of them at least partially financed by the project '*Flora iberica*' (Martínez-Ortega, 2004; Andrés-Sánchez et al., 2009; Martínez-Ortega et al., 2009); these works mainly included species and intraspecific taxa from the western Mediterranean.

The first phylogenetic analysis based on DNA sequences that included representatives for the whole subsection was performed by Rojas-Andrés et al. (2015) (Fig. 4). This work provided evidence for the monophyly of *V.* subsection *Pentasepalae*, which comprises five main clades: a central Asian clade (*V. krylovii* Schischk.), a North African clade (*V. rosea* Desf.), two clades corresponding to species endemic to the Iberian Peninsula (the '*V. tenuifolia* complex' and *V. aragonensis* Stroh) and a core clade, which comprises the remaining species mostly from northern and central Europe, Italy and the Balkans. Most of the diploid species traditionally recognized were recovered as monophyletic in this work, together with *V. aragonensis* —the only polyploid species recovered as monophyletic.

Reticulate evolutionary patterns within *V.* subsect. *Pentasepalae* may have led to the absence of phylogenetic resolution observed in the polyploid taxa (Rojas-Andrés et al., 2015). However, the origin of some of these polyploid species by different processes has been proposed. For example, evidence for autopolyploidization has been found in four of the five clades (the Asian clade is strictly diploid). Tetraploid individuals within *V. tenuifolia* Asso and *V. rosea* populations cluster together with the rest of diploid individuals of its species, suggesting an autopolyploid event occurring within the populations (Martínez-Ortega et al., 2004; Padilla-García et al., 2018). Multiple autopolyploidization events were proposed for the tetraploid *V. satureiifolia* Poit. & Turpin leading to the formation of octoploids that have been identified as *V. sennenii* (Pau) M.M.Mart.Ort. & E.Rico in the Iberian Peninsula and *V. teucrium* var. *angustifolia* Vahl in France (Padilla-García et al., 2018). Allopolyploidization and hybridization could also be important evolutionary processes occurring in the group. For instance, homoploid hybridization was confirmed for *V.* × *gundisalvi* (Martinez-Ortega et al., 2004).

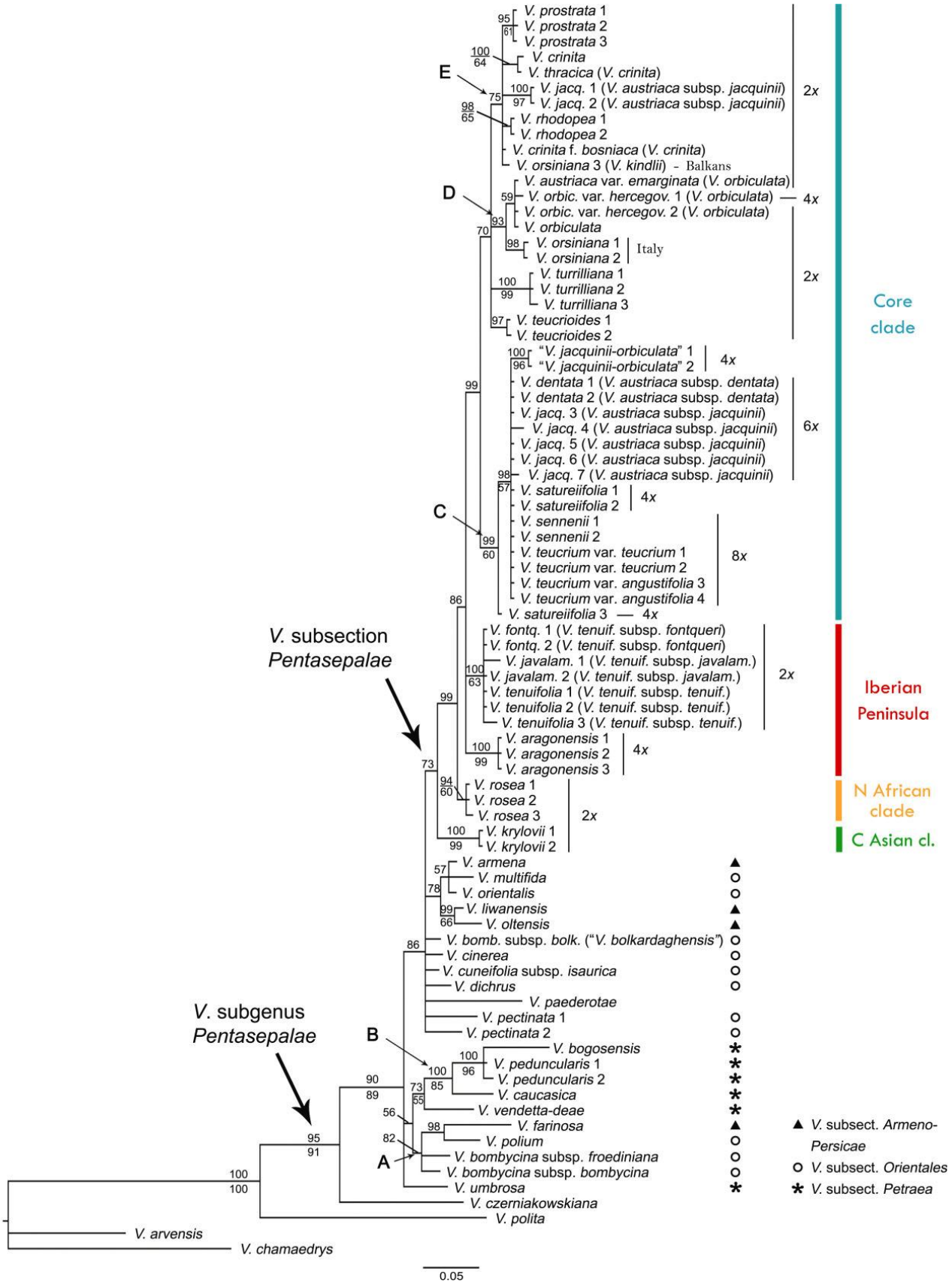



**Fig. 4.** Majority rule consensus tree obtained from the Bayesian analyses of the ITS region for *V.* subgenus *Pentasepalae*. Numbers above the branches indicate Bayesian posterior probabilities. Bootstrap support values > 50% from the parsimony analyses are indicated below the branches. Obtained from Rojas-Andrés et al. (2015).

Hybridization and/or introgression have been suggested for a tetraploid population of *V. aff. kindlii* located in Greece (Padilla-García et al., 2018). Different evolutionary processes can also occur within the same taxa, tetraploid individuals belonging to the diploid-tetraploid species *V. orbiculata* A. Kern have been proposed to be the result of autopoliploidization and allopolyploidization events (Rojas-Andrés et al., 2015; Padilla-García et al., 2018).

Preliminary studies on the western Mediterranean members of the subsection together with those on the whole subsection provided informative insights on the evolutionary mechanisms and relationships among diploid and polyploid members of *V.* subsect *Pentasepalae*. Nevertheless, some questions concerning particular taxa still remained unresolved.

### *2.4 Going deeper into the 'traditional V. austriaca complex': Definition of the study groups 'V. dalmatica – V.orbiculata complex' and 'V. austriaca complex' (= V. austriaca s.l.).*

One of the most taxonomically intricate cases within *V.* subsect. *Pentasepalae* is the traditionally recognized '*V. austriaca* complex'. The number of biological entities included and its respective taxonomic status have varied a lot along history (Rojas-Andrés and Martínez-Ortega, 2016). Some authors have accepted them at the specific rank (e.g., Fischer, 1991; Martínez-Ortega, 1999), while others have treated these taxa as subspecies or varieties (e.g., Watzl, 1910; Walters & Webb, 1972; Fischer, 2011). Morphologically intermediate populations between the taxa usually included within the complex are relatively frequent in nature and botanists have interpreted them either as transitional forms caused by phenotypic plasticity of one large, variable species (Watzl, 1910) or as the result of hybridization between species (Lehmann, 1937; Scheerer,

1949). However contemporary studies have shown that at least a couple of diploid representatives from these complex (i.e., "diploid *V. austriaca* subsp. *jacquinii*" and the diploid cytotype of *V. orbiculata*) are recovered as monophyletic unrelated taxonomic entities and that the relationship of several other taxa —sometimes considered members of the complex— with *V. austriaca* L. is not supported by genetic data (Rojas-Andrés et al., 2015, Padilla-García et al., 2018). On its side, the polyploid populations of *V. austriaca* were recovered together with a polyploid group that includes the following species and subspecies: *V. austriaca* subsp. *austriaca* (6*x*), *V. austriaca* subsp. *dentata* (F. W. Schmidt) Watzl (6*x*), *V. austriaca* subsp. *jacquinii* (Baumg.) Watzl (6*x*), *V. satureiifolia* (4*x*), *V. sennenii* (8*x*), and *V. teucrium* L. (8*x*) both according to the phylogenetic tree based on nuclear sequences (Rojas-Andrés et al., 2015) (Fig. 4) and the results obtained through AFLP markers (Padilla et al., 2018) (Fig. 5). *Veronica austriaca* subsp. *austriaca*, *V. austriaca* subsp. *dentata* and *V. austriaca* subsp. *jacquinii* were recognized together in the large (mostly hexaploid) '*V. austriaca* complex' (i.e., *V. austriaca* s.l., see below).

These studies confirmed the existence of two separate phylogenetic lineages differentiated by their ploidy levels within *V. austriaca* (diploids *vs*. hexaploids). The polyphyletic origin of *V. austriaca* had already been stated by Rojas-Andrés et al. (2015) but the decision to separate these lineages was postponed until further detailed analyses were conducted. In Padilla-García et al. (2018), an exhaustive revision of herbarium specimens allowed finding morphological characters to identify the diploid phylogenetic lineage. Thus, the "diploid individuals of *V. austriaca* subsp. *jacquinii*" were recognized at the specific rank as *V. dalmatica* N.Pad.Gar., Rojas-Andrés, López-González and M.M.Mart.Ort.

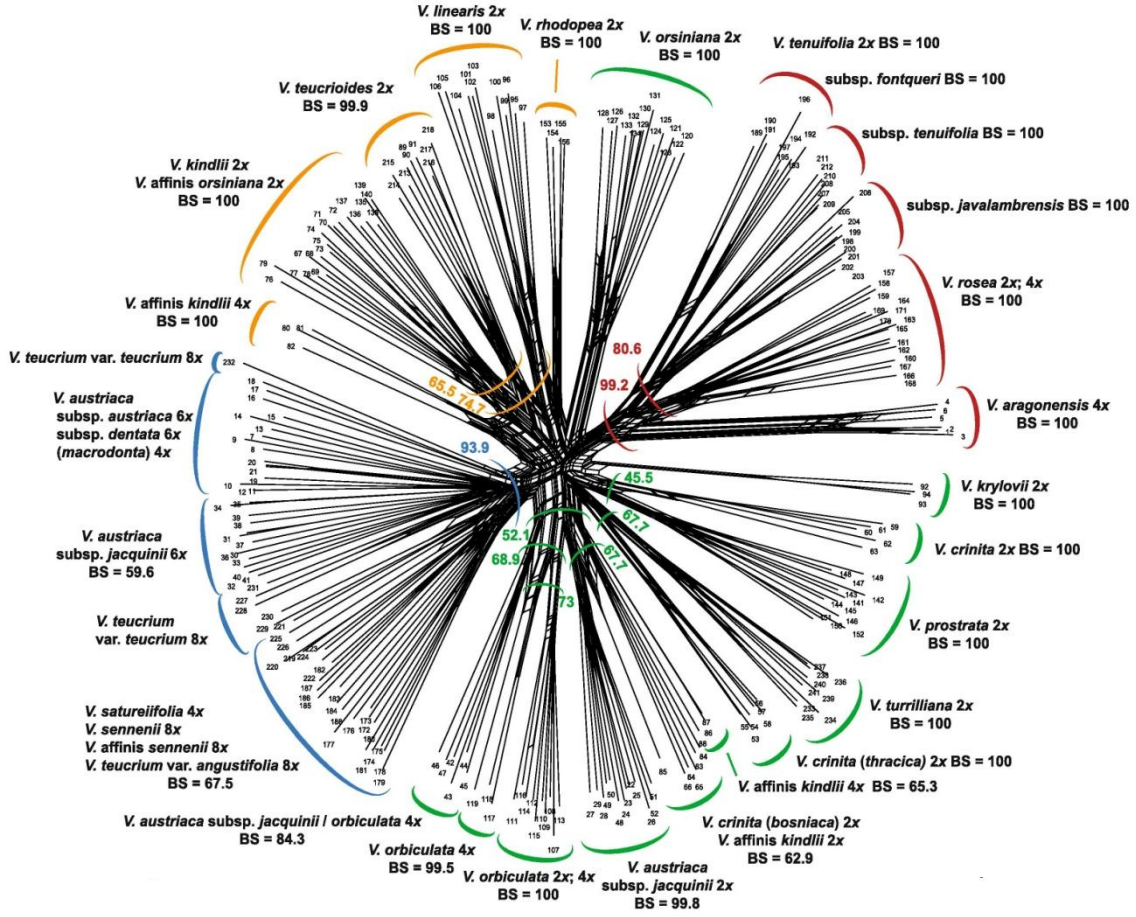



**Fig. 5.** Neighbor-Net network based on 1127 AFLP scored fragments of 241 individuals of *V.* subsect. *Pentasepalae* using Jaccard's genetic distances. Arcs delimit taxa whose initial taxonomic determination and ploidy are shown. Range of colors of the arcs differentiates the four clusters identified by Bayesian clustering analyses. Bootstrap values (BS) > 50% are shown. Obtained from Padilla-García et al. (2018).

According to these results it has been suggested that the populations of variable ploidy level (from di- to hexaploids) that are morphologically intermediate between *V. dalmatica*, *V. austriaca s.l.* (i.e., including the previously mentioned three subspecies), and *V. orbiculata* might be the result of interspecific hybridization and introgression, rather than transitional forms distributed along a gradient of ecological conditions (Rojas-Andrés et al., 2015; Padilla-García et al., 2018). In consequence, in this PhD

Thesis will be considered that *V. austriaca s.l.*, *V. dalmatica*, *V. orbiculata*, and the populations of uncertain taxonomic identity and variable ploidy level constitute a complex that will be called '*V. austriaca – V. orbiculata* complex'. The phylogenetic relationships within this complex, which represents a mixture of taxonomic entities and cytotypes are not yet fully understood.

Moreover, from now onwards in the present study we restrict the name '*V. austriaca* complex' or *V. austriaca* s. l. for the infraspecific taxa within the variable species *V. austriaca* [i.e., *V. austriaca* subsp. *austriaca*, *V. austriaca* subsp. *dentata* (F. W. Schmidt) Watzl, and *V. austriaca* subsp. *jacquinii* (Baumg.) Watzl] according to the last available taxonomic treatment (Rojas-Andrés and Martínez-Ortega., 2016) with the modification proposed by Padilla-García et al. (2018) (i.e., excluding the diploid *V. dalmatica*).

*Veronica austriaca* s. l. is variable regarding ploidy level, although previous studies have demonstrated that it is predominantly hexaploid (see supplementary data associated with Albach et al., 2008; available at http://www.researchgate.net/ publication/258769259_cariologia2013). This complex is also highly variable regarding morphology; according to the general description it is composed by suffrutex perennial herbs that show a gradient of variation regarding leaf incision (from deeply dentate to pinnatifid to pinnatisect leaves, in some cases even pinnatipartite or bipinnatipartite) (Fig. 6). *Veronica austriaca* s. l. displays wide ecological amplitude and represents one of the most widely distributed "species" of the subsection. It extends to a wide variety of altitudinal (100-2000 m) and climatic conditions. It is more commonly represented in grassland areas (from mesic meadows to Mediterranean dry steppe-like environments), although it can occur in grassy forest glades or forest edges.


**Fig. 6.** From left to right: *V. austriaca* subsp. *austriaca*, *V. austriaca* subsp. *dentata*, *V. austriaca* subsp. *jacquinii*.

## 3. GENERAL AIMS

The specific goals of this PhD thesis are:

1. To establish an automated protocol to find diagnostic characters for taxonomically complicated plant groups which could be readily implemented in identification keys. To this aim, *V.* subsect. *Pentasepalae* will be used as a

model and an exhaustive morphometric analysis of the taxa belonging to this group will be performed.

2. To formulate a morphologically optimal classification scheme by assigning cases to groups (i.e. populations to taxa) starting from the most recent available taxonomic hypothesis for *V.* subsect *Pentasepalae* and based on previous knowledge. This initial scheme will be used as a reference.

3. To analyse the morphometric dataset of *V.* subsect *Pentasepalae* using two data mining tools and an unsupervised systematic multivariate approach (assuming no previous knowledge), in order to search for diagnostic leaf features that allow unequivocal determination of species and infraspecific taxa of *V.* subsect. *Pentasepalae*. The three different approaches will be evaluated and compared with the optimal classification scheme.

4. To determine the chromosome numbers of several species within *Veronica* subsect. *Pentasepalae* combining information provided by direct chromosome counts and ploidy level estimations. Some of them will represent the first chromosome counts for the species.

5. To develop a set of polymorphic microsatellite markers focusing on the '*V. austriaca* – *V. orbiculata* complex'. Test these markers on species from different clades of *V.* subsect. *Pentasepalae* (*V. orsiniana* Ten. [core clade], *V. tenuifolia* subsp. *javalambrensis* Pau [Iberian clade], and *V. rosea* Desf. [North African clade]), and check their transferability to other subgenera.

6. To investigate the genetic relationships among different taxa and cytotypes within the '*V. austriaca* – *V. orbiculata* complex' to try to understand their evolutionary histories. Assess the genetic affinities of those populations that show intermediate morphological character states (i.e., between those shown by

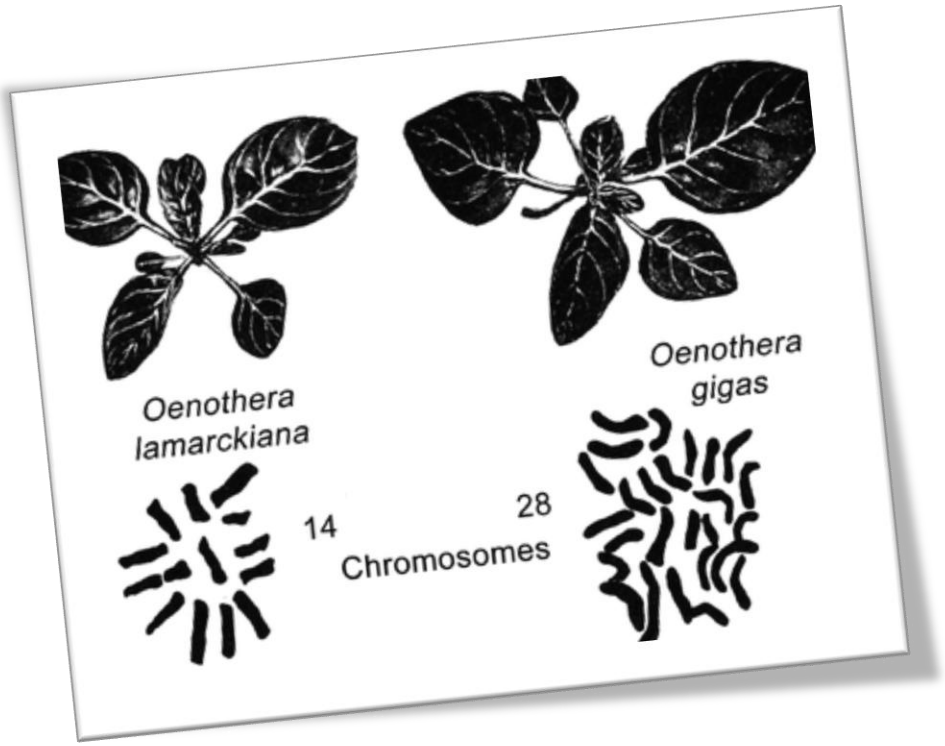
traditionally well-established taxa) and/or of uncertain taxonomic identity and variable ploidy level.

7. To establish hypotheses on the possible origins and ways of formation of the tetra- and hexaploid cytotypes within the '*V austriaca – V. orbiculata* complex', and reconstruct possible contact zones among taxa and cytotypes.

8. To assess the genetic diversity and population structure, and to describe the morphological and ecological variability of the intraspecific genetic units that are found within the '*V. austriaca* complex'.

9. To investigate the historical processes that may have been responsible for the contemporary geographic distribution of the populations of *V. austriaca* s.l.

10. To comment on the impact that allopolyploidy may have had in the colonization abilities and expansion of a species nowadays widely distributed in vast grassland areas of Europe.




# CAPÍTULO 1

## Introducción general

## 1. POLIPLOIDÍA E HIBRIDACIÓN

Los poliploides se definen como organismos que poseen más de dos juegos completos de cromosomas. El proceso que media en dicha adquisición se llama poliploidización o duplicación total del genoma (*whole-genome duplication*; WGD).

Teniendo en cuenta el origen de los genomas parentales, habitualmente se reconocen dos tipos de poliploides: autopoliploides y alopoliploides (Fig. 1). La autopoliploidía se define como la duplicación del genoma que proviene de una única especie parental, lo que daría como resultado conjuntos de cromosomas homólogos (el diploide AA duplica sus cromosomas para formar el tetraploide AAAA). La alopoliploidía en cambio, se define como la duplicación del genoma asociada a la unión de dos genomas (más o menos divergentes) tras una hibridación interespecífica. Este segundo proceso daría como resultado conjuntos de cromosomas homeólogos (AA × BB ⇨ AABB). Por lo tanto la poliploidización implica hibridación intra- o interespecífica.

Sin embargo, la hibridación no implica necesariamente poliploidización. La hibridación puede ocurrir sin que se produzca necesariamente la duplicación del genoma. En este caso se obtendrían como resultado organismos híbridos con el mismo número de cromosomas que sus correspondientes especies parentales (hibridos homoploides).

La poliploidía fue documentada por primera vez en plantas en los inicios del s. XX. Basándose en los trabajos previos de Hugo de Vries (1905), Lunt (1907) y Gates (1909) descubrieron de manera independiente que *Oenothera lamarckiana* mut. Gigas era un organismo tetraploide. Al mismo tiempo Kuwanda (1911) propuso que, en la historia evolutiva del maíz, se podía haber producido una duplicación cromosómica ancestral.

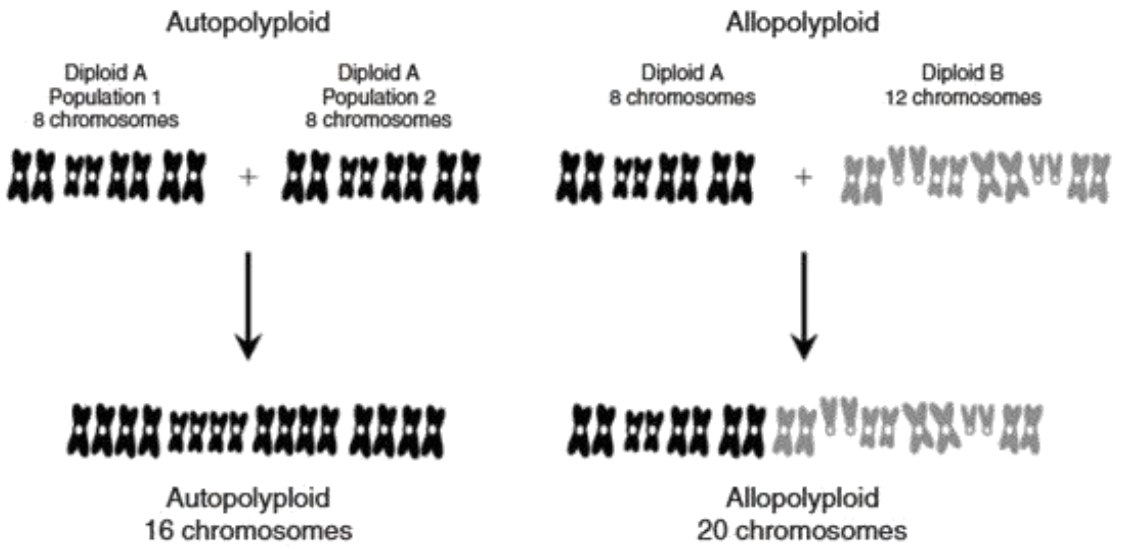



**Fig. 1.** Formación de auto- y alopoliploides. Los autopoliploides se forman por hibridación intraespecífica, mientras que los alopoliploides se forman por hibridación interespecífica (Tate et al., 2005).

Algunos años después, Winge (1917) introduce por primera vez el término "poliploide" y propone la explicación más influyente para el origen de estas plantas con sets dobles de cromosomas. Según Winge, estos organismos conseguían recuperar su fertilidad a través de la duplicación del genoma al restaurar el correcto apareamiento de cromosomas de híbridos previamente estériles. Esta hipótesis se vio apoyada por el trabajo posterior de Müntzing (1932) que generó "poliploides experimentales" (poliploides sintéticos), copiando especies poliploides naturales. Müntzing observó que los híbridos de *Galeopsis pubescens* Besser and *G. speciosa* Mill. eran mayoritariamente estériles. Sin embargo, los cruces de uno de estos híbridos (un individuo triploide) con su parental *G. speciosa* tenían como resultado una progenie tetraploide fértil. De esta manera, concluyó que los poliploides naturales debían haber seguido un proceso similar para *esquivar* los mecanismos postcigóticos de aislamiento reproductivo. Este estudio (entre otros, p.ej. Clausen & Goodspeed, 1925) supuso la

demostración de uno de los primeros mecanismos de formación de una nueva especie como consecuencia de cambios drásticos en el genoma de otra especie pre-existente.

Durante la década de los 30, la poliploidía era ya un tema de estudio relativamente bien conocido. La poliploidía fue identificada como la mayor diferencia en cuanto a la especiación de plantas y animales en la síntesis moderna de la evolución (Dobzhansky, 1937). De hecho, fue la corrección más importante a la teoría de Darwin y Wallace del origen de las especies (Haldane, 1959).

A pesar de este reconocimiento inicial, en la segunda mitad del s. XX, la poliploidía empezó a ser considerada como irrelevante por muchos investigadores. La posibilidad de que la poliploidía ancestral podría haber jugado un importante papel en la historia evolutiva de algunos organismos estaba aceptado (Stebbins, 1985) y el origen poliploide de muchas especies de plantas era inequívoco (Klekowski and Baker, 1966; Ohno, 1970). Lo sorprendente es que, pese a ello, la contribución de la poliploidía a la evolución comenzó a ponerse en duda (Stebbins, 1950, 1971). Refiriéndose a los poliploides y a los híbridos interespecíficos Wagner (1970) escribió: "En el marco general de la evolución, ¿no pueden estos fenómenos ser simplemente aberraciones triviales y defectos de la biología de las plantas, producidos por meras transformaciones genéticas o fertilizaciones ocasionales, que las plantas pueden tolerar, pero que tienen poca o ninguna importancia desde un punto de vista global?". Para algunos autores, la prevalencia y persistencia de diploides indicaba que todo el impacto evolutivo significativo recaía sobre los diploides, mientras que los poliploides no tenían ningún potencial evolutivo en sí mismos (Stebbins, 1971; Grant, 1981). Tanto híbridos como poliploides fueron descritos como "callejones sin salida", "vías muertas evolutivas" o simplemente "ruido evolutivo" (Wagner 1970; Stebbins 1971). Durante este tiempo, investigadores como A. Mützing, C. D. Darlington y A. Löve, mantuvieron que la

poliploidía y la hibridación eran mecanismos comunes y evolutivamente importantes para las plantas (Darlington, 1963; Löve, 1964; Müntzing, 1965), pero evidentemente fueron menos influyentes que Stebbins (Ramsey & Ramsey, 2014).

La idea de la poca relevancia evolutiva de la poliploidización empezó a ser refutada durante la era genómica (2000 – presente) (Adams & Wendel, 2005). Varios estudios en los que se secuenciaron organismos eucariotas encontraron rastros de eventos de poliploidización ancestral, aportando pruebas del potencial evolutivo de la poliploidía (p.ej. Gu et al., 2002; Wong et al., 2002; Becak & Kobashi, 2004; Paterson, 2005; Van de Peer & Meyer, 2005; Yu et al., 2005) e indicando que la poliploidía podría conferir flexibilidad evolutiva a largo plazo (Comai, 2005). Actualmente, existe consenso en cuanto a que la poliploidía ha ocurrido al menos una vez en la evolución de todas las plantas con semilla (Jiao et al., 2011; Wendel, 2015). De hecho, el éxito evolutivo de las angiospermas se ha asociado a duplicaciones de cromosomas (Fawcett et al., 2009; Van de Peer et al., 2009).

La pliploidía no es sólo importante a escala ancestral, es capaz de alterar la morfología, fenología, fisiología y/o ecología de las plantas en unas pocas generaciones (Levin, 2002). La naturaleza dinámica de los genomas poliploides —con capacidad para modificar el contenido genético, el número de genes, la disposición de los genes en cromosomas, la expresión génica y la actividad de los transposones— representa una fuente de innovación para la biodiversidad (Soltis et al., 2014). Aunque algunas combinaciones no tendrán capacidad para prosperar, ciertas combinaciones únicas de variación nuevamente generada pueden dar como resultado organismos con un potencial evolutivo y capacidades adaptativas enormes (Soltis, 2013). Por ello, la poliploidía representa un motor evolutivo de increíble fuerza en la evolución del reino vegetal, en

escalas temporales tanto ancestrales como actuales y con profundas consecuencias que van desde el nivel molecular al ecológico (Adams & Wendel, 2005).

Actualmente la investigación en poliplodía está más vigente que nunca (Barker et al., 2016). Se han hecho considerables avances en la predicción de resultados tras un evento de poliploidización (p.ej. Hollister & Gaut, 2009; Xiong & Pires, 2011; Wendel, 2015; Steige & Slotte, 2016). Sin embargo, los patrones evolutivos tras dichos eventos son inconsistentes y, por tanto, establecer un paradigma único para la poliploidía sigue suponiendo un desafío (Soltis et al. 2016).

Aunque aún existen muchas cuestiones abiertas sobre las consecuencias de la duplicación cromosómica, estudios moleculares basados en complejos poliploides han establecidos dos hallazgos de particular relevancia.

El primero es que tanto la poliploidía como la hibridación son comunes y están ampliamente representados en la naturaleza (Soltis et al., 2007; Spoelhof et al., 2017) y ambos procesos representan los extremos de un continuo. Es importante remarcar que el origen evolutivo de todos los poliploides naturales (auto- y alopoliploides) implica hibridación entre genomas más o menos relacionados (Parisod & Senerchia, 2012; Abbott et al., 2013). Por lo tanto, poliploidización e hibridación son las designaciones empleadas para dos procesos evolutivos complejos que se encuentran profundamente interrelacionados (Fig. 2) (Marques et al., 2018).

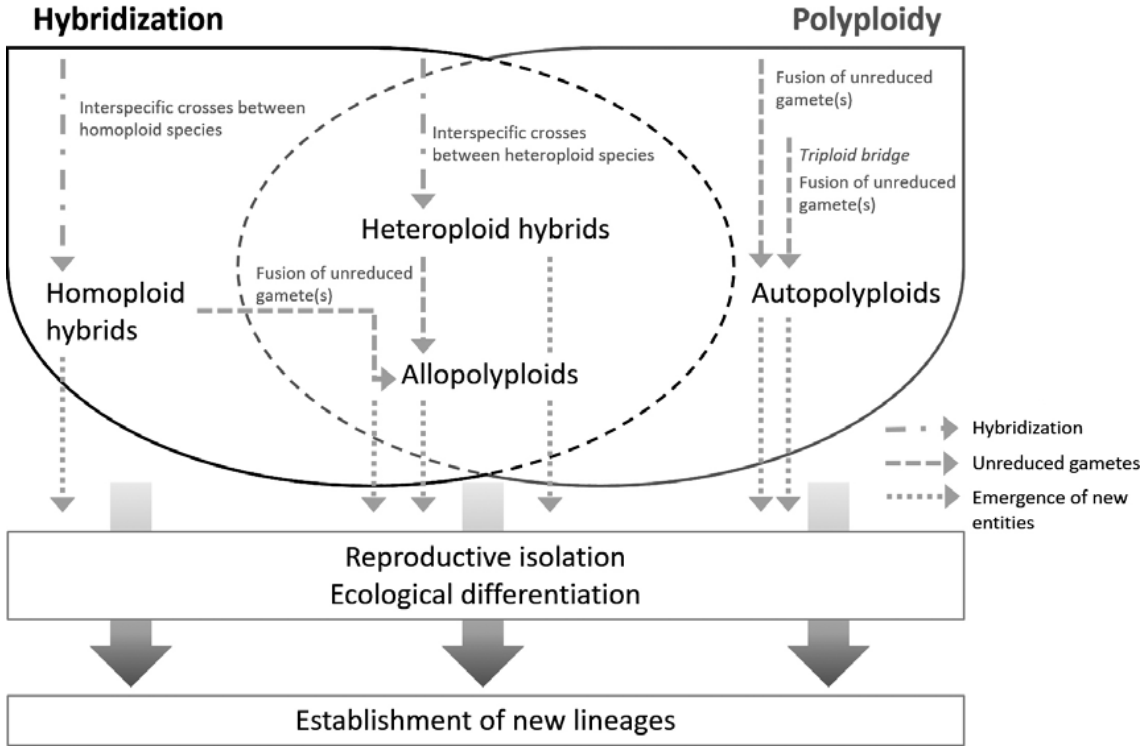



**Fig. 2.** Representación esquemática de la intricada conexión entre la hibridación y la poliploidización en la naturaleza (Marques et al., 2018).

El segundo es la confirmación, a través de numerosos estudios moleculares, de los múltiples orígenes de poliploides dentro de géneros de plantas o incluso en especies concretas (Parisod & Besnard, 2007; Soltis et al., 2016); de hecho, la formación recurrente de poliploides en la naturaleza parece ser la regla en vez de la excepción (Barker et al., 2016).

La poliploidía y la hibridación serán los *leitmotive* de los capítulos que componen esta Tesis Doctoral, cuyo objetivo principal es contribuir al entendimiento de la historia evolutiva y las relaciones entre las entidades que forman parte del complejo diploide-poliploide *Veronica* subsección *Pentasepalae* Benth.

## 2. DEFINIENDO Y AFINANDO EL GRUPO DE ESTUDIO: DESDE *VERONICA* SUBSECT. *PENTASEPALAE* HASTA EL 'COMPLEJO *VERONICA AUSTRIACA-ORBICULATA*'

### *2.1. Marco taxonómico*

El género *Veronica* L. ha sido incluido de manera tradicional dentro de la familia Scrophulariaceae. En la primera clasificación APG (1998), fue transferido a la familia Plantaginaceae, convirtiéndose en el género con mayor número de especies en dicha familia. *Veronica* incluye más de 450 especies distribuidas en su mayoría en el hemisferio norte y Australasia, con unas pocas especies presentes en África y Sudamérica. De acuerdo a los estudios más recientes, este enorme género de plantas se halla dividido en in 12 subgéneros, todos ellos monofiléticos (Garnock-Jones et al., 2007). Entre aquellos subgéneros con representación en el hemisferio norte, *V.* subgen. *Pentasepalae* (Benth.) M. M. Mart. Ort., Albach & M. A. Fisch. está constituido por 70-75 especies de hierbas perennes distribuidas en Eurasia y el norte de África. Las especies de este subgénero se caracterizan por un cáliz que consta de cinco sépalos (raramente cuatro) con el quinto sépalo significativamente más pequeño (Albach et al., 2004), y semillas con ornamentación reticulada-verrucosa (Muñoz-Centeno et al., 2006). El número base cromosómico, un caracter con importancia filogenética en el género, es $x = 8$. Se reconocen cuatro subsecciones dentro del subgénero (Albach et al., 2008): *V.* subsect. *Armeno-Persicae* Stroh, *V.* subsect. *Orientales* (Wulff) Stroh, *V.* subsect. *Petraea* Benth., and *V.* subsect. *Pentasepalae* Benth. Esta última, *V.* subsect. *Pentasepalae,* cuyo origen monofilético ha sido ya demostrado (Rojas-Andrés et al., 2015), representa el grupo de estudio de la presente Tesis Doctoral.

*Veronica* subsect. *Pentasepalae* fue inicialmente descrita por Bentham (1846) como el "grupo Pentasepalae" incluido dentro de *V.* sect. *Chamaedrys* W. D. J. Koch. En un principio este grupo estaba compuesto por unas pocas especies, pero en los años

subsiguientes muchos autores reevaluaron el grupo e hicieron nuevas propuestas taxonómicas con variaciones considerables (ver Rojas-Andrés & Martínez-Ortega [2016] para una revisión completa de los distintos tratamientos taxonómicos). En cada revisión han sido incluidos distintos taxones y muchos nombres han sido utilizados en sentidos diferentes de una revisión a otra (p.ej. Watzl, 1910; Walters & Webb, 1972; Fischer, 1982; Peev, 1995; Martínez-Ortega et al., 2009; Tison & Foucault, 2014), haciendo la taxonomía del grupo sumamente complicada. Esta frenética actividad taxonómica ha tenido como consecuencia la acumulación de más de 200 nombres para ca. 20 taxa.

Rojas-Andrés y Martínez-Ortega, tras un exhaustivo estudio del grupo, han propuesto un tratamiento nomenclatural y taxonómico para *V.* subsect. *Pentasepalae* (Rojas-Andrés & Martínez-Ortega, 2016; Rojas-Andrés et al., 2016). De acuerdo a esta minuciosa revisión, *V.* subsect. *Pentasepalae* es un complejo diploide-poliploide compuesto por 22 taxones perennes (17 especies, cuatro subespecies y una variedad). Las especies y taxones infraespecíficos se hallan distribuidos principalmente en Europa, con la excepción de *V. krylovii* Schischk., presente en Siberia y Kazajistán; y *V. rosea* Desf., que se extiende por el norte de Argelia y Marruecos. Este tratamiento taxonómico ha sido ligeramente modificado en base a marcadores de huella genética (AFLPs) y estimaciones de nivel ploidía (Padilla-García et al., 2018).

La exploración de *Veronica* subsect. *Pentasepalae* mediante sucesivos estudios ha demostrado que se trata de un complejo de divergencia reciente, en el que las barreras de aislamiento reproductivo no se han establecido de forma definitiva (Martínez-Ortega et al., 2004; Rojas-Andrés et al., 2015; Padilla-García et al., 2018). De manera adicional, tanto la poliploidía como la hibridación han sido identificados como procesos implicados en la evolución de la subsección, lo que puede representar un fuente de

homología (Martínez-Ortega et al., 2004). Se ha constatado que la poliploidía y la hibridación son mecanismos evolutivos que pueden causar alteraciones morfológicas que difuminan los límites entre especies e impiden el reconocimiento o la identificación precisa de taxones íntimamente relacionados (Stace, 2000). En casos extremos, la delimitación morfológica de las especies se convierte, simplemente, en una tarea imposible (p.ej. Nielsen & Olrik, 2001; Li et al., 2017). Como consecuencia de esta "tormenta perfecta" con barreras reproductivas débiles, incidencia de poliploidía e hibridación, los límites entre especies dentro de *V.* subsect *Pentasepalae* son, en ocasiones, realmente confusos. Sin embargo, en el caso de la subsección que nos ocupa, contamos con la ventaja de tener a nuestra disposición una hipótesis taxonómica sólida y a escala detallada (individuos y poblaciones) que se basa en datos genéticos, información sobre el nivel de ploidía, morfología general y rango geográfico (Rojas-Andrés & Martínez-Ortega, 2016). Para proponer este tratamiento taxonómico se han utilizado múltiples líneas de evidencia siguiendo el enfoque de la taxonomía integrativa (Dayrat, 2005) y el concepto general de especie como linaje evolutivo de De Queiroz (2005, 2007). Por ello, las entidades taxonómicas establecidas pueden ser utilizadas como punto de referencia para una selección precisa de rasgos morfológicos discriminantes.

### *2.2 Investigando los patrones de variación de los caracteres morfológicos en V. subsect. Pentasepalae: minería de datos versus análisis discriminante multivariante y la estrategia "divide y vencerás"*

*Veronica* subsect. *Pentasepalae* comprende hierbas perennes, subfrútices, normalmente leñosas en la base, con tallos de erectos a decumbentes que terminan en un renuevo apical vegetativo. Las hojas son generalmente opuestas, de sésiles a cortamente pecioladas, de enteras a bipinnatisectas, y de glabras a densamente pilosas cubiertas de

pelos glandulares. Las hojas del renuevo apical son normalmente opuestas. La inflorescencia es racemosa y generalmente axilar. Las flores son hermafroditas, con el cáliz más pequeño que la cápsula, persistente, glabro, subglabro o piloso. La corola consta de cuatro pétalos, ligeramente zigomórfica, más larga que el cáliz, violeta, violeta claro o azul oscuro, raramente blanquecina o rosa, normalmente con venación marcada. El fruto se dispone en una cápsula bilocular, dehiscente y comprimida lateralmente en la mayoría de los casos.

Los rasgos morfológicos más discriminantes están relacionados con la morfología foliar, pero existen otros caracteres útiles en la identificación como son: el tamaño de la planta, la forma de crecimiento, el número y pilosidad de los sépalos y la morfología del fruto (Martínez-Ortega et al., 2004). De manera específica los tratamientos taxonómicos mencionados en el epígrafe previo consideran las hojas como el principal rasgo distintivo para la identificación de especies, quizá porque los caracteres relacionados con las flores apenas presentan variación en esta subsección y las flores son, además, bastante efímeras.

En esta situación, para mejorar el conocimiento morfológico de *Veronica* subsect. *Pentasepalae,* parece necesario aplicar un enfoque innovador que sea también integrativo en términos de metodología. Por un lado, técnicas modernas como las herramientas de minería de datos han sido aplicadas con gran éxito en campos tan diversos como el marketing, la química o las ciencias sociales (p.ej. Delen et al., 2005; Fischer et al., 2006; Paliwal & Kumar, 2009; Baker, 2010; Michel et al., 2011; Strelcov et al., 2014). La minería de datos es el paso clave del Descubrimiento de Conocimiento en Bases de Datos (*Knowledge Discovery in Databases*; KDD) que está orientado a identificar y describir patrones en grandes bases de datos (Maimon & Rokach, 2010). Por otro lado, análisis estadísticos multivariantes clásicos como los análisis

discriminantes (morfometría multivariante) representan una herramienta muy robusta para evaluar patrones de variación y siguen siendo aplicados con excelentes resultados (p.ej. Henderson, 2006; Marhold, 2011; Lorenz et al., 2014).

La combinación de las técnicas expuestas se lleva a cabo bajo uno de los métodos de resolución de problemas más recurrente: la estrategia (o el algoritmo) "divide y vencerás". Esta estrategia consiste sencillamente en dividir un problema particular en problemas más pequeños que sean más fáciles de resolver, y luego utilizar estas soluciones parciales para construir una solución compleja al problema inicial (Fig. 3).

