Botanical Journal of the Linnean Society, 2020, **XX**, 1–17. With 6 figures.

Cytogenetic diversity in the polyploid complex *Linum suffruticosum s.l.* (Linaceae)

ANA AFONSO[1,*], JOÃO LOUREIRO[1], JUAN ARROYO[2], ERIKA OLMEDO-VICENTE[2] and SÍLVIA CASTRO[1]

[1]*Centre for Functional Ecology, Department of Life Sciences, University of Coimbra, Calçada Martim de Freitas, 3000–456 Coimbra, Portugal*
[2]*Departamento de Biología Vegetal y Ecología, Universidad de Sevilla, 41004 Sevilla, Spain*

Received 21 January 2020; revised 20 May 2020; accepted for publication 17 June 2020

Polyploidy plays a significant role in the evolution and diversification of flowering plants. In several polyploid complexes, high morphological variability and plasticity coupled with cytogenetic diversity make it difficult to disentangle their evolutionary history. The main goal of this study was to gain insights into the role of whole genome duplications as one of the factors shaping the evolution of flowering plants. *Linum suffruticosum s.l.* has been described as a polyploid complex, with high morphological variability, but nothing is known about current cytogeographical patterns. We investigated cytotype diversity and distribution patterns in 151 populations covering most of the distribution range, in the Iberian Peninsula, south-eastern France, north-western Italy and Morocco, using flow cytometric analyses complemented with chromosome counts. A high cytogenetic diversity was found with five major cytotypes being detected (diploids, tetraploids, hexaploids, octoploids and decaploids) and with new ploidy levels being reported for the first time. The different ploidies were distributed parapatrically, with geographical structure and several contact zones. Most of the populations comprised one cytotype, but a few mixed-ploidy populations were observed. Our results suggest that whole genome duplications are one of the key mechanisms, alone or together with hybridization, governing the diversification of *L. suffruticosum s.l.* Genome size and/or chromosome counts might be useful tools for identifying specimens of *L. suffruticosum s.l.* Also, geographical overlap and high cytogenetic diversity suggest multiple origins of the polyploids. The diversity observed here has been mostly neglected to date and should be accounted when studying the biosystematics of this complex.

ADDITIONAL KEYWORDS: chromosome counts – contact zones – evolutionary history – genome size – Mediterranean plants – ploidy.

INTRODUCTION

Polyploidy plays a significant role in the evolution and diversification of flowering plants (Ramsey & Schemske, 1998; Soltis & Soltis, 1999; Otto & Whitton, 2000; Blanc & Wolfe, 2004; Madlung, 2013). This widespread phenomenon is observed in the evolutionary history of virtually all flowering plants, being frequent in many plant lineages (Soltis, 2005) and correlated with explosions in species diversity (Soltis *et al.*, 2009). Estimates suggest that a high percentage of speciation events in angiosperms has been associated with ploidy increases (Wood *et al.*, 2009), and there is evidence that some polyploid taxa have multiple origins (e.g. Soltis, Doyle & Soltis, 1992; Kolář *et al.*, 2009; Chelaifa, Monnier & Ainouche, 2010; Castro *et al.*, 2018; Wan, Guo & Rao, 2019). Estimates of the incidence of polyploidy in current floras also reveal high levels of polyploid taxa in certain regions (e.g. 37% in the Mediterranean region and 49% in the Iberian Peninsula; Marques *et al.*, 2018). The Mediterranean Basin is considered a cradle where polyploidy has frequently occurred through the evolutionary history of plants groups thriving in these territories, linked to its dynamic palaeogeographic and climatic history (e.g. Late Miocene Salinity Crisis, spread of a Mediterranean-type climate at the Pliocene and Pleistocene Ice Ages; Thompson, 2005). Additionally, the detection of taxa with multiple ploidies (e.g. Buggs & Pannell, 2007; Balao *et al.*, 2009; Kolář *et al.*, 2009; Castro *et al.*, 2012, 2018, 2019; Kim *et al.*, 2012; Muñoz-Pajares *et al.*, 2018;

*Corresponding author. E-mail: ana.s.s.afonso@gmail.com

Prančl *et al.*, 2018) supports the idea that polyploidy is a dynamic and ongoing process in nature (Ramsey & Schemske 1998, 2002; Soltis, 2005; Wood *et al.*, 2009; Marques *et al.*, 2018).

Whole-genome duplications generate a new entity reproductively isolated from the progenitor(s) and, thus, have been described as an important mechanism of sympatric speciation (Otto & Whitton, 2000; Soltis *et al.*, 2010). Polyploids arise through the duplication of genomes from the same species (autopolyploidy) or by the combination of genomes from two species (allopolyploidy; Ramsey & Schemske, 1998). The recognition of the origin of polyploids is, however, difficult in many occasions. Although allopolyploids typically have phenotypes differentiated from their progenitors and may be more easily detected as hybrids, autopolyploids may be nearly indistinguishable from their progenitors (Brochmann *et al.*, 2004; Doyle *et al.*, 2004; Soltis *et al.*, 2010; Spoelhof *et al.*, 2017). Multiple origins and recurrent hybridization and introgression may also increase the complexity of certain taxa in natural populations and generate intricate series of polyploids (e.g. Segraves *et al.*, 1999; Soltis & Soltis, 1999; Sampson & Byrne, 2012). Additionally, in several plant groups, the taxonomic identification of polyploids is problematic due to the lack of reliable diagnostic characters, high morphological variability and phenotypic plasticity (Brochmann *et al.*, 2004; Doyle *et al.*, 2004; Prančl *et al.*, 2018). These traits allied with polyploidization events significantly increase the difficulty to understand the evolutionary history of certain plant taxa. In this context, genome size can be an additional diagnostic character, helpful in recognizing polyploid series, with potential for being a tool for identifying different evolutionary histories and/or independent polyploidization events (Balao *et al.*, 2009; Kolář *et al.*, 2009). However, we know little about many polyploid complexes with major taxonomic problems and that have not been studied systematically throughout their entire distribution range, even though polyploidy has already been identified to have played an important role in their evolution.

Linum suffruticosum L. *s.l.* (Linaceae) is composed of perennial, variable woody plants with a complex floral dimorphism and breeding system (Nicholls, 1985a; Armbruster *et al.*, 2006), occurring in the Iberian Peninsula, south-eastern France, north-western Italy and north-western Africa (Fig. 1). The taxonomy of the group is complex, and it has been subjected to different taxonomic treatments over the years. This is mainly due to the high morphological variability observed in the group and the lack of strong/unambiguous diagnostic characters. The most recent treatment of the group of *L. suffruticosum s.l.* records high levels of variability and recognizes morpho-geographical divisions with numerous transitional areas, leading to a taxonomic treatment comprising > 20 taxa for the Iberian Peninsula alone (Martínez-Labarga & Garmendia, 2015). However, although this exhaustive division could be functional on a regional scale, it does not always work at a wider scale given the continuum of morphological variability. Among the previously available taxonomic treatments, three taxa have been consensually accepted as distinct species (*L. salsoloides* Lam., *L. appressum* Caball. and *L. suffruticosum*), with some varieties being described in the last species (e.g. Jahandiez & Maire, 1932; Ockendon & Walters, 1968; López González, 1979). Additionally, the group is monophyletic with uncertain phylogenetic relationships between all its entities (Ruiz-Martín *et al.*, 2018). *Linum suffruticosum s.l.* has been described as a polyploid complex with a basic chromosome number of 9, bearing diploids ($2n = 2x = 18$ chromosomes), tetraploids ($2n = 4x = 36$) and octoploids ($2n = 8x = 72$), all in the Iberian Peninsula (Supporting Information, Table S1). The available chromosome counts describe *L. salsoloides* and *L. appressum* as diploids, and *L. suffruticosum s.s.* as a polyploid complex including tetraploid and octoploid individuals (Supporting Information, Table S1). The available records already pointed for some segregation among cytotypes at a regional scale, namely tetraploids occurring in southern Spain and octoploids in northern regions (Nicholls, 1985a, b, 1986). However, information about the prevalence of each cytotype, its diversity and distribution patterns across the distribution range of the group is scarce due to poor sampling and limited cytogenetic information.

The main goal of this study was to explore whole-genome duplications as one of the factors shaping the evolution of the apparently self-incompatible, style dimorphic *L. suffruticosum s.l.* complex (Nicholls, 1985b, 1986). For that, we explored in detail the cytotype diversity and distribution patterns in *L. suffruticosum s.l.* throughout most of its distribution range (north-western Italy, southern France, Iberian Peninsula, northern Morocco). Flow cytometric analyses complemented with chromosome counts were used to address the following specific objectives: (1) describe the diversity of chromosome numbers, ploidies and genome sizes within the group; (2) explore the geographical distribution and variation of ploidy levels, including dominant and rare cytotypes, across the entire distribution range of the group; (3) explore the potential existence of contact zones among ploidies and mixed-ploidy populations and (4) evaluate the potential of genome size and ploidy as additional diagnostic characters for future