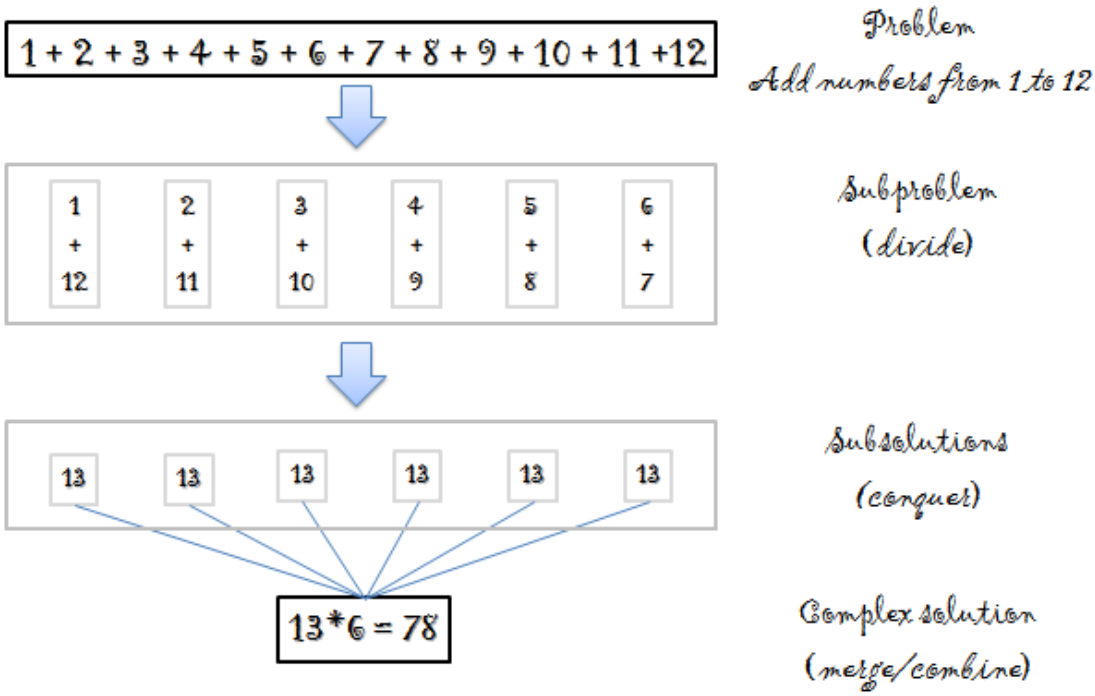



**Fig. 3.** Ejemplo simplificado de la estrategia "Divide y vencerás". Modificado de Magnus y Berner (2000).

### *2.3. Estudios filogenéticos y evolutivos previos en V. subsect. Pentasepalae*

Varios estudios han abordado previamente los mecanismos evolutivos y han investigado las relaciones filogenéticas subyacentes entre los taxones que forman parte de *V.* subsect. *Pentasepalae*. El primer trabajo que utiliza la combinación de datos moleculares y morfométricos aparece en la Tesis Doctoral de Martínez-Ortega (1999) presentada en la Universidad de Salamanca, y se continúa con los trabajos subsiguientes, parcialmente financiados con el proyecto de investigación '*Flora iberica*' (Martínez-Ortega, 2004; Andrés-Sánchez et al., 2009; Martínez-Ortega et al., 2009); estos trabajos incluyen mayoritariamente especies y taxones infraespecíficos del Mediterráneo occidental.

El primer análisis filogenético basado en secuencias de AND que incluye representantes de la subsección completa fue realizado por Rojas-Andrés et al. (2015) (Fig. 4). En este trabajo se aportan evidencias para la monofilia de *V.* subsection *Pentasepalae,* que comprende cinco clados: un clado en Asia central (*V. krylovii*), un clado norteafricano (*V. rosea*), dos clados que corresponden a especies endémicas de la península ibérica (el complejo '*V. tenuifolia*' y *V. aragonensis* Stroh) y un clado principal o central que agrupa al resto de especies distribuidas por el norte y centro de Europa, Italia y Balcanes. La mayoría de las especies diploides reconocidas tradicionalmente fueron recuperadas como monofiléticas en este trabajo, junto con *V. aragonensis* —la única especie poliploide identificada como monofilética.

Se han propuesto los patrones de evolución reticulada encontrados como los causantes de la falta de resolución filogenética observada para los taxones poliploides de *V.* subsect. *Pentasepalae* (Rojas-Andrés et al., 2015). Pese a la falta de resolución, sí se ha podido proponer el origen de algunas de estas especies poliploides como resultado de distintos mecanismos evolutivos. Por ejemplo, existen evidencias de

autopoliploidización en cuatro de los cinco clados (el clado asiático es estrictamente diploide). Individuos tetraploides encontrados en poblaciones de *V. tenuifolia* Asso y *V. rosea* se agrupan junto con el resto de individuos diploides de sus respectivas especies, lo que sugiere que estos poliploides podrían haber surgido a raíz de eventos de autopoliploidización en las propias poblaciones (Martínez-Ortega et al., 2004; Padilla-García et al., 2018). Se han propuesto eventos múltiples de autopoliploidización en el caso de la especie tetraploide *V. satureiifolia* Poit. & Turpin que han resultado en la formación de de octaploides identificados como *V. sennenii* (Pau) M.M.Mart.Ort. & E.Rico en la península ibérica y *V. teucrium* var. *angustifolia* Vahl en Francia (Padilla-García et al., 2018). También se han sugerido la alopoliploidización y la hibridación como procesos evolutivos importantes para la subsección. Se ha confirmado, por ejemplo, la hibridación homoploide en el caso de *V.* × *gundisalvi* (Martinez-Ortega et al., 2004). Eventos de hibridación y/o introgresión se han propuesto para una población tetraploide de *V. aff. kindlii* localizada en Grecia (Padilla-García et al., 2018). Incluso se ha constatado que distintos procesos evolutivos pueden ocurrir dentro de un mismo taxón: los individuos tetraploides de *V. orbiculata* A. Kern parecen ser el resultado de procesos de auto- y alopoliploidía (Rojas-Andrés et al., 2015; Padilla-García et al., 2018).

Los estudios iniciales centrados en las especies del Mediterráneo occidental junto con los trabajos que abordan la subsección completa, han aportado una perspectiva muy informativa acerca de los mecanismos evolutivos y las relaciones entre los miembros diploides y poliploides de *V.* subsect *Pentasepalae*. Sin embargo, algunas cuestiones respecto a taxones particulares de la subsección que nos ocupa, permanecen aún sin resolver.

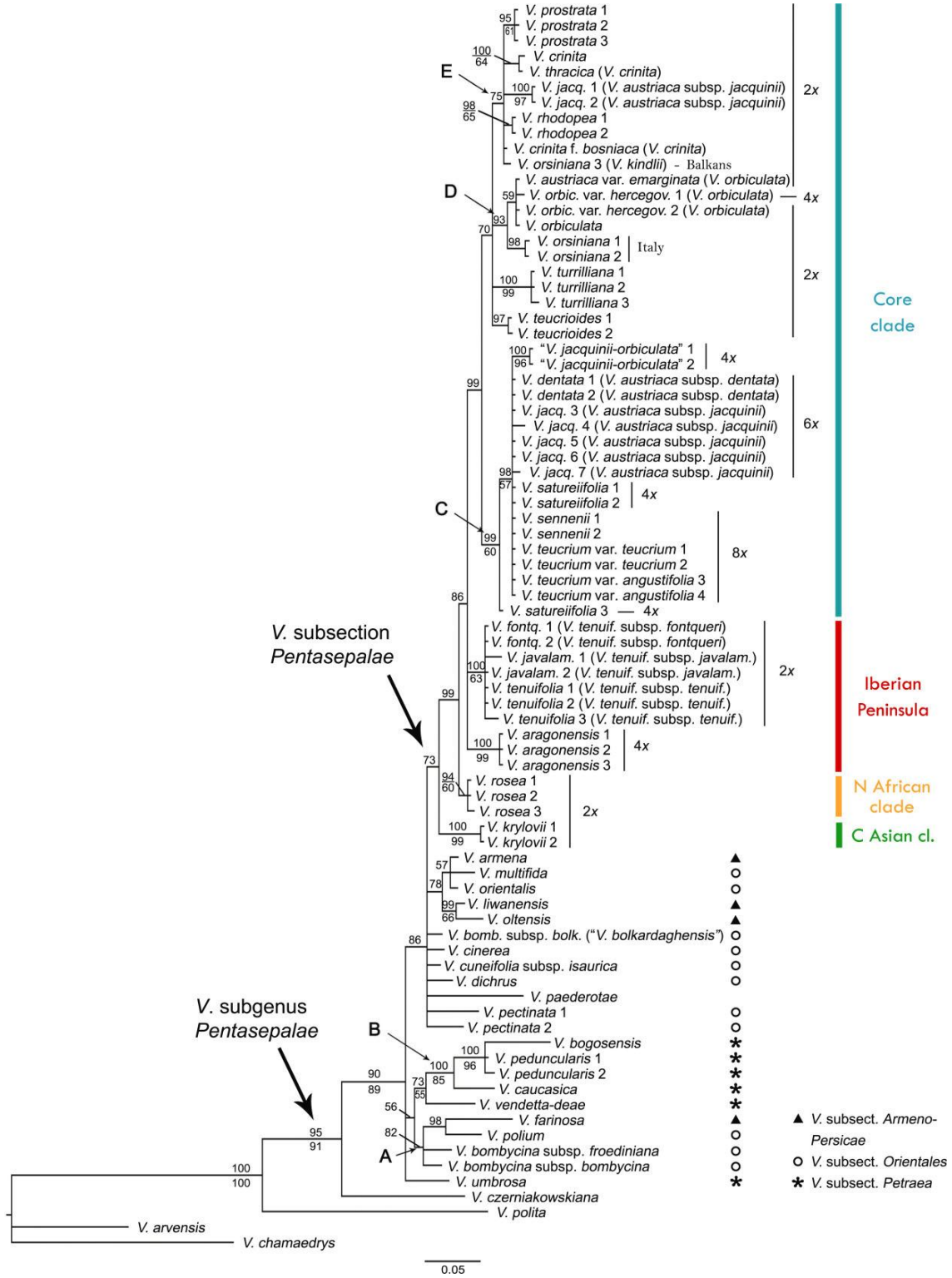



**Fig. 4.** Árbol de consenso obtenido de los análisis bayesianos de la región ITS de *V.* subgen. *Pentasepalae*. Los números que aparecen sobre las ramas indican probabilidades posteriores de acuerdo al análisis bayesiano. Los valores de soporte de *bootstrap* (> 50%) de los análisis de parsimonia se indican bajo las ramas. Obtenido de Rojas-Andrés et al. (2015).

### *2.4 Profundizando en el tradicionalmente reconocido 'complejo V. austriaca': definición de los grupos de estudio 'complejo V. dalmatica – V.orbiculata' y 'complejo V. austriaca' (= V. austriaca s.l.).*

Uno de los casos taxonómicamente más complicados dentro de *V.* subsect. *Pentasepalae* es el 'complejo *V. austriaca*', que ha sido tradicionalmente reconocido por numerosos autores. El número de entidades que lo componen y sus respectivas categorías taxonómicas han variado considerablemente a lo largo de la historia (Rojas-Andrés & Martínez-Ortega, 2016). En algunos tratamiento taxonómicos, se han aceptado dichas entidades a nivel específico (p.ej. Fischer, 1991; Martínez-Ortega, 1999), mientras que en otros han sido consideradas como subespecies o variedades (p.ej, Watzl, 1910; Walters & Webb, 1972; Fischer, 2011). No es raro encontrar en la naturaleza poblaciones de morfología intermedia entre los taxones habitualmente incluidos dentro del complejo. Los botánicos han interpretado estas poblaciones bien como formas de transición que derivan de la plasticidad fenotípica de una especie amplia y muy variable (Watzl, 1910), bien como el resultado de procesos de hibridación interespecífica (Lehmann, 1937; Scheerer, 1949). Sin embargo, los estudios más recientes han demostrado que, al menos un par de representantes diploides de este complejo (el citotipo diploide de *V. austriaca* subsp. *jacquinii* y *V. orbiculata*), se recuperan como entidades taxonómicas monofiléticas independientes, y que la relación de otras entidades —a veces consideradas como miembros del complejo— con *V. austriaca* L. no tiene respaldo bajo el punto de vista genético (Rojas-Andrés et al., 2015, Padilla-García et al., 2018). Por otro lado, las poblaciones poliploides de *V. austriaca* se recuperaron junto con un grupo poliploide que incluye las siguientes especies y taxones infraespecíficos: *V. austriaca* subsp. *austriaca* (6*x*), *V. austriaca* subsp. *dentata* (F. W. Schmidt) Watzl (6*x*), *V. austriaca* subsp. *jacquinii* (Baumg.) Watzl (6*x*), *V. satureiifolia* (4*x*), *V. sennenii* (8*x*), and *V. teucrium* L. (8*x*), de acuerdo

tanto al árbol filogenético basado en secuencias de nucleares de ADN (Rojas-Andrés et al., 2015) (Fig. 4) como a los estudios basados en marcadores de huella genética tipo AFLPs (Padilla-García et al., 2018) (Fig. 5). *Veronica austriaca* subsp. *austriaca*, *V. austriaca* subsp. *dentata* y *V. austriaca* subsp. *jacquinii* fueron agrupados dentro del amplio y mayoritariamente hexaploide 'complejo *V. austriaca*' (*V. austriaca* s.l., ver más abajo).

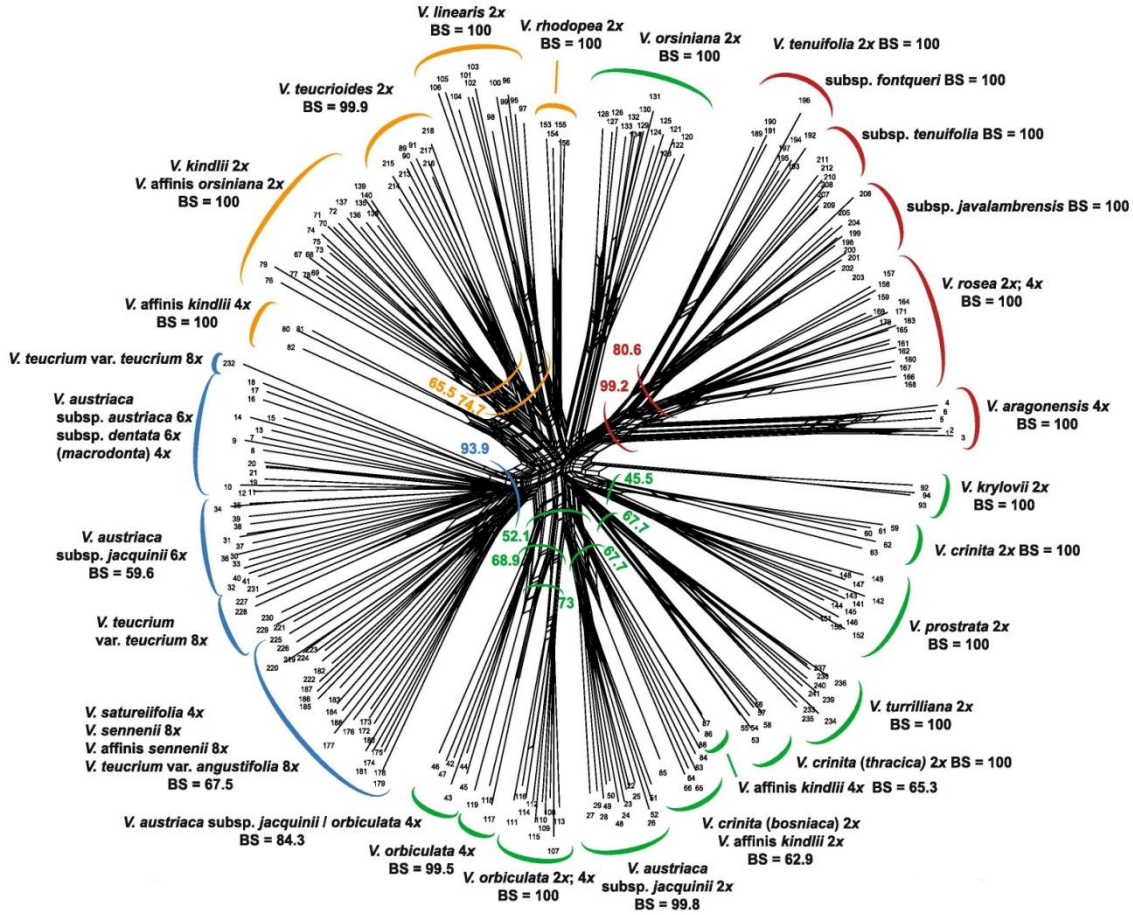



**Fig. 5.** *Neighbor-Net* basada en la evaluación de 1127 fragmentos AFLP de 241 individuos de *V.* subsect. *Pentasepalae* utilizando la distancia genética de Jaccard. Los arcos delimitan taxones mostrando nivel de ploidía e identificación inicial. Los colores de los arcos se corresponden con los cuatro grupos encontrados como resultado del análisis bayesiano. Se indican los valores de *bootstrap* (BS) > 50%. Obtenido de Padilla-García et al. (2018).

Estos trabajos confirmaron la existencia, dentro de *V. austriaca,* de dos linajes filogenéticos separados y diferenciados por su nivel de ploidía (diploides *vs.* hexaploides). El origen polifilético de *V. austriaca* ya había sido propuesto por Rojas-Andrés et al. (2015), pero la decisión de separar estos linajes fue pospuesta hasta que se realizaran análisis más detallados. En el trabajo de Padilla-García et al. (2018) se llevó a cabo una concienzuda revisión de especímenes de herbario, lo que permitió encontrar caracteres morfológicos para identificar el linaje diploide. Debido a ello, los "individuos diploides de *V. austriaca* subsp. *jacquinii*" fueron elevados a categoría de especie bajo el nombre de *V. dalmatica* N.Pad.Gar., Rojas-Andrés, López-González and M.M.Mart.Ort. De acuerdo a estos resultados se ha sugerido que las poblaciones de nivel de ploidía variable (de di- a hexaploide) que son morfológicamente intermedias entre *V. dalmatica*, *V. austriaca* s.l. (que incluye las tres subespecies previamente mencionadas) y *V. orbiculata* podriían ser el resultado de eventos de hibridación interespecífica e introgresión, en vez de formas de transición que responden a un gradiente de condiciones ecológicas (Rojas-Andrés et al., 2015; Padilla-García et al., 2018). Como consecuencia, para el presente trabajo se considera que *V. austriaca* s.l., *V. dalmatica, V. orbiculata,* y las poblaciones de ploidía variable e identidad taxonómica incierta, constituyen el 'complejo *V. austriaca – V. orbiculata'.* Las relaciones filogenéticas entre los miembros de este complejo, que representa una amalgama de entidades taxonómicas y citotipos, despiertan todavía numerosos interrogantes. Además, de aquí en adelante para el presente trabajo, se utiliza el nombre de 'complejo *V. austriaca*' o *V. austriaca* s.l. para referirnos a los taxones infraespecíficos incluidos dentro de la especie de gran variabilidad *V. austriaca* [*V. austriaca* subsp. *austriaca*, *V. austriaca* subsp. *dentata* (F. W. Schmidt) Watzl, y *V. austriaca* subsp. *jacquinii*

(Baumg.) Watzl], de acuerdo al último tratamiento taxonómico disponible (Rojas-Andrés & Martínez-Ortega., 2016) considerando las modificaciones propuestas por Padilla-García et al. (2018) (exclusin del diploide *V. dalmatica*).


**Fig. 6.** Aspecto general de las subspecies incluidas en *V. austriaca* s.l. De izquierda a derecha: *V. austriaca* subsp. *austriaca*, *V. austriaca* subsp. *dentata* y *V. austriaca* subsp. *jacquinii*.

*Veronica austriaca* s.l. es variable atendiendo al nivel de ploidíaa, aunque estudios previos han demostraado que es predominantemente hexaploide (ver material supplementario en Albach et al., 2008; disponible en: http://www.researchgate.net/publication/258769259_cariologia2013). Este complejo es también enormemente variable en cuanto a morfología; de acuerdo a la descripción general se compone de hierbas perennes subfrútices que muestran un amplio gradiente de variación respecto a la incisión de la lámina foliar (desde profundamente dentadas a hojas pinnatífidas o pinnatisectas, en algunos casos incluso pinnatipartidas o bipinnatipartidas) (Fig. 6). *Veronica austriaca* s.l. muestra una vasta amplitud ecológica y representa una de las "especies" con mayor rango de distribución de la subsección. Se extiende a lo largo de una gran variedad de condiciones climáticas y un amplio rango altitudinal (100-2000 m), se encuentra en pastizales o herbazales (desde prados mésicos a zonas de carácter estepario), aunque puede aparecer también en bordes o claros de bosque.

## 3. OBJETIVOS GENERALES

Los objetivos específicos de esta Tesis Doctoral son:

1. Establecer un protocolo automatizado para la búsqueda de caracteres diagnósticos en grupos de plantas taxonómicamente complicados que pueda ser implementado en claves de identificación. Para ello se utilizará *V.* subsect. *Pentasepalae* como modelo, y se realizará un exhaustivo análisis morfmétrico de los taxones pertenecientes a esta subsección.
2. Formular un esquema de clasificación óptimo mediante la asignación de observaciones a grupos (es decir, poblaaciones a taxones), basado en el tratamiento

taxonómico más reciente disponible para *V.* subsect *Pentasepalae* y en el conocimiento previo que se posee del grupo. Este esquema inicial será usado como punto de referencia.

3. Analizar el set de datos morfométrico de *V.* subsect *Pentasepalae* empeando para ello dos téecnicas de minería de datos y un análisis multivariante no supervisado (asumienndo ausencia de conocimiento previo en el grupo), con el objetivo de buscar rasgos foliares identificativos que permitan la inequívoca identificación de especies y taxones infraespecíficos de *V.* subsect. *Pentasepalae*. Evaluar estas tres aproximaciones y comparar con el esquema de clasificación óptimo desarrollado inicialmente.

4. Determinar el número de cromosomas de varias especies pertenecientes a *Veronica* subsect. *Pentasepalae* combinando la información proporcionada por los conteos directos de cromosomas con las estimaciones de nivel de ploidía realizadas mediate citometría de flujo.

5. Desarrollar un set de marcadores polimórficos de tipo microsatélite tomando como referencia el 'complejo *V. austriaca – V. orbiculata*'. Evaluar estos marcadores en especies pertenecientes a diferentes clados de *V.* subsect. *Pentasepalae* (*V. orsiniana* Ten. [clado principal], *V. tenuifolia* subsp. *javalambrensis* Pau [clado ibérico], and *V. rosea* Desf. [clado norteafricano]), y comprobar su transferibilidad a otros subgéneros.

6. Inveestigar las relaciones gennéticasa de los diferentes taxones y citotipos incluidos en el 'complejo *V. austriaca – V. orbiculata*' con la intención de entender sus respectivas historias evolutivas. Evaluar las afinidades genéticas de las poblaciones que muestran estados de carácter morfológicos intermedios

(intermedios entre aquellos taxones tradicionalmente bien establecidos) y/o de identidad taxonómica incierta y nivel de ploidía variable.

7. Establecer hipótesis sobre los posibles orígenes y vías de formación de los citotipos tetra- y hexaploide del 'complejo *V. austriaca – V. orbiculata*' y reconstruir las posibles zonas de contacto entre taxones y citotipos.

8. Evaluar la diversidad genética y estructura poblacional y describir la variabilidad ecológica y morfológica de las unidades genéticas infraespecíficas que aparecen en el 'complejo *V. austriaca*'.

9. Investigar los procesos históricos que pueden haber sido los responsables de la distribución geográfica contemporánea de las poblaciones de *V. austriaca* s.l.

10. Valorar el impacto que ha podido tener la alopoliploidía en las capacidades de colonización y expansión de una especie hoy en día ampliamente distribuida en vastas áreas de pastizales de Europa.

## BIBLIOGRAFÍA (REFERENCES)

**Abbott R, Albach DC, Ansell S, Arntzen JW, Baird SJ, Bierne N., Boughman J, Brelsford A, Buerkle CA, ..., Butlin RK**. 2013. Hybridization and speciation. *Journal of Evolutionary Biology*, 26: 229–246.

**Adams KL, Wendel JF**. 2005. Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology*, 8: 135–141.

**Albach DC, Martínez-Ortega MM, Fischer MA, Chase MW**. 2004. A new classification of the tribe *Veroniceae* – problems and a possible solution. *Taxon*, 53: 429–452.

**Albach DC, Martínez-Ortega MM, Delgado L, Weiss-Schneeweiss H, Özgökce F. Fischer MA**. 2008. Chromosome numbers in Veroniceae (Plantaginaceae): Review and several new counts. *Annals of the Missouri Botanic Garden*, 95: 543–566.

**Andrés-Sánchez S, Rico E, Herrero, A, Santos-Vicente M, Martínez-Ortega MM.** 2009. Combining traditional morphometrics and molecular markers in cryptic taxa: towards an updated integrative

taxonomic treatment for *Veronica* subgenus *Pentasepalae* (Plantaginaceae sensu APG II) in the western Mediterranean. *Botanical Journal of the Linnean Society*, 159: 68–87.

**Angiosperm Phylogeny Group.** 1998. An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden*, 531–553.

**Baker RSJd**. 2010. Data mining for education, in: Peterson P, Baker E, McGaw B (eds.), International encyclopedia of education. Oxford: Elsevier, pp.112–118.

**Barker MS, Husband BC, Pires JC**. 2016. Spreading Winge and flying high: the evolutionary importance of polyploidy after a century of study. *American Journal of Botany*, 103: 1139–1145.

**Becak ML, Kobashi LS**. 2004. Evolution by polyploidy and gene regulation in Anura. *Genetic and Molecular Research*, 3: 195–2121.

**Bentham G.** 1846. Scrophulariaceae, in: de Candolle A (ed.), *Prodromus Systematis Naturalis Regni Vegetabilis* vol.10. Victoris Masson, Paris, pp. 180–586.

**Clausen RE, Goodspeed TH.** 1925. Interspecific hybridization in *Nicotiana*. II. A tetraploid *glutinosa* – *tabacum* hybrid, an experimental verification of Winge's hypothesis. *Genetics*, 10: 278–284.

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

**Darlington CD**. 1963. Chromosome botany and the origins of cultivated plants**,** second ed. Hafner, New York.

**Dayrat B.** 2005. Towards integrative taxonomy. *Biological Journal of the Linnean Society*, 85:407–415.

**De Queiroz K.** 2005. Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences of the United States of America*, 102: 6600–6607.

**De Queiroz K.** 2007. Species concepts and species delimitation. *Systematic Biology*, 56: 879–886.

**De Vries H.** 1905. Species and Varieties: Their Origin by Mutation. Open Court Publishing Company, Chicago.

**Delen D, Walker G, Kadam A**. 2005. Predicting breast cancer survivability: a comparison of three data mining methods. *Artificial Intelligence in Medicine*, 34:113–127.

**Dobzhansky TG.** 1937. Genetics and the origin of species. Columbia University Press, New York.

**Fawcett A, Maere S, Van de Peer Y**. 2009. Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. *Proceedings of the National Academy of Sciences of the United States of America*, 106: 5737–5742

**Fischer CC, Tibbetts KJ, Morgan D, Ceder G**. 2006. Predicting crystal structure by merging data mining with quantum mechanics. *Nature Materials*, 5: 641–646.

**Fischer MA.** 1982. *Veronica* L., in: Pignatti S (ed.), Flora d'Italia, vol. 2. Edagricole, Bologna, pp. 558–573.

**Fischer MA.** 1991. *Veronica* L., in: Strid A, Tan K (eds.), The mountain flora of Greece. Edinburgh University Press, Edinburgh, pp. 209–234.

**Fischer MA**. 2011. *Veronica* L., in: Jäger EJ (ed.), Rothmaler – Exkursionsflora von Deutschland: Gefäßpflanzen; Grundband. Spektrum Akademischer Verlag, Heidelberg, pp. 647–657.

**Garnock-Jones P, Albach DC, Briggs BG**. 2007. Botanical names in southern hemisphere *Veronica* (Plantaginaceae): sect. *Detzneria*, sect. *Hebe*, and sect. *Labiatoides*. *Taxon*, 56: 571–582.

**Grant V.** 1981. Plant speciation, second ed. Columbia University Press, New York.

**Gates RR.** 1909. The stature and chromosomes of *Oenothera* gigas De Vries. *Archiv für Zellforschung*, 3: 525–552.

**Gu Z, Cavalcanti A, Chen FC, Bouman P, Li WH.** 2002. Extent of gene duplication in the genomes of *Drosophila*, nematode, and yeast. *Molecular Biology and Evolution*, 19: 256–262.

**Haldane JBS.** 1959. Natural selection, in: Bell PR (ed.), Darwin's biological work: Some aspects reconsidered. Cambridge University Press, Cambridge, pp. 101–149.

**Henderson A.** 2006. Traditional morphometrics in plant systematics and its role in palm systematics. *Botanical Journal of the Linnean Society*, 151:103–111.

**Hollister JD, Gaut BS.** 2009. Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Research*, 19: 1419–1428.

**Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, ..., Soltis DE**. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature*, 473: 97–100.

**Klekowski EJ Jr, Baker HG.** 1966. Evolutionary significance of polyploidy in the Pteridophyta. *Science*, 153*:* 305–307.

**Kuwada Y**. 1911. Meiosis in the pollen mother cells of *Zea mays* L. (with Plate V.). *Shokubutsugaku Zasshi*, 25: 163–181.

**Lehmann E**. 1937. Die Gattung *Veronica* in Entwicklungsgeschichtlicher Betrachtung. *Cytologia* vol. Fujii Jubilaee: 903–919.

**Levin DA**. 2002. The role of chromosomal change in plant evolution. Oxford University Press, Oxford.

**Li FW, Rushworth CA, Beck JB, Windham MD.** 2017. *Boechera* microsatellite website: An online portal for species identification and determination of hybrid parentage. *Database:The Journal of Biological Databases and Curation*, 1: baw169.

**Lorenz C, Marques TC, Sallum MAM, Suesdek L**. 2014. Altitudinal population structure and microevolution of the malaria vector *Anopheles cruzii*. *Parasite Vectors*, 7: 581.

**Löve Á**. 1964. The biological species concept and its evolutionary structure. *Taxon*, 13: 33–45.

**Lutz AM**. 1907. A preliminary note on the chromosomes of *Oenothera lamarckiana* and one of its mutants, *O. gigas*. *Science*, 26: 151–152.

**Magnus HE, Berner RS**. 2000. El diablo de los números. Siruela Ediciones, Madrid.

**Maimon O, Rokach L**. 2010. Data mining and knowledge discovery handbook, second ed. Springer, New York.

**Marhold K**. 2011. Multivariate morphometrics and its application to monography at specific and infraspecific levels, in: Stuessy TF, Lack HW (eds.), Monographic plant systematics: fundamental assessment of plant biodiversity. Gantner, Ruggell, pp. 73–99.

**Marques I, Loureiro J, Draper D, Castro M, Castro S**. 2018. How much do we know about the frequency of hybridisation and polyploidy in the Mediterranean region? *Plant Biology*, 20: 21–37.

**Martínez-Ortega MM**. 1999. Revisión taxonómica de *Veronica* sect. *Veronica* L. y *V.* sect. *Veronicastrum* W.D.J. Koch (Scrophulariaceae) en el Mediterráneo Occidental. PhD Dissertation, University of Salamanca, Spain.

**Martínez-Ortega MM, Delgado L, Albach D, Rico E, Elena-Rossello JA**. 2004. Species boundaries and phylogeographic patterns in cryptic taxa inferred from AFLP markers: *Veronica* subgen. *Pentasepalae* (Scrophulariaceae) in the Western Mediterranean. *Systematic Botany*, 29: 96–986.

**Martínez Ortega MM, Sánchez Agudo JÁ, Rico E**. 2009. *Veronica* L., in. Benedó C, Rico E, Güemes J, Herrero A (eds.), Flora iberica v. XIII Plantaginaceae-Scrophulariaceae. Real Jardín Botánico-CSIC, Madrid, pp. 360–434.

**Michel JB, Shen YK, Aiden AP, Veres A, Gray MK, The Google Books Team, Pickett JP, Hoiberg D, …, Aiden EL.** 2011. Quantitative analysis of culture using millions of digitized books. *Science*, 331: 176–182.

**Muñoz-Centeno LM, Albach DC, Sanchez-Agudo JA, Martínez-Ortega MM.** 2006. Systematic significance of seed morphology in *Veronica* (Plantaginaceae): a phylogenetic perspective. *Annals of Botany*, 98: 335–350.

**Müntzing A**. 1932. Cytogenetic investigations on synthetic *Galeopsis Tetrahit*. *Hereditas*, 16: 105–154.

**Müntzing A**. 1965. Cytotaxonomy. II. Synthesis. *Proceedings of the XI International Congress of Genetics* 2, 417–421.

**Nielsen J, Olrik DC.** 2001 A morphometric analysis of *Prunus spinosa*, *P. domestica* ssp. *insititia*, and their putative hybrids in Denmark. *Nordic Journal of Botany*, 21: 349–363.

**Ohno S**. 1970. Evolution by gene duplication. Springer Science & Business Media, Berlin.

**Padilla-García N, Rojas-Andrés BM, López-González N, Castro M, Castro S, Loureiro J, Albach DC, Machon N, Martínez-Ortega MM**. 2018. The challenge of species delimitation in the diploid-polyploid complex *Veronica* subsection *Pentasepalae*. *Molecular Phylogenetics and Evolution*, 119: 19–209.

**Paliwal M, Kumar UA.** 2009. Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications*, 36: 2–17.

**Parisod C, Besnard G**. 2007. Glacial *in situ* survival in the Western Alps and polytopic autopolyploidy in *Biscutella laevigata* L. (Brassicaceae). *Molecular Ecology*, 16: 2755–2767.

**Parisod C, Senerchia N**. 2012. Responses of transposable elements to polyploidy, in: Grandbastien MA, Casacuberta JM (eds.), Plant transposable elements: Impact on Genome Structure and Function. Springer, Heidelberg, pp. 147–168.

**Paterson AH**. 2005. Polyploidy, evolutionary opportunity, and crop adaptation. *Genetica*, 123: 191–196.

**Peev D**. 1995. *Veronica* L., in: Kozuharov SI, Kuzmanov BA (eds.), Flora Reipublicae Bulgaricae. Editio Academica "Professor Marin Drinov", Sofia, pp. 142–189.

**Ramsey J, Ramsey TS**. 2014. Ecological studies of polyploidy in the 100 years following its discovery. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369: 20130352.

**Rojas-Andrés BM, Albach DA, Martínez-Ortega MM**. 2015. Exploring the intricate evolutionary history of the diploid-polyploid complex *Veronica* subsection *Pentasepalae* (Plantaginaceae). *Botanical Journal of the Linnean Society*, 179: 67–692.

**Rojas-Andrés BM, Martínez-Ortega MM**. 2016. Taxonomic revision of *Veronica* subsection *Pentasepalae* (*Veronica*, Plantaginaceae sensu APG III). *Phytotaxa*, 285: 1–100.

**Rojas-Andrés BM, Rico E, Martínez-Ortega MM**. 2016. A nomenclatural treatment for *Veronica* subsect. *Pentasepalae* (Plantaginaceae sensu APG III) and typification of several names. *Taxon*, 65: 617–627.

**Scheerer H**. 1949. Zur Polyploidie und Genetik der *Veronica* - gruppe *Pentasepala. Planta*, 37: 293–298.

**Soltis DE, Soltis PS, Schemske DW, Hancock JF, Thompson JN, Husband BC, Judd WS.** 2007. Autopolyploidy in angiosperms: have we grossly underestimated the number of species? *Taxon*, 56: 13–30.

**Soltis DE, Visger CJ, Marchant DB, Soltis PS**. 2016. Polyploidy: Pitfalls and paths to a paradigm. *American Journal of Botany*, 103: 1146–1166.

**Soltis PS**. 2013. Hybridization, speciation and novelty. *Journal of Evolutionary Biology*, 26: 291–293.

**Soltis PS, Liu X, Marchant DB, Visger CJ, Soltis DE**. 2014. Polyploidy and novelty: Gottlieb's legacy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369: 20130351.

**Spoelhof JP, Soltis PS, Soltis DE**. 2017. Pure polyploidy: closing the gaps in autopolyploid research. *Journal of Systematics and Evolution*, 55: 340–352.

**Stace CA**. 2000. Cytology and cytogenetics as a fundamental taxonomic resource for the 20th and 21st centuries. *Taxon*, 49: 451–477.

**Stebbins GL**. 1950. Variation and evolution in plants. Columbia University Press, New York.

**Stebbins GL**. 1971. Chromosomal evolution in higher plants. Edward Arnold, London.

**Stebbins GL.** 1985. Polyploidy, hybridization, and the invasion of new habitats. *Annals of the Missouri Botanical Garden*, 72*:* 824–832.

**Steige KA, Slotte T**. 2016. Genomic legacies of the progenitors and the evolutionary consequences of allopolyploidy. *Current Opinion in Plant Biology*, 30: 88–93.

**Strelcov E, Belianinov A, Hsieh YH, Jesse S, Baddorf AP, Chu YH, Kalinin SV.** 2014. Deep data analysis of conductive phenomena on complex oxide interfaces: physics from data mining. *ACS Nano*, 8: 6449–6457.

**Tate JA, Soltis DE, Soltis PS.** 2005. Polyploidy in plants, in: Gregory TR (ed.), The evolution of the genome. Academic Press, London, pp. 371–426.

**Tison JM, Foucault B**. 2014. Flora Gallica–Flore de France. Biotope Editions, Mèze.

**Van de Peer Y, Meyer A.** 2005. Large-scale gene and ancient genome duplications, in Gregory TR (ed.), The Evolution of the Genome. Academic Press, London, pp. 329–368.

**Van de Peer Y, Maere S, Meyer A**. 2009. The evolutionary significance of ancient genome duplications. *Nature Reviews Genetics*, 10: 725–732.

**Wagner WH**. 1970. Biosystematics and evolutionary noise. *Taxon*, 19*:* 146–151.

**Walters SM, Webb DA**. 1972. *Veronica* L., in: Tutin TG, Heywood VH, Burges NA (eds.), Flora Europaea. Cambridge University Press, Cambridge, pp. 242–251.

**Watzl B**. 1910. *Veronica prostrata* L., *teucrium* L. und *austriaca* L. Nebst einem anhange Èber deren na Èchsteverwante. *Abhandlungen der kaiserlich-königlichen zoologisch-botanischen Gesellschaft in Wien*, 5:1–94.

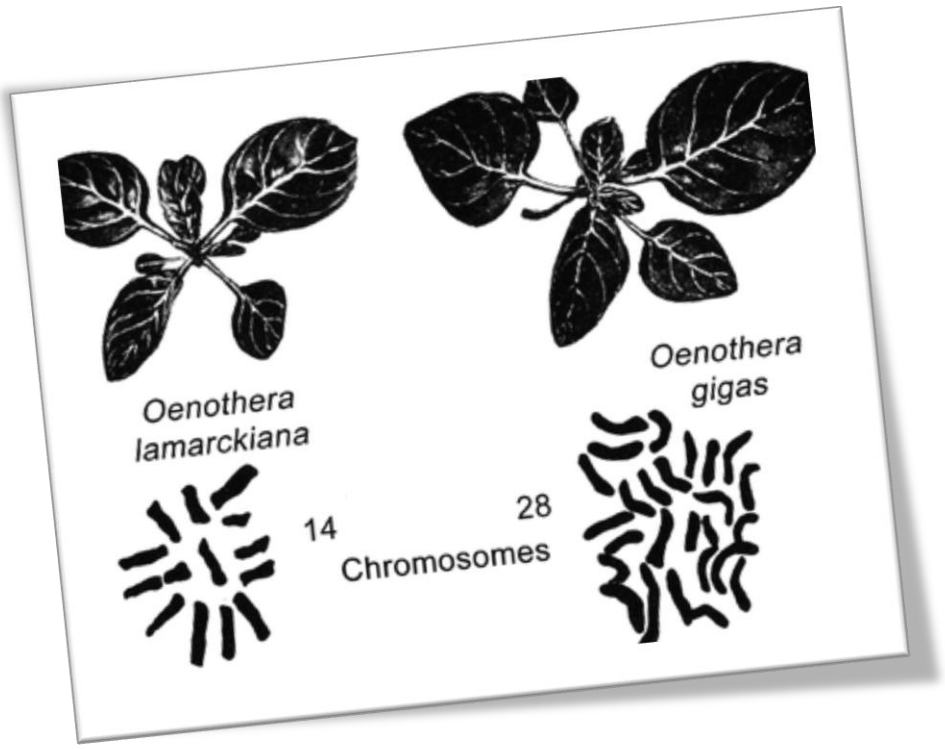
**Wendel JF**. 2015. The wondrous cycles of polyploidy in plants. *American journal of botany*, 102: 17531756.

**Winge O**. 1917. The chromosomes. Their numbers and general importance. *Comptes Rendus des Travaux du Laboratoire Carlsberg*, 13: 131–175.

**Wong S, Butler G, Wolfe KH**. 2002. Gene order evolution and paleopolyploidy in hemiascomycete yeasts. *Proceedings of the National Academy of Sciences of the United States of America*, 99: 9272–9277.

**Xiong Z, Pires JC**. 2011. Karyotype and identification of all homoeologous chromosomes of allopolyploid *Brassica napus* and its diploid progenitors. *Genetics*, 187: 37–49.

**Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Peixiang N, Dong W, Hu S, …, Yang H.** 2005. The Genomes of *Oryza sativa*: a history of duplications. *PLoS Biology*, 3: e38.




# CHAPTER 2

**Divide and conquer! Data-mining tools and sequential multivariate analysis to search for diagnostic morphological characters within a plant polyploid complex (*Veronica* subsect. *Pentasepalae,* Plantaginaceae)**

***¡Divide y vencerás! Herramientas de minería de datos y análisis multivariante secuencial para la búsqueda de caracteres morfológicos diagnósticos en un complejo poliploide de plantas (*Veronica *subsect.* Pentasepalae*, Plantaginaceae)***

PLOS ONE

# Divide and conquer! Data-mining tools and sequential multivariate analysis to search for diagnostic morphological characters within a plant polyploid complex (*Veronica* subsect. *Pentasepalae,* Plantaginaceae).