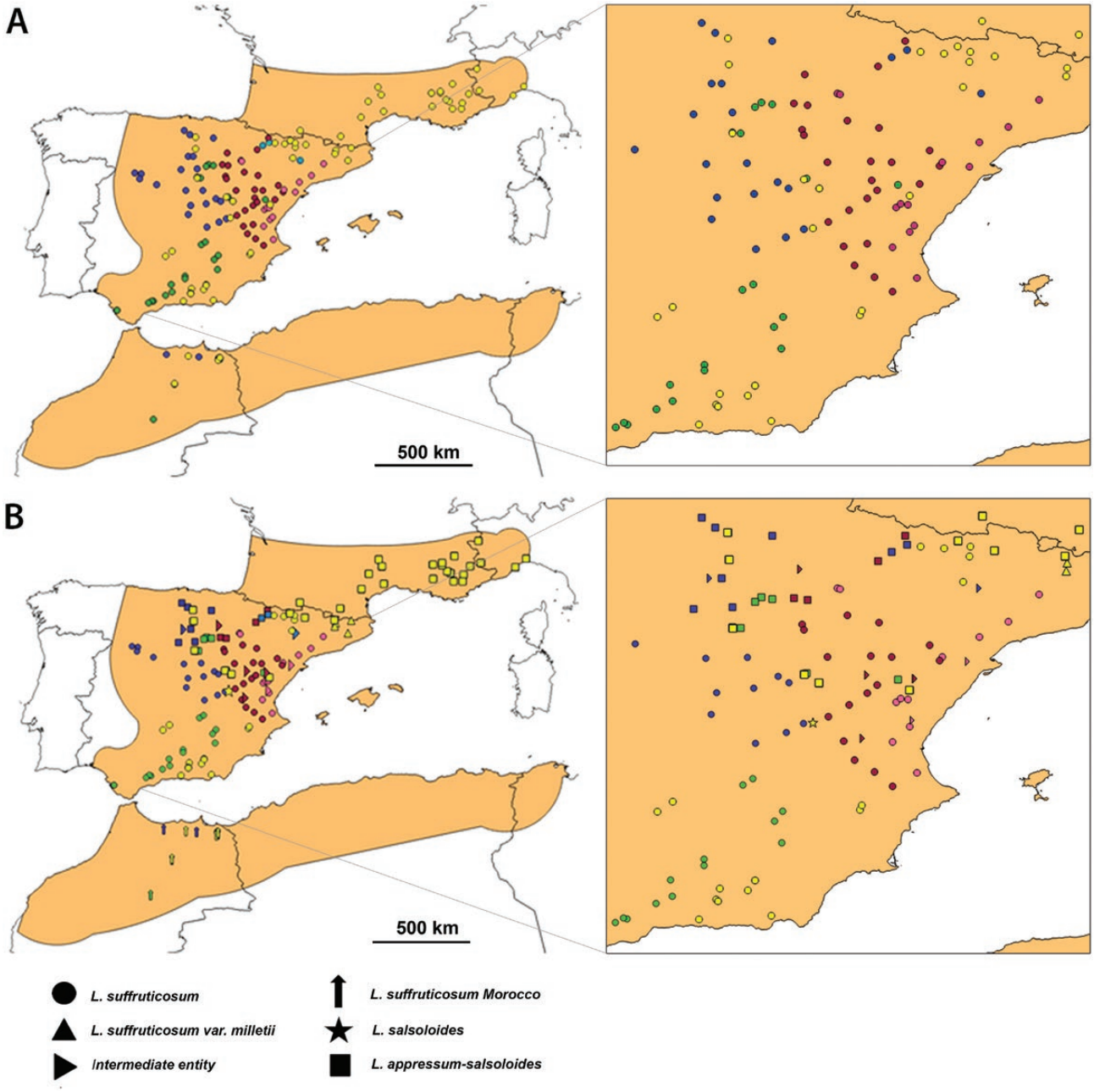

Figure 1. A, Geographical distribution of *L. suffruticosum s.l.* (orange) and of all populations sampled in the study area depicted with the respective ploidy (diploid: yellow; tetraploid: green; hexaploid with low genome size: dark blue; hexaploid with high genome size: light blue; octaploid: dark red; decaploid: pink). B, Geographical distribution of taxonomic entities of *L. suffruticosum s.l.* (*L. salsoloides*: star; *L. suffruticosum*: circles; *L. suffruticosum* var. *milletii*: upright triangles; *L. appressum-salsoloides*: squares; intermediate entity: triangles on their sides, *L. suffruticosum* Morocco: arrows) with ploidies presented with different colours as in A. The base map was downloaded from https://www.diva-gis.org/gdata.

taxonomic treatments. This study offers novel insights in the cytogenetic diversity of this complex with new key diagnostic characters, namely ploidy and genome size, and opens up new avenues for understanding the complex evolutionary pathways in *L. suffruticosum s.l.*

MATERIAL AND METHODS

FIELD SAMPLING

In south-western Europe, field sampling was carried out in two periods, in the flowering and fruiting seasons (May to July in the southerly most locations,

and August in most of the northern locations) of 2016 and 2017. Sampling in each population included the collection of herbarium vouchers for taxonomic confirmation and flower buds or recently open flowers for flow cytometric analyses of the petals. We used this tissue because, in contrast to other plant organs, petals did not have mucilaginous compounds, which hampered flow cytometric analyses as samples clogged the flow cytometer. Fresh flower buds or petals were collected in individual plastic bags and stored in a portable or a conventional refrigerator (for up to 7 days) until flow cytometric analyses. Up to 30 individuals (mean ± SD: 16 ± 11) were sampled per population. In the fruiting season, targeted populations encompassing all morphological and cytogenetic entities were revisited for the collection of seeds to be used for chromosome counting. Seeds from 30 individuals were collected in selected populations and stored in individual paper bags. In Morocco, field sampling was made in the flowering season of 2018 and included the collection of herbarium vouchers and petals. Geographical coordinates of all sampled localities were obtained and detailed information about all sites is provided in Supporting Information, Table S2. In total, 151 populations were sampled throughout most of the distribution range of the group (Fig. 1). Voucher specimens are deposited in COI and SEV herbaria.

Field sampling was designed to record most of regions were the group is present and the morphological variability described by taxonomic treatments. All specimens collected in the field were identified according to López González (1979) and Fennane *et al.* (2007) and assigned to four taxa: *L. suffruticosum* var. *milletii* (Sennen & Gonzalo) G.López, *L. suffruticosum*, *L. salsoloides* and *L. appressum-salsoloides*, the last including plants that could not be clearly assigned to either of these two species. With exception of *L. suffruticosum* var. *milletii* (a easily distinguishable variety from Catalonia), it was not possible to determine unambiguously the lower rank categories (varieties) of *L. suffruticosum* due to the occurrence of intermediate characters. Additionally, we were unable to use the taxonomic treatment of López González (1979) to identify the plants from Morocco, and thus these plants were classified as *L. suffruticosum* following the available literature for this region (Jahandiez & Maire, 1932; Emberger & Maire, 1941; Quézel & Santa, 1962; Fennane *et al.*, 2007; Valdés *et al.*, 2007). In one locality, only, the individuals were unambiguously identified as *L. salsoloides*, following López González (1979). For the remaining populations, individuals had intermediate characters between *L. appressum* [considered by López González (1979) as an Iberian endemic] and *L. salsoloides* (distributed in France and Italy; Ockendon & Walters, 1968), and thus we treated those specimens as *L. appressum-salsoloides*. Also, in a few populations in Spain, the identification of some specimens was dubious due to the occurrence of intermediate morphological characters between *L. suffruticosum*, *L. salsoloides* and *L. appressum* and, thus, these individuals were classified as intermediate entities.

Genome size and DNA ploidy estimates using flow cytometry

Genome size and DNA ploidy were assessed using flow cytometry. The methodology of Galbraith *et al.* (1983) was used to obtain nuclear suspensions. In brief, *c.* 50 mg of petal tissue of *Linum* L. was chopped together with 50 mg of leaves of an internal reference standard (*Solanum lycopersicum* L. 'Stupické', hereafter *S.l.*, with 2C = 1.96 pg; Doležel, Sgorbati & Lucretti, 1992) using a sharp razor blade in a glass Petri dish with 1 mL of WPB buffer (0.2 M Tris-HCl, 4 mM $MgCl_2.6H_2O$, 1% Triton X-100, 2 mM EDTA $Na_2.2H_2O$, 86 mM NaCl, 10 mM sodium metabisulphite, 1% PVP-10, pH adjusted to 7.5 and stored at 4 °C; Loureiro *et al.*, 2007). The nuclear suspension was filtered through a 50 µm nylon filter and 50 µg mL^{-1} of propidium iodide (PI; Fluka, Buchs, Switzerland) and 50 µg mL^{-1} RNAse (Fluka) were added to stain the DNA and avoid the staining of dsRNA, respectively (Doležel *et al.*, 2007). After 5 min of incubation, the samples were analysed in a Partec CyFlow Space flow cytometer (532-nm green solid-state laser, operating at 30 mW; Partec GmbH., Görlitz, Germany). The results were acquired using Partec FloMax software v.2.4d (Partec GmbH, Münster, Germany) in the form of four graphics: a histogram of the fluorescence pulse integral on the linear scale (FL); forward light scatter (FS) vs. side light scatter (SS), both in logarithmic (log) scale; FL vs. time and FL vs. SS in log scale. To remove debris, a polygonal region was defined in the FL vs. SS histogram and subsequently applied to all graphics. At least 1300 nuclei in both sample and standard G1 peaks were analysed per sample (Suda *et al.*, 2007). Only samples with coefficient of variation values of 2C peaks > 5% were accepted (data not shown); otherwise, a new sample was prepared and analysed until such quality standards were achieved (Greilhuber *et al.*, 2007).

In all populations, three to six individuals were analysed individually, enabling the estimation of the genome size of the population. For the remaining individuals, a pooled sample strategy was followed (five or six *Linum* individuals plus the reference standard) enabling to access the DNA ploidy. The holoploid genome size (2C in pg; *sensu* Greilhuber *et al.*, 2007) was calculated using the formula:

$$\textit{Linum}\ 2\text{C nuclear DNA content (pg)} = \frac{\textit{Linum}\ G_1\ \text{peak mean} \times \textit{S.l.}\ \text{genome size}}{\textit{S.l.}\ G_1\ \text{peak mean}}$$

The DNA ploidy of each individual was inferred from chromosome counts (see the section 'Chromosome count's) and genome size estimates obtained for the individuals in the population. The monoploid genome size (1C*x*; *sensu* Greilhuber *et al.*, 2005) was calculated in mass values (pg) by dividing the holoploid genome size (2C) by the assigned DNA ploidy. Populations were characterized according to the ploidy of their individuals and mapped.

Chromosome counts

Seeds from the selected populations (at least one for each genome size category; Supporting Information, Table S2) were germinated in Petri dishes. Actively growing root tips were harvested and pre-treated with ice at 4 ˚C in the dark for 24 h, and then root tips were fixed in a solution of 3:1 95% ethanol:glacial acetic acid for 48 h at room temperature. Root tips were then washed twice for 5 min with distilled water and incubated in acetic carmine for at least 48 h at room temperature. Finally, chromosomes were squashed under a glass cover in 45% acetic acid. Chromosome spreads were observed using a Nikon Eclipse 80i light microscope and photographed using a Nikon Plan Apo VC 100×/1.40 oil-immersion lens, with a Q Imaging Retiga 2000R Fast 1394 digital camera and Q-Capture Pro v.7 software. Chromosome counts were assigned to a genome size category, enabling the DNA ploidy of the remaining populations analysed to be estimated using flow cytometry.