## *¡Divide y vencerás! Herramientas de minería de datos y análisis multivariante secuencial para la búsqueda de caracteres morfológicos diagnósticos en un complejo poliploide de plantas (*Veronica *subsect.* Pentasepalae*, Plantaginaceae)*

Noemí López-González[1]*, Santiago Andrés-Sánchez[1,2], Blanca M. Rojas-Andrés[1¤], M. Montserrat Martínez-Ortega[1]

[1]*Departamento de Botánica y Biobanco de ADN vegetal, Universidad de Salamanca, Salamanca, Spain,*

[2]*Departamento de Didáctica de las Matemáticas y Didáctica de las Ciencias Experimentales, Universidad de Salamanca, Salamanca, Spain*

[¤]*Current address: Department of Molecular Evolution and Plant Systematics & Herbarium (LZ), Institute of Biology, Leipzig University, Leipzig, Germany.*

PLOS ONE

OPENACCESS

**Citation:** López-González N, Andrés-Sánchez S, Rojas-Andrés BM, Martínez-Ortega MM (2018) Divide and conquer! Data-mining tools and sequential multivariate analysis to search for diagnostic morphological characters within a plant polyploid complex (*Veronica* subsect. *Pentasepalae*, Plantaginaceae). PLoS ONE 13(6): e0199818. https://doi.org/10.1371/journal. pone.0199818

**Editor:** Tzen-Yuh Chiang, National Cheng Kung University, TAIWAN

**Received:** January 25, 2018

**Accepted:** June 14, 2018

**Published:** June 29, 2018



**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This research was financially supported by the Spanish Ministry of Research and Innovation through the projects [CGL2012-32574], [CGL2009-07555] and [CGL2014-52787C3-2-P]; Spanish Ministry of Research, Development and Innovation through PhD scholarships to NLG [AP2010-2968] and BMRA [AP2008-03434]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

**ABSTRACT**

This study exhaustively explores leaf features seeking diagnostic characters to aid the classification (assigning cases to groups, i.e. populations to taxa) in a polyploid plant-species complex. A challenging case study was selected: *Veronica* subsection *Pentasepalae*, a taxonomically intricate group. The "divide and conquer" approach was implemented—that is, a difficult primary dataset was split into more manageable subsets. Three techniques were explored: two data-mining tools (artificial neural networks and decision trees) and one unsupervised discriminant analysis. However, only the decision trees and discriminant analysis were finally used to select diagnostic traits. A previously established classification hypothesis based on other data sources was used as a starting point. A guided discriminant analysis (i.e. involving manual character selection) was used to produce a grouping scheme fitting this hypothesis so that it could be taken as a reference. Sequential unsupervised multivariate analysis enabled the recognition of all species and infraspecific taxa; however, a suboptimal classification rate was achieved. Decision trees resulted in better classification rates than unsupervised multivariate analysis, but three complete taxa were misidentified (not present in terminal nodes). The variable selection led to a different grouping scheme in the case of decision trees. The resulting groups displayed low misclassification rates when analyzed using artificial neural networks. The decision trees as well as the discriminant analysis are recommended in the search of diagnostic characters. Due to the high sensitivity that artificial neural networks have to the combination of input/output layers, they are proposed as evaluation tools for morphometric studies. The "divide and conquer" principle is a promising strategy, providing success in the present case study.

## RESUMEN

Este studio hace una exploración exhaustiva de caracteres foliares en la búsqueda de rasgos diagnósticos que sirvan para la clasificación (asignación de casos a grupos, o poblaciones a taxones en este caso) en un complejo de plantas poliploides. Para tal fin se selecciona un caso de estudio desafiante: *Veronica* subsección *Pentasepalae*, un grupo de gran complejidad taxonómica. Se implementa la estrategia "divide y vencerás" que consiste en dividir el complicado conjunto de datos inicial en subconjuntos más manejables. Se exploran tres técnicas: dos herramientas de minería de datos (redes neuronales artificiales y árboles de decisión) y un análisis discriminante no supervisado. Sinn embargo, únicamente los árboles de decisión y el análisis discriminante han sido empleados para la selección de caracteres diagnósticos. Como punto de partida se utiliza una hipótesis de clasificación previa basada en otras fuentes de datos. Para construir un esquema de agrupación que sea congruente con la hipótesis de clasificación previa y que pueda ser utilizado como referencia, se utiliza un análisis discriminante guiado (con selección manual de caracteres). El análisis discriminante secuencial no supervisado permitió la identificación de todas las especies y taxones infraespecíficos; sin embargo este análisis tuvo una tasa de clasificación subóptima. Los árboles de decisión obtuvieron mejores resultados en cuanto a tasa de clasificación, pero tres taxones no fueron identificados (no aparecen en los nodos terminales). La selección de variables mediante los árboles d decisión dio como resultado un esquema de clasificación diferente. Los grupos resultantes mostraron bajas tasas de error en la clasificación al analizarlos mediante redes neuronales artificiales. Para la búsqueda de caracteres diagnósticos se rcomiendan tanto los árboles de decisión como los análisis discriminantes. Debido a la alta sensibilidad de las redes neuronales artificiales a la combinación de variables de entrada y de salida, se proponen como herramientas de evaluación para estudios morfométricos. El principio"divide y vencerás" supone una estrategia prometedora que ha aportado resultados exitosos en el presente caso de estudio.

## 1. 1. INTRODUCTION

Polyploidization is known to have occurred at least once during the evolutionary history of all angiosperms (Jiao et al., 2002; Soltis et al., 2015) and it is widely thought to play an important role in plant evolution and ecology (Fawcett et al., 2013; Ramsey and Ramsey, 2014; Zozoma-Lihova et al., 2015; Soltis et al., 2016). Also, interspecific hybridization may have occurred over plant evolution more frequently than previously suspected (Rieseberg, 1997; Seehausen, 2004) and in fact involves at least one-quarter of plant speciation events (Mallet, 2005). Hybridization (including allopolyploidization) and introgression are complex processes that may blur species boundaries in hybrid zones if isolating factors are not definitely established (Suehs et al., 2004; Tovar-Sánchez and Oyama, 2004) and may even end up merging species that were formerly separated (Raudnitschka, 2007). These processes affect species delimitation, giving rise to intermediate phenotypes between the parents (Levin, 1983; Abott and Lowe, 2004; Ramsey and Ramsey, 2014), leading to overlapping character states, and many gradual phenotypic transitions (e.g. in related subgenera of *Veronica*, [Bardy et al., 2011]; or in other genera, e.g. Koutecký [2007]; Horändl et al. [2009] among many others) or results in high intraspecific variation (Lavergne et al., 2010; Balao et al., 2011; Li et al., 2012).

*Veronica* subsect. *Pentasepalae* is a recently diversified complex in which genetic isolation barriers are not definitely established (Martínez-Ortega et al., 2004; Rojas-Andrés et al., 2015; Padilla-García et al., 2018). In addition, both polyploidy and hybridization have been identified as processes causing morphological alterations that make species boundaries indistinct and avoid clear-cut recognition of closely related taxa (Stace, 2000). Consequently, some key aspects remain controversial and/or poorly studied, mainly the determination of species boundaries and the accurate selection of morphological traits to identify them. The complex taxonomy of the study group is

reflected in the existence of c. 230 names for just 22 accepted taxa (Rojas-Andrés and Martínez-Ortega, 2016). Although most of the Eurasian species of this group have been reviewed throughout history in partial monographs or taxonomic treatments within several *Floras* (e.g. Watzl, 1910; Walters and Webb, 1972; Martínez-Ortega et al., 2009), Rojas-Andrés and MartínezOrtega (2016) have proposed the most recent taxonomic treatment for the whole subsection. This taxonomic proposal is based on the results of DNA sequence-based phylogenetic analyses that included all the taxa belonging to the subsection known at that time (Rojas-Andrés et al., 2015), which are considered together with information on ploidy level, phenotypic characters, habitat preferences, and species distributions. The subsection contains 17 species, four subspecies and one variety (Rojas-Andrés and Martínez-Ortega, 2016) and is represented in the temperate regions of Eurasia and in North Africa (only one species). This taxonomic treatment has recently been revised and slightly modified based on AFLP fingerprinting and DNA ploidy-level estimations (Padilla-García et al., 2018). This latest taxonomic proposal based on several data sources and on the general lineage concept is taken here as a starting hypothesis (see Materials and Methods). The members of the subsection are characterized by a pentapartite calyx (rarely tetrapartite) with the fifth sepal significantly smaller (Rojas-Andrés et al., 2015). Within this subsection some of the taxa are registered on the International Union for the Conservation of Nature Red List (http://www.iucnredlist.org/) and regional catalogues (Peñas de Giles et al., 2004), because they are threatened plants with narrow distribution areas and low numbers of known populations (Petrova and Vladimirov, 2009). It is necessary to define species boundaries and provide tools to recognize taxa (i.e. useful discriminant characters to be implemented in identification keys), but this is even more important when endangered species are involved.

For identification keys, leaf-lamina shape is one of the most relevant characters; it is remarkably informative for woody plants (Muir et al., 2000; Kafkas and Perl-Treves, 2001; Jensen et al. 2002), but it is also useful to identify species belonging to many other plant groups (e.g. Ackerfield and Wen, 2002; Plotze et al., 2005; Andrade et al., 2008). Specifically, the taxonomic treatments available for *V.* subsect. *Pentasepalae* thoroughly consider and use leaves as a primary source of characters for species identification (Watzl, 1910; Walters and Webb, 1972; Peev, 1995; Martínez-Ortega et al., 2009; Tison and Foucault, 2014; Rojas-Andrés and Martínez-Ortega, 2016), mainly because floral features show little variation in *Veronica* and they are quite ephemeral in comparison to leaf attributes. A previous work that examined leaf variation in eight taxa from the Iberian Peninsula and North Africa demonstrated that an overall separation of taxonomic units was possible based on a set of morphological characters despite some particular cases in which unequivocal identification through these features alone was not accomplished (Andrés-Sánchez et al., 2009).

At present, different methods are available to analyze morphometric data. The classical data analysis through multivariate discriminant analyses (hereafter DAs) are still being successfully applied (Henderson, 2006; Lorenz et al., 2014). Multivariate morphometrics represents a robust tool for evaluating variation patterns at the specific and infraspecific levels (Marhold, 2011), but new techniques are being implemented and show noteworthy outcomes. Data mining is the core step in the Knowledge Discovery in Databases (KDD), and data-mining tools find and describe structural patterns in data (Maimon and Rokach, 2010). Data-mining tools have been successfully applied to a broad range of fields such as marketing, chemistry or social studies (Delen et al., 2005; Fischer et al., 2006; Paliwal and Kumar, 2009; Baker, 2010; Michel et al., 2011; Strelcov et al., 2014). Although these methods have not been widely used in

morphometrics, some examples can be found (see below). Specifically, two well-known data-mining techniques have been previously applied in morphometric studies: “Decision Trees” and “Artificial Neural Networks” (hereafter DTs and ANNs, respectively). DTs are designed to identify patterns defining a given number of different groups, using direct information about the membership of the units (Quinlan, 1985), which results in classification trees providing decisions at each branch point or node. This technique makes direct use of the “divide and conquer” principle and generates groups automatically while the tree is constructed. DTs have been used in taxonomic and palaeoecological studies involving plant species (Lindbladh et al., 2002; Joly and Bruneau, 2007). ANNs, computational models inspired by biological systems, are formed by a number of elements (neurons) organized in layers. Each neuron in a layer is connected with each neuron in the next one by weights, and these weights are adjusted through a learning process (i.e. they are "trained" with respect to specific data until they "learn" the underlying hidden patterns). This technique has lately been used to identify organisms on the basis of morphological traits, mostly in animals (Dobigny et al., 20052; Lorenz et al., 2015) but not exclusively (Pandolfi et al., 2009; Clark et al., 2012). Also some studies have explored the usefulness of the three previously mentioned approaches in different areas of knowledge and with different objectives, such as species distribution (Manel et al., 1999), medical data analysis (Dreiseitl and Ohno-Machado, 2002), prediction accuracy (Razi and Athappilly, 2005) or disease prediction (Kurt et al., 2008; Huang and Hsu, 2012). There is a wide range of data-mining techniques (such as support vector machines, methods based on the K-nearest neighbor algorithm, rule induction, etc.) and statistical methods (e.g. Bayesian approaches, regression-based approaches), but these have been less used for

morphometric studies and therefore, are not considered here and thus lie beyond the scope of this work.

The purpose of the present work was to compare the performance of three classification techniques, using the morphologically highly heterogeneous diploid-polyploid complex *V.* subsect. *Pentasepalae* as a case study, applying a "divide and conquer" approach (i.e. a dataset that was difficult to handle was split into more manageable subsets). For this, a search was made for discriminant morphological characters to allow accurate taxon identification in taxonomically intricate species groups. The "divide and conquer" approach has been successfully used for example to align high numbers of DNA sequences (Liu et al., 2009; Liu et al., 2011) and phylogenetic analyses using parsimony (Goloboff et al., 2008). The selection of the study group is based on two main criteria that make the case both challenging and robust. First, despite the knowledge acquired after years working on this group, species identification remains problematical; and, second, enough molecular, cytological, biogeographic, and phylogenetic information is available, ensuring a solid starting taxonomic working hypothesis for the reference taxa. Morphometric data have been partially gathered from a previous work by Andrés-Sánchez et al. (Andrés-Sánchez et al., 2009), but this dataset has been substantially augmented (threefold) with information on virtually all the species included in *V.* subsect *Pentasepalae* and, whenever possible, from the entire distribution area of each taxon.

For the aim, this work involves the following:

1) Formulation of an optimal classification scheme by assigning cases to groups (i.e. populations to taxa) in accordance with the available taxonomic starting hypothesis. The separation of the entities is forced with the help of subsequent guided DAs. From the leaf features with importance in each DA, the final selection is based on previous

knowledge (i.e. manual character selection). This character selection and the initial scheme are used as a reference to be compared with the results found using other techniques (see point 3).

2) Analysis of the morphometric dataset through three techniques at the same level: two datamining tools [DTs and ANNs, currently available under GNU-GPL license (General Public License)] and an unsupervised systematic multivariate approach. For these methods no previous knowledge is assumed. The analyses are focused on the search for leaf features that are diagnostic for the species (many of them narrowly distributed and with a few known populations) that comprise a recently diversified and morphologically highly heterogeneous plant group affected by hybridization and polyploidization.

3) Assessment of the pros and cons of each approach plus an evaluation of the diagnostic features resulting from each technique. Use of ANNs to determine the suitability of the variables (input layers) over the groupings established (output layers) and comparison with the optimal classification scheme.

4) Verification of whether it is possible to establish an automated protocol to find out diagnostic characters to be readily used in taxon identification keys.

It should be remarked that the purpose of this study is not to achieve automated plant recognition. As stated above, within *V.* subsect. *Pentasepalae*, multiple lines of evidence have previously been used to propose a taxonomic starting hypothesis, following an integrative taxonomic approach (Dayrat, 2005) and the general lineage species concept of De Queiroz (2005, 2007); see Rojas-Andrés et al. (2015), as well as Padilla-García et al. (2018). Here, well-established taxonomic entities were used as a reference to carry out the main points mentioned here.

## 2. MATERIALS AND METHODS

A total of 605 specimens (individuals) from 209 populations were studied, either on loan from 19 herbaria—B, BC, BCF, BM, DR, E, FCO, G, GDA, JACA, K, MA, MAF, MGC, RNG, SEST, SEV, VAB and VIT—or collected during the present study and deposited in SALA (herbarium acronyms according to Thiers, continuously updated). The selection of the material measured was based on the species distribution. The initial attempt was to evaluate all the species and subspecies currently included in *V.* subsect. *Pentasepalae,* but finally three taxa could not be studied for lack of available material (*V. krylovii* Schischk., *V. thracica* Velen., and *V. dalmatica* N.Pad.Gar., Rojas-Andrés, López-González & M.M.Mart.Ort.). Therefore 20 of the 23 species and subspecies comprising the subsection according to Rojas-Andrés and Martínez-Ortega (2016), and Padilla-García et al. (2018) were examined. Details about the plant material are given in S1 Table ordered according to the initial identification. The number of individuals and populations studied is summarized in Table 1, and the abbreviation assigned to each operational taxonomic unit (OTU) is indicated. The taxonomic starting hypothesis follows Rojas-Andrés et al. (2015) and Padilla-García et al. (2018), which is based on the results from previous molecular, cytological, biogeographic, phylogenetic and morphological studies (Fig 1 and Table 1). The spatial distribution of the specimens selected is displayed in Fig 2.

The 30 quantitative characters (abbreviations shown in Table 2) already used in Andrés-Sánchez et al. (2009) were measured for the additional taxa and populations included here. Except for cases in which the available material was insufficient, each character was measured in three specimens per population and the arithmetic mean was calculated. The matrices containing raw data and all the average values per population are available on GitHub (https:// github.com/NoeLG4/morpho.dataset).

**Table 1. Plant material.**

| Operational taxonomic unit (OTU) | Number of individuals | Number of populations |
|---|---|---|
| *V. aragonensis* Stroh. (ARA) | 21 | 7 |
| *V. austriaca* L. | - | - |
| (1)*V. austriaca* L. ssp. *austriaca* (AUS) | 15 | 5 |
| (2)*V. austriaca* ssp. *dentata* (F. W. Schmidt) Watzl (DEN) | 36 | 12 |
| (3)*V. austriaca* ssp. *jacquinii* (Baumg.) Watzl (JCQ) | 55 | 19 |
| *V. crinita* Kit. (CRI) | 25 | 9 |
| *V. kindlii* Adamović (KIN) | 24 | 8 |
| *V. linearis* (Bornm.) Rojas-Andrés & M. M. Mart. Ort. (LIN) | 6 | 2 |
| *V. orbiculata* A. Kern. (ORB) | 33 | 11 |
| *V. orsiniana* Ten. (ORS) | 72 | 24 |
| *V. prostrata* L. (PRO) | 43 | 15 |
| *V. rhodopea* Degen. ex Stoj. & Stef (RHO) | 14 | 5 |
| *V. rosea* Desf. (ROS) | 40 | 14 |
| *V. satureiifolia* Poit. & Turp. (SAT) | 42 | 15 |
| *V. senneni* (Pau) M. M. Mart. Ort. & E. Rico (SEN) | 41 | 15 |
| *V. tenuifolia* Asso | - | - |
| (1)*V. tenuifolia* ssp. *fontqueri* (Pau) M. M. Mart. Ort. & E. Rico (FON) | 14 | 5 |
| (2)*V. tenuifolia* ssp. *javalambrensis* (Pau) Molero & J. Pujadas (JAV) | 34 | 12 |
| (3)*V. tenuifolia* Asso ssp. *tenuifolia* (TEN) | 29 | 10 |
| *V. teucrioides* Boiss. & Heldr. (TCR) | 9 | 3 |
| *V. teucrium* L. (TEU) | 43 | 15 |
| *V. turrilliana* Stoj. & Stef. (TUR) | 9 | 3 |
| Total | 605 | 209 |

Summary of individuals and populations included in the morphometric study. The abbreviations of the 20 operational taxonomic units (OTUs) corresponding to the taxonomic starting hypothesis are indicated in brackets.

* The species marked with an asterisk comprise several subspecies; those belonging to *V. austriaca* have been highlighted in blue, while those of *V. tenuifolia* appear in red.

The measurements were taken from a leaf situated in the central segment of the stem (medium leaf) (Fig 2 in Andrés-Sánchez et al., 2009) and from one on the apical shoot (Fig 3 in Andrés-Sánchez et al., 2009). The measurements were taken with a digital electronic caliper Digimatic 500 (Mitutoyo American Corporation, Aurora, USA). Characters related to the indumentum were calculated only in the medium leaves. One measurement was made for each variable except for hair length, for which five trichomes per leaf were considered. "Density" was indirectly estimated by counting the number of hairs on a 1-cm-long linear transect at the leaf margin. Hair length and "density" were determined by means of a stereoscopic zoom microscope NIKON SMZ-U (Nikon Corporation, Tokyo, Japan) equipped with a video camera SONY 3CCD DXC-930P (Sony Corporation, Tokyo Japan). The photos taken were transferred to a computer and examined through the image-analysis software Image-Pro Plus version 1.0 (Media Cybernetics Inc., Rockville, USA).

In an effort to avoid the size effect, some characters were considered as quotients (LLM/MLWM, LLM/WMPM, LLM/DBMWM, FTLM/FTWM, STLM/STWM, DLAUM/TLWM, LLM/DLAUM, LLS/MLWS, LLS/WMPS, LLS/DBMWS, FTLS/FTWS, STLS/STWS, DLAUS/ TLWS and LLS/DLAUS).

The absence of normality was checked and the Spearman correlation coefficients were determined from the original matrix of descriptors in order to test for correlation between primary variables. The primary matrix was reduced by removing one of the variables shown to be correlated for all subsequent analyses; the threshold applied was 0.95. Statistical analyses were performed using the open-source R platform (descriptive statistics, Spearman correlation) (R Core Team, 2015).

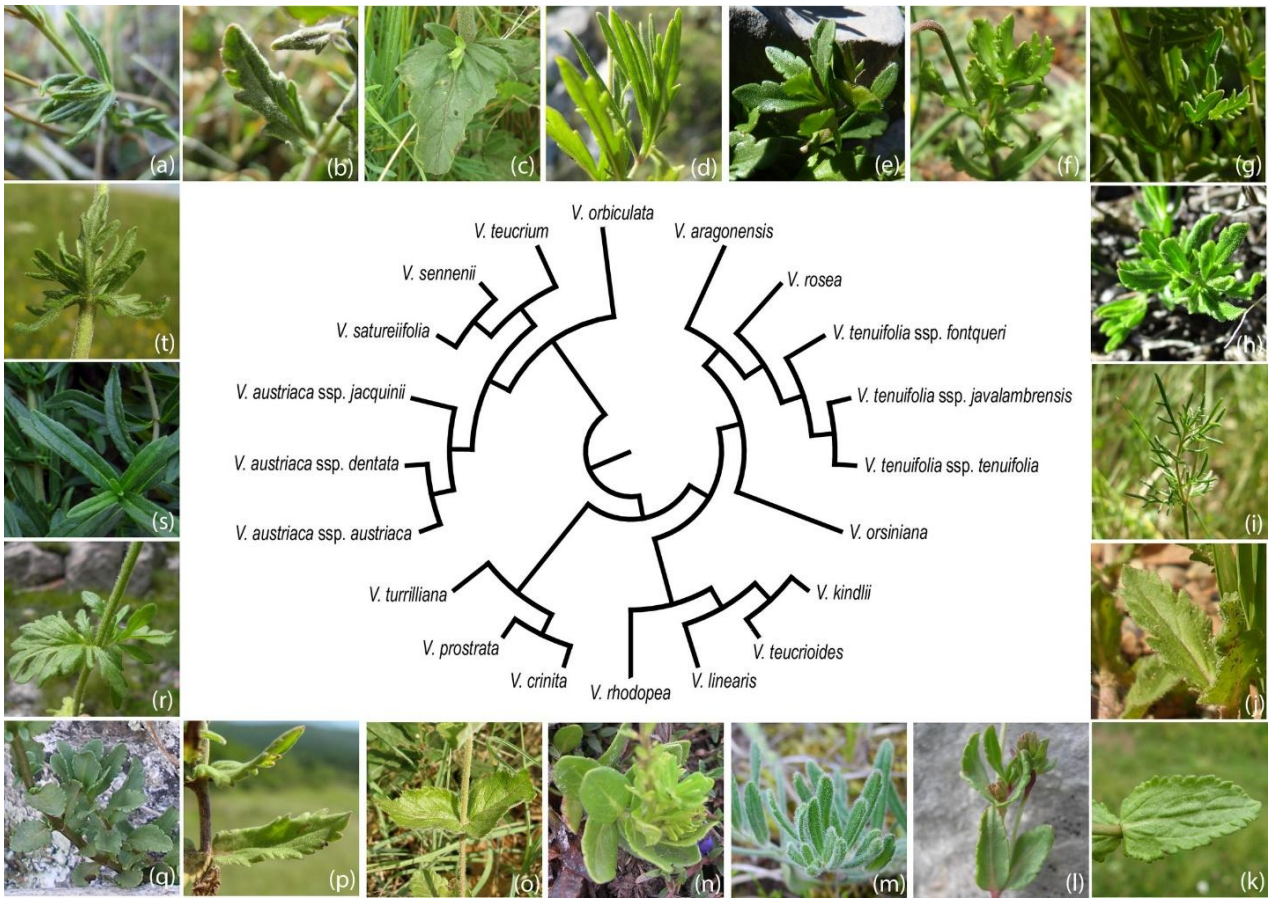



**Fig 1. Starting taxonomic hypothesis.** Simplified neighbour joining of the taxa examined for *V.* subsect *Pentasepalae*; modified from Padilla et al. 2017. a) *V. satureiifolia*, Borau, Spain. Photo: N. Padilla-García; b) *V. senneni*, Borau, Spain. Photo: N. Padilla-García; c) *V. teucrium*, Novi Sad, Serbia. Photo: S. Andrés-Sánchez. d) *V. orbiculata*, Makarska, Croatia. Photo: S. Andrés-Sánchez; e) *V. aragonensis*, Mount Baziero, Spain. Photo: N. Padilla-García; f) *V. rosea*, Djebel Lakra, Marruecos. Photo: S. Andrés-Sánchez; g) *V. tenuifolia* ssp. *fontqueri*, Sierra de las Nieves, Spain. Photo: J. Peñas de Giles; h) *V. tenuifolia* ssp. *javalambrensis*, Valdeajos, Spain. Photo: N. Padilla-García; i) *V. tenuifolia* ssp. *tenuifolia*, Bordón, Spain. Photo: M. M. Martínez-Ortega; j) *V. orsiniana*, Iglesuela del Cid, Spain. Photo: M. M. Martínez-Ortega; k) *V. kindlii*, Pljevlja, Montenegro. Photo: S. Andrés-Sánchez; l) *V. teucrioides*, Mount Olimpus, Greece. Photo: B. M. Rojas-Andrés; m) *V. linearis*, Kozjak Lake, FYROM. Photo: N. LópezGonzález; n) *V. rhodopea*, Belmeken, Bulgaria. B. M. Rojas-Andrés; o) *V. crinita*, Popovitsa, Bulgaria. Photo: M. M. Martínez-Ortega; p) *V. prostrata*, Pirot, Serbia. Photo: S. Andrés-Sánchez; q) *V. turrilliana*, Veleka river, Bulgaria. Photo: B. M. Rojas-Andrés; r) *V. austriaca* ssp. *austriaca*, Cerna Mountains, Romania. Photo: A. Badarau; s) *V. austriaca* ssp. *dentata*, Botanical Garden (Univerzity Karlovy, Prague), Czech Republic. Photo: M. Kesl; t) *V. austriaca* ssp. *jacquinii*, Josipdol, Croatia. Photo: S. Andrés-Sánchez.

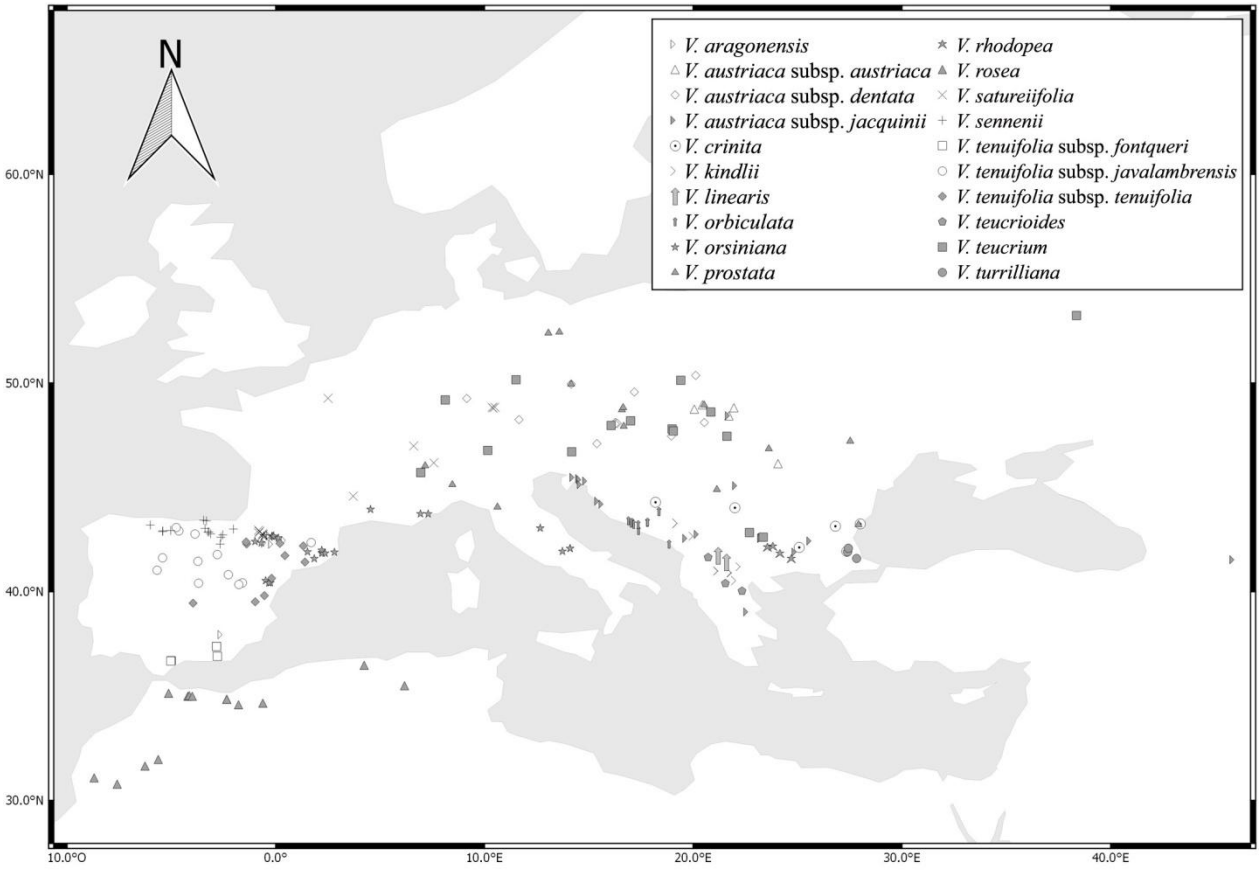



**Fig 2.** Distribution map of the populations included in this study.

A Euclidean coefficient was used to compute the secondary distance matrix after standardization of the characters in the primary matrix. Then, a principal component analysis (PCA) was performed with no *a priori* knowledge of the population groupings (i.e. ordination of the OTUs as revealed by leaf characters). Computations were made with the software NTSYSpc 2.21n (Rohlf, 2006).

### *2.1 Building an initial classification scheme based on the taxonomic starting hypothesis: Guided discriminant analyses*

Several canonical discriminant analyses (DAs) were performed using the software SPSS v. 15 for Windows (SPSS, Chicago, USA) over the standardized variables, which were selected manually to force the separation of each of the previously accepted taxonomic units and to provide an initial reference classification scheme.

Four sequential DAs were conducted for the division of the initial data set into smaller subsets and therefore simplify its complexity. This was done by selecting one of the most discriminant characters derived from the original and subsequent DAs (i.e. those based on the initial data set and different subsets established in further steps; see Results section). Character selection was manual, based on previous knowledge of the species group. The variables finally employed were: STLM/STWM (which divides the taxa into specimens with medium leaves entire to pinnatifid vs. pinnatipartite to bipinnatisect; *sensu* Beentje, 2010), DI (densely hairy leaves vs. subglabrous to glabrous leaves), LT (short vs. long trichomes) and LLM (to distinguish taxa showing large medium leaves from those with small medium leaves). By this character selection, some phenotypic groups arise. Within these final groups, several characters were further used, forcing taxon classification. To show the variability of the selected characters within each species in a comparable way, graphic tests (i.e. box-plot with indication of median values) were conducted.

**Table 2. Characters measured and abbreviations.**

| Abbreviation | Morphological character | | |
|---|---|---|---|
| **LT** | Medium leaf | Length of trichomes | |
| **DI** | | Density of indumentum | |
| **MLWM** | | Width | Maximum width |
| **WMPM** | | | Middle part |
| **TLWM** | | | Entire terminal part |
| **FTWM** | | | First tooth |
| **STWM** | | | Second tooth |
| **LLM** | | Length | Total |
| **FTLM** | | | First tooth/segment |
| **LFFM** | | | First division/segment (bipinnatisect leaf) |
| **STLM** | | | Second tooth/segment |
| **LFSM** | | | First tooth of the second segment (bipinnatisect leaf) |
| **PLM** | | | Petiole |
| **DBMWM** | | Distance between the leaf base and the maximum width line | |
| **DLAUM** | | Distance between the leaf apex and the uppermost teeth | |
| **NTM** | | Number of teeth per hemilimb | |
| **MLWS** | Leaf of the apical shoot | Width | Maximum width |
| **WMPS** | | | Middle part |
| **TLWS** | | | Entire terminal part |
| **FTWS** | | | First tooth |
| **STWS** | | | Second tooth |
| **LLS** | | Length | Total |
| **FTLS** | | | First tooth/segment |
| **LFFS** | | | First division/segment (bipinnatisect leaf) |
| **STLS** | | | Second tooth/segment |
| **LFSS** | | | First tooth of the second segment (bipinnatisect leaf) |
| **PLS** | | | Petiole |
| **DBMWS** | | Distance between the leaf base and the maximum width line | |
| **DLAUS** | | Distance between the leaf apex and the uppermost teeth | |
| **NTS** | | Number of teeth per hemilimb | |

The box-plots were generated using the "ggplot2" package in R (Wickham, 2010). Following this procedure, some particular observations (populations) were classified as

belonging to a taxon that did not match the initial identification. These observations were considered errors. The misclassification rate (MCR: number of misclassified cases regarding the total) was calculated as the sum of errors (i.e. the misclassified cases) in each division. A misclassification in a superior division forces an observation to be misguided and never reach correct classification.

### *2.2 Searching for diagnostic characters assuming no previous knowledge*

The purpose of the data analysis was to search for leaf features that would be diagnostic for the species. This search was assumed to be uninfluenced by prior knowledge of the group, meaning that manual intervention or decisions based on previous knowledge should be ruled out. For the implementation of the "divide and conquer" approach, the character selection should reduce the complexity of the initial dataset, recurrently dividing this initial matrix into subsequent subgroups (i.e. generating a grouping scheme).

**(1) Unsupervised discriminant analyses (unsupervised multivariate analysis).** Unsupervised multivariate analysis discarding manual intervention was carried out. For this, several canonical discriminant analyses (DAs) at different scales were performed using the software SPSS v. 15 for Windows (SPSS, Chicago, USA) following a systematic, sequential approach. The procedure was unsupervised, assuming some artificial criteria to rule out manual intervention and thus decisions based on previous knowledge. This was done by selecting the variable showing the highest percentage of variance explained in the first discriminant function in each DA. This character was then represented in a box-plot, allowing the separation of the dataset into two subsets. Once the variable was chosen, the threshold for splitting the data was established according to two conditions: (1) the main bodies of the box-plots could not overlap, and

(2) the threshold should minimize the number of misclassified cases for each step. This procedure was recurrently applied until every species and subspecies was individually classified. The misclassification rate was calculated as explained in the previous section. The box-plots were generated using the "ggplot2" package in R (Wickham, 2010).

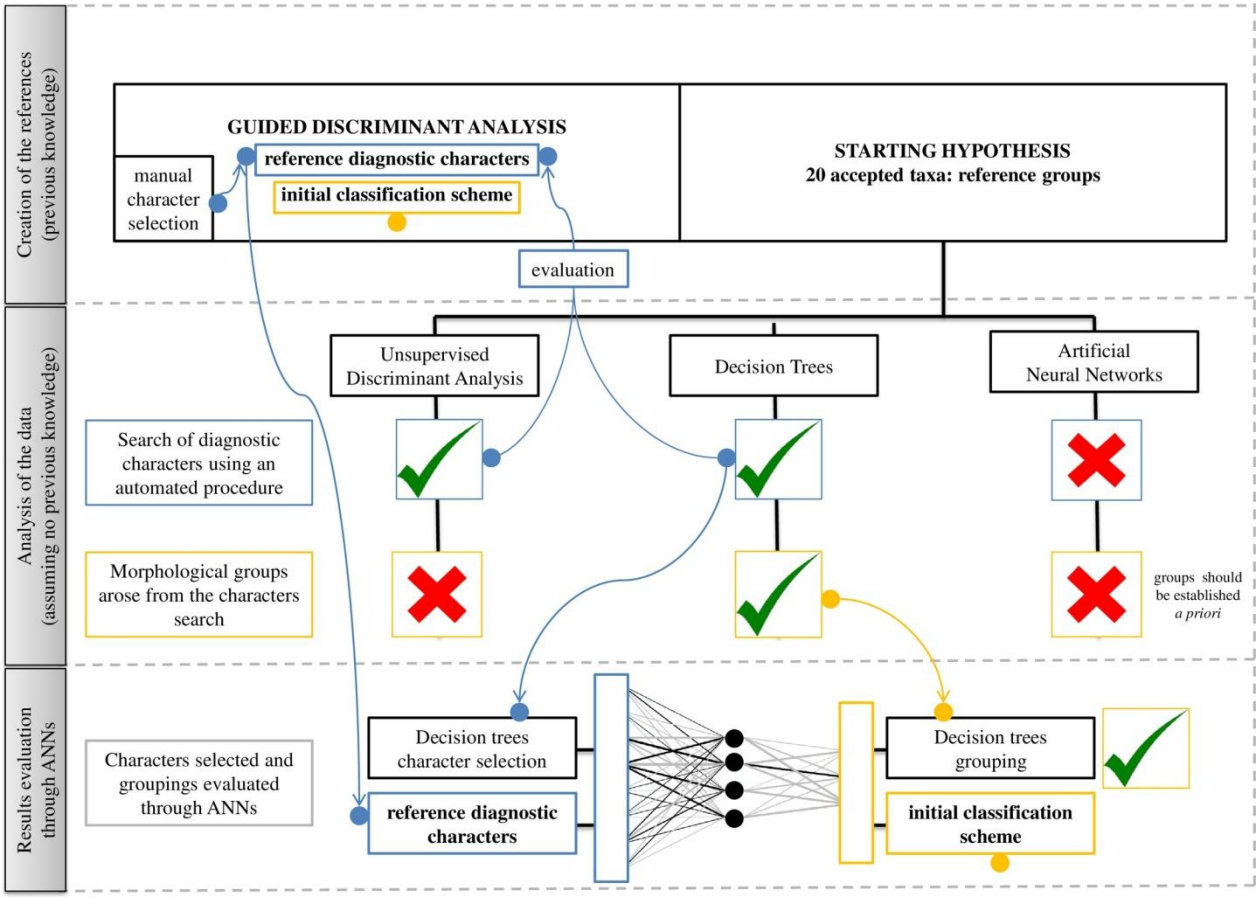



**Fig 3. Workflow.** The workflow involves the following steps (separated in the image by dashed gray lines): creation of the references, data-analysis approaches and evaluation of the results. The green ticks mean optimal outcomes while red crosses mean suboptimal ones. Processes related to the search of diagnostic characters are indicated in blue, while those corresponding to the groupings are indicated in light orange.

**(2) Decision Trees.** DTs have a built-in mechanism for performing variable selection (Breiman et al., 1984). This technique explicitly focuses on relevant features while ignoring irrelevant ones (Hall, 2000), so that there is no need of prior feature selection.

Together with feature selection, the treatment of missing data is a key issue to be considered during the pre-processing of the data when working with data-mining tools. Due to the low number of missing cases in the present study, the only population presenting them was removed from both DTs and ANNs. First of all, a perfect tree that fits the data was produced, setting the minimum size of the terminal node to the minimum number of observations in the dataset (two, as it is indicated below) and the minimum residual deviance to zero. These parameters enable the tree to detect taxa even with only two cases (populations) available (e.g. *V. linearis*, see Table 1) and classify all the observations (if the limit on tree depth allowed it), but this tree is clearly over-fitted (see Results) and therefore useless. The tree was grown by binary recursive partitioning.

The splitting criterion is the division that maximizes the reduction in deviance; splitting continues until the terminal nodes are too small or too few to be split (Ripley, 2015). These kinds of trees lead to a large number of terminal nodes and are usually over-fitted, so in a second step the tree was simplified by "pruning" (Chen et al., 2011). This technique reduces the initial size by removing the least important splits. The classification trees and the parameters to evaluate them (residual mean deviance and misclassification rate) were taken directly from the package "tree" in R (Ripley, 2015). The procedure for calculating the misclassification rate is analogous to that of DAs: it results from the sum of the misclassified cases in each node. However, in this case a misclassification in a higher division does not necessarily force an observation to be misguided because some taxa appear in more than one final node. The script used to analyze the data is available on GitHub (https://github.com/NoeLG4/ morpho.DT).

**(3) Artificial neural networks.** Feature selection when working with ANNs is a critical step (Maimon an Rokach, 2010). Perfectly correlated variables are truly redundant,

meaning that no further information is gained by adding them (Guyon and Elisseeff, 2003). Therefore, correlated variables were removed from the dataset and all remaining features were initially considered. Most of the variables considered in the present study were leaf measurements so that some degree of correlation was expected. Furthermore, some of them were highly correlated with each other (>0.8), making the task of selecting sufficient independent variables especially difficult. With this taken into account, the determination of the best conditions for the ANN was performed by a preliminary test among several ANNs with different configurations of variables in combination with inspections of time-series plots of potential inputs and outputs (Maier and Dandy, 2000). Max-Min standardization was carried out to ensure that each input variable received the same attention (Maier and Dandy, 2000). The output layers (representing the taxa) were transformed into binary variables through effect coding. The algorithm used by the ANN for its training was designated by "rprop+" (resilient backpropagation with weight backtracking (Riedmiller, 1994). All neural networks were performed using the "neuralnet" package (Fritsch and Guenther, 2012) included in R.

Once the input layers were established, several networks were performed with 50%, 60%, and 70% of the cases randomly chosen as training data (and the rest reserved for testing the models), with different number of hidden layers (1, 2), and different number of neurons within each hidden layer (from 8 to 16). Because ANNs are sensitive to subtle changes (Breiman, 1996) three different training datasets were generated for each analysis. With the use of these three datasets, the parameters were established (percentage of training set, number of hidden layers and number of neurons). With the parameters fixed, 10 different training and test sets were created and the total and per species misclassification rates were then calculated as the average of incorrectly assigned examples in the distinct test sets. Analyses including the 20 taxa resulted in

high misclassification rates (see Results) so that 10 random groups were generated for four categories of output layers: 4, 8, 12, and 16 (i.e. 4, 8, 12, and 16 species or infraspecific taxa) to evaluate the performance by number of species. The parameters 'percentage of training set' and 'number of hidden layers' remained constant, the number of neurons changed in each case to optimize the outcomes. A neural interpretation diagram (hidden layers = 1; neurons = 8; output layers = 8) is shown in S1 Fig. This general scheme of a typical three-layered ANN architecture was produced using the R function plot.nnet (Beck, 2013). The graphic displaying a misclassification rate by the number of species (S2 Fig) was calculated through "ggplot2" package (Wickham, 2010). The script used for analyzing the data and generating the graphics is available on GitHub (https://github.com/NoeLG4/morpho.ANN).

### *2.3 Artificial neural networks as a tool to evaluate morphological groups established through a set of specific features*

Since only suboptimal results were found when the whole taxonomic group was considered, ANNs were finally not used for the initial aim. However, taking advantage of the high sensitivity of this technique to the combination of input and output layers, they were used to assess the capacity of the selected variables to classify the taxa within the final groups established by the best technique (see Results). This procedure was also used with the variables selected with the help of the guided DAs and the corresponding groupings (initial identification scheme) to be used as reference. The variables used in guided DAs and DTs were selected as input layers for ANNs, and the different final groups established with these techniques (see Results) were treated as output layers. The number of neurons was set and the misclassification rate calculated as explained in the

previous section. The analyses were made using the "neuralnet" package (Fritsch and Guenther, 2012) included in R.

## 3. RESULTS

The results of the PCA (Table 3) indicate that the variance of the data is explained mostly by the selected morphological variables. The first, second, and third components accounted for 53.57%, 17.04%, and 7.86%, respectively, of the total variation among populations.

Nevertheless, due to the high number of observations a clear structure is not evident in the corresponding graphic (figure not shown).

**Table 3. Principal component analysis.**

| Axis | Eigenvalue | Percent | Cumulative |
|---|---|---|---|
| 1 | 696.91 | 53.57 | 53.57 |
| 2 | 221.66 | 17.04 | 70.61 |
| 3 | 102.31 | 7.86 | 78.47 |
| 4 | 69.06 | 5.31 | 83.78 |
| 5 | 52.20 | 4.01 | 87.79 |
| 6 | 31.60 | 2.43 | 90.22 |
| 7 | 22.05 | 1.70 | 91.92 |
| 8 | 21.46 | 1.65 | 93.57 |
| 9 | 16.39 | 1.26 | 94.83 |
| 10 | 15.23 | 1.17 | 96.00 |

Eigenvalues and percentages of the data variance accounted by each axis.

The Spearman correlation coefficients calculated from the original matrix of descriptors showed that some of the primary characters were highly correlated (> 0.95). The pairs MLWM-WMPM, LLM/WMPM-LLM/MLWM, LLS/MLWS-LLS/WMPS, and

MLWS-WMPS displayed the following values: 0.971, 0.961, 0.955, and 0.953, respectively. Therefore, the variables MLWM, LLM/WMPM, LLS/MLWS, and MLWS were excluded from all subsequent analyses.

### *3.1 Initial classification scheme through guided discriminant analyses*

An initial DA performed using the original data matrix showed that a set of variables contributed highly to the discriminant functions and therefore could be selected to delimit the first two sub datasets. Some of these features (i.e. STLM, FTLM, STLM/STWM, FTLM/FTWM, STLS, and FTLS) were related to leaf division (Fig 4A). There was another set of variables not related to leaf division (i.e. NTM, WMPM, and LLM) that could also be used to delimit the first two sub-datasets. Among these two sets of variables, those related with leaf division were considered more informative, and consequently STLM/STWM was finally selected. Following this procedure, subsequent DAs applied to different subsets of species showed sets of features that could be selected for the recursive partitioning of the dataset. Among these variables DI, LT, and LLM were chosen. The selection of these variables reduced the complexity of the dataset even if these features did not contribute the most to the discriminant functions (all discriminant functions, standard coefficients, and structure matrix tables are shown in S2 Table). The subsequent partitions of the original dataset into subsets of species (Groups I to VIII) are displayed in Fig 4, together with box-plots for the chosen leaf characters corresponding to sequential DAs that maximize differences between subsets. These eleven variables used for species and infraspecific taxon classification constitute the reference diagnostic characters (STLM/STWM, DI, LT, LLM, STWM, LLS/WMPS, NTS, STLM, PLS, FTLM and WMPM).

This initial classification scheme has a misclassification rate of 0.18 (38/209).

Group I holds the taxa with pinnatipartite to bipinnatisect medium leaves, while Group II contains the species with entire to pinnatifid medium leaves. Within Group I the quotient STLM/STWM (Fig 4A) helps to differentiate between [JCQ+TEN] and [AUS+JAV+ORB]. STWM further helps to distinguish JCQ from TEN, while LLM and LLS/WMPS differentiate among AUS, JAV, and ORB [all box-plots corresponding to Group I are shown in S1 File, from (a) to (d)].

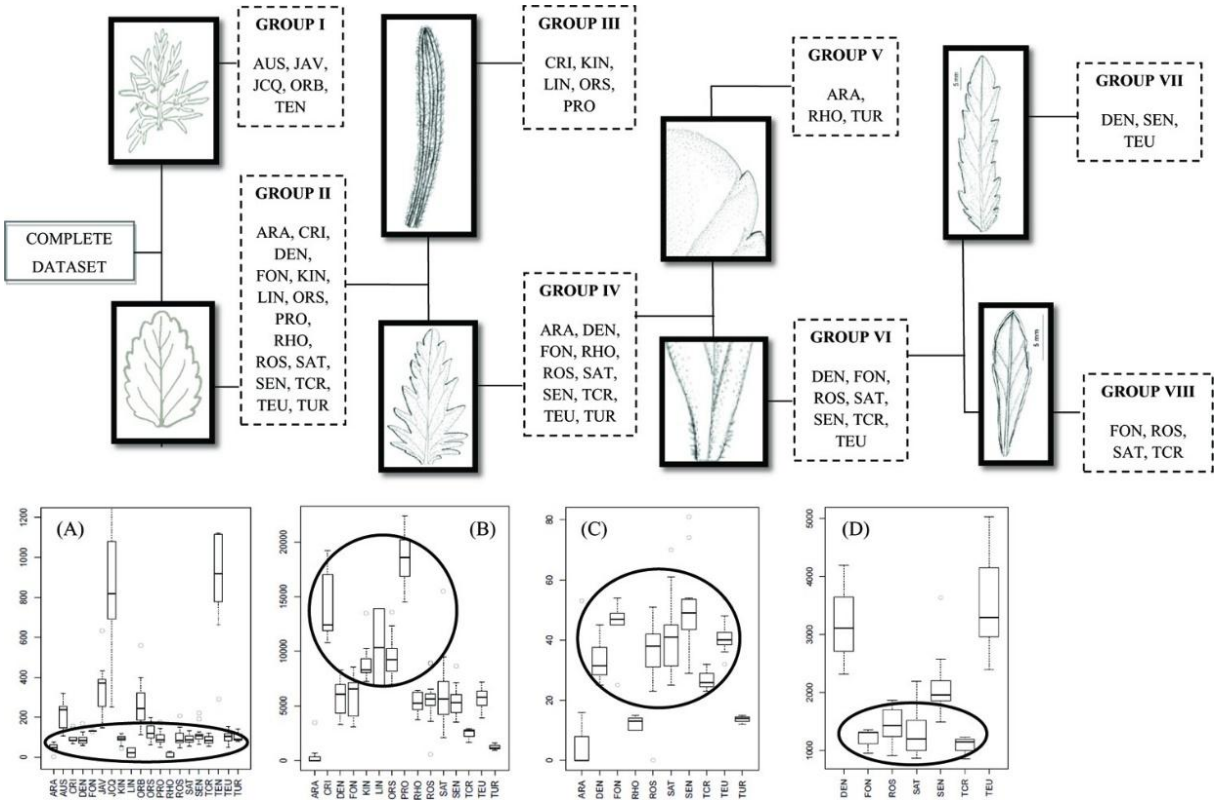



**Fig 4. Initial classification scheme through guided DAs.** Partition of the original dataset in accordance with the starting hypothesis. Box-plots for (A) STLM/STWM, (B) DI, (C) LT, and (D) LLM. See Table 1_for abbreviations. The circles indicate Group II, Group IV, Group VI, and Group VIII, respectively.

The next partition within Group II was made using the character DI, which renders Group III (taxa showing pubescent leaves), and Group IV (taxa with subglabrous leaves) (Fig 4B). Within Group III, NTS is most helpful to distinguish among CRI and [LIN+KIN+ORS+PRO] and STLM values do not overlap between LIN and the rest. Furthermore, PRO is easily distinguishable from [KIN+ORS] based on the character DI, and KIN can be differentiated from ORS based on the character LT [all box-plots corresponding to Group III are shown in S1 File, from (e) to (h)].

Within Group IV, the next partition is based on the length of the trichomes (LT; Fig 4C). Thus, two further groups resulted: Group V, which contains the species bearing short trichomes on their leaves, while Group VI includes taxa having long ones. Within Group V, ARA can be distinguished from [RHO+TUR] using the character PLS. Moreover, FTLM could be used to separate RHO from TUR [box-plots corresponding to Group V are shown in S1 File, (i) and (j)].

Finally, the taxa included in Group VI could be separated into two further subgroups based on the character LLM (Fig 4D): While Group VII included the entities having medium-sized and long leaves, Group VIII comprised the taxa with small leaves. Within Group VII, SEN could be differentiated from [DEN+TEU] (and even from the remaining taxa within Group VI) based on this character [LLM; S1 File, (k)]. Additionally, TEU differed from DEN in the width of their medium leaves (WMPM) [box-plots corresponding to Group VII are shown in S1 File, (k) and (l)]. Within Group VIII, the characters STWM help distinguish ROS from [FON+SAT+TCR]; moreover, the variable STLM registered values that did not overlap between FON and [SAT+TCR], and these two species could be easily differentiated using the variables DI or DLAUM [all box-plots corresponding to Group VIII are shown in S1 File, from (m) to (o)].

### *3.2 Searching for diagnostic characters assuming no previous knowledge*

**(1) Unsupervised discriminant analyses.** Ten variables were used for species and infraspecific taxon classification through unsupervised DAs (STLM, WMPM, NTM, FTLM/FTWM, DI, LT, LLM, STWM, LLM/DBMWM and STLS/STWS). All the divisions performed with the most discriminant characters determined by DAs are available in S2 File. The first DA carried out indicated that the variable STLM gave the highest percentage of variance explained in the first discriminant function according to the structure matrix (all discriminant functions, standard coefficient, and structure matrix data are shown in S3 Table). In this first step [JCQ+TEN] were separated from the rest of the taxa; additionally, the variable identified to distinguish JCQ from TEN was WMPM. The best character according to the next pair of DAs was NTM, which separated in one step [ARA+LIN+RHO] from the rest of taxa and in the second one, the remaining species from CRI. Within the above-mentioned group, FTLM/FTWM helped to distinguish [ARA+LIN] from RHO, and DI separates ARA from LIN. The fourth DA applied revealed WMPM as the best character and led to the distinction of TEU from the other taxa. The leaf feature showing the highest percentage of variance explained in the first discriminant function suggested by the next discriminant analysis was DI, which applied to the subset separates [AUS+KIN+ORS+PRO] from the remaining taxa (nine at this point). DI arose again as the best variable to differentiate PRO from [AUS+KIN+ORS]; LT allowed the separation of ORS from [AUS+KIN], finally AUS could be distinguished from KIN based on LLM.

LLM was used again in the subsequent DA to distinguish [DEN+SEN] from the remaining taxa. This character was also useful to differentiate DEN from SEN. In the next DA the best variable found was STWM, and the threshold with the fewest misclassified cases split the current subset into [TCR+TUR] and the rest of the taxa;

additionally, LT was found to be the best feature to differentiate between TCR and TUR. At this point, only five taxa remained; the DAs performed to distinguish among them gave as a result STWM, LLM/DBMWM, and STLS/STWS as the most discriminant characters, separating ROS, JAV, and SAT from [FON+ORB], in this order. The best variable emerged for the last DA was DI, with slightly higher values in FON. Through this variable selection in most of the steps, just one or two species were separated from the rest. This process generated a greater number of groups than did guided DAs, with a low number of species in each group. The complete sequence of DAs had a misclassification rate of 0.33 (68/209).

**(2) Decision trees.** The perfect tree correctly classified all the observations (residual mean deviance = 0; misclassification error rate = 0), but led to 48 final nodes (S3 Fig). Some of the species (e.g. ROS, SAT, SEN) appeared more than three times, highly over-fitted, considering that the dataset contained a total of 20 taxa.

The pruned tree (Fig 5) showed 20 terminal nodes, which did not correspond exactly to the 20 taxa: 14 entities were identified corresponding to a single terminal node, three were found in two terminal nodes of separate branches (ARA, DEN, SEN), and three complete taxa were misidentified (AUS, LIN, TUR). The residual mean deviance was 0.998 and the misclassification rate 0.19 (40/208; one observation was eliminated due to lack of information; see Materials and Methods). In tree construction, 11 variables were actually used (STLM/STWM, STWM, STLM, DI, LT, NTM, WMPM, LLS, FTLM, LLM, and FTWS). According to this feature selection, the taxonomic entities can be classified into three main groups (characters and exact values that lead to this classification are shown in Fig 5).

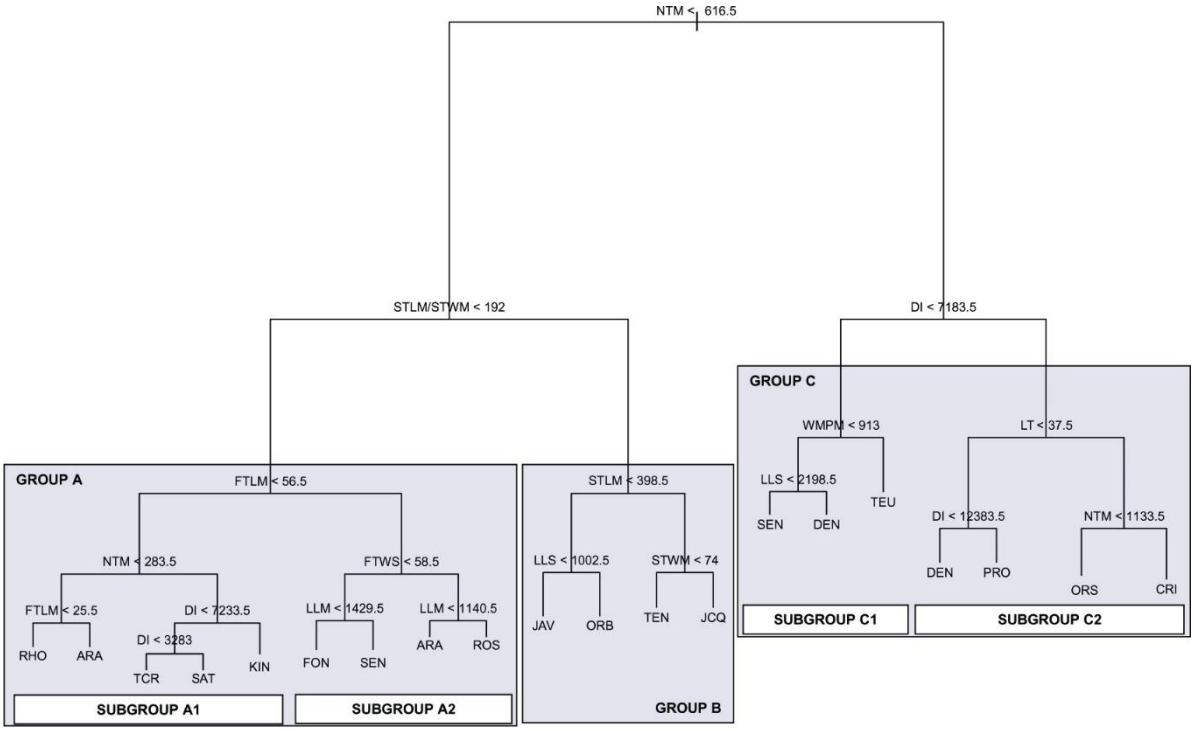



**Fig. 5. Pruned Tree**

1) Group A [ARA+FON+KIN+RHO+ROS+SAT+SEN+TCR] has low values of NTM and the ratio STLM/STWM. This first group can be subdivided by the character FTLM into Subgroup A1 [ARA+KIN+RHO+SAT+TCR], with shorter teeth, and Subgroup A2 [ARA+FON +ROS+SEN], with longer teeth. 2) Group B [JAV+JCQ+ORB+TEN], has low values of NTM and high for the quotient STLM/STWM, which means pinnatipartite to bipinnatisect leaf. 3) Group C [CRI+DEN+ORS+PRO+SEN+TEU], has high values of NTM; this group can also be subdivided into two subgroups based on DI: Subgroup C1 [DEN+SEN+TEU], with low values; and Subgroup C2 [CRI+DEN+ORS+PRO], with high values of DI.

**(3) Artificial neural networks.** From the initial set of variables, 20 were finally selected for ANNs (LT, DI, PLM, STLM/STWM, LLM, WMPM, DBMWM, DLAUM/TLWM, NTM, PLS, STWS, DLAUS/TLWS, LLS/ DLBMWS, LLM/DBMWM, TLWS, NTS, LLS/WMPS, LFFM, LFFS, and LFSS). The different percentages used for splitting the data into training and testing sets showed no significant differences in terms of misclassification rate. In each scenario, one hidden layer displayed better outcomes than two or more and heavily reduced the computing time. The number of neurons which resulted in higher predictive capacities varied between 8 and 16, depending on the number of output layers. The analyses including all taxa led to high misclassification rates (mean value of 0.41, see S2 Fig). Even using the best model, nine of the 20 taxa displayed values over 0.40 (Table 4). Reducing the number of output layers resulted in better values. As shown in S2 Fig, ANNs perform well with groups of less than eight taxa (values of correctly classified cases above 0.75, i.e. a misclassification rate below 0.25), even though it depended largely on the combination of species and infraspecific taxa.

### *3.2 Comparison of the variables and groups resulting from guided DAs and DTs. Evaluation of the classification power with ANNs*

Of the three techniques applied in the present study: ANNs gave suboptimal results when applied to the total dataset, unsupervised DA failed to generate manageable groups, and DTs offered good results regarding both aspects. Therefore, the leaf characters selected through DTs (and the corresponding groupings) were compared to those manually selected with the help of guided DAs (and the initial classification scheme) in terms of misclassified cases using ANNs. Some groups displayed high misclassification rates (Group VIII for DAs: 0.36; Group A and Subgroup A1 for DTs:

0.33 and 0.39, respectively), but the overall misclassification rate was below 0.2 in most cases, as shown in Table 5 (guided DAs variables and groups) and Table 6 (DTs variables and groups).

**Table 4. ANNs per species results.**

| Species | MCR | NC |
|---|---|---|
| **ARA** | 0.39 | 7 |
| **AUS** | 1.00 | 5 |
| **DEN** | 0.38 | 12 |
| **JCQ** | 0.17 | 19 |
| **CRI** | 0.89 | 9 |
| **KIN** | 0.56 | 8 |
| **LIN** | 0.33 | 2 |
| **ORB** | 0.49 | 11 |
| **ORS** | 0.18 | 24 |
| **PRO** | 0.10 | 15 |
| **RHO** | 0.33 | 5 |
| **ROS** | 0.56 | 14 |
| **SAT** | 0.30 | 15 |
| **SEN** | 0.55 | 15 |
| **FON** | 0.44 | 5 |
| **JAV** | 0.05 | 12 |
| **TEN** | 0.17 | 10 |
| **TCR** | 1.00 | 3 |
| **TEU** | 0.10 | 15 |
| **TUR** | 1.00 | 3 |

(Hidden layer = 1, number of neurons = 16, output layers = 20). MCR = misclassification rate. NC = number of cases. Values over 0.4 indicated with an asterisk.

**Table 5. MCR calculated through ANN.**

| | Guided DAs final groups; hidden layer = 1 | | | | |
|---|---|---|---|---|---|
| **Neurons** | **8** | **10** | **10** | **8** | **12** |
| **Group** | I | III | V | VII | VIII |
| **Error** | 0.14 | 0.053 | 0.024 | 0.038 | 0.089 |
| **Reached threshold** | 0.0086 | 0.0089 | 0.0085 | 0.0086 | 0.0086 |
| **Steps** | 81.00 | 64.80 | 27.36 | 47.34 | 86.38 |
| **MCR** | **0.14** | **0.12** | **0.17** | **0.09** | **0.36** |

In this case the input layers are the variables manually selected with the help of guided DAs and the output layers the entities within the final groups (initial classification scheme).

**Table 6. MCR calculated through ANN.**

| | DTs final groups; hidden layer = 1 | | | | | | |
|---|---|---|---|---|---|---|---|
| **Neurons** | **18** | **10** | **8** | **12** | **8** | **10** | **12** |
| **Group** | A | A1 | A2 | B | C | C1 | C2 |
| **Error** | 0.153 | 0.177 | 0.05 | 0.109 | 0.057 | 0.037 | 0.041 |
| **Reached threshold** | 0.0089 | 0.009 | 0.0088 | 0.0089 | 0.0087 | 0.0089 | 0.0089 |
| **Steps** | 321.04 | 161.49 | 106.33 | 101.53 | 107.64 | 53.02 | 50.62 |
| **MCR** | **0.33** | **0.39** | **0.21** | **0.16** | **0.10** | **0.09** | **0.03** |

The input layers are the variables selected in DT analysis and the output layers the entities within the final groups obtained through the pruned tree.

## 4. DISCUSSION

This work presents an extensive morphometric analysis focused on relevant leaf characters that covers the complete geographical range of *V.* subsection *Pentasepalae*. A classical multivariate technique (DA) and two data-mining techniques (DTs and ANNs) are used to facilitate the search for discriminant characters and formulation of morphological groups within taxonomically complex plants in which polyploidy and hybridization are involved. A phenotypic species concept is of crucial importance for taxon identification, especially in the field, where other kinds of evidence such as genetic data is still difficult to use. However, when allopolyploidy is involved, morphological data could easily disagree with other sources of information, such as genetic, biogeographic or cytological data, thus complicating the implementation of integrative taxonomy. In this common situation, the methodology used when working with angiosperm species could help to transfer a taxonomic hypothesis based on several lines of evidence to the description of "morphological groups".

### *4.1 Generating morphological groups through feature selection*

**(1) Morphological groups based on manual selection.** As an initial step in this work, a classic multivariate analysis combined with guided recursive partitioning (guided DAs) was used to establish "morphological groups" as an optimal classification scheme, based directly on prior knowledge of the subsection and on an initial taxonomic working hypothesis (Rojas-Andrés and Martínez-Ortega, 2016) (Fig 1). DA is a powerful technique for examining differences among groups with respect to determining whether meaningful differences exist between them (Marhold, 2011). DA finds discriminant functions that best differentiate predefined groups by maximizing the

differences between groups while minimizing variation within groups (Klecka, 1980). In the present case, a "divide and conquer" strategy was used through DAs in a directional way. For the implementation of this strategy and for the establishment of "morphological groups" (i.e. I to VIII, Fig 4), the original dataset was sequentially split into subgroups using the most informative variable among those with high absolute correlation within any discriminant function. The misclassification rate was low, but not zero, due to the particularities of the data set, mainly because it included taxa having high levels of phenotypic variability (e.g. allopolyploid taxa).

**(2) Using DTs to establish morphological groups automatically.** For establishing morphological groups in an automated way, DTs are appropriate. DTs (and ANNs) learning methods are part of the KDD process (knowledge discovery on databases), whose final aim is the search for patterns in huge databases (Maimon and Rokach, 2010). This technique implements the "divide and conquer" principle itself, and the generation of groups based on features is completely automatic. Furthermore, the built-in mechanism of the trees automatically selects the most important variables (Breiman et al., 1984). DTs therefore are highly useful for effective and, at least, fast initial approximations, because no previous knowledge on the group is required and successful results are achieved even with closely related species (Barciová and Macholán, 2009; Depypere et al., 2009). The resulting trees display the root at the top. Each sequential division shows an annotation in the graphic output representing the splitting criterion. Cases meeting the criterion go left and those failing to do so go right. The size of the branch above each division shows the decrease in deviance associated with that split. Therefore, the first divisions have longer branches than do the last ones and the branch length diminishes with the depth of the division.

This approach has many advantages: feature selection is intrinsic to the methodology, data transformation is unnecessary, classification success does not depend on the data meeting normality conditions or covariance homogeneity, and the non-linear effect of explanatory variables can be handled (Steinberg and Colla, 1995). Moreover, some studies reveal that DT analyses perform better on data-sets with incomplete records (Karels et al., 2004).

Decision trees also have drawbacks. They can be unstable and small changes in the training data can result in alterations in the final tree (Sutton, 2005), but this is not a serious disadvantage, being easily solved by bootstrapping (Mendoza, 2007; May and Lacourse, 2012). Another problem is the method's inability to manage groups with low numbers of cases unless a perfect tree is produced. This generates an over-fitted tree which leads to a useless classification. When a large reference sample is available, DTs are an appropriate choice (Lindbladh et al., 2002; Aitkenhead, 2008). Consequently, extending the observations in the initial dataset could provide more accurate results with this technique. The main limitation arises when information on these species cannot be added due, as in this particular case, to the small number of existing localities and individuals for some species.

**(3) Using ANNs for evaluating the "morphological groups" through the combination of variables selected.** The ANNs used in this study are based on adaptive learning algorithms (backpropagation algorithms) and are the most widely used type. They consist of an input layer (with neurons representing input variables), an output one (with neurons representing the dependent variables), and one or more hidden layers containing neurons intended to capture the nonlinearity in the data (Basheer and Hajmeer, 2000). These networks are versatile and can be used for data modelling, classification, forecasting, control, data and image compression, and pattern recognition

(Hassoun, 1995). They can handle a great array of data types and integrate them into categorized outputs which can represent nearly anything, from medical diagnoses (Amato, 2013) to echolocation calls in bats (Parsons and Jones, 2000).

ANNs also have limitations. They can handle various types of data, but for modelling data of low dimensionality, ANNs perform worse than do conventional statistics. On the other hand, they may be used when higher accuracy is required (Basheer and Hajmeer, 2000). The data pre-processing is not straightforward and represents a critical step (Maimon and Rokach, 2010), and consequently it has a significant effect on the final model performance (Maier and Dandy, 2000). Another drawback is that using ANNs does not allow to direct selection of the most important variables and does not provide p-values for testing the significance of the parameter estimate (Tso and Yau, 2007). However, in this case it should be taken into account that there are some approaches which allow the assessment of the contribution of variables to the model (Olden et al., 2004). Another disadvantage usually attributed to the traditional ANNs is a limitation on the generalization of the results that can over-fit the data (Sugumaran et al., 2007).

In any case, their power as classification tools is beyond doubt. The reason why ANNs are not an appropriate approach in this case is related to the characteristics of the dataset generated by the study group (i.e. too many entities to classify, too few observations per taxon in some cases and the fact that some are highly polymorphic). Probably, ANNs combined with another kind of initial dataset would provide better outcomes. For example, computer-based image analysis has excellent potential even for identification at varietal levels in some plant groups (wheat: Dubey et al., 2006; *Camellia sinensis*: Pandolfi et al., 2009) or when sufficient information about each output case is available (onion varieties: Rodriguez Galdon et al., 2010).

It bears highlighting that ANNs were not used in this work as a classification method, but rather to assess each of the final morphological groups with respect to the variables leading to these groups. ANNs were employed to compare the groups established using DTs and groups formed by a guided DA using the set of characters previously defined by each technique. The properties of this construction (ability to capture hidden patterns in data, good results when accuracy is needed) were taken advantage of together with the good results displayed when dealing with small groups of species (the sensibility to the set of input/output layers observed during this work). ANNs easily adjust to any set of input-output patterns and through a robust training process perform a model function with the minimum possible error. For all these reasons, a novel use of ANNs is proposed here to evaluate the adequacy of an input set of variables to classify the dependent variables.

### *4.2 Searching automatically for discriminant characters: DTs vs. unsupervised DAs*

Looking for a set of discriminant characters to distinguish taxonomic entities represents a primary objective in taxonomic studies (Dayrat, 2005). In this study, DTs (11 variables selected, see Results) and unsupervised DAs (10 variables selected, see Results) were both appropriate to this aim. Despite the misclassification rates–suboptimal results using unsupervised DAs, as compared to DTs–the characters selected through DTs and unsupervised DAs are overall consistent with each other and the reference diagnostic characters (11 variables selected, see Results). That is, six variables coincided between guided DAs and DTs (DI, LT, LLM, STLM, STWM, and STLM/STWM), and a slightly different set of six features were shared between the two approaches that use DAs (DI, LT, LLM, STLM, STWM and WMPM).

The differences in the misclassification rate depend in part on the order of selection of diagnostic variables which eventually lead to the groupings. The variables which account for a greater amount of the variance were identified through unsupervised DAs. However, each step usually separated one or two species, but did not really split the dataset. This situation differed in the case of manual selection (guided DAs) or DTs, but neither the variable selection nor the order has one valid solution. Notably, the misclassification rate here was almost optimal using DTs. DTs achieved nearly the same results as guided DAs using a different combination of 11 variables from the initial set of 44. The similarity between these two techniques leads to some groupings that are equivalent (Group VII and C1) or quite comparable (Group I and B; Group III and C2).

DTs use specific classification rules allowing the direct and automatic creation of dichotomous keys to distinguish the different OTUs (Chakaravarthy et al., 2007). They also provide clear information on the importance of significant factors for prediction or classification (Tso and Yau, 2007). In fact, DTs have been used in the pre-processing of data for the feature-selection step (Sugumaran et al., 2007). DTs (as data-mining techniques) can deal with sets having considerably high levels of incomplete data in several ways (Brown and Kros, 2003), but as mentioned above, they are not the most appropriate tools in the search of discriminant variables to classify species with low numbers of cases. By contrast, an important advantage of DAs is the ability to differentiate all the entities regardless of the number of observations. For particular scenarios involving endangered species or narrowly distributed taxa that, moreover, occur sympatrically with closely related species, DA may be the best choice (Lambert et al., 2006; Reichenbacher et al., 2009). In the present study, the species with low numbers of observations (e.g. LIN, TUR) were identified better by the combination of features implemented by DAs than by DTs.

For their part, DAs constitute an extraordinarily robust technique that in all cases should be considered in the search of morphological evidence for classification purposes (often in combination with molecular studies) when complete data sets are available (Baker and Johnson, 2000; Feldesman, 2002; Mandáková and Münzbergová, 2008). Even with geometric morphometrics, multivariate analyses provide a good strategy for testing population differences (Viscosi and Cardini, 2012). With respect to the unsupervised DAs, the results would probably improve maintaining the sequential approach, but eliminating the restrictive rule of choosing the best variable in statistical terms and considering instead the set of the most useful variables with the help of graphic analysis (e.g. using box-plots).

### *4.3 A "divide and conquer" strategy applied to the polyploid complex V. subsect Pentasepalae*

The methodological approach followed in this study to search for diagnostic characters that could aid taxon classification is based on the "divide and conquer" principle. With this procedure, the most informative diagnostic characters were used to divide the initial dataset and progressively decrease its complexity. The methods that use recursive partitioning (i.e. splitting the initial task into various subtasks until they become simple enough to be easily solved) successfully address different kinds of intricate problems (Brinkmann et al., 1997; Hu et al., 2008; Bai et al., 2011). The combination of different techniques that are based on this principle (i.e. sequential DAs, DTs) seems to have been an excellent approach at least in the case of study. *Veronica* subsect. *Pentasepalae* is particularly challenging due not only to the high intraspecific morphological variability of some taxa, which makes species identification difficult, but also to the low number of populations known for several narrow endemics. However, this approach

reduced the initial complexity, generating smaller subsets of data and avoiding the loss of information concerning the OTUs with low numbers of observations. It is noteworthy that in this case, the DTs, which implement a recursive partitioning method (Breiman et al., 1984), provided satisfactory results. These recursive partitioning methodologies therefore seem to be reliable for assigning a population to taxa, either by conventional multivariate analysis techniques such as DAs, by implementing data-mining approaches, or by combining the two methods, which should not be considered mutually exclusive (Mendoza, 2007).

## 5. CONCLUSIONS

In summary, the present study used the following workflow: 1) If possible, consider only individuals/populations which can be identified by other sources of information. Take detailed measurements corresponding to all the morphometric characters that *a priori* show variability (leaf features in this case, but any organ should be considered, and if that organ has a threedimensional structure, geometric morphometrics should be considered as well). 2) Perform a PCoA or PCA (depending on the type of data) to verify whether there is enough variance present in the variables to explain the cases. 3) Implement a "divide and conquer" approach through the DT technique as a fast, easy, and effective solution in the search of diagnostic characters. In the case of species with low numbers of populations (or scarce data for any other reason), take advantage of the properties of the DAs to determine whether there is a sequence of characters that allows their classification aside from DTs and afterwards apply DTs without these entities. 4) Assess through ANNs the capacity of the variables to classify the taxa included in the final groups. Consider the variables selected as input layers and the taxa as output layers, divide the corresponding subset into several training/testing groups, and

calculate the misclassification rate. If the rates show consistently high values or the different results are too unstable, the search of other characters would be recommended.

Establishing a general protocol based on this particular example seems of course too bold. However, these methodological guidelines may be of use to find robust morphological characters to differentiate among closely related taxa which have been taxonomically recognized as different entities based on multiple lines of evidence. Morphological data has its limitations in that it can disagree with phylogenetic data or can be misinterpreted due to homoplasy. Thus, gathering as many data as possible about species or infraspecific taxa (i.e. genetic, cytological, biogeographic, etc.) appears to be the most appropriate way to achieve classification. Integrative taxonomy appears to be the most suitable way to inspect biodiversity, the selection of the most appropriate combination of characters to identify each group of organisms is crucial, and morphological features should not be ruled out (Dayrat, 2005; Padial et al., 2010).

## 6. APPLICATIONS

As mentioned above, each group of organisms has its particularities (and hopefully its appropriate solutions). There are scenarios in which the extreme morphological and ecological variation among species (Turgeon and Bernatchez, 2003) or the existence of cryptic taxa has never allowed the identification of diagnostic phenotypic characters (Bickford et al., 2007) (by definition no morphological character would be found in the latter example) and cases in which genetic approaches show promising results (Schander and Willassen, 2005; Sass et al., 2007).

In practical terms, conservationists cannot protect organisms that cannot be identified (Mace, 2004). Adequate knowledge and description are needed to develop the necessary plans and mechanisms for species conservation (Rojas, 1992; Samper, 2004). For

species complexes that are difficult to determine, it is recommended to perform careful morphometric studies on previously established taxa, which may allow finding robust characters in order to achieve proper identification. The adequate determination of endangered species and their distinctiveness with respect to their closest relatives is required, mainly when distribution areas are sympatric (Pillon et al., 2007; Pedersen et al., 2016). Studies based on the recursive partitioning or the "divide and conquer" principle are easily implementable to identification guides or even mobile apps (e.g., ArbolApp: http://www.arbolapp.es/; IPflanzen: http://www.ipflanzen.ch/; NatureGate: http://www.luontoportti.com/suomi/en/), which would increase the knowledge of species outside the academic sphere, thus facilitating their protection. The applications of these approaches may therefore facilitate the necessary dialogue with practitioners, communication that needs urgent improvement (Taylor et al., 2017), even in order to avoid the imminent extinction of taxonomists, an additional endangered species (Pearson et al., 2011).

**ACKNOWLEDGMENTS**

The authors wish to thank the curators of the following herbaria that provided material for this study: B, BC, BCF, BM, E, FCO, G, GDA, GDAC, JACA, K, MA, MAF, MGC, RNG, SALA, SEST, VAB, and VIT. We also wish to thank X. Giráldez, M. Santos-Vicente, N. Padilla-García, D. Albach and all other colleagues who helped to collect material for this study. Thank goes to Julio Peñas, Alexandru Badarau and Michael Kesl for their kindness and their valuable photographs. We also wish to thank two anonymous reviewers for their comments on a previous version of the manuscript.

## Author Contributions

*Conceptualization:* Noemí López-González, Santiago Andrés-Sánchez, Blanca M. Rojas Andrés, M. Montserrat Martínez-Ortega.

*Formal analysis:* Noemí López-González.

*Funding acquisition:* M. Montserrat Martínez-Ortega.

*Investigation:* Noemí López-González, Santiago Andrés-Sánchez.

*Methodology:* Noemí López-González.

*Resources:* Santiago Andrés-Sánchez, Blanca M. Rojas-Andrés, M. Montserrat Martínez Ortega.

*Supervision:* Santiago Andrés-Sánchez, Blanca M. Rojas-Andrés, M. Montserrat Martínez Ortega.

*Visualization:* Noemí López-González.

*Writing – original draft:* Noemí López-González.

*Writing – review & editing:* Santiago Andrés-Sánchez, Blanca M. Rojas-Andrés, M. Montserrat Martínez-Ortega.

## REFERENCES

**Abbott RJ, Lowe AJ.** 2004. Origins, establishment and evolution of new polyploid species: *Senecio cambrensis* and *S. eboracensis* in the British Isles. *Biological Journal of the Linnean Society*, 82: 467–474.

**Ackerfield J, Wen J**. 2002. A morphometric analysis of *Hedera* L. (the ivy genus, Araliaceae) and its taxonomic implications. *Adansonia*, 24: 197–212.

**Aitkenhead MJ.** 2008. A co-evolving decision tree classification method. *Expert System with Applications*, 34: 18–25.

**Amato F, López A, Peña-Méndez EM, Vaňhara P, Hampl A, Havel J.** 2013. Artificial neural networks in medical diagnosis. *Journal of Applied Biomedicine*, 11: 47–58.

**Andrade IM, Mayo SJ, Kirkup D, Van den Berg C**. 2008. Comparative morphology of populations of *Monstera* Adans. (Araceae) from natural forest fragments in Northeast Brazil using elliptic Fourier Analysis of leaf outlines. *Kew Bulletin*, 63: 193–211.

**Andrés-Sánchez S, Rico E, Herrero A, Santos-Vicente M, Martínez-Ortega MM.** 2009. Combining traditional morphometrics and molecular markers in cryptic taxa: towards an updated integrative taxonomic treatment for *Veronica* subgenus *Pentasepalae* (Plantaginaceae *sensu* APG II) in the western Mediterranean. *Botanical Journal of the Linnean Society*, 159: 68–87.

**Bai Y, Craiu RV, Di Narzo AF**. 2011. Divide and conquer: a mixture-based approach to regional adaptation for MCMC. *Journal of Computational and Graphical Statistics*, 20: 63–79.

**Baker MA, Johnson RA.** 2000. Morphometric Analysis of *Escobaria sneedii* var. *sneedii*, *E. sneedii* var. *leei*, and *E. guadalupensis* (Cactaceae). *Systematic Botaty*, 25: 577–587.

**Baker RSJd**. 2010. Data mining for education, in: Peterson P, Baker E, McGaw B, (eds.), International encyclopedia of education. Elsevier, Oxford, pp. 112–118.

**Balao F, Herrera J, Talavera S.** 2011. Phenotypic consequences of polyploidy and genome size at the microevolutionary scale: a multivariate morphological approach. *New Phytologist*, 192: 256–265.

**Barciová L, Macholán M.** 2009. Morphometric key for the discrimination of two wood mice species, *Apodemus sylvaticus* and *A. flavicollis*. *Acta Zoologica Academiae Scientiarum Hungaricae*, 55: 31–38.

**Bardy KE, Schönswetter P, Schneeweiss GM, Fischer MA, Albach DC.** 2011. Extensive gene flow blurs species boundaries among *Veronica barrelieri*, *V. orchidea* and *V. spicata* (Plantaginaceae) in southeastern Europe. *Taxon*, 60: 108–121.

**Basheer IA, Hajmeer M.** 2000. Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods*, 43: 3–31.

**Beck MW.** 2013. Visualizing neural networks in R (plot_nnet). Available at: https://gist.github.com/fawda123/7471137

**Beentje HJ.** 2010. The Kew plant glossary: an illustrated dictionary of plant terms. Royal Botanic Gardens, Kew.

**Bickford D, Lohman DJ, Sodhi NS, Peter KL, Meier R, Winker K, Ingram KK, Das I.** 2007. Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, 22: 148–155.

**Breiman L, Friedman JH, Olshen RA, Stone CJ.** 1984. Classification and Regression Trees. CRC Press, London.

**Breiman L.** 1996. Heuristics of instability and stabilization in model selection. *Annals of Statistics*, 24: 2350– 2383.

**Brinkmann G, Dress AW, Perrey SW, Stoye J.** 1997. Two applications of the Divide&Conquer principle in the molecular sciences. *Mathematical Programming*, 79: 71–97.

**Brown ML, Kros JF.** 2003. Data mining and the impact of missing data. *Industrial Management* & *Data Systems*, 103: 611–621.

**Chakaravarthy VT, Pandit V, Roy S, Awasthi P, Mohania M.** 2007. Decision trees for entity identification: Approximation algorithms and hardness results, in: Proceedings of the twenty-sixth ACM SIGMODSIGACT-SIGART symposium on Principles of database systems. ACM, New York, pp. 53–62.

**Chen X, Wang M, Zhang H.** 2011. The use of classification trees for bioinformatics. *WIREs Data Mining and Knowledge Discovery*, 1: 55–63.

**Clark JY, Corney DPA, Tang HL.** 2012. Automated plant identification using artificial neural networks, in: IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). San Diego, California: IEEE, pp. 343–348.

**Dayrat B.** 2005. Towards integrative taxonomy. *Biological Journal of the Linnean Society*, 85: 407–415.

**De Queiroz K**. 2005. Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences of the United States of America*, 102: 6600–6607.

**De Queiroz K**. 2007. Species concepts and species delimitation. *Systematic Biology*, 56: 879–886.

**Delen D, Walker G, Kadam A.** 2005. Predicting breast cancer survivability: a comparison of three data mining methods. *Artificial Intelligence in Medicine*, 34: 113–127.

**Depypere L, Chaerle P, Vander Mijnsbrugge K, Goetghebeur, P.** 2009. Classification trees and plant identification: a case study of European *Prunus* section *Prunus* taxa. *Belgian Journal of Botany*, 142: 163–176.

**Dobigny G, Baylac M, Denys C.** 2002. Geometric morphometrics, neural networks and diagnosis of sibling *Taterillus* species (Rodentia, Gerbillinae). *Biological Journal of the Linnean Society*, 77: 319–327.

**Dreiseitl S, Ohno-Machado L**. 2002. Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Informatics*, 35: 352–359.

**Dubey BP, Bhagwat SG, Shouche SP, Sainis JK.** 2006. Potential of artificial neural networks in varietal identification using morphometry of wheat grains. *Biosystems Engineering*, 95: 61–67.

**Fawcett JA, Van de Peer Y, Maere S**. 2013. Significance and biological consequences of polyploidization in land plant evolution. In: Greilhuber J, Dolezel J, Wendel JF, editors. Physical structure, behaviour and evolution of plant genomes. Vienna: Springer. pp. 277–293.