Statistical analyses

Descriptive statistics of holoploid genome size were calculated for each cytotype (mean, standard deviation of the mean, maximum and minimum values) based only on individual flow cytometric estimates. Coefficient of variation (CV, in %) was calculated for each ploidy and taxon/entity as the ratio between standard deviation and the mean. To assess differences among cytotypes in holoploid and monoploid genome sizes, generalized linear models were used (Bolker *et al.*, 2009), with a Gaussian distribution and an identity link function to model the responses. Cytotype was used as factor and genome size as response variable. Statistical analyses were performed in R software v.3.6.1 (R Core Development Team, 2019), using the packages *car* for Type-III analysis of variance (Fox *et al.*, 2005), *glm* for generalized linear models (Hastie & Pregibon, 1992) and *multcomp* for multiple comparisons after Type-III analysis of variance (Hothorn *et al.*, 2017). Spatial correlation analysis for all cytotypes was evaluated with a Mantel test using the package *ade4* (Dray & Dufour, 2007). The same analysis was performed for each cytotype individually. The Mantel test provides a correlation coefficient between the two data matrices, namely the geographical distance matrix and the genome size distance matrix, with $P < 0.05$ indicating significant correlation between them and positive r values indicating positive association, i.e. more similar genome sizes are found geographically together.

RESULTS

Cytogenetic diversity in *Linum suffruticosum s.l.*

A large variation in genome size was observed, with 2C values ranging from 1.33 to 7.76 pg (Table 1, Fig. 2, $N = 729$ individuals, $N = 151$ populations). However, with a few exceptions, the variation was not continuous (Fig. 2; Table 1) and, with data on chromosome counts ($N = 134$ individuals, 53 populations), we were able to clearly assign the DNA ploidy to most genome size ranges (Figs 3–5, Supporting Information, Figs S1, S2; Table S3). Five main cytotypes were detected, namely diploids (2*x*), tetraploids (4*x*), hexaploids (6*x*), octoploids (8*x*) and decaploids (10*x*), with occasional triploids (3*x*) and aneuploids (an.) also being found (Table 2). The holoploid (2C) and monoploid (1C*x*) genome sizes of the dominant cytotypes differed significantly ($2C - F_{4,751} = 5123.2$, $P < 0.001$; $1Cx - F_{4,751} = 211.12$, $P < 0.001$; Fig. 2).

Variation in chromosome number and genome size was also observed within most of the ploidies (Table 1). Diploids comprised individuals with 16 or 18 chromosomes and genome sizes with 9% variation (ranging from 1.33 to 1.78 pg; Figs 2, 3, 4A, 5A, S2A). Tetraploids had $2n = 36$ chromosomes, rarely 32 or 38, and genome size showed a 25% variation (ranging from 2.63 to 3.63 pg; Figs 2, 4C, 5B, S2B). The genome size variation in the hexaploids was even higher, 51% (ranging from 3.64 to 5.51 pg), with $2n = 54$ chromosomes, and occasionally 48 chromosomes (Figs 2, 4D, 5C, S1A, S2C). Some of the genome size values from hexaploids fall within the range observed for octoploids, but their ploidy was confirmed with chromosome counts (Figs 2, 5D, S1B). Octoploids had a lower genome size variation when compared with some of the other ploidies (22%; ranging from 4.61 to 5.67 pg), and all

Table 1. Ploidies detected and observed chromosome numbers in *L. suffruticosum s.l.* Abbreviations: 2*x*, diploid; 3*x*, triploid; 4*x*, tetraploid; 6*x*, hexaploid; 8*x*, octoploid; 10*x*, decaploid; *N*, number of individuals analysed; 1C*x* G.s, mean and standard deviation of the mean of the monoploid genome size in picograms (pg); 2C G.s. mean and standard deviation of the mean of the holoploid genome size in picograms (pg); range, maximum and minimum values in pictograms (pg); CV, coefficient of variation calculated as the ratio of the standard deviation to the mean (in %). Chromosome numbers (individual chromosome numbers are separated by commas; rare chromosome counts are presented in parentheses).