**Feldesman MR.** 2002. Classification trees as an alternative to linear discriminant analysis. *American Journal of Physical Anthropology*, 119: 257–275.

**Fischer CC, Tibbetts KJ, Morgan D, Ceder G.** 2006. Predicting crystal structure by merging data mining with quantum mechanics. *Nature Materials*, 5: 641–646.

**Fritsch S, Guenther F, following earlier work by Suling M.** 2012. neuralnet: Training of neural networks. R package version 1.33. Available at: https://cran.r-project.org/package=neuralnet

**Goloboff PA, Farris JS, Nixon KC.** 2008. TNT, a free program for phylogenetic analysis. *Cladistics*, 24: 774–786.

**Guyon I, Elisseeff A.** 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3: 1157–1182.

**Hall MA**. 2000. Correlation-based feature selection of discrete and numeric class machine learning, in: Proceedings of the Seventeenth International Conference on Machine Learning (ICML'00). Morgan Kaufmann Publishers, San Francisco, pp. 359–366.

**Hassoun MH.** 1995. Fundamentals of artificial neural networks. MIT press, Cambridge (Massachusetts).

**Henderson A.** 2006. Traditional morphometrics in plant systematics and its role in palm systematics. *Botanical Journal of the Linnean Society*, 151: 103–111.

**Hörandl E, Greilhuber J, Klímová K, Paun O, Temsch E, Emadzade K.** 2009. Reticulate evolution and taxonomic concepts in the *Ranunculus auricomus* complex (Ranunculaceae): insights from analysis of morphological, karyological and molecular data. *Taxon*, 58: 1194–1216.

**Hu W, Qu Y, Cheng G.** 2008. Matching large ontologies: A "divide and conquer" approach. *Data & Knowledge Engineering*, 67: 140–160.

**Huang ML, Hsu YY.** 2012. Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network. *Journal of Biomedical Science and Engineering*, 5: 525–533.

**Jensen RJ, Ciofani KM, Miramontes LC.** 2002. Lines, outlines, and landmarks: morphometric analyses of leaves of *Acer rubrum*, *Acer saccharinum* (Aceraceae) and their hybrid. *Taxon*, 51: 475–492.

**Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, …, dePamphilis CW**. 2002. Ancestral polyploidy in seed plants and angiosperms. *Nature*, 473: 97–100.

**Joly S, Bruneau A.** 2007. Delimiting species boundaries in *Rosa* sect. *Cinnamomeae* (Rosaceae) in eastern North America. *Systematic Botany*, 32: 819–836.

**Kafkas S, Perl-Treves R.** 2001. Morphological and molecular phylogeny of *Pistacia* species in Turkey. *Theoretical and Applied Genetics*, 102: 908–915.

**Karels TJ, Bryant AA, Hik DS.** 2004. Comparison of discriminant function and classification tree analyses for age classification of marmots. *Oikos*, 105: 575–587.

**Klecka WR.** 1980. Discriminant analysis. Sage, Thousand Oaks.

**Koutecký P**. 2007. Morphological and ploidy level variation of *Centaurea phrygia* agg. (Asteraceae) in the Czech Republic, Slovakia and Ukraine. *Folia Geobotanica*, 42: 77–102.

**Kurt I, Ture M, Kurum AT.** 2008. Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. *Expert Systems with Applications*, 34: 366–374.

**Lambert SM, Borba EL, Machado MC.** 2006. Allozyme diversity and morphometrics of the endangered *Melocactus glaucescens* (Cactaceae), and investigation of the putative hybrid origin of *Melocactus* × *albicephalus* (*Melocactus ernestii* × *M. glaucescens*) in north-eastern Brazil. *Plant Species Biology*, 21: 93–108.

**Lavergne S, Muenke NJ, Molofsky J.** 2010. Genome size reduction can trigger rapid phenotypic evolution in invasive plants. *Annals of Botany*, 105: 109–116.

**Levin DA**. 1983. Polyploidy and novelty in flowering plants. *The American Naturalist*, 122: 1–25.

**Li X, Yu E, Fan C, Zhang C, Fu T, Zhou Y.** 2012. Developmental, cytological and transcriptional analysis of autotetraploid *Arabidopsis*. *Planta*, 236: 579–596.

**Lindbladh M, O'Connor R, Jacobson GL.** 2002. Morphometric analysis of pollen grains for paleoecological studies: classification of *Picea* from eastern North America. *American Journal of Botany*, 89: 1459–1467.

**Liu K, Raghavan S, Nelesen S, Linder CR, Warnow T.** 2009. Rapid and accurate large-scale coestimation of sequence alignments and phylogenetic trees. *Science*, 324: 1561–1564.

**Liu K, Warnow TJ, Holder MT, Nelesen SM, Yu J, Stamatakis AP, Linder CR.** 2011. SATe-II: very fast and accurate simultaneous estimation of multiple sequence alignments and phylogenetic trees. *Systematic Biology*, 61: 90–106.

**Lorenz C, Ferraudo AS, Suesdek L.** 2015. Artificial Neural Network applied as a methodology of mosquito species identification. *Acta Tropica*, 152: 165–169.

**Lorenz C, Marques TC, Sallum MAM, Suesdek L**. 2014. Altitudinal population structure and microevolution of the malaria vector *Anopheles cruzii*. *Parasite Vectors*, 7: 581.

**Mace GM.** 2004. The role of taxonomy in species conservation. *Philosophical Transactions of The Royal Society B: Biological Sciences*, 359: 711–719.

**Maier HR, Dandy GC**. 2000. Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. *Environmental Modelling and Software*, 15: 101–124.

**Maimon O, Rokach L.** 2010. Data mining and knowledge discovery handbook. 2nd ed. Springer, New York.

**Mallet J**. 2005. Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, 20: 229–237.

**Mandáková T, Münzbergová Z.** 2008. Morphometric and genetic differentiation of diploid and hexaploid populations of *Aster amellus* agg. in a contact zone. *Plant Systematics and Evolution*, 274: 155.

**Manel S, Dias JM, Ormerod SJ.** 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling*, 120: 337–347.

**Marhold K.** 2011. Multivariate morphometrics and its application to monography at specific and infraspecific levels, in: Stuessy TF, Lack HW, (eds.), Monographic plant systematics: fundamental assessment of plant biodiversity. Ruggell, Gantner, pp. 73–99.

**Martínez Ortega MM, Sánchez Agudo JÁ, Rico E**. 2009. *Veronica* L., in. Benedó C, Rico E, Güemes J, Herrero A (eds.), Flora ibérica v. XIII Plantaginaceae-Scrophulariaceae. Real Jardín Botánico-CSIC, Madrid, pp. 360–434.

**Martínez-Ortega MM, Delgado L, Albach D, Rico E, Elena-Rossello JA.** 2004. Species boundaries and phylogeographic patterns in cryptic taxa inferred from AFLP markers: *Veronica* subgen. *Pentasepalae* (Scrophulariaceae) in the Western Mediterranean. *Systematic Botany*, 29: 965–986.

**May L, Lacourse T.** 2012. Morphological differentiation of *Alnus* (alder) pollen from western North America. *Review of Palaeobotany and Palynology*, 180: 15–24.

**Mendoza M.** 2007. Decision trees: a machine learning methodology for characterizing morphological patterns resulting from ecological adaptations, in: MacLeod N, (ed.), Automated taxon identification in systematics: Theory, approaches and applications. CRC Press, Florida, pp. 261–276.

**Michel JB, Shen YK, Aiden AP, Veres A, Gray MK, The Google Books Team, Pickett JP, Hoiberg D, …, Aiden EL.** 2011. Quantitative analysis of culture using millions of digitized books. *Science*, 331: 176–182.

**Muir G, Fleming CC, Schlötterer C.** 2000. Species status of hybridizing oaks. *Nature*, 405: 6790.

**Olden JD, Joy MK, Death RG.** 2004. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling*, 178: 389–397.

*ovatus* (Compositae, Senecioneae) along an altitudinal gradient in Harz National Park (Germany).

**Padial JM, Miralles A, De la Riva I, Vences M.** 2010. The integrative future of taxonomy. *Frontiers in Zoology*, 7: 16.

**Padilla-García N, Rojas-Andrés BM, López-González N, Castro M, Castro S, Loureiro J, Albach DC, Machon N, Martínez-Ortega MM.** 2018. The challenge of species delimitation in the diploid-polyploid complex *Veronica* subsection *Pentasepalae*. *Molecular Phylogenetics & Evolution*, 119: 196–209.

**Paliwal M, Kumar UA.** 2009. Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications*, 36: 2–17.

**Pandolfi C, Mugnai S, Azzarello E, Bergamasco S, Masi E, Mancuso S.** 2009. Artificial neural networks as a tool for plant identification: a case study on Vietnamese tea accessions. *Euphytica*, 166: 411– 421.

**Parsons S, Jones G.** 2000. Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks. *The Journal of Experimental Biology*, 203: 2641–2656.

**Pearson DL, Hamilton AL, Erwin TL.** 2011. Recovery plan for the endangered taxonomy profession. *BioScience*, 61: 58–63.

**Pedersen HÆ, Srimuang KO, Watthana S.** 2016. Strengthening the taxonomic backbone of Thai orchid conservation: genetic fingerprinting and morphometry applied to a species complex in *Geodorum*. *Annals of Botany*, 118: 125–133.

**Peev D.** 1995. Veronica L., in: Kozuharov SI, Kuzmanov BA, (eds.), Flora Reipublicae Bulgaricae. Editio Academica "Professor Marin Drinov", Sofia, pp. 142–189.

**Peñas De Giles J, Martínez-Ortega MM, Pérez LaTorre AV, Cabezudo Artero B.** 2004. *Veronica tenuifolia* subsp. *fontqueri* (Pau) M. M. Mart. Ort. and E. Rico, in: Bañares A, Blanca G, Güemes J, Moreno JC, Ortiz S, (eds.), Atlas y Libro Rojo de la Flora Vascular Amenazada de España. Madrid: Dirección General de Conservación de la Naturaleza, Madrid, pp. 564–565.

**Petrova A, Vladimirov V.** 2009. Red List of Bulgarian vascular plants. *Phytologia Balcanica*, 15: 63–94.

**Pillon Y, Qamaruz-Zaman F, Fay MF, Hendoux F, Piquot Y.** 2007.Genetic diversity and ecological differentiation in the endangered fen orchid (*Liparis loeselii*). *Conservation Genetics*, 8: 177–184.

**Plotze RDO, Falvo M, Pádua JG, Bernacci LC, Vieira MLC, Oliveira GCX, Martínez Bruno O.** 2005. Leaf shape analysis using the multiscale Minkowski fractal dimension, a new morphometric method: a study with *Passiflora* (Passifloraceae). *Canadian Journal of Botany*, 83: 287–301.

**Quinlan JR.** 1985. Introduction to decision trees. *Machine Learning*, 1: 81–106.

**R Core Team.** 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna: r-project.org.

**Ramsey J, Ramsey TS.** 2014. Ecological studies of polyploidy in the 100 years following its discovery. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369: 20130352.

**Raudnitschka D, Hensen I, Oberprieler C**. 2007. Introgressive hybridization of *Senecio hercynicus* and *S*.

**Razi MA, Athappilly K.** 2005. A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. *Expert Systems with Applications*, 29: 65–74.

**Reichenbacher B, Kamrani E, Esmaeili HR, Teimori A.** 2009. The endangered cyprinodont *Aphanius ginaonis* (Holly, 1929) from southern Iran is a valid species: evidence from otolith morphology. *Environmental Biology of Fishes*, 86: 507.

**Riedmiller M.** 1994. Rprop-Description and implementation details: technical report. Institut für Logik, Komplexität und Deduktionssysteme, Karlsruhe.

**Rieseberg LH.** 1997. Hybrid origins of plant species. *Annual Review of Ecology, Evolution, and Systematics*, 1: 359–389.

**Ripley B.** 2015. tree: Classification and Regression Trees. R package version 1.0–36. Available at: https://cran.r-project.org/package=tree

**Rodriguez Galdon B, Penña-Meéndez E, Havel J, Rodriéguez Rodriéguez EM, Diaz Romero C.** 2010. Cluster analysis and artificial neural networks multivariate classification of onion varieties. *Journal of Agricultural and Food Chemistry*, 58: 11435–11440.

**Rohlf FJ**. 2006. NTSYS-pc Numeral Taxonomy and Multivariate Analysis System, version 2.21. Exeter Software, New York.

**Rojas M.** 1992. The species problem and conservation: what are we protecting? *Conservation Biology*, 6: 170–178.

**Rojas-Andrés BM, Albach DA, Martínez-Ortega MM.** 2015. Exploring the intricate evolutionary history of the diploid–polyploid complex *Veronica* subsection *Pentasepalae* (Plantaginaceae). *Botanical Journal of the Linnean Society*, 179: 670–692.

**Rojas-Andrés BM, Martínez-Ortega MM.** 2016. Taxonomic revision of *Veronica* subsection *Pentasepalae* (*Veronica*, Plantaginaceae sensu APG III). *Phytotaxa*, 285: 1–100.

Samper C. Taxonomy and environmental policy. 2004. *Philosophical Transactions of The Royal Society B: Biological Sciences*, 359: 721–728.

**Sass C, Little DP Stevenson DW, Specht CD.** 2007. DNA barcoding in the cycadales: testing the potential of proposed barcoding markers for species identification of cycads. *PloS One*, 2: e1154.

**Schander C, Willassen E.** 2005. What can biological barcoding do for marine biology? *Marine Biological Research*, 1: 79–83.

**Seehausen O.** 2004. Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, 19: 198–207.

**Soltis DE, Visger CJ, Marchant DM, Soltis PS.** 2016. Polyploidy: Pitfalls and paths to a paradigm. *American Journal of Botany*, 103: 1146–1166.

**Soltis PS, Marchant DB, Van de Peer Y, Soltis DE**. 2015. Polyploidy and genome evolution in plants. *Current Opinion in Genetics & Development*, 35: 119–125.

**Stace CA.** 2000. Cytology and cytogenetics as a fundamental taxonomic resource for the 20th and 21st centuries. *Taxon*, 49: 451–477.

**Steinberg D, Colla P.** 1995. CART: tree-structured non-parametric data analysis. Salford Systems, San Diego.

**Strelcov E, Belianinov A, Hsieh YH, Jesse S, Baddorf AP, Chu YH, Kalinin SV.** 2004. Deep data analysis of conductive phenomena on complex oxide interfaces: physics from data mining. *ACS Nano*, 8: 6449–6457.

**Suehs CM, Affre L, Médail F**. 2004. Invasion dynamics of two alien *Carpobrotus* (Aizoaceae) taxa on a Mediterranean island: I. Genetic diversity and introgression. *Heredity*, 92: 31–40.

**Sugumaran V, Muralidharan V, Ramachandran KI.** 2007. Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. *Mechanical Systems and Signal Processing*, 21: 930–942.

**Sutton DC.** 2005. Classification and regression trees, bagging and boosting, in: Rao CR, Wegman EJ, Solka JL, (eds.), Handbook of Statics. Elsevier, Amsterdam, pp. 303–329.

*Systematic Biodiversity*, 5: 333–344.

**Taylor HR, Dussex N, van Heezik Y.** 2017. Bridging the conservation genetics gap by identifying barriers to implementation for conservation practitioners. *Global Ecology and Conservation*, 10: 231–242.

**Thiers B.** Index Herbariorum: A global directory of public herbaria and associated staff. New York Botanical Garden's Virtual Herbarium. [Continuosly updated]. Available at: http://sweetdum.nybg.org/ ih/

**Tison JM, Foucault B**. 2014. Flora Gallica–Flore de France. Biotope Editions, Mèze.

**Tovar-Sánchez E, Oyama K**. 2004. Natural hybridization and hybrid zones between *Quercus crassifolia* and *Quercus crassipes* (Fagaceae) in Mexico: Morphological and molecular evidence. *American Journal of Botany*, 91: 1352–1663.

**Tso GK, Yau KK**. 2007. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, 32: 1761–1768.

**Turgeon J, Bernatchez L.** 2003. Reticulate evolution and phenotypic diversity in North American ciscoes, *Coregonus* ssp. (Teleostei: *Salmonidae*): implications for the conservation of an evolutionary legacy. *Conservation Genetics*, 4: 67–81.

**Viscosi V, Cardini A.** 2012. Leaf Morphology, Taxonomy and Geometric Morphometrics: A Simplified Protocol for Beginners. *PloS One*, 6: e25630.

**Walters SM, Webb DA**. 1972. *Veronica* L., in: Tutin TG, Heywood VH, Burges NA (eds.), Flora Europaea. Cambridge University Press, Cambridge, pp. 242–251.

**Watzl B**. 1910. *Veronica prostrata* L., *teucrium* L. und *austriaca* L. Nebst einem anhange Èber deren na Èchsteverwante. *Abhandlungen der kaiserlich-königlichen zoologisch-botanischen Gesellschaft in Wien*, 5:1–94.

**Wickham H.** 2010. ggplot2: elegant graphics for data analysis. *Journal of Statistical Software*, 35: 65–88.

**Zozomova-Lihova J, Malanova-Krasna I, Vit P, Urfus T, Senko D, Svitok M, Kempa M, Marhold K.** 2015. Cytotype distribution patterns, ecological differentiation, and genetic structure in a diploid–tetraploid contact zone of *Cardamine amara*. *American Journal of Botany*, 102: 1380–1395.

## SUPPORTING INFORMATION

Supporting information includes:

**S1 Table. Voucher information.** Voucher information for the *Veronica* samples used in this study.

**S1 Fig. Example of the architecture of an artificial neural network.** (I) input layers = 15; (H) hidden layers = 1; number of neurons = 8; (O) output layers = 8. Output layers correspond to taxa (see Table 1 for abbreviations), input layers correspond to variables (see Table 2 for abbreviations). Positive and negative connections are represented by black and grey lines, respectively. Line width indicates the strength of the connection.

**S2 Fig. MCR vs. number of species.** Misclassification rate in relation to the number of output layers (i.e., number of species and subspecies). Each point represents a different combination of randomly chosen OTUs.

**S3 Fig. Perfect tree.**

Supporting information not included (available online at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0199818#sec025):

**S2 Table. Guided DA information.** Discriminant functions, standard coefficients and structure matrix tables for the sequential guided discriminant analysis. (XLSX)

**S3 Table. Systematic DA information.** Discriminant functions, standard coefficients and structure matrix tables for the systematic unsupervised discriminant analysis. (XLSX)

**S1 File. Box-plots corresponding to terminal groups (I to VIII) established through guided DAs.** The black line in each box-plot indicates the threshold applied to perform the division. (DOCX)

**S2 File. Box-plots showing the divisions performed with the most discriminant characters found through unsupervised DAs.** The black line in each box-plot indicates the threshold that minimizes the misclassification rate. (DOCX)

**S1 Table. Voucher information.** Voucher information for the *Veronica* samples used in this study.

| Accession number[a] (Collectors[b], date) | Species | Collection country and locality (Latitude, Longitude) | No. of individuals |
|---|---|---|---|
| BCF41230 (unknown collector, VI-1947) | *V. aragonensis* | Spain. Huesca, Guara. (42.28142N, 00.14620W) | 3 |
| GDA18086 (Negrillo, 2-VII-1983) | *V. aragonensis* | Spain. Granada, La Sagra mountains. (37.94617N, 02.56752W) | 3 |
| JACA256379 (Montserrat, 22-VII-1979) | *V. aragonensis* | Spain. Huesca, La Estiva. (42.47367N, 00.26289E) | 3 |
| JACA349179 (Villar, 7-VIII-1979) | *V. aragonensis* | Spain. Huesca, Yésero-Gavín. (42.67128N, 00.24202W) | 3 |
| JACA473871 (Montserrat, 15-VII-1971) | *V. aragonensis* | Spain. Huesca, Ordesa. (42.65647N, 00.01089E) | 3 |
| JACA52774 (Montserrat, 12-V-1974) | *V. aragonensis* | Spain. Huesca, Bentué de Rasal. (42.33473N, 00.48740W) | 3 |
| JACA665787 (Sesé & Montserrat, 4-VII-1988) | *V. aragonensis* | Spain. Huesca, Montañeta de Gabas. (42.45913N, 00.46255E) | 3 |
| SALA153002 (Rojas-Andrés et al., 17-VII-2014) | *V. austriaca* subsp. *austriaca* | Slovakia. Spisšká Nová Ves. Letanovce. P.N. Slovensky raj. (48.95525N, 20.43775E) | 3 |
| SALA157057 (Martínez-Ortega et al., 16-VII-2014) | *V. austriaca* subsp. *austriaca* | Slovakia. Muráň, P.N. Muránska Planina. (48.75112N, 20.04861E) | 3 |
| SALA157051 (Martínez-Ortega et al., 20-VII-2014) | *V. austriaca* subsp. *austriaca* | Rumania. Valea Lungă, Tauri monastery. (46.14058N, 24.05728E) | 3 |
| SALA155836 (López-González et al., 18-VII-2014) | *V. austriaca* subsp. *austriaca* | Slovakia. Michalovce, Vinné, Viniansky Hrad. (48.81709N, 21.94442E) | 3 |
| SALA157058 (López-González et al., 18-VII-2014) | *V. austriaca* subsp. *austriaca* | Slovakia. Černochov. (48.43022N, 21.70586E) | 3 |
| B147196-138 (Bartha, 20-V-1929) | *V. austriaca* subsp. *dentata* | Hungary. Budapest, Ruppehegy mountain. (47.46330N, 18.95022E) | 3 |
| B147196-23 (Bässler et al., 6-VI-1967) | *V. austriaca* subsp. *dentata* | Hungary. Bükk mountains in direction to Ómassa. (48.10868N, 20.52962E) | 3 |
| B147196-6 (Preissmann, 19-V-1887) | *V. austriaca* subsp. *dentata* | Austria. Gösting in Grar. (47.10000N, 15.38333E) | 3 |
| BM67715 (Krebs, VII-1902) | *V. austriaca* subsp. *dentata* | Austria. Perchstoldorf. (48.12044N, 16.26894E) | 3 |

| Accession number[a] (Collectors[b], date) | Species | Collection country and locality (Latitude, Longitude) | No. of individuals |
|---|---|---|---|
| BM68162 (Sillinger, V-1930) | *V. austriaca* subsp. *dentata* | Czech Republic. Distr. Beroun, Doutnác hill near the village of Srbsko. (49.95078N, 14.14976E) | 3 |
| BM68170 (Braun, date unknown) | *V. austriaca* subsp. *dentata* | Austria. Laxenburg. (48.06865N, 16.35624E) | 3 |
| BM68180 (Sychowa & Jasiewicz, 17-V-1968) | *V. austriaca* subsp. *dentata* | Poland. Distr. Miechów. Kalina Mala. (50.36648N, 20.11656E) | 3 |
| BM68606 (Weiss,16-VI-1883) | *V. austriaca* subsp. *dentata* | Germany. Bavaria, München, Garching. (48.24883N, 11.65112E) | 3 |
| MA112519 (Skrenbergen, 18-V-1876) | *V. austriaca* subsp. *dentata* | Germany. Between Pebüsch and Calvarienberg. (49.26333N, 9.148519E) | 3 |
| MA112520 (Laus, VI-1939) | *V. austriaca* subsp. *dentata* | Czech Republic. Olmütz. (49.57155N, 17.17745E) | 3 |
| SALA124613 (Martínez-Ortega, 15-V-2001) | *V. austriaca* subsp. *dentata* | Austria. Niederösterreich, NW Alpenostrand. (48.04522N, 16.26287E) | 3 |
| SALA124612 (Martínez-Ortega, 3-VI-2001) | *V. austriaca* subsp. *dentata* | Slovakia. Bratislava, Thebener Kogel. (48.18926N, 16.99543E) | 3 |
| SALA124579 (Albach, 12-VI-2001) | *V. austriaca* subsp. *jacquinii* | Bulgaria. Central Rhodopes, Asenograv. (41.89500N, 29.92944E) | 3 |
| SALA124576 (Albach, 17-VI-2001) | *V. austriaca* subsp. *jacquinii* | Bulgaria. Vitosha mountain, mid-station of góndola. (42.56724N, 23.28578E) | 3 |
| SALA124581 (Albach, 17-VI-2001) | *V. austriaca* subsp. *jacquinii* | Bulgaria. Vitosha mountain, near point where road comes closest to góndola. (42.58861N, 23.31194E) | 3 |
| SALA124582 (Albach, 20-VI-2001) | *V. austriaca* subsp. *jacquinii* | Bulgaria. Vitosha mountain. Ljulimin, beneath ski slope. (42.56724N, 23.28578E) | 3 |
| SALA124583 (Albach, 20-VI-2001) | *V. austriaca* subsp. *jacquinii* | Bulgaria. Vitosha mountain. Ljulimin, beneath ski slope. (42.56724N, 23.28578E) | 3 |
| SALA124584 (Albach, 20-VI-2001) | *V. austriaca* subsp. *jacquinii* | Bulgaria. Eastern Stara mountain, 2 km south of Gindzi. (42.43278N 25.64194E) | 3 |
| E32508 (Bujorean, 22-V-1922) | *V. austriaca* subsp. *jacquinii* | Romania. Banatus, distr. Caras-Severin. (45.08333N 22.08333E) | 3 |
| K (Cook et al., 21-VII-1959) | *V. austriaca* subsp. *jacquinii* | Kosovo. Metohija, Rusolija Massif. (42.75056N 20.24528E) | 3 |
| K (Rechinger & Scheffer, 26-VII-1933) | *V. austriaca* subsp. *jacquinii* | Albania. Bertiscus, Greben mountain. (42.55750N, 19.69556E) | 2 |
| SALA124610 (Martínez-Ortega, 18-V-2001) | *V. austriaca* subsp. *jacquinii* | Georgia. Kakheti, Shirakis vake, between Kvemo Bodve and Gamardzhveba. (41.53528N, 45.90500E) | 3 |

| Accession number[a] (Collectors[b], date) | Species | Collection country and locality (Latitude, Longitude) | No. of individuals |
|---|---|---|---|
| SALA124607 (Martínez-Ortega & Tribsch, 16-VI-2001) | *V. austriaca* subsp. *jacquinii* | Croatia. Velika Kapela, Gorska Kotar, P.N. Risnjak. (45.42083N, 14.56306E) | 2 |
| SALA124604 (Martínez-Ortega & Tribsch, 16-VI-2001) | *V. austriaca* subsp. *jacquinii* | Croatia. Velika Kapela, Gornje Jelenje, in the crossroad to Crikvanica. (45.36583N, 14.61639E) | 3 |
| SALA124605 (Martínez-Ortega & Tribsch, 16-VI-2001) | *V. austriaca* subsp. *jacquinii* | Croatia. Between Mrkopajl and Begovo Razdolje. (45.30361N, 14.89750E) | 3 |
| SALA124603 (Martínez-Ortega & Tribsch, 17-VI-2001) | *V. austriaca* subsp. *jacquinii* | Croatia. Opatija, road to Ljubljana between Rupa and Lipa. (45.47833N, 14.28639E) | 3 |
| SALA124609 (Martínez-Ortega & Tribsch, 17-VI-2001) | *V. austriaca* subsp. *jacquinii* | Croatia. Licausko, Polje. (45.13667N, 14.64222E) | 3 |
| SALA124606 (Martínez-Ortega, 18-VI-2001) | *V. austriaca* subsp. *jacquinii* | Croatia. Velebit, Mali Alan pass, between Obrovac and Sveti Rok. (44.20056N, 15.68222E) | 3 |
| SALA124608 (Martínez-Ortega et al., 4-VII-2002) | *V. austriaca* subsp. *jacquinii* | Slovakia. Cernochov, Vychodoslovenska lowlands. (48.43750N, 21.73833E) | 3 |
| SALA124564 (Schneeweiss et al., 21-VI-2001) | *V. austriaca* subsp. *jacquinii* | Croatia. Zadar, Velebit, N. P. Paklenica. (44.34040N, 15.47480E) | 3 |
| RNG (Stamatiadou, 13-VI-1971) | *V. austriaca* subsp. *jacquinii* | Greece. Sterea Ellas, distr. Fthiotis, Othris mountain. SE side of Jerakovouni. (39.03333N, 22.61667E) | 3 |
| SALA149288 (Rojas-Andrés et al., 26-VII-2010) | *V. crinita* | Serbia. Between Zlot and Brestovac. (44.02367N, 21.99328E) | 3 |
| SALA149289 (Rojas-Andrés et al., 23-VI-2009) | *V. crinita* | Turkey. Dereköy, road to Geçitagzi. (41.93761N, 27.35306E) | 3 |
| SALA149038 (Santos et al., 18-VI-2009) | *V. crinita* | Bulgaria. Around Popovitsa. (42.12294N, 25.07567E) | 3 |
| SALA149290 (Santos et al., 21-VI-2009) | *V. crinita* | Bulgaria. Around Veliki Preslav. (43.13978N, 26.80806E) | 3 |
| SALA149037 (Santos et al., 22-VI-2009) | *V. crinita* | Bulgaria. Varna, between Vinitsa and Aladza monastery. (43.25097N, 28.00242E) | 3 |
| SALA149244 (Frajman & Schönswetter, 23-V-2009) | *V. crinita* | Bosnia-Herzegovina. Ravan mountain. Summit of Mt. Tajan. (44.28861N, 18.19028E) | 3 |
| PRC (Fiala, 1892) | *V. crinita* | Bosnia-Herzegovina. Klek mountain close to Foça. (45.26083N, 15.14528E) | 3 |
| PRC (Fiala, 1892) | *V. crinita* | Bosnia-Herzegovina. Klek mountain. (45.26083N, 15.14528E) | 1 |

| Accession number[a] (Collectors[b], date) | Species | Collection country and locality (Latitude, Longitude) | No. of individuals |
|---|---|---|---|
| PRC455004 (Beck, date unknown) | *V. crinita* | Bosnia-Herzegovina. Stražica in Vranica mountains. (43.95417N, 17.7711E) | 3 |
| SALA157013 (Rojas-Andrés et al., 9-VII-2014) | *V. kindlii* | Greece. Above Boras ski resort. (40.91902N, 21.80960E) | 3 |
| SALA149277 (Martínez-Ortega et al.,19-VII-2010) | *V. kindlii* | Montenegro. Cakor, next to Kosovo border. (42.66072N, 19.99247E) | 3 |
| SALA149278 (Martínez-Ortega et al., 25-VII-2010) | *V. kindlii* | FYROM. Gevgelija, in the way to Kozuf ski resort. (41.20006N, 22.24369E) | 3 |
| SALA157011 (Martínez-Ortega et al., 8-VII-2014) | *V. kindlii* | FYROM. P. N. Pelister, Pelister mountain. (40.99456N, 21.17847E) | 3 |
| SALA157012 (Martínez-Ortega et al., 8-VII-2014) | *V. kindlii* | FYROM. P. N. Pelister, Pelister mountain. (40.99456N, 21.17847E) | 3 |
| SALA149346 (Santos et al., 25-VI-2009) | *V. kindlii* | Greece. Vermion mountain. (40.53606N, 22.02403E) | 3 |
| SALA149279 (Santos et al., 27-VI-2009) | *V. kindlii* | Greece. In the ascent to Kaïmaktsalán mountain. (40.90708N, 21.82561E) | 3 |
| SALA149282 (Santos et al., 27-VI-2009) | *V. kindlii* | Greece. In the ascent to Kaïmaktsalán mountain. (40.90708N, 21.82561E) | 3 |
| DR041947 (Behr, 1937) | *V. linearis* | FYROM. Sivec at Prilep. (41.40944N, 21.59194E) | 3 |
| SALA153001 (Martínez-Ortega et al., 11-VII-2014) | *V. linearis* | FYROM. Between Modrište and Zdunje, next to Jezero Kozjak. (41.72449N, 21.19320E) | 3 |
| B147196-192 (Maly, VI-1909) | *V. orbiculata* | Bosnia-Herzegovina. Sarajevo, below Mrkovic. (43.84864N, 18.35644E) | 3 |
| BM68701 (Maly, 26-V-1907) | *V. orbiculata* | Bosnia-Herzegovina. Sarajevo. (43.84864N, 18.35644E) | 3 |
| SALA149336 (Rojas-Andrés et al., 11-VII-2010) | *V. orbiculata* | Bosnia-Herzegovina. Mostar, in the ascent to Hum mountain. (43.32728N, 17.79939E) | 3 |
| SALA149294 (Rojas-Andrés et al., 14-VII-2010) | *V. orbiculata* | Croatia. Peljesak Peninsula, between Trstenik and Pijavicino. (42.93728N, 17.37764E) | 3 |
| K (Raap, 1895) | *V. orbiculata* | Bosnia-Herzegovina. Mostar. (43.34333N, 17.80806E) | 3 |
| SALA124585 (Martínez-Ortega, 19-VI-2001) | *V. orbiculata* | Croatia. Brela, between Omis and Makarska. (43.33000N, 17.07000E) | 3 |
| SALA124587 (Martínez-Ortega, 20-VI-2001) | *V. orbiculata* | Croatia. Makarska, between Tucepi and G. Igrane. (43.23222N, 17.14361E) | 3 |
| SALA124586 (Martínez-Ortega & Solic, 20-VI-2001) | *V. orbiculata* | Croatia. Makarska, Osejaba. (43.29694N, 17.01778E) | 3 |

| Accession number[a] (Collectors[b], date) | Species | Collection country and locality (Latitude, Longitude) | No. of individuals |
|---|---|---|---|
| SALA149337 (Martínez-Ortega et al., 10-VI-2010) | *V. orbiculata* | Croatia. Brela, crossroad towards Gornja Brela. (43.40383N, 16.89364E) | 3 |
| SALA149295 (Martínez-Ortega et al., 10-VI-2010) | *V. orbiculata* | Croatia. Prapatnice, in the ascent to Matokit. (43.22114N, 17.35972E) | 3 |
| RNG (Halliday, 15-IV-1972) | *V. orbiculata* | Montenegro. Budva, behind Podostrog monastery. (42.28639N, 18.84000E) | 3 |
| SALA149297 (Rico et al., 6-VII-2002) | *V. orsiniana* | Italy. P. N. d'Abruzzo, La Majella. (42.07556N, 14.09694E) | 3 |
| SALA149298 (Rico et al., 6-VII-2002) | *V. orsiniana* | Italy. P. N. d'Abruzzo, La Majella. (42.07556N, 14.09694E) | 3 |
| CAME (Conti, 30-V-1995) | *V. orsiniana* | Italy. P. N. d'Abruzzo,Colle Biferno. (41.93794N, 13.73200E) | 3 |
| SALA110621 (Delgado et al., 11-V-1999) | *V. orsiniana* | Spain. Riudellots de la Selva. (41.89500N, 02.80694E) | 3 |
| SALA110622 (Delgado et al., 9-V-1999) | *V. orsiniana* | Spain. Malla, Alto del Closcar. (41.88152N, 02.21663E) | 3 |
| SALA110626 (Delgado et al., 11-V-1999) | *V. orsiniana* | Spain. Montserrat, Ermita de San Miguel. (41.58667N, 01.84139E) | 3 |
| MA180274 (Gavelle, 28-VI-1957) | *V. orsiniana* | France. Nice, Gros mountain. (43.72916N, 07.29851E) | 3 |
| MA425880 (Tallon, 14-V-1931) | *V. orsiniana* | France. Gard, Pont du Gard. (43.94683N, 04.53606E) | 3 |
| SALA93492 (Martínez-Ortega & Martín Ballesteros, 20-VI-1996) | *V. orsiniana* | Spain. Huesca, Sobrarbe, Chisagüés valley. (42.66489N, 00.18595W) | 3 |
| SALA93483 (Martínez-Ortega & Delgado, 8-VII-1997) | *V. orsiniana* | Spain. Huesca, Borau, As Blancas. (42.67833N, 00.570000W) | 3 |
| SALA93482 (Martínez-Ortega & Delgado, 10-VII-1997) | *V. orsiniana* | Spain. From Torla to Bujaruelo, Los Navarros bridge. (42.65333N, 00.102780W) | 3 |
| SALA110623 (Martínez-Ortega et al., 10-V-1999) | *V. orsiniana* | Spain. Barcelona, Tona, Virgen de Lourdes sanctuary. (41.84549N, 02.217070W) | 3 |
| SALA124591 (Martínez-Ortega et al., 10-V-1999) | *V. orsiniana* | Spain. Santa Cecilia de Voltregá, km 3 in the road to Manlleu. (41.98950N, 02.203240E) | 3 |
| SALA110624 (Martínez-Ortega et al., 10-V-1999) | *V. orsiniana* | Spain. Road from Vic to Villadrau. (41.84639N, 02.36160E) | 3 |
| SALA110625 (Martínez-Ortega et al., 12-V-1999) | *V. orsiniana* | Spain. El Miracle, road Cardona-El Miracle. (41.910640N, 01.51697E) | 3 |
| SALA124590 (Martínez-Ortega & Delgado, 11-VI-1999) | *V. orsiniana* | Spain. Luesia, Luesia mountains. (42.40702N, 00.99497W) | 3 |
| SALA110619 (Martínez-Ortega & Delgado, 13-VI-1999) | *V. orsiniana* | Spain. Loarre, Loarre mountains, signal repeater. (42.33832N, 00.69371W) | 3 |

| Accession number[a] (Collectors[b], date) | Species | Collection country and locality (Latitude, Longitude) | No. of individuals |
|---|---|---|---|
| SALA124588 (Martínez-Ortega & Delgado, 2-VIII-1999) | *V. orsiniana* | Spain. Huesca, Puértolas, Castillo Mayor. (42.57583N, 00.12222W) | 3 |
| SALA93486 (de Retz, 1-VI-1990) | *V. orsiniana* | France. Dép. Alpes-Maritimes, Caussols. (43.73528N, 06.92836E) | 3 |
| SALA72239 (Biondi & Amor, 29-V-1992) | *V. orsiniana* | Italy. Umbria, Perugia, Spello, Monte Subasio. (43.05000N, 12.66667E) | 3 |
| VAB932949 (Fabregat & Lopez, 19-VI-1993) | *V. orsiniana* | Spain. Teruel, Iglesuela del Cid. (40.48312N 00.31938W) | 3 |
| VAB933759 (Fabregat & Lopez, 19-VI-1993) | *V. orsiniana* | Spain. Castellón, Villafranca. (40.42853N, 00.25771W) | 3 |
| VAB944259 (Mercadal, 12-VII-1993) | *V. orsiniana* | Spain. Teruel, Cantavieja. (40.53295N, 00.48806W) | 3 |
| VAB947095 (Fabregat & López, 8-VI-1991) | *V. orsiniana* | Spain. Castellón, Villafranca, La Moleta. (40.42196N, 00.33607W) | 3 |
| B (Damboldt, 30-V-1970) | *V. prostrata* | Germany. Berlin, Glienicker Volkspark meadow. (52.44250N 13.58222W) | 3 |
| BC661136 (Negri, VII-1877) | *V. prostrata* | Italia. Casale Monferrato hill. (45.13333N, 08.45694E) | 3 |
| BM68672 (Garllorph, V-1905) | *V. prostrata* | Germany. Potsdam. (52.39667N, 13.05836E) | 3 |
| BM68674 (Mayer, 26-IV-1967) | *V. prostrata* | Serbia. Banat, Dliblatska peščara, Devojacki bunar. (44.89874N, 21.12988E) | 3 |
| BM68702 (Schneider, 16-V-1907) | *V. prostrata* | Bulgaria. Varna. (43.21667N, 27.91667E) | 3 |
| BM68710 (Filarszky & Kümmerle, 9-V-1916,) | *V. prostrata* | Slovakia. Comit Szepes, Smižany. (48.95525N, 20.51774E) | 3 |
| BM68720 (Toma, 10-V-1962) | *V. prostrata* | Romania. Iasi Breazu. (47.21667N, 27.51667E) | 2 |
| BM68839 (Lacaita, 6-V-1906) | *V. prostrata* | Italy. Toscana, Monte di prato. (44.05894N, 10.61664E) | 3 |
| SALA124616 (Martínez-Ortega, 28-V-2001) | *V. prostrata* | Austria. Leithagebirge, Donnerskirchen. (47.93321N, 16.66752E) | 3 |
| SALA124614 (Martínez-Ortega, 3-VI-2001) | *V. prostrata* | Austria. Hainburg, Braunsberg. (48.15406N, 16.95782E) | 2 |
| SALA124619 (Martínez-Ortega, 10-VI-2001) | *V. prostrata* | Czech Republic. Srbsko. (49.95219N, 14.14736E) | 3 |
| SALA124615 (Martínez-Ortega, 11-VI-2001) | *V. prostrata* | Austria. Falkenstein. (48.72445N, 16.57919E) | 3 |
| SALA124617 (Martínez-Ortega, 24-VI-2001) | *V. prostrata* | Czech Republic. Between Mikulov and Klentnice. (48.82603N, 16.64077E) | 3 |
| SALA124620 (Martínez-Ortega et al., 30-VI-2000) | *V. prostrata* | Romania. Fanatele Clujului, 8 Km from Cluj-Napoca. (46.85247N, 23.61688E) | 3 |