Ploidy	*N*	1C*x* G.s. (pg)	2C G.s. (pg)	Range (pg)	CV (%)	Chromosome number(s)
2*x*	1144	0.80 ± 0.04	1.59 ± 0.09	1.33–1.78	5.57%	16, 18
3*x*	3	0.86 ± 0.04	2.59 ± 0.00	2.58–2.59	0.16%	27
4*x*	496	0.78 ± 0.06	3.11 ± 0.25	2.63–3.63	8.06%	(32), 36 (38)
6*x*	642	0.70 ± 0.09	4.21 ± 0.51	3.64–5.51	12.11%	(48) 54
8*x*	417	0.64 ± 0.03	5.13 ± 0.22	4.61–5.67	4.28%	72
10*x*	219	0.66 ± 0.05	6.64 ± 0.52	5.75–7.76	7.83%	90

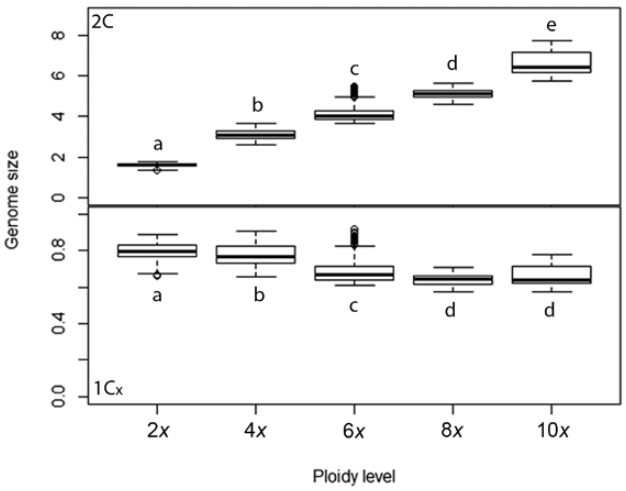

Figure 2. Holoploid and monoploid genome size range and mean (black line) of diploid, tetraploid, hexaploid, octoploid and decaploid populations. Abbreviations: 2C, holoploid genome size; 1C*x*, monoploid genome size; 2*x*, diploid; 4*x*, tetraploid; 6*x*, hexaploid; 8*x*, octoploid; 10*x*, decaploid. Outliers are given as white circles. Different letters correspond to statistically significant differences at $P < 0.05$.

analysed individuals had $2n = 72$ chromosomes (Figs 2, 4E, 5E, S1C). Finally, decaploids exhibited the highest genome size (52%) and intrapopulation variation (ranging from 5.75 to 7.76 pg), with individuals having $2n = 90$ chromosomes (Figs 2, 5D, S1D). Clearly, two basic chromosome numbers were observed, $n = 8$ and 9, with the former being rare.

Cytotype distribution patterns in *Linum suffruticosum s.l.*

Cytotypes had different distribution patterns across the range. Diploids were scattered through the entire distribution area (Fig. 1A). In contrast, polyploids were found mostly in the Iberian Peninsula (4*x*, 6*x*, 8*x* and 10*x*) and Morocco (4*x* and 6*x*), although complex contact zones are observed in central and northern Spain (Fig. 1A). In the Pyrenees, France and Italy, all populations were found to be diploid; in Spain, diploids were detected mostly in mountainous regions (one cluster in the Pyrenees and another in southern regions of Spain; Fig. 1A), with only a few populations being found at low elevations. Tetraploids were found mostly in southern Spain (Fig. 1A). Hexaploids have a more northern distribution in Spain, some populations reaching the western Pyrenees (Fig. 1A). Octoploids occur at lower elevation and in hotter regions in eastern Spain (Fig. 1A). Finally, decaploids mostly occur in areas nearer the north-east coast of Spain (not necessarily at low elevations), but some populations were also found in more inland areas in north-eastern Spain (Fig. 1A). Although this differential pattern of distribution of cytotypes, there is a non-significant spatial correlation of genome size for the whole sample of populations across the cytotypes (Fig. 1A, Mantel test: $r = 0.0252$, $P > 0.05$). This is probably due to the weight of non-significant correlation of diploid populations scattered through the entire range (Fig. 1A, Mantel test: $r = -0.032$, $P > 0.05$). In contrast, most polyploid populations showed significant positive correlations within their more restricted ranges (Fig. 1A, Mantel tests: tetraploids, $r = 0.845$, $P < 0.05$; hexaploids, $r = 0.180$, $P < 0.05$; octoploids: $r = 0.165$, $P < 0.05$). The non-significant correlation of decaploid populations (Fig. 1A, Mantel test: $r = -0.054$, $P > 0.05$) is probably due to the low number of populations found.

Most populations comprised only one cytotype (85% populations had one ploidy), but some mixed-ploidy populations were also detected (Table 2, Supporting Information, Fig. S3). Most mixed-ploidy populations are characterized by a dominant cytotype growing with another cytotype occurring at low frequency, e.g. a few triploids or tetraploids

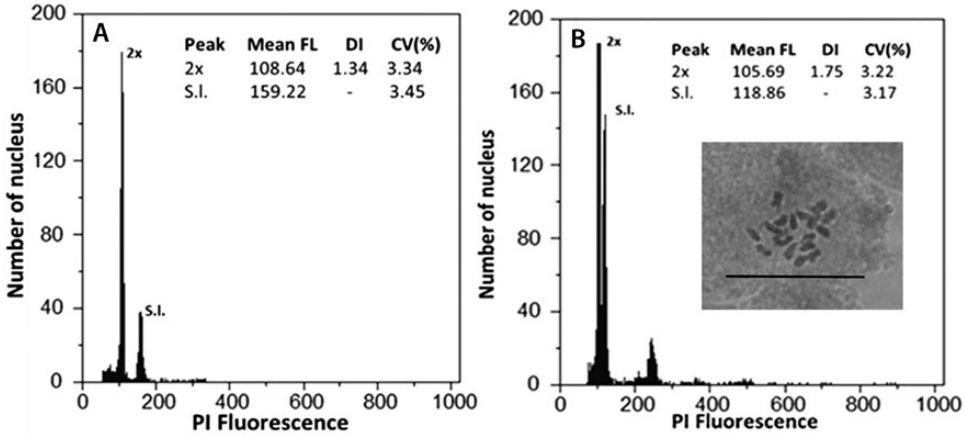

Figure 3. Genome size estimation and chromosome number of A, *L. salsoloides* and B, *L. suffruticosum* var. *milletii*. Abbreviations: 2*x*, diploid; *S.l.*, *Solanum lycopersicum*; Mean FL, mean relative fluorescence in picograms; DI, DNA index; CV(%), coefficient of variation of the peak in percent. Scale bar: 20 μm (black line).

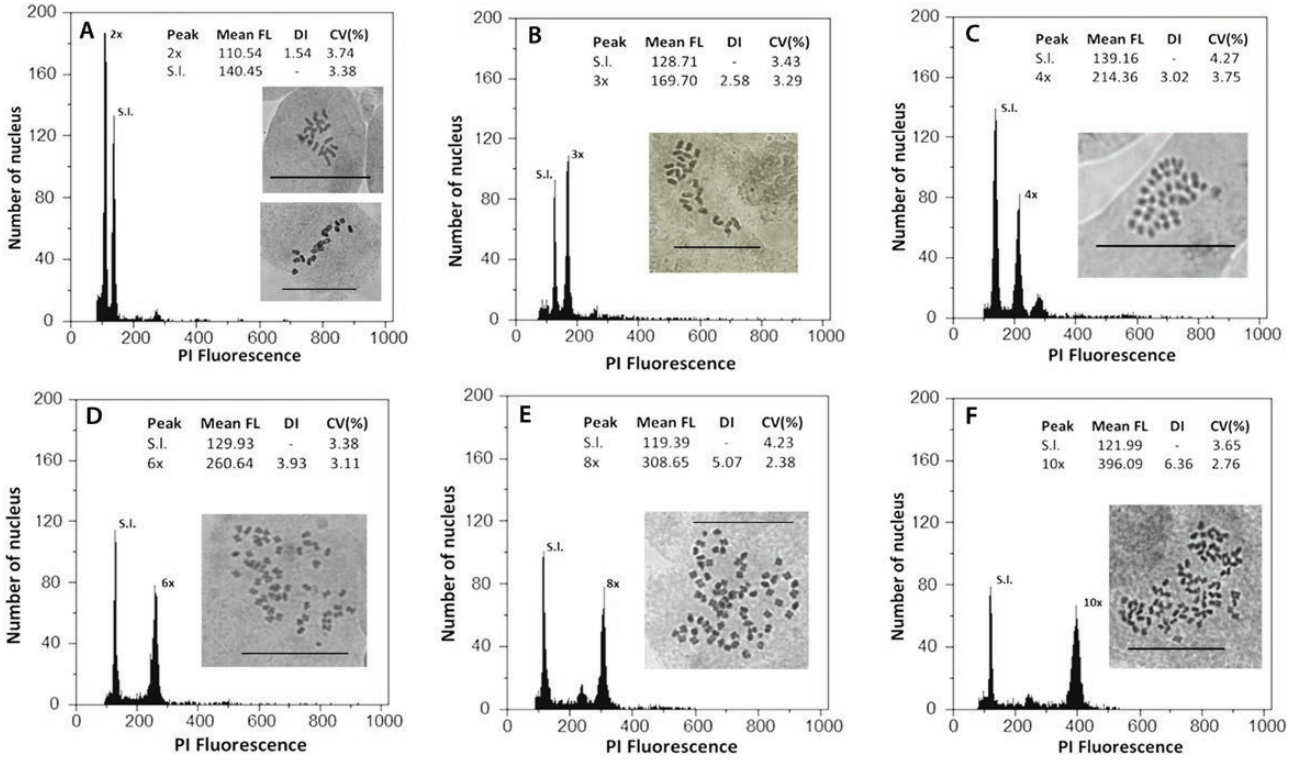

Figure 4. Genome size estimation and chromosome number of *L. suffruticosum* A, diploid; B, triploid; C, tetraploid; D, hexaploid; E, octoploid and F, decaploid. Abbreviations: 2*x*, diploid; 3*x*, triploid; 4*x*, tetraploid; 6*x*, hexaploid; 8*x*, octoploid; 10*x*, decaploid; *S.l.*, *Solanum lycopersicum*; Mean FL, mean relative fluorescence in picograms; DI, DNA index; CV(%), coefficient of variation of the peak in percent. Scale bar: 20 μm (black line).

growing in diploid populations (three localities each), a few tetraploids growing in hexaploid populations (six localities), a few hexaploids growing in octoploid populations (five localities) and a few octoploids in decaploid populations (two localities). The exceptions were one diploid-tetraploid population (AA87, with 19 2*x* individuals and 11 4*x* individuals) and one tetraploid-hexaploid population (DP1980, with five