| Accession number[a] (Collectors[b], date) | Species | Collection country and locality (Latitude, Longitude) | No. of individuals |
|---|---|---|---|
| SEV16475 (Bührer, 24-V-1950) | *V. prostrata* | Switzerland. Kt. Wallis, Gemeinde Orsières, Westhaang. (46.03850N, 07.15628E) | 2 |
| SALA149321 (Rojas-Andrés et al., 13-VI-2009) | *V. rhodopea* | Bulgaria. Pazardzhik, Belmeken. (42.17653N, 23.80769E) | 3 |
| C (Bondev, 1955) | *V. rhodopea* | Bulgaria. Rila mountain. (42.18333N, 23.74361E) | 2 |
| PR806998 (Klásterský & Deyl, date unknown) | *V. rhodopea* | Bulgaria. Kara Balkan mountains. (41.58528N, 24.69194E) | 3 |
| B100217752 (Markgraf & Markgraf, 1973) | *V. rhodopea* | Bulgaria. Beglika southwards Pestera. (41.81512N, 24.12809E) | 3 |
| PRC455001 (Mrkvicka, 1916) | *V. rhodopea* | Bulgaria. Dobro Pole. (N/A) | 3 |
| BM67493 (Ibrahim, 5-VI-1884) | *V. rosea* | Morocco. Lalla-Aziza. (31.08505N, 08.70431W) | 3 |
| BM67502 (Davis, 21-VI-1975) | *V. rosea* | Algeria. Djurdjura, between Tizi-N'Kouilal pass and Tikdja. (36.47669N, 04.23206E) | 3 |
| BM67505 (Faure, 29-VI-1930) | *V. rosea* | Algeria. Bossuet. (34.66213N, 00.62089W) | 3 |
| BM67508 (Harley, 5-VII-1966) | *V. rosea* | Morocco. Beni Mellal, below summit of Irhil Ouaougoulzate. (31.65000N, 06.26667W) | 3 |
| BM67522 (Font Quer, 22 VII-1929) | *V. rosea* | Morocco. Djebel Lakraa. (35.13658N, 05.13711W) | 3 |
| E32376 (Alexander & Kupicha, 26-V-1972) | *V. rosea* | Morocco. Ketama-El Hoceima. (34.99332N, 04.21485W) | 2 |
| G8263-1733 (Bourgeau, 21-V-1856) | *V. rosea* | Algeria. Wilaya Oran, Ghar Rouban mountain. (34.58590N, 01.78783W) | 3 |
| G8263-1736 (Podlech, 7-VI-1984) | *V. rosea* | Algeria. Wilaya Batna, Awras region, Djebel Chélia. (35.49916N, 06.16661E) | 3 |
| G8263-1746 ( Wilczek et al, 20-IV-1928) | *V. rosea* | Morocco. Beni Suassene, refuge Zegzel. (34.83501N, 02.353421W) | 3 |
| G8263-1767 (Charpin et al., 31-V-1980) | *V. rosea* | Morocco. Ouarzazate, next to Tizi n'Melloul. (30.78333N, 07.60000W) | 3 |
| GDA27834 (Font Quer, 29-VI-1927) | *V. rosea* | Morocco. El Ferrah, Beni Hadifa. (35.02060N, 04.14370W) | 3 |
| K (Wilczek et al., 20-IV-1928) | *V. rosea* | Morocco. Beni Suassene, Tamcojoutan mountain. (34.85001N, 02.34996W) | 3 |
| MA302831 (Blanché et al., 1-VI-1985) | *V. rosea* | Morocco. Ouarzazate, Tizi n'Ouaro. (31.96666N, 05.63333W) | 3 |
| MA429804 (Sennen, date unknown) | *V. rosea* | Morocco. Riff. (34.99269N, 04.00250W) | 2 |

| Accession number[a] (Collectors[b], date) | Species | Collection country and locality (Latitude, Longitude) | No. of individuals |
|---|---|---|---|
| BC654127 (Don de Fritz, 18-VII-1872) | *V. satureiifolia* | Switzerland. Jura Neuchâteloise, Vallon de la Brèvine. (46.98685N, 06.60921E) | 3 |
| BC832014 (Despaty, 20-V-1917) | *V. satureiifolia* | France. Seine et Oise, Champareil. (49.27036N, 02.50791E) | 3 |
| G8263/1680 (Beckerer, 4-VII-1938) | *V. satureiifolia* | Austria. Valais, colline northwards d'Orsiéres. (46.17798N, 07.57521E) | 2 |
| G8263/1718 (Vautier, 22-V-1954) | *V. satureiifolia* | Switzerland. Dep. Lozère Aveu Armand, Causse méjean. (44.57987N, 03.71332E) | 2 |
| JACA10062571 (Villar, 2-VI-1971) | *V. satureiifolia* | Spain. Navarra, Isaba, Belagua. (42.89472N, 00.78111W) | 3 |
| JACA680871 (Montserrat & Villar, 2-IX-1971) | *V. satureiifolia* | Spain. Huesca, Hecho, Alanos, La Renclusa. (42.81950N, 00.83000W) | 3 |
| JACA71965 (Montserrat, date unknown) | *V. satureiifolia* | Spain. Navarra, Belagua, Roncal. (42.94508N, 00.83041W) | 3 |
| MA185334 (Paul, 3-VI-1933) | *V. satureiifolia* | Germany. Schwaben, Marienhöhe to Nördlingen. (48.83791N, 10.49653E) | 3 |
| SALA124594 (Martínez-Ortega, 9-V-2001) | *V. satureiifolia* | Germany. Baden Württhemberg, Schwäbische. (48.83556N, 10.37028E) | 2 |
| SALA124595 (Martínez-Ortega, 9-V-2001) | *V. satureiifolia* | Germany. Baden Württhemberg, Schwäbische. (48.80917N, 10.40528E) | 3 |
| SALA93470 (Martínez-Ortega & Delgado, 7-VII-1997) | *V. satureiifolia* | Spain. Huesca, Aragüés del Puerto, Llanos de Lizara. (42.76944N, 00.63361W) | 3 |
| SALA93472 (Martínez-Ortega & Delgado, 7-VII-1997) | *V. satureiifolia* | Spain. Huesca, Aragües del Puerto, Collado de Mesola. (42.74028N, 00.61278W) | 3 |
| SALA93471 (Martínez Ortega & Delgado, 10-VII-1997) | *V. satureiifolia* | Spain. Huesca, Hoz de Jaca, in the ascending to El Mandilar. (42.69333N, 00.28083W) | 3 |
| SALA124592 (Martínez Ortega & Delgado, 5-VIII-1999) | *V. satureiifolia* | Spain. Huesca, Aragües del Puerto, Pico Cucuruzuelo. (42.72472N 00.62361W) | 3 |
| SALA124593 (Martínez Ortega & Delgado, 6-VIII-1999) | *V. satureiifolia* | Spain. Huesca, Ansó, Paso del Onso. (42.89472N, 00.78111W) | 3 |
| FCO04428 (Navarro Andrés, 15-VII-1973) | *V. senneni* | Spain. Asturias, Valdemurrio water reservoir. (43.19836N, 06.01512W) | 2 |
| SALA110642 (Delgado & Rico, 7-VI-2001) | *V. senneni* | Spain. León, Valdepiélago, Valdorria. (42.88517N, 05.42452W) | 3 |
| SALA110637 (Delgado & Rico, 7-VI-2001) | *V. senneni* | Spain. León, Valdelugueros, Las Majadas. (42.92141N, 05.41369W) | 3 |
| SALA110634 (Delgado & Rico, 7-VI-2001) | *V. senneni* | Spain. León, Riaño. (42.93779N, 05.03442W) | 3 |

| Accession number[a] (Collectors[b], date) | Species | Collection country and locality (Latitude, Longitude) | No. of individuals |
|---|---|---|---|
| MA532538 (Gil de Zúñiga & Alejandre, 23-VI-1991) | *V. senneni* | Spain. La Rioja, Pedroso, Camero Nuevo mountains. (42.28194N, 02.66599W) | 3 |
| SALA93480 (Martínez-Ortega, 2-V-1997) | *V. senneni* | Spain. Álava, Salinas de Añana (Gesaltza Añana), trail from Sobrón to Rastrilla hill. (42.787872N, 3.092367W) | 3 |
| SALA93495 (Martínez-Ortega & Martín Ballesteros, 23-VI-1996) | *V. senneni* | Spain. Alava, Salinas de Añana (Gesaltza Añana). (42.77571N, 03.12224W) | 3 |
| SALA93479 (Martínez-Ortega, 2-V-1997) | *V. senneni* | Spain. Alava, Valdegobía, Villamardones. (42.86694N, 03.25377W) | 3 |
| SALA93478 (Martínez-Ortega, 3-V-1997) | *V. senneni* | Spain. Alava, Pipaón. (42.61309N, 02.63424W) | 3 |
| SALA93477 (Martínez-Ortega, 26-IV-1999) | *V. senneni* | Spain. Cantabria, Castro Urdiales, Oriñón. (43.39664N, 03.32104W) | 3 |
| SALA93475 (Martínez-Ortega, 26-IV-1999) | *V. senneni* | Spain. Cantabria, Laredo, El Puntal. (43.43220N, 03.45714W) | 3 |
| SALA110638 (Martínez-Ortega et al., 18-VI-1999) | *V. senneni* | Spain. Navarra, Aralar mountains. (43.00580N, 02.03073W) | 2 |
| SALA93485 (Uribe-Echebarría, 21-VI-1991) | *V. senneni* | Spain. Alava, Valdegobía, Lalastra. (42.867214N, 03.20272W) | 2 |
| SALA93487 (Martínez & Morante, 17-VI-1984) | *V. senneni* | Spain. Alava, Arlucena. (42.72678N, 02.54325W) | 2 |
| SEST38691 (Patino & Valencia, 27-V-1991) | *V. senneni* | Spain. Burgos, Junta de Trasloma. (43.03776N, 03.38948W) | 3 |
| GDA26404 (Torres et al., 9-VI-1983) | *V. tenuifolia* subsp. *fontqueri* | Spain. Granada, Baza mountains. (37.37903N, 02.84187W) | 3 |
| MA389465 (Fernández Casas, 19-VII-1974) | *V. tenuifolia* subsp. *fontqueri* | Spain. Granada, Baza mountains, Los Tejos pass. (37.37903N, 02.84187W) | 2 |
| MGC36827 (Cabezudo et al., 4-VII-1991) | *V. tenuifolia subsp. fontqueri* | Spain. Málaga, Yunquera, las Nieves mountains, Peña de los Enamorados. (36.69500N, 05.01130W) | 3 |
| MGC46659 (Cabezudo & Martínez-Ortega, 9-VI-1998) | *V. tenuifolia* subsp. *fontqueri* | Spain. Málaga, Ronda, las Nieves mountains, Los Quejigales. (36.68611N, 05.03028W) | 3 |
| SALA95041 (Martínez-Ortega, 11-VI-1998) | *V. tenuifolia* subsp. *fontqueri* | Spain. Almería, Fondón, Gador mountains, El Boliche plain. (36.91023N, 02.79794W) | 3 |
| B147196 (Preissmann, 19-V-1887) | *V. tenuifolia* subsp. *javalambrensis* | Spain. Teruel, Tramacastilla mountains. (40.42611N, 01.60000W) | 3 |
| E32503 (Brummit et al., date unknown) | *V. tenuifolia* subsp. *javalambrensis* | Spain. Cuenca, Valdeminguete mountains. (40.34586N, 01.77017W) | 2 |

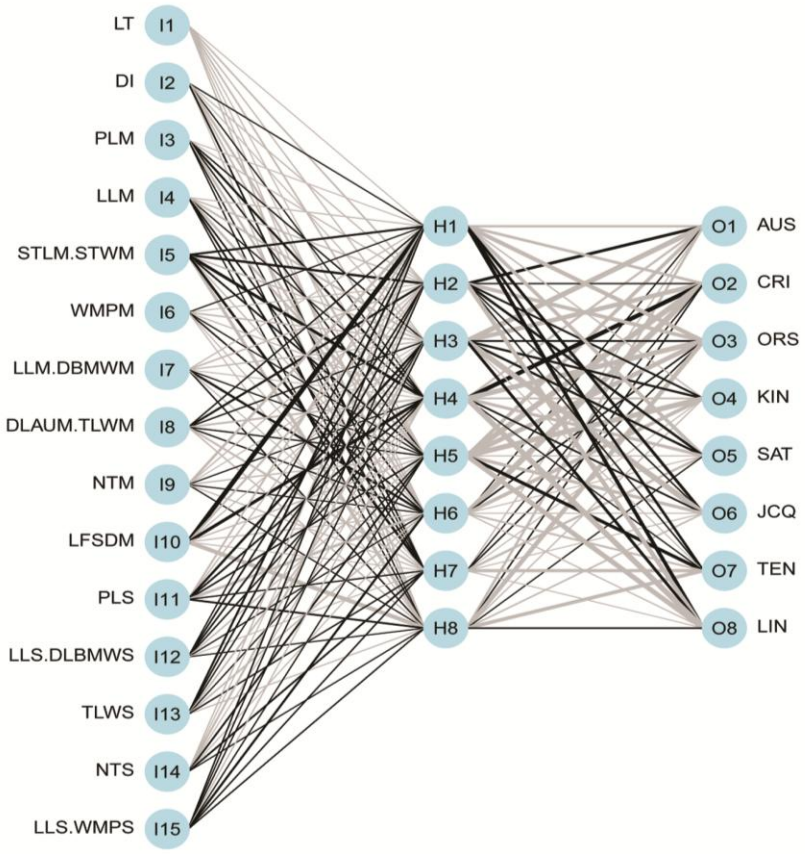
| Accession number[a] (Collectors[b], date) | Species | Collection country and locality (Latitude, Longitude) | No. of individuals |
|---|---|---|---|
| JACA463686 (Montserrat & Montserrat, 6-VII-1986) | *V. tenuifolia* subsp. *javalambrensis* | Spain. Lérida, Puigcerda-Martinet. (42.35956N, 01.69584E) | 3 |
| MA339646 (Fz. de Betoño & Alejandre, 6-VI-1985) | *V. tenuifolia* subsp. *javalambrensis* | Spain. Burgos, Sargentes de la Lora. (42.76917N, 03.87278W) | 3 |
| MA468044 (Gil de Zúñiga & Alejandre, 11-VI-1988) | *V. tenuifolia* subsp. *javalambrensis* | Spain. Soria, Villaciervos, Cabrejas mountains. (41.78333N, 02.80000W) | 3 |
| MA532710 (Alejandre, 30-V-1990) | *V. tenuifolia* subsp. *javalambrensis* | Spain. Palencia, Cervera de Pisuerga, Peña de Santa Lucía. (42.91590N, 04.64302W) | 3 |
| SALA93463 (Martínez-Ortega, 14-V-1996) | *V. tenuifolia* subsp. *javalambrensis* | Spain. Salamanca, La Mata de la Armuña. (41.03317N, 05.68832W) | 3 |
| SALA1411 (Rivas & Fernandez Galiano, 18-V-1952) | *V. tenuifolia* subsp. *javalambrensis* | Spain. Madrid, Cabrizos de Chozas. (40.41045N, 03.70175W) | 2 |
| SALA49322 (Casaseca et al., 11-VI-1990) | *V. tenuifolia* subsp. *javalambrensis* | Spain. Zamora, Abezames. (41.62642N, 05.42577W) | 3 |
| SALA93456 (Martínez-Ortega, 12-VII-1996) | *V. tenuifolia* subsp. *javalambrensis* | Spain. Cantabria, Camaleño, Mediana mountains. (43.07727N, 04.76881W) | 3 |
| SALA93468 (Martínez-Ortega, 7-VI-1996) | *V. tenuifolia* subsp. *javalambrensis* | Spain. Segovia, Navares de las Cuevas, Pradales mountains. (41.45927N, 03.73041W) | 3 |
| VIT37034 (Alejandre et al., 20-VI-1986) | *V. tenuifolia* subsp. *javalambrensis* | Spain. Guadalajara, from Buenafuente to Huertahernan. (40.81410N, 02.27383W) | 3 |
| BC113724 (Font Quer, 9-V-1926) | *V. tenuifolia* subsp. *tenuifolia* | Spain. Cataluña, close to Almacelles. (41.73216N, 0.437220E) | 3 |
| JACA61188 (Gómez & Aseginolaza, 16-V-1980) | *V. tenuifolia* subsp. *tenuifolia* | Spain. Zaragoza, Bardenas Reales, next to Tres Mugas. (42.29040N, 01.39557W) | 3 |
| JACA77571 (Montserrat, 4-V-1971) | *V. tenuifolia* subsp. *tenuifolia* | Spain. Lérida, Organya. (42.19681N, 01.31644E) | 3 |
| JACA9857 (Montserrat, 29-IV-1957) | *V. tenuifolia* subsp. *tenuifolia* | Spain. Huesca, Mediano. (42.31899N, 00.19680E) | 2 |
| MAF108794 (Velasco, 10-VI-1977) | *V. tenuifolia* subsp. *tenuifolia* | Spain. Toledo, Montes de Toledo, Rebollarejo mountains. (39.45444N, 03.96480W) | 3 |
| SALA95040 (Martínez-Ortega, 7-VI-1998) | *V. tenuifolia* subsp. *tenuifolia* | Spain. Navarra, Cáseda. (42.38666N, 01.42078W) | 3 |
| RNG (Heywood, 16-V-1971) | *V. tenuifolia* subsp. *tenuifolia* | Spain. Barcelona, La Panadella. (41.42293N, 01.39684E) | 3 |

| Accession number[a] (Collectors[b], date) | Species | Collection country and locality (Latitude, Longitude) | No. of individuals |
|---|---|---|---|
| K (Reverchon, VI-1891) | *V. tenuifolia* subsp. *tenuifolia* | Spain. Valencia, Segorbe mountains, Montemalo. (39.81436N, 00.54648W) | 3 |
| SALA33052 (Peris & Stübing, 20-VI-1984) | *V. tenuifolia* subsp. *tenuifolia* | Spain. Valencia, Requena, Pico Tejo. (39.51882N, 00.98775W) | 3 |
| VAB80719 (Mansanet & Mateo, VI-1980) | *V. tenuifolia* subsp. *tenuifolia* | Spain. Castellón, Forcall. (40.64542N, 00.19992W) | 3 |
| SALA149270 (Herrero, 1-VII-2007) | *V. teucrioides* | Greece. Kozani, Askio, Siniatsikon mountain. (40.40333N, 21.53722E) | 3 |
| SALA149329 (Rojas-Andrés et al., 23-VII-2010) | *V. teucrioides* | FYROM. P.N. Mavrovo, Bistra mountain. (41.64972N, 20.71075E) | 3 |
| SALA149330 (Rojas-Andrés et al., 26-VI-2009 | *V. teucrioides* | Greece. Olimpo mountain. (40.03872N, 22.33369E) | 3 |
| SALA124600 (Schneeweiss et al., 23-VII-2001) | *V. teucrium* | Italy. Aosta, Valle di la Thuile. (45.70889N, 06.94056E) | 3 |
| BM68135 (Filarszky , 28-VI-1814) | *V. teucrium* | Slovakia. Szepes, Igló. (47.45281N, 21.61058E) | 3 |
| BM68159 (Sag, 2-X-1875) | *V. teucrium* | Hungary. Visegrad. (47.80000N, 18.98333E) | 2 |
| BM68550 (Palkowa & Necka, 27-VI-1974) | *V. teucrium* | Poland. Distr. Chrzanów. (50.13317N, 19.40050E) | 3 |
| SALA124597 (Albach, 15-VI-2001) | *V. teucrium* | Bulgaria. Znepole region, Tran. (42.83778N, 22.69083E) | 3 |
| SALA124578 (Albach, 17-VI-2001) | *V. teucrium* | Bulgaria. Vitosa mountain, Simeonovo, base station. (42.61556N, 23.34750E) | 3 |
| WU (Scheneider, 14-VI-1937) | *V. teucrium* | Austria. Kärnten, Schneeweiss. (46.70907N, 14.17363E) | 3 |
| MA112464 (unknown collector, 20-V-1929) | *V. teucrium* | Germany. Palatinat. (49.19198N, 08.11498E) | 3 |
| MA185323 (Putchler, 8-VI-1918) | *V. teucrium* | Germany. Oberfranken Stadtsteinach. (50.16039N, 11.50514E) | 3 |
| MA333222 (Andrsovszky, 4-VI-1916) | *V. teucrium* | Hungary. Szent, Stara-Vod valley. (47.70529N, 19.04510E) | 3 |
| SALA93469 (Castroviejo et al., 17-VII-1990) | *V. teucrium* | Switzerland. Grisons, Engadina Bassa, Guarda. (46.77585N, 10.15289E) | 3 |
| SALA124599 (Martínez-Ortega, 3-VI-2001) | *V. teucrium* | Slovakia. Bratislava, Thebener Kogel. (48.19250N, 16.99833E) | 3 |
| SALA124602 (Martínez-Ortega & Fischer, 28-VI-2001) | *V. teucrium* | Austria. Niederösterreich between Pottenstein and Weissenbach. (47.96944N, 16.06333E) | 3 |
| SALA124598 (Martínez-Ortega et al., 3-VII-2002) | *V. teucrium* | Slovakia. Dvorníky-Vceláre, Zádiel. (48.61278N, 20.84222E) | 3 |
| SALA124562 (Schanzer & Majorov, 7-VII-2001) | *V. teucrium* | Russia. Tula, S valley of Krasivaya, Mecha Riv. (53.23750N, 38.35944E) | 2 |

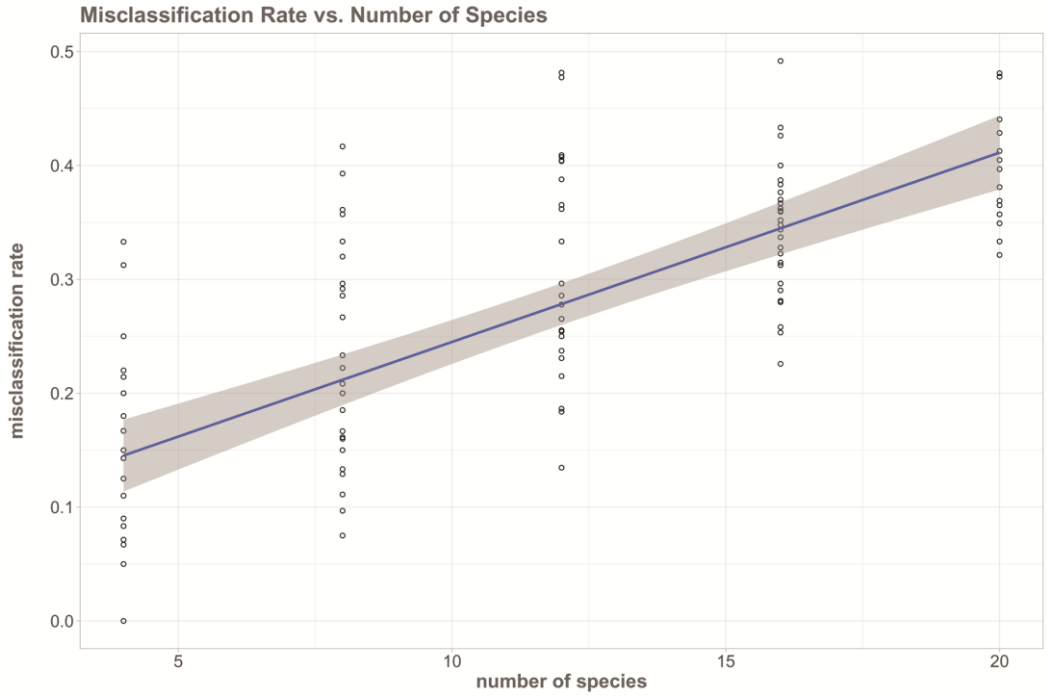
| Accession number[a] (Collectors[b], date) | Species | Collection country and locality (Latitude, Longitude) | No. of individuals |
|---|---|---|---|
| SALA149331 (Rojas-Andrés et al., 23-VI-2009) | *V. turrilliana* | Turkey. Dereköy to Armagan. (41.90808N, 27.38706E) | 3 |
| SALA149333 (Rojas-Andrés et al., 23-VI-2009) | *V. turrilliana* | Turkey. 6 Km. from Vize, in the way to Kömürköy-Alkpinar. (41.59472N, 27.82417E) | 3 |
| SALA149334 (Santos et al., 22-VI-2009) | *V. turrilliana* | Bulgaria. 15 Km northwards Malko Turnovo, western from the bridge upon Veleka river. (42.08506N, 27.42903E) | 3 |

[a] Herbarium and herbarium codes when available
[b] When there are more than two collectors only the first one is displayed followed by *et al.*

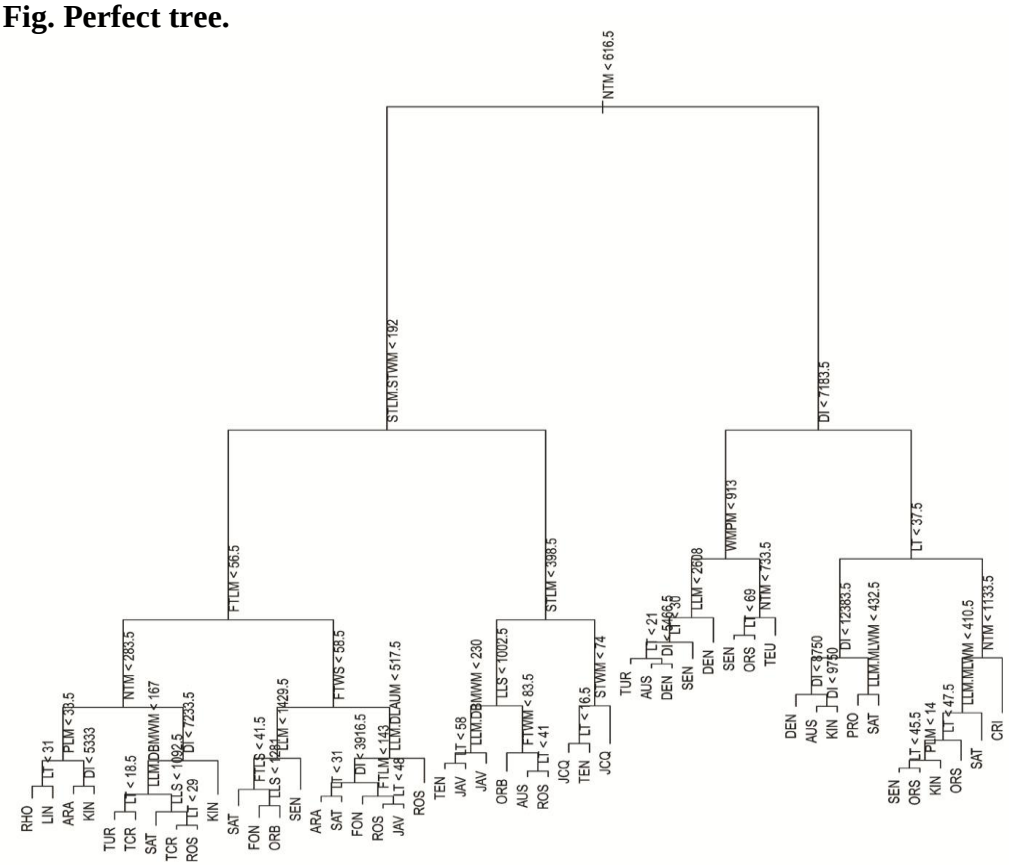
**S1 Fig. Example of the architecture of an artificial neural network.** (I) input layers = 15; (H) hidden layers = 1; number of neurons = 8; (O) output layers = 8. Output layers correspond to taxa (see Table 1 for abbreviations), input layers correspond to variables (see Table 2 for abbreviations). Positive and negative connections are represented by black and grey lines, respectively. Line width indicates the strength of the connection.

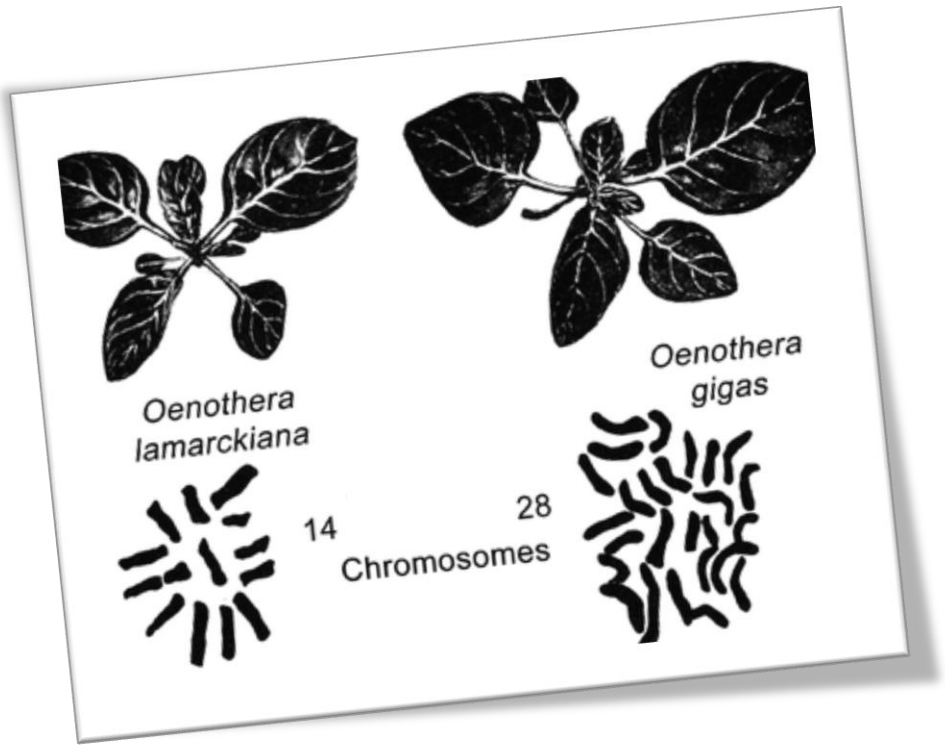


**S2 Fig. MCR vs. number of species.** Misclassification rate in relation to the number of output layers (i.e., number of species and subspecies). Each point represents a different combination of randomly chosen OTUs.




**S3 Fig. Perfect tree.**






# CHAPTER 3

Plantaginaceae. In: Karol Marhold & Jaromír Kučera (eds.), IAPT/IOPB chromosome data 28.

*Plantaginaceae. Editores: Karol Marhold & Jaromír Kučera. Asociación Internacional para la Taxonomía Vegetal. Datos cromosómicos 28*

**IAPT chromosome data 28**

**Edited by Karol Marhold & Jaromír Kučera**

**Luis Delgado,[1*] Blanca M. Rojas-Andrés,[1,2,3] Noemí López-González,[1,2] Nélida Padilla-García[1,2] & M. Montserrat Martínez-Ortega[1,2]**

1 *Departamento de Botánica y Fisiología Vegetal, University of Salamanca, 37007 Salamanca, Spain*

2 *Biobanco de ADN Vegetal, University of Salamanca, Edificio Multiusos I+D+i, Calle Espejo s/n, 37007 Salamanca, Spain*

3 Present address: *Department of Molecular Evolution and Plant Systematics & Herbarium (LZ), Institute of Biology, Leipzig University, Leipzig, Germany*

* Author for correspondence: *ldelsan@usal.es*

▲ First chromosome count for the species.

**DOI** https://doi.org/10.12705/676.39

**Funding**: This work has been supported by the Spanish Ministerio de Economía y Competitividad (projects CGL2009-07555, CGL2012-32574, *Flora iberica* VIII [CGL2008-02982-C03-02/CLI], *Flora iberica* IX [CGL2011-28613-C03-03], *Flora iberica* X [CGL2014-52787-C3-2-P]); the Spanish Ministerio de Ciencia e Innovación (Ph.D. grants to BR and NLG), and the University of Salamanca (Ph.D. grant to NPG financed jointly by Banco Santander). The authors are very grateful to all colleagues who helped with plant location and sampling.

**RESUMEN**

Este estudio presenta los recuentos validados de cromosomas y las estimaciones de nivel de ploidía de varias especies y taxones infraespecificos pertenecientes a *Veronica* subsect. *Pentasepalae*. Mediante análisis citológico (conteos cromosómicos directos), citometría de flujo o la combinación de ambas técnicas se aporta información para un total de 12 especies y taxones infraespecíficos con los requisitos básicos de calidad que garantiza la Asociación Internacional para la Taxonomía Vegetal. El material ha sido recolectado durante sucesivas campañas de muestreo e incluye especímenes de 14 países de Europa, además de Marruecos y Turquía.

En el caso de las especies *V. dalmatica* y *V. kindlii* la información en cuanto al número de cromosomas del presente estudio supone el primer conteo disponible para la especie. Asimismo, para *V. thracica* y *V. prostrata* se aportan, por primera vez, conteos cromosomicos o medidas de nivel de ploidia para estas especies en Turquía y Rumanía respectivamente.

Chromosome numbers counted by L. Delgado and ploidy level estimated by B. Rojas-Andrés and N. López-González; collectors: *AA* = Antonio Abad, *AT* = Andreas Tribsch, *BR* = Blanca Rojas-Andrés, *DGL* = David Gutiérrez Larruscain, *DP* = Daniel Pinto, *JASA* = José Ángel Sánchez Agudo, *JPG* = Julio Peñas de Giles, *LMC* = Luz Mª Muñoz Centeno, *MO* = M. Montserrat Martínez-Ortega, *MS* = María Santos Vicente, *NLG* = Noemí López-González, *NPG* = Nélida Padilla-García, *SA* = Santiago Andrés, *SB* = Sara Barrios, *VL* = Víctor Lucía, *XG* = Ximena Giráldez.

### CHROMOSOME COUNTS

Material CHN, collected in Bosnia and Herzegovina, Bulgaria, Montenegro, Macedonia (F.Y.R.O.M.) and Switzerland. All cytological investigations have been carried out on anthers and gynoecia. Material was fixed in 3:1 absolute ethanol-glacial acetic acid and stained in 2% acetic orcein (Cour, 1945).

### DNA PLOIDY LEVEL ESTIMATES

Material FCM, collected in Austria, Czech Republic, Croatia, France, Morocco, Romania, Serbia, Slovakia, Spain, Switzerland and Turkey. DNA ploidy levels were estimated from silica-gel-dried leaves. Samples were prepared according to the procedure described in Rojas-Andrés & al. (2015). Leaf samples of *Pisum sativum* L. 'Ctirad' (2C DNA = 9.09 pg; Dolezel & al., 1998), *Solanum pseudocapsicum* L. (2C DNA = 2.589 pg; Temsch & al., 2010) and *Zea mays* L. 'CE-777' (2C DNA = 5.43 pg; Lysak & Dolezel, 1998) were used as internal reference standards according to sample C-value and standard availability. Results on DNA ploidy level were acquired using a CyFlow SL system (Partec, Münster, Germany) equipped with a blue 488 nm solid state laser. For each individual, the ratio of the G0/G1 peak positions of samples and internal standards were recorded.

DNA ploidy level was estimated based on the genome size value found and the available chromosome number for each particular taxon. In many cases FCM measurements were directly compared with CHN corresponding to the same samples. The coefficient of variation (CV) was calculated for each sample and the standard used. All data were suitable for DNA ploidy level estimation because CV values did not exceed the 10% threshold.

## TAXONOMIC TREATMENT

Material determination and nomenclatural treatment follow Rojas-Andrés & Martínez-Ortega (2016) and Rojas-Andrés & al. (2016), respectively, with the modifications proposed by Padilla-García & al. (2018).

## PLANTAGINACEAE

### *1. Veronica angustifolia* (Vahl) Bernh.

2*n* = 64, CHN. Switzerland, Valais, on the ascent to Tanay lake, 46°20′36.0″N, 06°50′33.4″E, 1418 m, 11 Jul 2011, *X. Giráldez, M.M. Martínez-Ortega 6038 & B. Rojas-Andrés* (SALA 149416) [Fig. 3I]. 2*n* ~ 8*x* ~ 64, FCM. France, Departement of Hautes-Alpes, Gap. Chorges, near Gap. On the ascent to Monte Chabrières from Les Andrieux, at the margins of the road, 44°34′35.3″N, 06°16′35.3″E, 1393 m, 10 Jul 2011, *X. Giráldez, M.M. Martínez-Ortega & B. Rojas-Andrés 167-1*, *X. Giráldez, M.M. Martínez-Ortega & B. Rojas-Andrés 167-2*, *X. Giráldez, M.M. Martínez-Ortega & B. Rojas-Andrés 167-3* (SALA 149398); Switzerland, Valais, on the ascent to Tanay lake, 46°20′36.0″N, 06°50′33.4″E, 1418 m, 11 Jul 2011, *X. Giráldez, M.M. Martínez-Ortega 6038-1 & B. Rojas-Andrés*, *X. Giráldez, M.M. Martínez-Ortega 6038-2 & B. Rojas-Andrés*, *X. Giráldez, M.M. Martínez-Ortega 6038-3 & B. Rojas-Andrés* (SALA 149416).

Based on AFLP data, this species name has been recently resurrected to name plants that were previously included under the variation of *V. teucrium* L. (Padilla-García & al., 2018). Simonet (1934) and Brandt (1952) counted the chromosomes of plants from this group from France and Switzerland, but we have not been able to revise the

corresponding herbarium vouchers and do therefore not know for sure whether they correspond to *V. angustifolia* or *V. teucrium*.

***2. Veronica austriaca* subsp. *jacquinii* (Baumg.) Watzl**

2*n* = 32, CHN. Bosnia and Herzegovina, Trebinje, between Gacko and Tjentiste, subalpine meadows on limestone soils, 43°11′04.92″N, 18°33′56.808″E, 1076 m, 10 Jun 2015, *M.M. Martínez-Ortega 6122, X. Giráldez, N. Padilla-García & N. López-González* (SALA 157023) [Fig. 3A], *M.M. Martínez-Ortega 6123, X. Giráldez, N. Padilla-García & N. López-González* (SALA 157024) [Fig. 3B]. 2*n* = 48, CHN. Bosnia and Herzegovina, Trebinje, between Gacko and Tjentiste, subalpine meadows on limestone soils, 43°11′04.92″N, 18°33′56.808″E, 1076 m, 10 Jun 2015, *M.M. Martínez-Ortega 6120, X. Giráldez, N. Padilla-García & N. López-González* (SALA 157021) [Fig. 3C]; Bulgaria, Pazardzhik, between Batak and Beglika, perennial grassland, 41°50′27″N, 24°08′48″E, 1522 m, 17 Jun 2009, *X. Giráldez, M.M. Martínez-Ortega, B. Rojas-Andrés & M. Santos Vicente* (SALA 149368); Bulgaria, Pernik, between Staychovtsi and Dolna Melna, 42°41′04″N, 22°31′57″E, 950 m, 14 Jun 2009, *X. Giráldez, M.M. Martínez-Ortega 4577, B. Rojas-Andrés & M. Santos Vicente* (SALA 149376) [Fig. 3D]. 2*n* ~ 6*x* ~ 48, FCM. Croatia, Licausko Polje, meadows on limestone soils, 704 m, 17 Jun 2001, *M.M. Martínez-Ortega 1404-1 & A. Tribsch, M.M. Martínez-Ortega 1404-2 & A. Tribsch* (SALA124609); Croatia, Primorje-Gorski Kotar, Velika Kapela, Gornje Jelenje pass, at the crossroad towards Crikvanica, meadows on limestone soils, 45°21′58″N, 14°37′09″E, 880 m, 16 Jun 2001, *M.M. Martínez-Ortega 1391-1 & A. Tribsch, M.M. Martínez-Ortega 1391-2 & A. Tribsch* (SALA 124604).

Numerous chromosome counts are known for this taxon, 2*n* = 32, 48, 64 (cf. Scheerer, 1937; Moore, 1973, 1977; Peev, 1978; Goldblatt & Johnson, 1979+; Goldblatt, 1981, 1988; Májovský & Murín, 1987), but 2*n* = 48 is the most frequent one. The tetraploid cytotype 2*n* = 32 is rare and has been reported only once from Bulgaria (Peev, 1972).

The population studied here from Bosnia and Hercegovina contains tetraploid (2*n* = 32; SALA 157023 and SALA 157024) and hexaploid individuals (2*n* = 48; SALA 157021). The morphology of the tetraploid individuals is also close to that of *Veronica orbiculata* A.Kern.

***3. Veronica dalmatica* N.Pad.Gar., Rojas-Andrés, López-González & M.M.Mart.Ort. ▲**

2*n* = 16, CHN. Montenegro, Kotor, Lovcen, shrub clearings on limestones, 42°25′04.9″N, 18°47′39″E, 904 m, 9 Jun 2015, *M.M. Martínez- Ortega, X. Giráldez, N. Padilla-García & N. López-González 137* (SALA 157018) [Fig. 3E]; Montenegro, Žabljak, Meždo, dry meadows on limestone soils with *Juniperus*, 43°09′49.824″N, 19°08′56.688″E, 1390 m, 12 Jun 2015, *M.M. Martínez-Ortega, X. Giráldez, N. Padilla-García & N. López-González 139* (SALA157030) [Fig. 3F].

It has not been possible to revise herbarium material corresponding to a previous chromosome count (2*n* = 16) from an Albanian population (Baltisberger, 1988) identified as *V. jacquinii*. Padilla-García & al. (2018) have recently shown that the diploid representatives traditionally assigned to *V. austriaca* subsp. *jacquinii* represent a distinct species named *V. dalmatica*. Here, material clearly assigned to this species collected in Montenegro is investigated for the first time.

***4. Veronica kindlii* Adam. ▲**

2*n* = 16, CHN. Macedonia (F.Y.R.O.M.), Bitola, on the ascent to Pelister summit, subalpine meadows on granite soils, 40°59′40.416″N, 21°10′42.492″E, 2355 m, 8 Jul 2014, *X. Giráldez, M.M. Martínez-Ortega 6090, B. Rojas-Andrés & N. López-González* (SALA 157011) [Fig. 3G, H].

The name *V. kindlii* has been recently resurrected to designate those populations from the Balkan Peninsula previously identified as *V. orsiniana* (Rojas-Andrés & al., 2015). A previous chromosome count for *V. kindlii* was carried out on material collected in the Macedonian locality of Barbaros (Makedonski Brod) (Sopova & al., 1983 in Goldblatt, 1988). However, AFLP data showed that the latter population corresponds to the closely related species *V. linearis* (Padilla-García & al., 2018). Thus, the chromosome count published here would represent the first one for *V. kindlii*.

***5. Veronica orsiniana* Ten.**

2*n* ~ 2*x* ~ 16, FCM. France, Department of Lozère, Cévennes, 5 kilometers from Aven Armand cave, at the crossroad towards Masde-la-Font, montane meadows with *Buxus* on acid soil, 44°12′22″N, 03°23′31″E, 945 m, 8 Jun 2012, *X. Giráldez, M.M. Martínez-Ortega & B. Rojas-Andrés 203-1, X. Giráldez, M.M. Martínez-Ortega & B. Rojas-Andrés 203-2, X. Giráldez, M.M. Martínez-Ortega & B. Rojas-Andrés 203-3* (SALA 149303); F rance, Department of Lozère, Cévennes, Aven Armand cave, meadows on limestone soils, 44°13′29″N, 03°21′34″E, 1006 m, 8 Jun 2012, *X. Giráldez, M.M. Martínez-Ortega & B. Rojas-Andrés 207-1, X. Giráldez, M.M. Martínez-Ortega & B. Rojas-Andrés 207-2, X. Giráldez, M.M. Martínez-Ortega & B. Rojas-Andrés 207-3* (SALA 149305); France, Department of Var, on the ascent to Sainte Baume massif, Braque pathway, meadows on limestone soils, 43°19′21″N, 05°42′09″E, 693 m, 9 Jun 2012, *X. Giráldez, M.M. Martínez-Ortega & B. Rojas-Andrés 210-1, X. Giráldez, M.M. Martínez-Ortega & B. Rojas-Andrés 210-2, X. Giráldez, M.M. Martínez-Ortega & B. Rojas-Andrés 210-3* (SALA 149308); Spain, Barcelona, Collsuspina, between Tona and Moyá, grazed pastures near the road, grassy slopes, 41°49′39″N, 02°10′45″E, 916 m, 14 Jun 2013, *M.M. Martínez-Ortega, B. Rojas-Andrés 235-1, A. Abad & N. López-González, M.M. Martínez-Ortega, B. Rojas-Andrés 235-2, A. Abad & N. López-González, M.M. Martínez-Ortega, B. Rojas-Andrés 235-3, A. Abad & N. López-González* (SALA 155093), *M.M. Martínez-Ortega, B. Rojas-Andrés 235-22*, A. Abad & N. López-González* (SALA 155067); Spain, Barcelona, Espunyola, towards Montclar village, clearings in *Quercus ilex* and *Q. faginea* forest on limestone soils, 42°01′15″N, 01°46′11″E, 757 m, 16 Jun 2013, *M.M. Martínez-Ortega 6064-1, B. Rojas-Andrés, X. Giráldez & N. López- González, M.M. Martínez-Ortega 6064-2, B. Rojas-Andrés, X. Giráldez & N. López-González, M.M. Martínez-Ortega 6064-3, B. Rojas-Andrés, X. Giráldez & N. López-González* (SALA 155119); Spain, Barcelona, Montserrat, on the road to the monastery El Bruc, to Manresa, Can Maçana, meadows on limestone soils behind the farmhouse, 41°36′36″N, 01°46′01″E, 724 m, 13 Jun 2013, *M.M. Martínez-Ortega, B. Rojas-Andrés, A. Abad & N. López-González 8-1, M.M. Martínez-Ortega, B. Rojas-Andrés, A. Abad & N. López-González 8-2, M.M. Martínez-Ortega, B. Rojas-Andrés, A. Abad & N. López-González 8-3* (SALA 155097); Spain, Barcelona, Tona, on the road to Castell and the Lourdes chapel, dry grassy meadows at the side of the pathway, marly limestones, 41°51′16″N, 02°13′18″E, 661 m, 14 Jun 2013, *M.M. Martínez-Ortega, B. Rojas-Andrés 234-1, A. Abad & N. López- González, M.M. Martínez-Ortega, B. Rojas-Andrés 234-2, A. Abad & N. López-González, M.M. Martínez-Ortega, B. Rojas-Andrés 234-3, A. Abad & N. López-González* (SALA 155094); Spain, Gerona, from Viladray to Sta. Fe del Montseny, at the crossroad to Mas el Martí, meadows near the road, 41°50′24″N, 02°24′52″E, 966 m, 16 Jun 2013, *M.M. Martínez-Ortega 6063-1, B. Rojas-Andrés, A. Abad & N. López-González, M.M. Martínez-Ortega 6063-2, B. Rojas-Andrés, A. Abad & N. López-González, M.M. Martínez-Ortega 6063-3, B. Rojas-Andrés, A. Abad & N. López-González* (SALA 155123); Spain, Huesca, Sta. Cruz de la Serós, San Juan de la Peña, meadows at the picnic area, near the interpretation center and the new Monastery, 42°30′26″N,

00°39′55″W, 1214 m, 20 Jun 2013, *M.M. Martínez-Ortega, B. Rojas-Andrés 242-1, X. Giráldez & N. López-González, M.M. Martínez-Ortega, B. Rojas-Andrés 242-2, X. Giráldez & N. López-González, M.M. Martínez-Ortega, B. Rojas-Andrés 242-3, X. Giráldez & N. López-González* (SALA 155109); Spain, Huesca, pathway between Binacua and the crossroad to the road N-240, meadows on limestone soils and slopes with *Buxus*, 42°33′09″N, 00°41′47″W, 718 m, 20 Jun 2013, *M.M. Martínez-Ortega, B. Rojas-Andrés 243-1, X. Giráldez & N. López-González, M.M. Martínez-Ortega, B. Rojas-Andrés 243-2, X. Giráldez & N. López-González, M.M. Martínez-Ortega, B. Rojas-Andrés 243-3, X. Giráldez & N. López-González* (SALA 155110).

### *6. Veronica prostrata* L.

$2n \sim 2x \sim 16$, FCM. Austria, Burgenland, Leithagebirge, between Donnerskirchen and Franz Josef-Warte, dry grasslands, Pannonian vegetation, 47°55′59″N, 16°40′03″E, 250 m, 28 May 2001, *M.M. Martínez- Ortega 1055-1, M.M. Martínez-Ortega 1055-2, M.M. Martínez-Ortega 1055-3* (SALA 124616); Austria, Niederösterreich, Falkenstein, near the castle ruins, 48°43′28″N, 16°34′45″E, 390 m, 11 Jun 2001, *M.M. Martínez-Ortega 1337-1, M.M. Martínez-Ortega 1337-2, M.M. Martínez-Ortega 1337-3* (SALA 124615); Czech Republic, South Moravia, Břeclav District, Mikulov, near the Austrian border, Šibeničník Nature reserve, dry meadows, Pannonian vegetation, 48°47′25″N, 16°37′56″E, 236 m, 23 Jul 2011, *X. Giráldez, M.M. Martínez- Ortega & B. Rojas-Andrés 184-1, X. Giráldez, M.M. Martínez- Ortega & B. Rojas-Andrés 184-2, X. Giráldez, M.M. Martínez- Ortega & B. Rojas-Andrés 184-3* (SALA 149310); Czech Republic, South Moravia, Břeclav District, between Mikulov and Klentnice, dry meadows on limestone soils, in an environment of Pannonian vegetation, 48°49′33″N, 16°38′26″E, 345 m, 24 Jul 2001, *M.M. Martínez-Ortega 1445-1* (SALA 124617); France, Departement of Hautes-Alpes, Gap, Col du Noyer, between Le Noyer and L'Enclus, subalpine meadows on limestone soils, 44°41′35″N, 05°59′13″E, 1671 m, 14 Jun 2012, *X. Giráldez, M.M. Martínez-Ortega & B. Rojas-Andrés 219-1, X. Giráldez, M.M. Martínez-Ortega & B. Rojas-Andrés 219-2, X. Giráldez, M.M. Martínez-Ortega & B. Rojas-Andrés 219-3* (SALA 149313); Romania, Cluj, Fanatele Clujului, 8 km from Cluj-Napoca, meadows on limestone soils, 46°51′08″N, 23°37′01″E, 550 m, 30 Jun 2000, *X. Giráldez, M.M. Martínez-Ortega 904-1 & J.A. Sánchez Agudo, X. Giráldez, M.M. Martínez-Ortega 904-2 & J.A. Sánchez Agudo* (SALA 124620); Serbia, Južnobanatski okrug (South Banat), Devojački Bunar, Vladimirovac, dry sandy meadows, 45°00′58″N, 20°57′17″E, 163 m, 27 Jul 2010, *S. Andrés 428-1, X. Giráldez, M.M. Martínez-Ortega & B. Rojas-Andrés, S. Andrés 428-2, X. Giráldez, M.M. Martínez-Ortega & B. Rojas-Andrés, S. Andrés 428-3, X. Giráldez, M.M. Martínez-Ortega & B. Rojas-Andrés* (SALA 149319).

Ploidy level estimations for the species in Romania are here given for the first time. No previous chromosome count or ploidy level estimation was available for *V. prostata* in this country (cf. Bolkhovskikh & al., 1969; Goldblatt & Johnson, 1979+; Albach & al., 2008).

***7. Veronica rosea* Desf.**

2*n* ~ 2*x* ~ 16, FCM. Morocco, Ifrane, Meknès-Tafilalet, Jebel Hebri, grazed pastures with *Thymelaea*, 33°21′13″N, 05°08′43″W, 1931 m, 18 Jul 2013, *D. Pinto, N. López-González 52-1 & V. Lucía García*, *D. Pinto, N. López-González 52-2 & V. Lucía García*, *D. Pinto, N. López-González 52-3 & V. Lucía García* (SALA 155072); Morocco, Ifrane, in front of Aquelmame de Si-Ali, road between Azrou and Midelt (between Timahdite and Ait-Oufella), scrubland in clearings of *Juniperus thurifera* subsp. *africana*, on limestone soils, 33°04′50.988″N, 05°01′13.094″W, 2168 m, 1 May 2013, *J. Peñas de Giles 1, J. Peñas de Giles 2, J. Peñas de Giles 3* (GDA 59936).

***8. Veronica sennenii* (Pau) M.M.Mart.Ort. & E.Rico**

2*n* ~ 8*x* ~ 64, FCM. Spain, León, Lois, near the road towards Liegos/Anciles, meadows on limestone soils, 42°59′04″N, 05°07′05″W, 1278 m, 2 Jul 2013, *M.M. Martínez-Ortega, B. Rojas-Andrés 248-1 & N. López-González, M.M. Martínez-Ortega, B. Rojas-Andrés 248-2 & N. López-González, M.M. Martínez-Ortega, B. Rojas-Andrés 248-3 & N. López-González* (SALA 155081).

***9. Veronica tenuifolia* subsp. *javalambrensis* (Pau) Molero & J.Pujadas**

2*n* ~ 2*x* ~ 16, FCM. Spain, Burgos, A maya, on t he a scent to Peña Amaya, meadows on limestone soils, 42°39′02″N, 04°10′41″W, 1167 m, 5 Jul 2013, *M.M. Martínez-Ortega, B. Rojas-Andrés 250-1 & N. López-González, M.M. Martínez-Ortega, B. Rojas-Andrés 250-2 & N. López-González, M.M. Martínez-Ortega, B. Rojas-Andrés 250-3 & N. López-González* (SALA 155086); Spain, Burgos, Ciruelos de Cervera, Briongos de Cervera, Raposa pathway, near the road BU-914, clearings of *Juniperus sabina*, limestone soil, 41°54′50″N, 03°29′48″W, 1085 m, 2 Jun 2013, *D. Gutiérrez-Larruscain, N. López-González & D. Pinto 1278-1, D. Gutiérrez-Larruscain, N. López-González & D. Pinto 1278-2, D. Gutiérrez-Larruscain, N. López-González & D. Pinto 1278-3* (SALA 150521); Spain, Burgos, Espinosa de Cervera, near the old football ground, short pastures in clearings of *Juniperus sabina* close to the pathway, 41°54′01.3″N, 03°27′53.9″W, 1074 m, 2 Jun 2013, *D. Gutiérrez-Larruscain, N. López-González & D.*

*Pinto 1280-1*, *D. Gutiérrez-Larruscain, N. López-González & D. Pinto 1280-2*, *D. Gutiérrez-Larruscain, N. López-González & D. Pinto 1280-3* (SALA 150519); Spain, Santander, road between Espinama and the Áliva refuge, subalpine meadows on limestone soils, 43°09′55.86″N, 04°46′25.92″W, 1565 m, 3 Jul 2013, *M.M. Martínez-Ortega 6074-1*, *B. Rojas-Andrés & N. López-González*, *M.M. Martínez-Ortega 6074-2*, *B. Rojas-Andrés & N. López-González, M.M. Martínez-Ortega 6074-3, B. Rojas-Andrés & N. López-González* (SALA 155084); Spain, Segovia, Moral de Hornuez, Sabinar de Hornuez (picnic area near the chapel), dry and short meadows on limestone soils, 41°29′00″N, 03°37′32.8″W, 25 May 2013, *M.M. Martínez-Ortega & N. López-González 3-1*, *M.M. Martínez-Ortega & N. López-González 3-2*, *M.M. Martínez-Ortega & N. López-González 3-3* (SALA 155104); Spain, Soria, El Burgo de Osma, Torralba del Burgo, Arroyo de los Barranquillos, in environments of *Juniperus thurifera*, on clay and gravel soils, 41°37′44.1″N, 02°55′00.2″W, 974 m, 8 Jun 2013, *D. Pinto 1315-1*, *D. Pinto 1315-2*, *D. Pinto 1315-3* (SALA 150484); Spain, Soria, Herrera de Soria, Camino del Oropar, Barranco de la Covatilla, streambed in *Juniperus thurifera* forest, 41°46′16.8″N, 03°02′05.0″W, 1093 m, 8 Jun 2013, *D. Pinto 1311-1*, *D. Pinto1311-2*, *D. Pinto 1311-3* (SALA 150488); Spain, Soria, Langa de Duero, Alcozar, near Cerro Hestilla, short meadows and scrublands of *Thymus* sp. on limestone soils, 41°37′29.9″N, 03°19′28.1″W, 904 m, 2 Jun 2013, *D. Gutiérrez-Larruscain, N. López-González & D. Pinto 1287-1*, *D. Gutiérrez-Larruscain, N. López-González & D. Pinto 1287-2, D. Gutiérrez-Larruscain, N. López-González & D. Pinto 1287-3* (SALA 150512); Spain, Soria, Recuerda, on the ascent to La Muela from the water source el Cepo, *Juniperus thurifera* forest on limestones, 41°28′01.0″N, 02°58′54.1″W, 950 m, 8 Jun 2013, *D. Pinto 1307-1*, *D. Pinto 1307-2*, *D. Pinto 1307-3* (SALA 150492); Spain, Soria, Villaciervos, El Santo, streambed on *Genista scorpius* scrublands in environment of *Juniperus sabina*, 41°46′08.1″N, 02°38′54.6″W, 1228 m, 8 Jun 2013, *D. Pinto 1322-1*, *D. Pinto 1322-2*, *D. Pinto 1322-3* (SALA 150477); Spain, Zamora, Vezdemarbán, hills near the village on limestone soils, table hills, 41°39′03.8″N, 05°21′50.0″W, 793 m, 19 Jun 2012, *S. Barrios, N. López-González, M.M. Martínez-Ortega & B. Rojas- Andrés 221-1*, *S. Barrios, N. López-González, M.M. Martínez-Ortega & B. Rojas-Andrés 221-2*, *S. Barrios, N. López-González, M.M. Martínez-Ortega & B. Rojas-Andrés 221-3* (SALA 149327).

### *10. Veronica tenuifolia* Asso subsp. *tenuifolia*

$2n \sim 2x \sim 16$, FCM. Spain, Barcelona, Collsuspina, meadows on limestones near the village, on the left side along the pathway to Sta. Coloma de Castellterçol (GR177), 41°49′24″N, 02°10′36.24″E, 905 m, 14 Jun 2013, *M.M. Martínez-Ortega, B. Rojas-Andrés 237-1, A. Abad & N. López-González, M.M. Martínez-Ortega, B. Rojas-Andrés 237-2, A. Abad & N. López-González, M.M. Martínez-Ortega, B. Rojas-Andrés 237-3, A. Abad & N. López-González* (SALA155065); Spain, Barcelona, on the ascent to Sta. Perpetua chapel, near the road, 41°59′54.9″N, 02°12′09.9″E, 663 m, 15 Jun 2013, *M.M.*

*Martínez-Ortega, B. Rojas-Andrés, A. Abad & N. López-González 16-1, M.M. Martínez-Ortega, B. Rojas-Andrés, A. Abad & N. López-González 16-2, M.M. Martínez-Ortega, B. Rojas-Andrés, A. Abad & N. López-González 16-3* (SALA 155125); Spain, Teruel, Bordón, on the road to Calanda, before the crossroad to Luco de Bordón and Bordón river, marly slopes, 40°41′36.6″N, 00°19′09.5″W, 769 m, 10 Jun 2013, *M.M. Martínez-Ortega 6059-1, B. Rojas-Andrés, X. Giraldez & N. López-González, M.M. Martínez-Ortega 6059-2, B. Rojas-Andrés, X. Giraldez & N. López-González, M.M. Martínez-Ortega 6059-3, B. Rojas-Andrés, X. Giraldez & N. López-González* (SALA 155099).

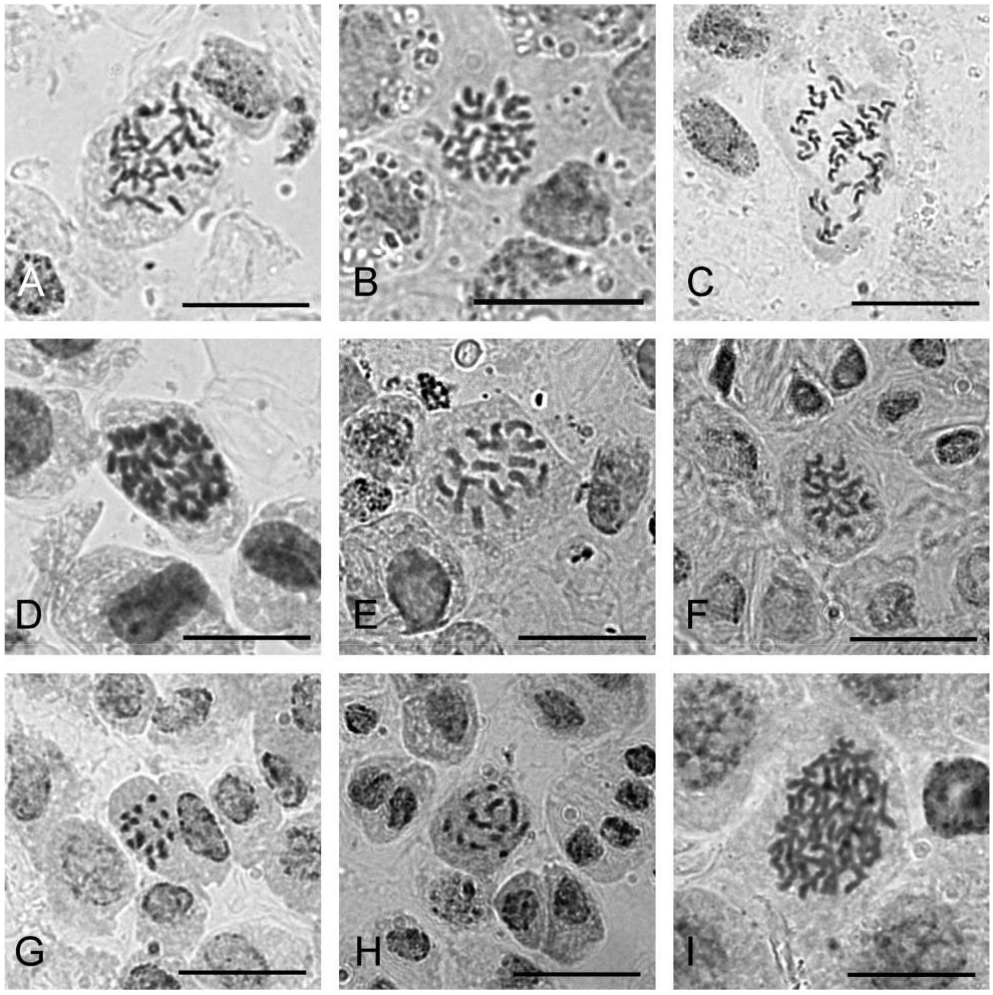



**Fig. 1.** Mitotic metaphases: **A–D,** *Veronica austriaca* subsp. *jacquinii*: **A,** 2*n* = 32 (SALA 157023); **B,** 2*n* = 32 (SALA 157024); **C,** 2*n* = 48 (SALA 157021); **D,** 2*n* = 48 (SALA 149376); **E & F,** *V. dalmatica*: **E,** 2*n* = 16 (SALA 157018); **F,** 2*n* = 16 (SALA 157030); **G & H,** *V. kindlii*, 2*n* = 16 (SALA157011); **I,** *V. angustifolia*, 2*n* = 64 (SALA 149416).

***11. Veronica teucrium* L.**

2*n* ~ 8*x* ~ 64, FCM. Slovakia, Košice, distr. Košice-okolie, Dvorníky-Včeláre, Zádiel, dry meadows on limestones, 48°36′46″N, 20°50′32″E, 300 m, 3 Jul 2002, *M.M. Martínez-Ortega 1551-1, X. Giráldez & Muñoz Centeno, M.M. Martínez-Ortega 1551-2, X. Giráldez & Muñoz Centeno, M.M. Martínez-Ortega 1551-3, X. Giráldez & Muñoz Centeno* (SALA 124598).

***12. Veronica thracica* Velen.**

2*n* ~ 2*x* ~ 16, FCM. Turkey, Kirklareli, Dereköy, road towards Geçitagzi, mixed forest and meadows on limestone soils, 41°56′15″N, 27°21′11″E, 462 m, 23 Jun 2009, *X. Giráldez, M.M. Martínez-Ortega, B. Rojas-Andrés 37-1 & M. Santos Vicente, X. Giráldez, M.M. Martínez-Ortega, B. Rojas-Andrés 37-2 & M. Santos Vicente* (SALA 149289).

*Veronica thracica* was previously included within the variation of *V. crinita* Velen. (Peev, 1972; Rojas-Andrés & al., 2015) but has been recently combined at the specific rank based on AFLP data (Padilla-García & al., 2018). Previous chromosome counts (Peev, 1975, 1976) and ploidy level estimations (Padilla-García & al., 2018) concern plants distributed in Bulgaria. Consequently, these are the first ploidy level estimations for this species from Turkey.

**AKNOWLEDGEMENTS**

This work has been supported by the Spanish Ministerio de Economía y Competitividad (projects CGL2009-07555, CGL2012-32574, *Flora iberica* VIII [CGL2008-02982-C03-02/CLI], *Flora iberica* IX [CGL2011-28613-C03-03], *Flora iberica* X [CGL2014-52787-C3-2-P]); the Spanish Ministerio de Ciencia e Innovación (Ph.D. grants to BR and NLG), and the University of Salamanca (Ph.D. grant to NPG financed jointly by Banco Santander). The authors are very grateful to all colleagues who helped with plant location and sampling.

## REFERENCES

**Albach DC, Martínez-Ortega MM, Delgado L, Weiss- Schneeweiss H, Özgökce F, Fischer MA.** 2008. Chromosome numbers in Veroniceae (Plantaginaceae): Review and several new counts. *Annals of the Missouri Botanical Garden*, 95: 543–566.

**Baltisberger M.** 1988. Chromosomenzahlen einiger Pflanzen aus Albanien. II = Chromosome numbers of some plants from Albania. II. *Berichte des Geobotanischen Institutes der Eidg. Techn. Hochschule, Stiftung Rübel*, 54: 42–50.

**Bolkhovskikh Z, Grif V, Matvejeva T, Zakharyeva O.** 1969. Chromosome numbers of flowering plants. Nauka, Leningrad.

**Brandt JP.** 1952. Contribution à la cytologie du genre *Veronica*. *Bulletin de la Société Neuchâteloise des Sciences Naturelles*, 75: 179–188.

**Cour LF.** 1945. Smear and squash techniques in plant cytology. *Laboratory Practices*, 3: 326–330.

**Dolezel J, Greilhuber J, Lucretti S, Meister A, Lysak MA, Nardi L, Obermayer R.** 1998. Plant genome size estimation by flow cytometry: Inter-laboratory comparison. *Annals of Botany*, 82: 17–26.

**Goldblatt, P.** 1981. Index to plant chromosome numbers 1975–1978. *Monographs in Systematic Botany from the Missouri Botanical Garden*, 5: 1–553.

**Goldblatt, P.** 1988. Index to plant chromosome numbers 1984–1985. *Monographs in Systematic Botany from the Missouri Botanical Garden*, 23: 1–264.

**Goldblatt P, Johnson DE.** 1979. Index to plant chromosome numbers. http://www.tropicos.org/Project/IPCN (accessed Jul 2018).

**Lysak MA, Dolezel J.** 1998. Estimation of nuclear DNA content in *Sesleria* (Poaceae). *Caryologia*, 51: 123–132.

**Májovský J, Murín A.** 1987. Karyotaxonomický prehľad flóry Slovenska. Veda Vydavateľsto Slovenskej Akadémie vied, Bratislava.

**Moore RJ.** 1973. Index to plant chromosome numbers for 1967– 1971. Regnum Vegetabile 90. Oosthoek, Scheltema & Holkema, Utrecht.

**Moore RJ.** 1977. Index to plant chromosome numbers for 1973−1974. Regnum Vegetabile 96. Oosthoek, Scheltema & Holkema, Utrecht.

**Padilla-García N, Rojas-Andrés BM, López-González N, Castro M, Castro S, Loureiro J, Albach DC, Machon N, Martínez-Ortega MM.** 2018. The challenge of species delimitation in the diploid-polyploid complex *Veronica* subsection *Pentasepalae*. *Molecular Phylogenetiics and Evolution*, 119: 196–209.

**Peev D.** 1972. New taxa and ploidy levels of some Bulgarian *Veronica* species. *Doklady Bolgarskoi akademii nauk*, 25: 811–814.

**Peev D.** 1975. Genus *Veronica* in Bulgaria – Some new taxa and chromosome numbers. *Fitologiya*, 2: 42–46.

**Peev D.** 1976. [Report], in: Löve, Á. (ed.), IOPB Chromosome number reports LIV. *Taxon*, 25: 631–649.

**Peev D.** 1978. Taxonomija i microevoljucija na divorastjastite predstaviteli na rod *Veronica* L. (velikdence) v Balgarija, in: Kozuharov SI, Kuzmanov BA (eds.), Evolution of flowering plants and florogenesis, vol. 1, *Fabaceae, Scrophulariaceae, Rubiaceae*. Bulgarian Academy of Sciences, Sofia, pp. 72–106.

**Rojas-Andrés BM, Martínez-Ortega MM.** 2016. Taxonomic revision of *Veronica* subsection *Pentasepalae* (*Veronica*, Plantaginaceae sensu APG III). *Phytotaxa*, 285: 1–100.

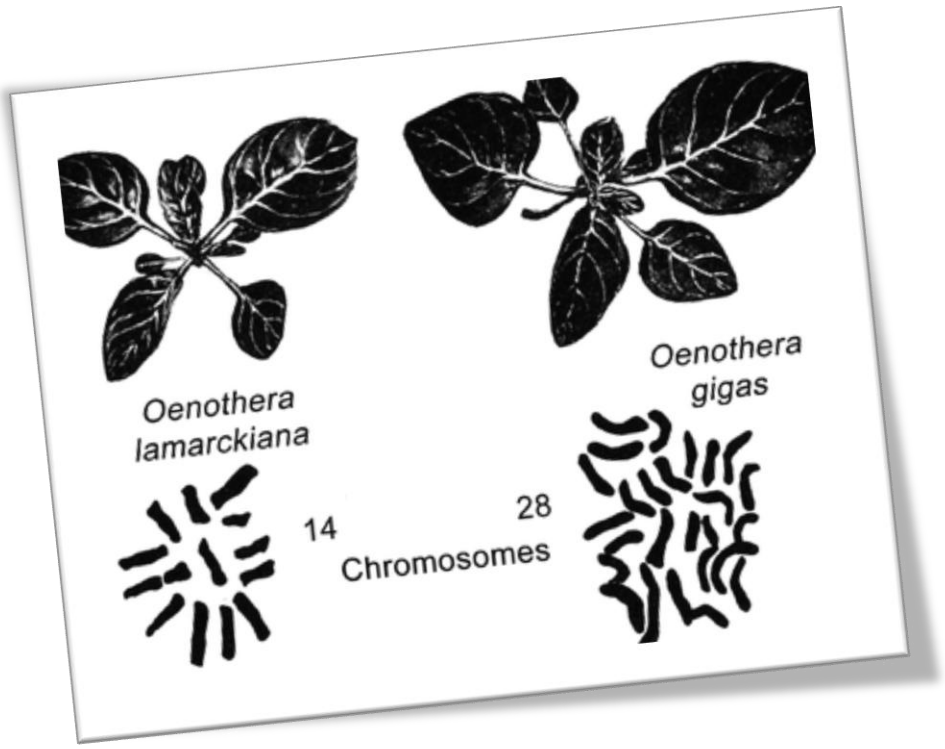
**Rojas-Andrés BM, Albach DC, Martínez-Ortega MM.** 2015. Exploring the intricate evolutionary history of the diploid–polyploid complex *Veronica* subsection *Pentasepalae* (Plantaginaceae). *Botanical Journal of the Linnean Societyy*, 179: 670–692.

**Rojas-Andrés BM, Rico E, Martínez-Ortega MM.** 2016. A nomenclatural treatment for *Veronica* subsect. *Pentasepalae* (Plantaginaceae sensu APG III) and typification of several names. *Taxon*, 65: 617–627.

**Scheerer H.** 1937. Experimentelle und zytologische Untersuchungen innerhalb der *Veronica*-Gruppe *Pentasepala*. *Flora*, 131: 287–323.

**Simonet M.** 1934. Contribution à l'étude caryologique et génétique des *Veronica*. *Comptes-Rendus Hebdomadaires des Séances et Mémoires de la Societé de Biologie*, 117: 1153–1156.

**Temsch E, Greilhuber J, Krisai R.** 2010. Genome size in liverworts. *Preslia*, 82: 63–80.




# CHAPTER 4

Characterization of 12 polymorphic SSR markers in *Veronica* subsect. *Pentasepalae* (Plantaginaceae) and cross-amplification in 10 other subgenera

*Caracterización de 12 marcadores microsatélites polimórficos en Veronica subsect. Pentasepalae (Plantaginaceae) y amplificación cruzada en otros 10 subgéneros*

Applications in Plant Sciences

PRIMER NOTE

# CHARACTERIZATION OF 12 POLYMORPHIC SSR MARKERS IN *VERONICA* SUBSECT. *PENTASEPALAE* (PLANTAGINACEAE) AND CROSS-AMPLIFICATION IN 10 OTHER SUBGENERA

NOEMÍ LÓPEZ-GONZÁLEZ [1,2,4], EIKE MAYLAND-QUELLHORST[3] DANIEL PINTO-CARRASCO [1,2] AND M. MONTSERRAT MARTÍNEZ-ORTEGA [1,2]

[1] Departamento de Botánica, Universidad de Salamanca, E-37007 Salamanca, Spain;

[2] Biobanco de ADN Vegetal, Banco Nacional de ADN, 37007 Salamanca, Spain;

[3] Institut für Biologie und Umweltwissenschaften, Carl von Ossietzky Universität, 26111 Oldenburg, Germany

*Applications in Plant Sciences* 2015 3 ( 10 ): 1500059; http://www.bioone.org/loi/apps

Applications in Plant Sciences

Manuscript received 18 May 2015; revision accepted 19 June 2015.

**Funding:** This research was financially supported by the Spanish Ministry of Science and Innovation through the projects CGL2012-32574 and CGL200907555. A predoctoral grant to N.L.G. from the Ministry of Education, Culture, and Sport (AP2010-2968) is also acknowledged. We are also deeply grateful to Blanca Rojas-Andrés and Dirk Albach for their continuous support.

Author for correspondence: noe_lg@usal.es

doi:10.3732/apps.1500059

*Applications in Plant Sciences* 2015 3 ( 10 ): 1500059; http://www.bioone.org/loi/apps

## ABSTRACT

**Premise of the study:** Microsatellite primers were developed in the perennial herbs of the diploid-polyploid complex *Veronica* subsect. *Pentasepalae* (Plantaginaceae) to investigate the role that hybridization has played in the evolution of the group, which includes several endangered species.

**Methods and Results:** Twelve pairs of primers leading to polymorphic and readable markers were identified and optimized from *V. jacquinii* and *V. orbiculata* using a microsatellite-enriched library method and 454 GS-FLX technique. The set of primers amplified dinucleotide to pentanucleotide repeats, and the number of alleles per locus ranged from one to six, one to 11, and one to nine for *V. orsiniana*, *V. javalambrensis*, and *V. rosea*, respectively. Transferability analyses were performed in 20 species corresponding to 10 different subgenera.

**Conclusions:** These results indicate the utility of the newly developed microsatellites across *Veronica* subsect. *Pentasepalae*, which will help in the study of gene flow patterns and genetic structure.


**Key words:** conservation; hybridization; Plantaginaceae; polyploid complex; *Veronica* subsect. *Pentasepalae.*

**RESUMEN**

**Objetivo del estudio:** se desarrollan un conjunto de marcadores moleculares de tipo microsatélite tomando como referencia las especies pertenecientes al complejo diploide-poliploide *Veronica* subsect. *Pentasepalae* (Plantaginaceae) para investigar el papel que la hibridación ha podido jugar en la evolución del grupo, que incluye varias especies en peligro de extinción.

**Métodos y resultados:** doce pares de *primers* que representan marcadores polimórficos legibles han sido identificados y optimizados para *V. jacquinii* y *V. orbiculata* utilizando el método de biblioteca enriquecida en secuencias microsatéllite mediante la técnica 454 GS-FLX. El conjunto final de *primers* amplifica repeticiones de dinucleótidos a pentanucleótidos. El número de alelos por locus tiene un rango de uno a seis, uno a 11 y uno a nueve para *V. orsiniana*, *V. javalambrensis* y *V. rosea*, respectivamente. Se realizaron análisis de transferibilidad en 20 especies pertenecientes a 10 subgéneros diferentes.

**Conclusiones:** estos resultados muestran la utilidad para *Veronica* subsect. *Pentasepalae* de estos marcadores microsatétite recién desarrollados. Estos marcadores podrán ser de ayuda en el estudio de patrones de flujo genético y estructura genética.

**Palabras clave:** conservación; hibridación; Plantaginaceae; complejo poliploide; *Veronica* subsect. *Pentasepalae*.

## 1. INTRODUCTION

The genus *Veronica* L. (Plantaginaceae) comprises ca. 450 species, which are grouped into 12 subgenera with between two and 180 species each (Albach et al., 2004; Garnock-Jones et al., 2007). It includes some perennials of relative economic importance in ornamental horticulture and others that are well-known widespread weeds. Additionally, several species of *Veronica* are registered on the International Union for Conservation of Nature Red List (http://www.iucnredlist.org/) and other regional catalogs of endangered plants (e.g., Peñas de Giles et al., 2004), or are threatened plants with narrow distribution areas (e.g., Petrova and Vladimirov, 2009).

*Veronica* subsect. *Pentasepalae* Benth. is a monophyletic diploid-polyploid complex and one of the four subsections currently recognized within the also monophyletic *Veronica* subgen. *Pentasepalae* M. M. Mart. Ort., Albach & M. A. Fischer (Albach et al., 2008). This subsection comprises ca. 20 perennial taxa and is represented in the temperate regions of Eurasia with one species in North Africa. The complex seems to be of recent origin and divergence, as many diploid representatives are still extant and short branches are found in the phylogenetic analyses based on ITS and plastid DNA sequence data (Rojas Andrés et al., 2015). Although the diploid species are characterized by subtle morphological differences, each has been recovered as monophyletic in previous studies. Hybridization and polyploidization are widespread in the group, and several authors (Lehmann, 1937; Scheerer, 1949; Rojas-Andrés et al., 2015) have concluded that gene flow and complex relationships among polyploids and their diploid relatives might exist. Interestingly, some of the diploid and polyploid species belonging to *Veronica* subsect. *Pentasepalae* are Mediterranean orophytes that face a high risk of extinction with climate warming and/ or grow in Important Plant Areas (IPAs; IPA online database: http://www.plantlifeipa.org/reports.asp), regions that

display exceptionally rich floras of biogeographic interest (Rojas-Andrés et al., 2015). Given that current gene flow and introgression may have blurred species limits, particularly in hybrid zones, accurate investigations of gene flow patterns within and among *Veronica* subsect. *Pentasepalae* populations are necessary for conservation and species delimitation purposes.

## 2. METHODS AND RESULTS

***2.1 Microsatellite development*** —F or the microsatellite library, silica gel–dried leaves of 12 diploid individuals of *V. jacquinii* Baumg. and *V. orbiculata* A. Kern. were selected from eight different populations (Appendix 1). Ploidy level was checked using flow cytometry. A microsatellite library was prepared by Genoscreen (Lille, France) using a 454 GS-FLX (Roche Diagnostics, Meylan, France) high-throughput DNA sequencer (Malausa et al., 2011). Genomic DNA was extracted using the cetyltrimethylammonium bromide method described in Doyle and Doyle (1987). The DNA was fragmented and enriched with TG, TC, AAC, AAG, AGG, ACG, ACAT, and ACTC motifs. A total of 32,052 high-quality sequences were obtained. Analyses of these sequences with QDD software (Mzeglécz et al., 2010) revealed 3010 sequences with microsatellite motifs, for which 195 pairs of primers were obtained. Given that it is too time consuming and not affordable to check all of the primer pairs obtained, 54 of them with low primer pair penalty and different lengths and repeat motifs were selected. These primers were ordered (Eurofins, Ebersberg, Germany) to evaluate polymorphic loci on 12 individuals from the complex *V. jacquinii–V. orbiculata*. PCRs were performed in a total volume of 15 μ L, which contained 1 × PCR Green GoTaq Buffer (Promega Corporation, Madison, Wisconsin, USA), 0.25 mM of each dNTP (Life Technologies, Carlsbad, California, USA), 0.33 mM of each primer, 0.5 units GoTaq

DNA Polymerase (Promega Corporation), and 18.2 ng of DNA template. PCRs used the following conditions: an initial step at 94 ° C for 2 min; followed by 35 cycles of 1 min at 94 °C , 1 min at 50–58 °C , and 50 s at 72 °C ; and a final extension of 15 min at 72 °C. All the reactions were conducted on a Mastercycler pro S thermocycler (Eppendorf, Hamburg, Germany). The PCR products were separated by electrophoresis on a 2.5% agarose gel and sent to Macrogen Europe sequencing service (Amsterdam, The Netherlands).

In a second step, those primers that were polymorphic in the *V. jacquinii*– *V. orbiculata* complex were tested in two individuals from three species, each from a different clade (*V. orsiniana* Ten. [core clade], *V. javalambrensis* Pau [Iberian clade], and *V. rosea* Desf. [North African clade]), using the same PCR conditions. Twelve polymorphic primer pairs were selected (see Appendix 2 for additional primers). Following the procedure developed by Schuelke (2000) , the sequence-specific forward primers were marked at the 5′ end with an M13 tail (5′ -TGTAAAACGACGGCCAGT-3′ ) (Eurofi ns), which was then labeled with 5-FAM, VIC, NED, or PET fluorescent dyes (Table 1) (Life Technologies). The PCR mix contained 1 × PCR Green GoTaq (Promega Corporation), 0.2 mM of each dNTP, 0.16 mM of each reverse and fluorescent-labeled M13 primer, 0.04 mM of forward primer, 0.75 units GoTaq DNA Polymerase, and 50 ng of DNA template in a total volume of 15 μ L. Conditions of the PCR amplification were as described above, adding 10 cycles of 1 min at 94 ° C, 1 min at 53 ° C, and 50 s at 72 ° C before the final extension. PCR products were analyzed with GeneMarker AFLP/Genotyping Software version 1.8 (SoftGenetics, State College, Pennsylvania, USA).

***2.2 Population genetics parameters in three further species from* Veronica *subsect. Pentasepalae*** — The first comprehensive phylogenetic analysis of *Veronica* subsect. *Pentasepalae* based on DNA sequence data revealed four main clades each corresponding to a broad geographic area (Rojas-Andrés et al., 2015). Thus, for the characterization of the microsatellite markers, diploid populations corresponding to species from different clades were selected (Appendix 1): *V. orsiniana* (core clade), *V. javalambrensis* (Iberian clade), and *V. rosea* (North African clade). The Central Asian clade was not considered because no material was available. The mean number of alleles per locus, observed and expected heterozygosities, possible deviations from Hardy–Weinberg equilibrium (HWE; Table 2), and tests for linkage disequilibrium between markers in each population were estimated using Arlequin version 3.5.1.2 (Excoffier and Lischer, 2010).

**TABLE 1.** Characterization of 12 polymorphic nuclear microsatellite loci isolated from *Veronica* subsect. *Pentasepalaea*

| Locus | Primer sequences (5′–3′) | Fluorescent dye | Repeat motif | Allele size range (bp) [b] | $T_a$ (°C) | GenBank accession no. |
|---|---|---|---|---|---|---|
| 8 | F: TGATGTGACTGATTGGGTCAG<br>R: TTACCTCCTCATCACTCCCC | 5-FAM | $(TGA)_5$ | 92–95 | 55 | KR698358 |
| 10 | F: TGAACAACACACAGGTTCAATTC<br>R: GGCTAGAAGTTGTGAAGAAGGG | 5-FAM | $(AG)_9$ | 113–119 | 55 | KR698359 |
| 13 | F: GCTTTTCTCGGTGAAAGGGT<br>R: CACCATAATCCACAGCCTGA | PET | $(TGAT)_5$ | 113–133 | 58 | KR698360 |
| 19 | F: TCGAAACTTATTCGGCAACG<br>R: GACTCACGAGTTTGGAAGCG | 5-FAM | $(ATT)_5$ | 133–157 | 55 | KR698361 |
| 20 | F: TGGAGACCAAAATTCAACCC<br>R: TCTTGTCTCCTACTCTCCTCCG | PET | $(AC)_{11}$ | 93–135 | 52 | KR698362 |
| 26 | F: ATGTCGACGTGTCAACTCCA<br>R: CACTTGTTTCCACAGCTGGC | NED | $(CAA)_6$ | 87–102 | 56 | KR698363 |
| 27 | F: TATGGGAGACGACATGGTCA<br>R: CTCCCTTTCGTAGCAACACC | PET | $(TTGTG)_6$ | 201–221 | 55 | KR698364 |
| 35 | F: CATTTAATGGTATCCGATGCG<br>R: TCGCTTTTCGATTTCTTCGT | NED | $(TATC)_7$ | 106–130 | 52 | KR698365 |

| Locus | Primer sequences (5′–3′) | Fluorescent dye | Repeat motif | Allele size range (bp) [b] | $T_a$ (°C) | GenBank accession no. |
|---|---|---|---|---|---|---|
| 49 | F: GGATGCTTTATTTTGTCTTGT<br>R: TGTTACGACATTTATGGTGATT | VIC | $(TGGA)_5$ | 222–242 | 52 | KR698366 |
| 50 | F: TGTGATGCACAGAGTTTTAGTT<br>R: TGAAAACATAACACCTCGATAA | VIC | $(AGA)_6$ | 400–460 | 50 | KR698367 |
| 52 | F: ATAAAAACATCCATACTTTCCG<br>R: GTTAACCGCCAGTCTAACTAAT | VIC | $(GTT)_5$ | 358–391 | 52 | KR698368 |
| 54 | F: CCAAATATCAAATGATACCACA<br>R: TCGTAAAATTACGTCATCAAGA | NED | $(AC)_{13}$ | 283–301 | 52 | KR698369 |

*Note*: $T_a$ = annealing temperature.
[a] All values are based on 90 samples from three *Veronica* populations.
[b] Range of fragment sizes does not include the M13

**TABLE 2.** Results of initial primer screening of polymorphic loci in three populations corresponding to three different taxa belonging to *Veronica* subsect. *Pentasepalae*[a]

| Locus | *V. orsiniana* (n=30) | | | | *V. javalambrensis* (n=30) | | | | *V. rosea* (n=30) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A | $H_o$ | $H_e$ | HWE[b] | A | $H_o$ | $H_e$ | HWE[b] | A | $H_o$ | $H_e$ | HWE[b] |
| 8 | 2 | 0.933 | 0.506 | 0.000*** | 2 | 0.167 | 0.155 | 1.000 ns | 1 | | | |
| 10 | 2 | 0.000 | 0.066 | 0.017* | 1 | | | | 3 | 0.033 | 0.097 | 0.017* |
| 13 | 2 | 0.167 | 0.440 | 0.001*** | 6 | 0.500 | 0.500 | 0.388 ns | 1 | | | |
| 19 | 2 | 0.333 | 0.488 | 0.125 ns | 4 | 0.700 | 0.697 | 0.852 ns | 4 | 0.233 | 0.298 | 0.968 ns |
| 20 | 4 | 0.700 | 0.525 | 0.140 ns | 10 | 0.767 | 0.818 | 0.077 ns | 9 | 0.690 | 0.736 | 0.144 ns |
| 26 | 1 | | | | 3 | 0.433 | 0.432 | 1.000 ns | 5 | 0.690 | 0.743 | 0.391 ns |
| 27 | 3 | 0.500 | 0.560 | 0.290 ns | 3 | 0.483 | 0.381 | 0.448 ns | 3 | 0.233 | 0.213 | 1.000 ns |
| 35 | 2 | 0.400 | 0.488 | 0.447 ns | 3 | 0.333 | 0.420 | 0.100 ns | 4 | 0.769 | 0.669 | 0.860 ns |
| 49 | 1 | | | | 6 | 0.633 | 0.742 | 0.061 ns | — | — | — | — |
| 50 | 3 | 0.233 | 0.216 | 1.000 ns | 11 | 0.567 | 0.785 | 0.017* | 4 | 0.037 | 0.240 | 0.000*** |
| 52 | 1 | | | | 1 | | | | 3 | 0.136 | 0.210 | 0.222 ns |
| 54 | 6 | 0.567 | 0.733 | 0.000*** | 3 | 0.367 | 0.310 | 0.632 ns | 4 | 0.600 | 0.494 | 0.399 ns |

*Note*: — = not amplified; $A$ = number of alleles; $H_e$ = expected heterozygosity; $H_o$ = observed heterozygosity; HWE = Hardy–Weinberg equilibrium probabilities; $n$ = number of individuals sampled.
[a] See Appendix 1 for locality and voucher information for each population .
[b] Deviations from HWE were not statistically significant (ns) and statistically significant at * $P < 0.05$, ** $P < 0.01$, and *** $P \leq 0.001$.

The number of alleles per locus ranged from one to six, one to 11, and one to nine in the *V. orsiniana*, *V. javalambrensis*, and *V. rosea* populations, respectively. Loci 26, 49, and 52 were monomorphic in *V. orsiniana*, loci 10 and 52 were monomorphic in *V.*

*javalambrensis*, and in *V. rosea*, loci 8 and 13 were monomorphic and locus 49 did not amplify. The observed and expected heterozygosities for all populations are shown in Table 2. Significant deviation from HWE ($P < 0.05$) was seen for loci 8, 10, 13, and 54 in *V. orsiniana*, for locus 50 in *V. javalambrensis*, and for loci 10 and 50 in *V. rosea*. Linkage disequilibrium showed significance levels below 0.05 after false discovery rate (FDR) correction in two pairwise comparisons (pair 20–52 in *V. rosea* and pair 27–54 in *V. orsiniana*).

***2.3 Cross-amplification in other species from Veronica subsect. Pentasepalae and 10 subgenera of Veronica —*** Cross-amplification performed for these 12 polymorphic loci showed successful results within the expected allele size in two additional species from *Veronica* subsect. *Pentasepalae*: *V. austriaca* L. and *V. dentata* F. W. Schmidt. Tests were also performed for 20 additional species from 10 different subgenera within the large genus *Veronica* (Table 3). The tests were carried out with the original PCR protocol. The 12 loci tested in agarose gel showed successful amplification of at least several bands. Six of these (8, 10, 13, 19, 26, and 35) showed good amplification results in most samples.

## 3. CONCLUSIONS

A set of polymorphic microsatellite markers for *Veronica* subsect. *Pentasepalae* is reported. Amplification success for these markers in the cross-transferability tests extends their potential usefulness to other subgenera. These markers will be useful for investigating genetic parameters, which may provide essential information for the conservation of threatened species, as well as data on the role of interspecific hybridization in the evolution of the genus.

TABLE 3. Amplification success of all microsatellite primers across 20 species from 10 subgenera of *Veronica*.

| Subgenera | Collector no.[a,b] | Species | 8 | 10 | 13 | 19 | 20 | 26 | 27 | 35 | 49 | 50 | 52 | 54 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Veronica* subg. *Beccabunga* (Hill) M. M. Mart. Ort., Albach & M. A. Fisch. | DCA350 | *V. gentianoides* | w | s | + | w | — | — | — | + | — | — | — | — |
| *Veronica* subg. *Beccabunga* | DCA297 | *V. gentianoides* | s | s | + | w | — | s | — | + | — | — | — | — |
| *Veronica* subg. *Beccabunga* | MO1598 | *V. gentianoides* | — | — | — | — | — | + | — | — | — | — | — | — |
| *Veronica* subg. *Chamaedrys* (W. D. J. Koch) M. M. Mart. Ort., Albach & M. A. Fisch. | KBch67 | *V. chamaedrys* subsp. *chamaedryoides* | s | s | w | + | + | + | w | + | — | — | — | s |
| *Veronica* subg. *Chamaedrys* | KBch54 | *V. vindobonensis* | s | + | w | + | + | s | + | + | — | — | — | s |
| *Veronica* subg. *Cochlidiosperma* (Rchb.) M. M. Mart. Ort. & Albach | DCA403 | *V. cymbalaria* | + | + | + | s | w | s | s | + | — | — | — | s |
| *Veronica* subg. *Cochlidiosperma* | HMM31 | *V. cymbalaria* | + | + | + | + | w | s | s | + | — | — | — | — |
| *Veronica* subg. *Cochlidiosperma* | HMM32 | *V. cymbalaria* | + | + | + | + | w | s | s | + | — | — | — | — |
| *Veronica* subg. *Cochlidiosperma* | HMM29 | *V. panormitana* | + | s | + | + | — | s | — | + | — | — | — | — |
| *Veronica* subg. *Cochlidiosperma* | HMM30 | *V. trichadena* | + | s | + | s | — | + | — | w | — | — | — | — |
| *Veronica* subg. *Pellidosperma* (E. B. J. Lehm.) M. M. Mart. Ort., Albach & M. A. Fisch. | DCAs434 | *V. triphyllos* | + | + | + | w | s | s | w | s | — | — | + | w |
| *Veronica* subg. *Pocilla* (Dumort.) M. M. Mart. Ort., Albach & M. A. Fisch. | DCA144 | *V. filiformis* | w | + | + | s | w | s | w | + | — | — | — | — |
| *Veronica* subg. *Pocilla* | DCA954 | *V. filiformis* | s | + | + | s | w | s | + | + | — | — | v | s |
| *Veronica* subg. *Pocilla* | DCA892 | *V. filiformis* | s | + | + | s | w | + | + | + | — | — | — | s |
| *Veronica* subg. *Pseudolysimachium* (W. D. J. Koch) M. M. Mart. Ort., Albach & M. A. Fisch. | KB847 | *V. orchidea* | s | + | w | s | + | s | + | + | — | — | — | s |
| *Veronica* subg. *Pseudolysimachium* | KBps54 | *V. orchidea* | + | s | + | + | — | + | — | + | — | — | — | w |
| *Veronica* subg. *Pseudolysimachium* | KBps57 | *V. orchidea* | w | s | + | w | — | + | — | + | — | — | — | w |
| *Veronica* subg. *Pseudolysimachium* | BF11726 | *V. incana* | w | s | + | w | — | + | — | + | — | — | — | — |
| *Veronica* subg. *Pseudoveronica* J. B. Armstr. | PGJ2878 | *V. speciosa* | s | s | + | s | + | s | s | s | — | — | — | — |
| *Veronica* subg. *Pseudoveronica* | HMM69 | *V. salicornioides* | s | s | + | s | + | s | s | s | — | — | — | — |
| *Veronica* subg. *Pseudoveronica* | HMM38 | *V. hectori* subsp. *coarctata* | w | s | + | s | + | s | s | w | — | s | — | s |
| *Veronica* subg. *Pseudoveronica* | HMM39 | *V. ochracea* | s | s | + | s | + | s | s | s | s | — | — | s |
| *Veronica* subg. *Pseudoveronica* | HMM40 | *V. planopetiolata* | s | + | + | s | + | s | s | s | — | s | — | s |
| *Veronica* subg. *Pseudoveronica* | HMM37 | *V. catarractae* | s | s | w | s | + | s | s | + | — | s | — | s |
| *Veronica* subg. *Stenocarpon* (Boriss.) M. M. Mart. Ort., Albach & M. A. Fisch. | LS1408 | *V. fruticans* | s | s | s | + | s | s | + | s | — | w | + | + |
| *Veronica* subg. *Stenocarpon* | DCA71 | *V. fruticulosa* | s | + | + | + | s | s | + | s | — | + | + | + |
| *Veronica* subg. *Synthyris* (Benth.) M. M. Mart. Ort., Albach & M. A. Fisch. | DCA124 | *V. missurica* | w | + | w | + | + | + | + | s | — | — | + | w |
| *Veronica* subg. *Veronica* | DCA114 | *V. officinalis* | w | w | s | w | w | + | w | w | — | — | + | w |

*Note*: + = successful amplification; — = no amplification; s = several bands; w = weak amplification.

[a]Abbreviations (collector numbers): BF = Bozo Frajman; DCA = Dirk C. Albach; HMM = Heidi M. Meudt; KB = Katharina E. Bardy; LS = Lena Struwe; PGJ = Phil Garnock-Jones.

[b]DNA samples are deposited at Carl von Ossietzky Universität Oldenburg (Germany).

## REFERENCES

**Albach DC, Martínez-Ortega MM, Fischer MA, Chase MW**. 2004. A new classification of the *Veroniceae*: Problems and possible solution. *Taxon*, 53: 429–452.

**Albach DC, Martínez-Ortega MM, Delgado L, Weiss-Schneeweiss H, Özgocke F, Fischer MA.** 2008. Chromosome numbers in Veroniceae (Plantaginaceae): Review and several new counts. *Annals of the Missouri Botanical Garden*, 95: 543–566.

**Doyle JJ, Doyle JL**. 1987. CTAB DNA extraction in plants. *Phytochemical Bulletin*, 19: 11 – 15.

**Excoffier L, Lischer HEL**. 2010. Arlequin suite version 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10: 564–567.

**Garnock-Jones P, Albach DC, Briggs G.** 2007. Botanical names in Southern Hemisphere *Veronica* (*Plantaginaceae*): sect. *Detzneria*, sect. *Hebe*, and sect. *Labiatoides. Taxon*, 56: 571–582.

**Lehmann E.** 1937. Die Gattung *Veronica* in entwicklungsgeschichtlicher Betrachtung. *Cytologia*, *(Fujii Jubilaei Volumen):* 903–919.

**Malausa T, Gilles A, Meglécz E, Blanquart H, Duthoy S, Costedoat C, Dubut V, Pech N, Castagnone-Sereno P, …, Martin JF**. 2011. High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Molecular Ecology Resources*, 11: 638–644.

**Meglécz E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, Martin JF**. 2010. QDD: A user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics (Oxford, England)* 26: 403–404.

**Peñas De Giles J, Martínez-Ortega MM, Pérez LaTorre AV, Cabezudo Artero B.** 2004. *Veronica tenuifolia* subsp. *fontqueri* (Pau) M. M. Mart. Ort. and E. Rico, in: Bañares A, Blanca G, Güemes J, Moreno JC, Ortiz S, (eds.), Atlas y Libro Rojo de la Flora Vascular Amenazada de España. Madrid: Dirección General de Conservación de la Naturaleza, Madrid, pp. 564–565.

**Petrova A, Vladimirov V.** 2009. Red List of Bulgarian vascular plants. *Phytologia Balcanica*, 15: 63–94.

**Rojas-Andrés BM, Albach DC, Martínez-Ortega MM**. 2015. Exploring the intricate evolutionary history of the diploid-polyploid complex *Veronica* subsection *Pentasepalae* Benth. (Plantaginaceae). *Botanical Journal of the Linnean Society* 179: 670–692

**Scheerer H**. 1949. Zur Polyploidie und Genetik der *Veronica*—Gruppe Pentasepala. *Planta*, 37: 293–298.

**Schuelke M**. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, 18: 233–234.

## SUPPORTING INFORMATION

Supporting information includes:

**Appendix 1.** Voucher information for the *Veronica* samples used in this study.

**Appendix 2.** Primers rejected during the study and reason for discarding

**Appendix 1**. Voucher information for the *Veronica* samples used in this study.

| Species | Collector no. (Herbarium code) [a,b] | Collection country and locality | Geographic coordinates |
|---|---|---|---|
| *V. austriaca* L (*n*=15) | BR94 (SALA) | Croatia. Gračac, Crnopac | 44º15'02.2"N, 15º48'35.5"E |
| *V. catarractae* G. Forst. (*n*=1) | HMM37 (OLD) | cult. Germany ex UK nursery "Botany Plants"; stock. Botanical Garden, Oldenburg | NA |
| *V. chamaedrys* L. subsp. *chamaedryoides* (Bory & Chaub.) M.A.Fisch (*n*=1) | KBch667 (WU) | Greece. Olympia | 37º51'47.0"N, 21º48'45.0"E |
| *V. cymbalaria* Bodard (*n*=1) | DCA403 (WU) | Greece. Vourakis | NA |
| *V. cymbalaria* (*n*=1) | HMM31 (OLD) | Turkey. Alanya Castle | 36º31'58.0"N, 31º59'25.0"E |
| *V. cymbalaria* (*n*=1) | HMM32 (OLD) | Turkey. Selge | 37º13'04.0"N, 31º07'45.0"E |
| *V. dentata* F. W. Schmidt (*n*=14) | BR178 (SALA) | Austria. Niederösterreich, Krems | 48º24'18.1"N, 15º31'04.4"E |
| *V. filiformis* Sm. (*n*=1) | DCA144 (WU) | Germany. Bonn-Venusberg | 50º41'43.0"N, 07º06'10.0"E |
| *V. filiformis* (*n*=1) | DCA954 (MJG) | Turkey. Cam Pass | 41º13'33.0"N, 42º27'44.0"E |
| *V. filiformis* (*n*=1) | DCA892 (MJG) | Turkey. Uzungoel | 40º35'00.0"N, 40º19'00.0"E |
| *V. fruticans* Jacq. (*n*=1) | LS1408 (WU) | USA. Seedling. Botanical Garden, New York. | NA |
| *V. fruticulosa* L. (*n*=1) | DCA71 (BONN) | Germany. Seedling. Botanical Garden, Bonn | NA |
| *V. gentianoides* Vahl. (*n*=1) | DCA350 (WU) | Georgia. Terek-Tal | 42º34'51.6"N, 44º25'12.0"E |
| *V. gentianoides* (*n*=1) | DCA297 (WU) | Georgia. Kreuzpass | 42º31'02.0"N, 44º28'00.0"E |
| *V. gentianoides* (*n*=1) | MO1598 (SALA) | Georgia. Great Caucasus, Monument Bidara | 42º29'33.0"N, 44º27'10.0"E |
| *V. hectorii* Hook. F. subsp. *Coarctata* (Cheeseman) Garn. Jones (*n*=1) | HMM38 (OLD) | cult. Germany ex New Zealand. Botanical Garden, Bonn | not available |
| *V. incana* L. (*n*=1) | BF11726 (WU) | Serbia. Grgurevci | 45º06'36.0"N, 19º40'05.0"E |
| *V. jacquinii* Baumg. (*n*=2)[c] | BR108 (SALA) | Bosnia-Herzegovina. Trebinje | 42º41'02.1"N, 18º17'49.2"E |
| *V. jacquinii* (*n*=2)[c] | BR112 (SALA) | Croatia. Dubrovnik, Gromača | 42º43'28.0"N, 18º01'4.0"E |
| *V. jacquinii* (*n*=1)[c] | SA389 (SALA) | Montenegro. Kotor, Lovćen | 42º25'04.9"N, 18º47'38.8"E |

| Species | Collector no. (Herbarium code) [a,b] | Collection country and locality | Geographic coordinates |
|---|---|---|---|
| *V. jacquinii* (*n*=2)[c] | SA390 (SALA) | Montenegro. Kotor, Lovćen | 42º25'04.9"N, 18º47'38.8"E |
| *V. jacquinii* (*n*=1)[c] | SA391 (SALA) | Montenegro. Žabljak | 43º09'49.6"N, 19º09'00.3"E |
| *V. javalambrensis* Pau (*n*=30)[c] | DP1278 (SALA) | Spain. Burgos. Ciruelos de Cervera | 41º54'50.4"N, 3º29'47.9"W |
| *V. missurica* Raf. subsp. *major* (Hook) M.M.Mart Ort & Albach (*n*=1) | DCA124 (K) | England. Seedling. Botanical Garden, Kew | NA |
| *V. ochracea* (Ashwin) Garn. Jones (*n*=1) | HMM39 (OLD) | cult. Germany ex New Zealand. Botanical Garden, Bonn | NA |
| *V. officinalis* L. (*n*=1) | DCA14 (K) | England. Seedling. Botanical Garden, Kew | NA |
| *V. orbiculata* A. Kern. (*n*=1)[c] | BR110 (SALA) | Croatia. Peljesak península | 42º56'14.2"N, 17º22'39.5"E |
| *V. orbiculata* (*n*=1)[c] | SA392 (SALA) | Montenegro. Žabljak | 43º09'49.6"N, 19º09'00.3"E |
| *V. orbiculata* (*n*=2)[c] | MO5547 (SALA) | Croatia. Prapatnice | 43º13'16.1"N, 17º21'35.0"E |
| *V. orchidea* Crantz (*n*=1) | KBps57 (WU) | Bulgaria. Lovech | 43º01'59.0"N, 24º18'09.0"E |
| *V. orchidea* (*n*=1) | KBps574(WU) | Bulgaria. Lovech | 43º10'49.0"N, 24º44'56.0"E |
| *V. orchidea* (*n*=1) | KB847 (WU) | Hungary. Szabolcs-Szatmár-Bereg | 47º45'02.0"N, 21º52'02.0"E |
| *V. orsiniana* Ten. (*n*=30)[c] | MO6056 (SALA) | Spain. Teruel. Iglesuela del Cid | 40º27'35.9"N, 0º18'46.5"W |
| *V. panormitana* Tineo ex Guss (*n*=1) | HMM29 (OLD) | Turkey. North of Paravallar | 36º40'02.0"N, 31º53'03.0"E |
| *V. planopetiolata* G. Simpson & J.S. Thompson (*n*=1) | HMM40 (OLD) | New Zealand. Shotover Saddle | 44º31'21.6"S, 168º40'24.0"E |
| *V. rosea* Desf. (*n*=30)[c] | DP1368 (SALA) | Morocco. Meknès-Tafilalet, Midelt | 32º36'21.1"N, 4º48'39.7''W |
| *V. salicornioides* Hook f. (*n*=1) | HMM69 (OLD) | cult. Kew ex New Zealand. Botanical Garden, Kew | NA |
| *V. speciosa* R. Cunn. ex A. Cunn. (*n*=1) | PGJ2878 (OLD) | cult. New Zealand ex cult. New Zealand. Wellington | NA |
| *V. trichadena* Jord & Fourr. (*n*=1) | HMM30 (OLD) | Spain. Mallorca, Camí des Raiguer | NA |
| *V. triphyllos* L. | DCAs434 (OLD) | Germany. Seedling botanical garden, Oldenburg | NA |
| *V. vindobonensis* M.A. Fisch. (*n*=1) | KKBch54 (WU) | Hungary. Heves megye | 47º50'19.0"N, 19º57'44.0"E |

*Note*: *n* = number of individuals used in the population genetic analyses; NA = not available.

[a]Abbreviations (collector numbers): BF = Bozo Frajman; BR = Blanca M. Rojas-Andrés; DCA = Dirk C. Albach; DP = Daniel Pinto-Carrasco; HMM = Heidi M. Meudt; KB = Katharina E. Bardy; LS = Lena Struwe; MO = M. Montserrat Martínez-Ortega; PGJ = Phil Garnock-Jones; SA = Santiago Andrés-Sánchez.

[b]Herbarium specimens are deposited at the herbaria of Universidad de Salamanca (SALA), Universität Wien (WU), University of Bonn (BONN), Royal Botanic Gardens, Kew (K), Johannes Gutenberg-

Universität (MJG), and Carl von Ossietzky Universität Oldenburg (OLD); DNA samples are deposited at Biobanco de ADN Vegetal (Universidad de Salamanca) and Carl von Ossietzky Universität Oldenburg (Germany).

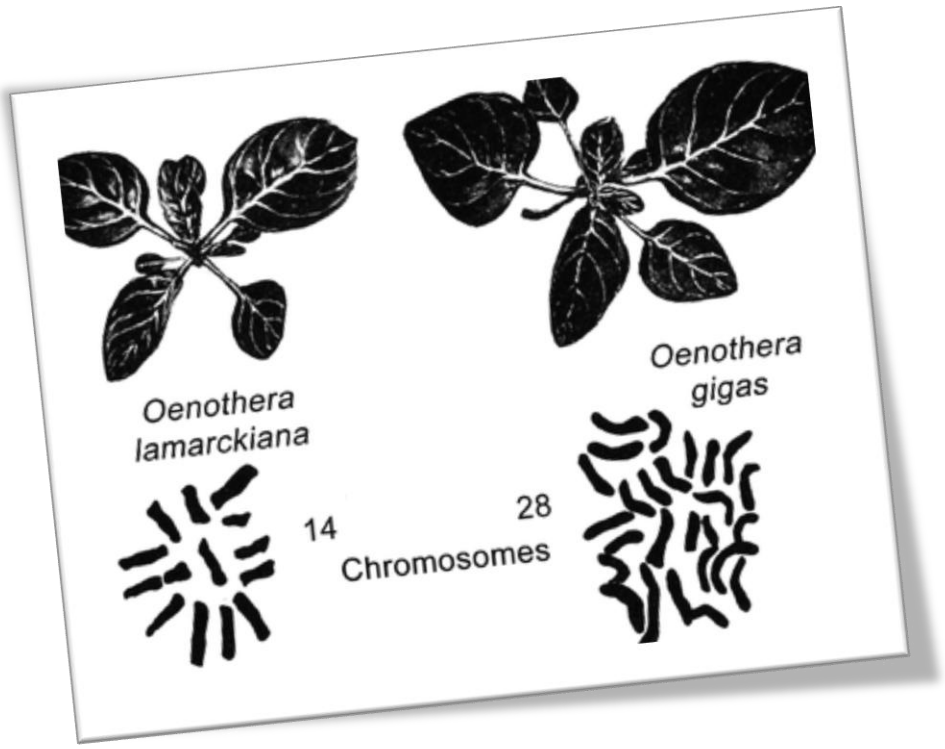
[c]Populations used to generate the data included in Appendix 2.

**Appendix 2.** Primers rejected during the study and reason for discarding

| Locus | Primer sequences (5'-3') | Repeat motif | PCR product size | GenBank accession no. | $T_a$ (ºC) | Discarding reason |
|---|---|---|---|---|---|---|
| 1 | F: TGATAGGGTTTGTGCGTGAG<br>R: TGTCGACCAAACCAAAACAA | $(TTG)_6$ | 146 | KT005181 | 52 | Suboptimal quality of the sequences |
| 2 | F: CCCTTTGGAGTTGTTATGATCG<br>R: GAATGAACGGTTTAAGTGGACA | $(AT)_5$ | 149 | — | — | Unsuccessful amplification |
| 3 | F: AACAAATCATAAGCAATGCCA<br>R: CGCTAGTGTCATCATGTTATGC | $(TA)_5$ | 208 | KT005182 | 58 | Monomorphic |
| 4 | F: AATTAAATTTCGCGGATCCTT<br>R: CGGTCTTACCAATGGCAGAT | $(TC)_{14}$ | 157 | — | — | Unsuccessful amplification |
| 5 | F: GCTGGAAAGAAAACCCAACA<br>R: TTGCATTGGATTTTGAACCA | $(ACA)_5$ | 104 | KT005183 | 50 | Suboptimal quality of the sequences |
| 6 | F: CGAAATCAGAATCAACACCAA<br>R: GAATCATCGATTGGGATCTTT | $(AAC)_6$ | 92 | KT005184 | 52 | Suboptimal quality of the sequences |
| 7 | F: CCCGAGTAGCGCTTGTTTTA<br>R: CACGAGTATGGGACGATTCA | $(TC)_8$ | 152 | — | — | Unsuccessful amplification |
| 9 | F: GCACGGAAACAACATGAACA<br>R: TCCCCATCATAATCACAATCA | $(AG)_8$ | 267 | KT005185 | 52 | Unsuccessful amplification on the Iberian clade |
| 11 | F: TTGTTGGTTTTGGTTTGTGG<br>R: GATGAACTCCAATCTACCCCA | $(CTT)_{12}$ | 91 | — | — | Unsuccessful amplification |
| 12 | F: GCCACGGAGACTCAGGTTAG<br>R: TGACGAATAGCAATAGACAACGA | $(GTT)_5$ | 132 | KT005186 | 55 | Suboptimal quality of the sequences |
| 14 | F: AAAGATAATTGTCCTAAAGTTAAGGGG<br>R: GCAGCATTATGCAGGTAGATT | $(ATGG)_6$ | 140 | — | — | Unsuccessful amplification |
| 15 | F: ACGCTTGAACGCGTCTAACA<br>R: AGATCCCCACTCACGATCTC | $(GT)_6$ | 144 | KT005187 | 54 | Monomorphic |
| 16 | F: ATCGAGGACGGATTTAGGCT<br>R: AAGTGCCCTTTCCTCCAAAC | $(GTA)_5$ | 113 | KT005188 | 56 | Monomorphic |
| 17 | F: GAGTGATCGAAAGATTGCATTAAG<br>R: TCCTCCCTAATTCCTCCGAC | $(GTG)_5$ | 148 | KT005189 | 54 | Suboptimal quality of the sequences |
| 18 | F: TTGAATATCAGGATCTTGTGCG<br>R: AAGTAATATGTCCATAAGTTCATCAGG | $(TCT)_6$ | 91 | KT005190 | 58 | Suboptimal quality of the sequences |
| 21 | F: AGAGGATGAAGACTCAGGCG<br>R: TGTCAGCTTTGGTGGAAGAA | $(GAA)_9$ | 140 | — | — | Unsuccessful amplification |
| 22 | F: GACGACGATCATCCAGATCC<br>R: CCGATTTCCTTTCGAATCAT | $(AGA)_6$ | 147 | KT005191 | 52 | Presence of indels |
| 23 | F: AAACTTGTGAAACTGTTTGAATGG<br>R: ATGCTCAGCGGAAGTATTTGA | $(CA)_5$ | 90 | — | — | Unsuccessful amplification |

| Locus | Primer sequences (5'-3') | Repeat motif | PCR product size | GenBank accession no. | $T_a$ (ºC) | Discarding reason |
|---|---|---|---|---|---|---|
| 24 | F: TTCCGATATTTCCGTTCTGC<br>R: CCATTCTACCCTCCGAACAA | $(GAG)_6$ | 142 | KT005192 | 52 | Presence of indels |
| 25 | F: GCACAAGGTAGCATTTGCATT<br>R: AGGGCGGGTAAAGGATAGAA | $(TTG)_9$ | 142 | — | — | Unsuccessful amplification |
| 28 | F: GTGTTCGTGTTTTAAATTTGCTT<br>R: TCACTCATATACCTAGTGACTGAACTG | $(GAG)_{11}$ | 141 | — | — | Unsuccessful amplification |
| 29 | F: TTGAATCCATTTCTTATTGGTTTG<br>R: CAATCGTGGTAACACATCATGG | $(TTC)_7$ | 90 | KT005193 | 53 | Unsuccessful amplification on the Iberian clade |
| 30 | F: CTTCCTTACCTCACCTCACTCTG<br>R: TGGTGTTTTGTTGATAGATTGATT | $(CAT)_5$ | 91 | KT005194 | 53 | Suboptimal quality of the sequences |
| 31 | F: GCCATTGCCTTGTTTTGAGT<br>R: CATCAACCATGATCCATCCA | $(GA)_9$ | 91 | — | — | Unsuccessful amplification |
| 32 | F: ATTGAGCGACACTCGTCAGA<br>R: CAATGGCTTTAAATGAATCCC | $(AC)_7$ | 140 | KT005195 | 52 | Monomorphic |
| 33 | F: TTCAGCTCATGACCAAGAACA<br>R: CAAATAGGGCATTCCGACAT | $(AAG)_6$ | 123 | KT005196 | 50 | Unsuccessful amplification on the Iberian clade |
| 34 | F: TAAACAAACAGATTGGTGGTCG<br>R: CCTTATGTCACTGAAAACCTACCT | $(TAA)_6$ | 190 | KT005197 | 54 | Unsuccessful amplification on the Iberian clade |
| 36 | F: CGGTGCCAAATTAAGATATTG<br>R: GCGGTGAAGAAAGGTTTTGA | $(ACTC)_5$ | 182 | — | — | Unsuccessful amplification |
| 37 | F: TGCACCCCTACTCGAGAAAT<br>R: TCCATTTAATTGTAAGCCCCA | $(CT)_8$ | 120 | — | — | Unsuccessful amplification |
| 38 | F: ACAGGTTGTGCGGAAGAAGT<br>R: GTGTGCCAACAAATCAAGGA | $(TGT)_9$ | 155 | KT005198 | 52 | Suboptimal quality of the sequences |
| 39 | F: GAAAAGAATTACCAACACGC<br>R: TTAAGGCCTAGCTAGCAGAA | $(AAAG)_6$ | 93 | — | — | Unsuccessful amplification |
| 40 | F: ATCTCCAAAACTCAGATCCA<br>R: TTAAGGCCTAGCTAGCAGAA | $(AAC)_6$ | 86 | — | — | Unsuccessful amplification |
| 41 | F: TCATAGCTTCTTCTCTTCGG<br>R: TATGATGGCCTTCAAAACAT | $(CTT)_5$ | 85 | — | — | Unsuccessful amplification |
| 42 | F: TGTATTATTCTATGAGACGCCA<br>R: GTGAGAAGACATATGAAAAGCA | $(TG)_{16}$ | 193 | KT005199 | 52 | Suboptimal quality of the sequences |
| 43 | F: ACGATAACTTTCCGGTGAA<br>R: CAACCATTTTCTTCATACACAG | $(GA)_8$ | 179 | — | — | Unsuccessful amplification |
| 44 | F: CTTTTAAATGTCTTTCTGGAGG<br>R: ATGTCCTTCATAGTAAACGTCC | $(TTG)_5$ | 179 | KT005200 | 52 | Monomorphic |
| 45 | F: CTTATCCTTGAATTTCATCTCC<br>R: GATTATTTTACGGTTAGACGGA | $(ACA)_6$ | 174 | KT005201 | 52 | Presence of indels |
| 46 | F: AAGCTTGAGTGGATTAAATGTT<br>R: AACTCTTACCACCTCAAATCAC | $(GTT)_6$ | 239 | KT005202 | 55 | Presence of indels |
| 47 | F: AGTAATCAATTCTCACTTGGCT<br>R: ACAACCCTAGTTCATACCAAAG | $(TC)_5$ | 236 | KT005203 | 53 | Monomorphic |
| 48 | F: TGAACAAATGTACAGCTAGAGG<br>R: GATGAGGAGAAGGAGTGTATGT | $(TG)_9$ | 246 | KT005204 | 54 | Presence of indels |
| 51 | F: ATTGTTGTATATGCGAATCTTG<br>R: TTCCATGTAAATTTCACTACCA | $(CA)_8$ | 303 | — | — | Unsuccessful amplification |
| 53 | F: GAATACATTCAGACCACGTCTT<br>R: AAACGATAGAGTCTCAAGAGGA | $(TC)_8$ | 301 | KT005205 | 52 | Unsuccessful amplification on the Iberian clade |

*Note* : — = no information available; *T* a = annealing temperature




# CHAPTER 7

## Main conclusions

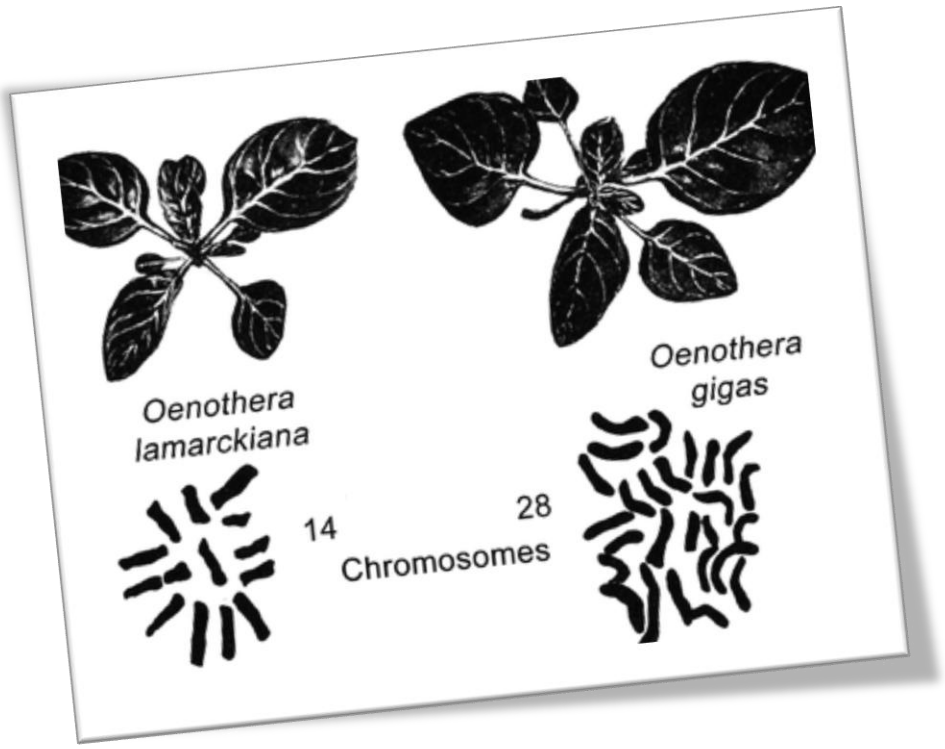
## MAIN CONCLUSIONS

1. Starting from individuals and populations belonging to *V.* subsect. *Pentasepalae*, that were determined according to the most recent taxonomic treatment available, the "divide and conquer" strategy was implemented, which led to the establishment of several "morphological groups" and, thus, to the creation of a reference taxonomic framework. The most informative diagnostic characters were used to divide the initial dataset (i.e., morphometric measurements) and progressively decrease its complexity. The misclassification rate was low, but not zero, due to the particularities of the data set, mainly because it included taxa that show high levels of morphological variability (e.g. allopolyploid taxa).
2. The recursive partitioning methodologies seem to be reliable for assigning a population to a taxon, either by conventional multivariate analysis, by implementing data-mining approaches, or by combining the two methods, which should not be considered mutually exclusive. Discriminant trees are highly useful for effective and fast initial approximations, because no previous knowledge on the group is required and successful results are achieved even with closely related species. Discriminant analysis constitutes an extraordinarily robust technique that should always be considered in the search of morphological evidence for classification purposes. A novel use of artificial neural networks is proposed to evaluate the adequacy of an input set of variables to classify the dependent variables. Finally, methodological guidelines to find robust morphological characters to differentiate among closely related taxa are proposed.
3. Chromosome numbers based on chromosome counts and/or ploidy level estimations for twelve species and infraspecific taxa within *V.* subsect *Pentasepalae* are

provided. Those referring to *V. dalmatica* and *V. kindlii* represent the first chromosome counts for the species. In the cases of *V. thracica* and *V. prostrata*, the first ploidy level estimations for these species in Turkey and Romania, respectively, are given.

4. A set of polymorphic microsatellite markers for *Veronica* subsect. *Pentasepalae* has been reported. Amplification success for these markers in the cross-transferability tests extends their potential usefulness to other subgenera. These markers can be used to investigate genetic parameters and evolutionary processes, such as the role of interspecific hybridization in the evolution of the genus.
5. Conclusive evidence on the formation of the '*V. austriaca – V. orbiculata* complex' as a result from both auto- and allopolyploid processes is provided. The results based on SSR markers and cpDNA data indicate a clear genetic differentiation between the diploid and hexaploid cytotypes traditionally included within the variation of *V. austriaca* subsp. *jacquinii*. Two well-established entities at the diploid level are recognized, namely *V. dalmatica* (i.e., traditionally considered to be the diploid cytotype of *V. austriaca* subsp. *jacquinii*) and *V. orbiculata*. Moreover, a cryptic allotetraploid lineage within the '*V. austriaca – V. orbiculata* complex' is here identified for the first time. The latter and *V. dalmatica* are shown to be two of the putative parental taxa of the allohexaploid *V. austriaca* subsp. *jacquinii*.
6. Different refugial areas are hypothesized for the survival of *V. orbiculata* and *V. dalmatica* during the Last Glacial Maximum. Different ecological conditions in northern and southern parts of the Western Balkans and posterior differentiation by isolation by distance might have been responsible for the divergence of these two species. Allopolyploidization events may have occurred afterwards, as a result of range expansions of these species during the postglacial periods.

7. *Veronica austriaca* s.l. displays high genetic diversity, ecological tolerances and morphological variability. The ecological tolerance of *V. austriaca* s.l. is the result of the sum of the notably divergent requirements shown by the three molecular lineages distributed across the different ecoregions represented in the Balkan Peninsula. The morphological variability of the leaf characters displayed by *V. austriaca* s.l. is suggested to be the result of adaptation to different environments.
8. The range expansion of *V. austriaca* s.l. across the Balkan Peninsula and beyond may be explained by one or several ecologically successful allopolyploid events. It is possible that one allopolyploidization event led to increased ecological amplitude resulting in a rapid colonization of new locations. Nevertheless, it cannot be ruled out that multiple allopolyploidization events may have promoted range expansion by generating different outcomes displaying distinct ecological tolerances.




# CAPÍTULO 7

## Conclusiones principales

## CONCLUSIONES PRINCIPALES

1. Tomando como punto de partida individuos y poblaciones pertenecientes a *V.* subsect *Pentasepalae*, que fueron determinados de acuerdo con el trataamiento taxonómico mas recente disponible, se implementó la estrategia "divide y vencerás", que llevó al establecimiento de varios grupos morfológicos" y, por lo tanto, a la creación de un marco taxonómico de referencia. Se utilizaron los caracteres diagnósticos más informativos para divider el conjunto inicial de datos (medidas morfométricas) para disminuir progresivamente su complejidad. La tasa de error en la clasificación fue baja, pero no cero, debido a las particularidades el set de datos que incluye taxones con un alto nivel de variabilidad morfológica (p.ej. taxones alopoliploides).
2. Las metodologías de particionamiento recursive parecen ser fiables para la asignación de poblacciones a taxones concretos, mediante análisis multivariantes convencionales, implementando técnicas de minería de datos o utilizando una combinación de ambas herramientas, que no deberíar ser consideradas excluyentes. Los árboles discriminantes son altamente efectivos para, al menos, las aproximaciones iniciales, ya que no requieren ningún tipo de conocimiento previo y aportan resultados considerablmente exitosos incluso en species intimamente relacionadas. Los análisis discriminantes se presentan como una herramienta extraordinariamente robusta que debería ser siempre considerada en la búsqueda de caracteres morfológicos distintivos para cuestiones de clasificación. Se propone un nuevo uso de las redes neuronales artificiales: utilizar esta herramienta para valorar el nivel adecuación de un conjunto de variables de entrada para clasificar las variables de salida. Finalmente, se aportan unas nociones metodológicas para buscar

caracteres morfológicos estadísticamente robustos que sirvan para discriminar entre especies íntimamente relacionadas.

3. Se aportan los numeros cromosómicos basados en conteos cromosómicos y/o estimaciones del nivel de ploidía para doce especies y taxones infraespecíficos pertenecientes a *V.* subsect *Pentasepalae*. Los referidos a *V. dalmatica* y *V. kindlii* representan los primeros conteos para dichas species. En el caso de *V. thracica* y *V. prostrata* se presentan las primeras estimaciones de nivel de ploidía para las especies en Turquía y Rumanía respectivamente.
4. Se proporciona un conjunto de marcadores polimórficos de tipo microsatélite para *Veronica* subsect. *Pentasepalae*. Su utilidad se extiende a otros subgéneros de acuerdo a los resultados positivos en los tests de transferibilidad. Estos marcadores pueden emplearse para investigar parámetros genéticos y procesos evolutivos como el papel que la hibridación interespecífica puede tener en la evolución del género.
5. Se aportan evidencias concluyentes de la formación del 'complejo *V. austriaca – V. orbiculata* complex' como resultado de eventos de auto- y allopoliploidía. Los resultados basados en marcadores microsatellites y secuencias de and del plasto indicant una clar diferenciación entre los citotipos diploide y hexaploide tradicionalmente incluidos en la variación de *V. austriaca* subsp. *jacquinii*. En el nivel diploide, se reconocen dos entidades bien establecidas: *V. dalmatica* (tradicionalmente considerada el citotipo diploide de *V. austriaca* subsp. *jacquinii*) y *V. orbiculata*. Además se identifica por primera vez un linaje críptico alopoliploide dentro del 'complejo *V. austriaca – V.orbiculata*'. Este linaje junto con *V. dalmatica* se proponen como las supuestas entidades parentales del alohexaploide *V. austriaca* subsp. *jacquinii*.

6. Se plantean distintas areas refugio implicadas en la supervivencia de *V. orbiculata* y *V. dalmatica* durante el último máximo glacial. Las distintas condiciones ecológicas en las partes norte y sur de la costa oeste de Balcanes y la posterior diferenciación mediante aislamiento por distancia podrían haber sido la causa de la divergencia de estas dos especies. Los eventos de alopoliploidazión podrían haber ocurrido después como resultado de las expansiones en el rango de distribución de estas especies durante el periodo postglacial.
7. *Veronica austriaca* s.l. presenta altos niveles de diversidad genetica, tolerancia ecológica y variabilidad morfológica. La tolerancia ecológica de *V. austriaca* s.l. es el resultado de la suma de los requerimientos, notablemente divergentes, mostrados por los tres linajes moleculares distribuidos a lo largo de las distintas ecoregiones presentes en la península balcánica. Se sugiere que la variabilidad morfológica de los caracteres foliares observada en *V. austriaca* s.l. podría ser resultado de la adaptación a los diferentes ambientes.
8. El rango de expansión of *V. austriaca* s.l. a lo largo la peninsula balcánica y fuera de esta área, podria ser explicado por uno o varios eventos de alopoliploidía exitosos en términos ecológicos. Es posible que un evento de alopoliploidización diese como resultado un organismo que contase con un incremento en su tolerancia ecológica, lo que le podría haber permitido una rápida colonización de nuevas localidades. Sin embargo, no puede ser descartado que el incremento en la tolerancia ecológica sea el resultado de varios eventos de alopoliploidización que hayan dado lugar a generar diferentes organismos poliploides con capacidades ecológicas distintas y divergentes.