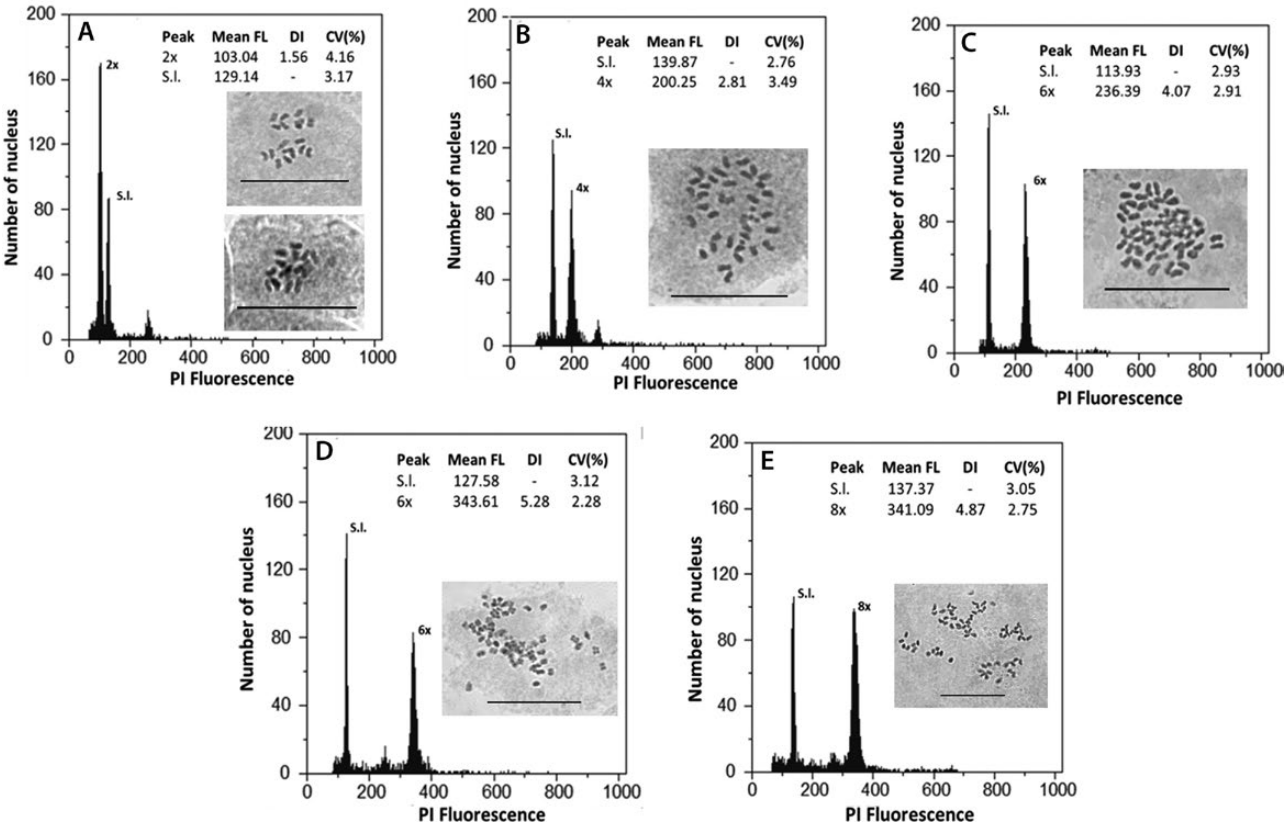

Figure 5. Genome size estimation and chromosome number of *L. appressum-salsoloides*. A, diploid; B, tetraploid; C and D, hexaploid and E, octoploid. Abbreviations: 2*x*, diploid;4*x*, tetraploid; 6*x*, hexaploid; 8*x*, octoploid; *S.l.*, *Solanum lycopersicum*; Mean FL, mean relative fluorescence in picograms; DI, DNA index; CV(%), coefficient of variation of the peak in percent. Scale bar: 20 μm (black line).

Table 2. Number of populations and individuals observed with different ploidies of *L. suffruticosum s.l.* Abbreviations: 2*x*, diploid; 3*x*, triploid; 4*x*, tetraploid; 6*x*, hexaploid; 8*x*, octoploid; 10*x*, decaploid; an., aneuploidy; *N* pop, number of populations analysed; *N* total, number of individuals analysed.

Ploidy level	*N* pop (*N* total)
2*x*	53 (962)
2*x* + 3*x*	3 (107 + 3)
2*x* + 4*x*	3 (75 + 16)
4*x*	27 (464)
6*x*	19 (428)
6*x* + 4*x*	6 (175 + 16)
6*x* + an.	1 (29 + 1)
8*x*	19 (238)
8*x* + 6*x*	5 (151 + 10)
8*x* + an.	1 (26 + 3)
10*x*	10 (160)
10*x* + 8*x*	2 (16 + 2)
10*x* + an.	2 (43 + 2)
Total	151 (2927)

6*x* individuals and five 4*x* individuals) (Supporting Information, Table S2) where the two cytotypes occur in more even proportions.

Cytogenetic diversity in taxonomic and geographical entities

Diversity was also observed within and among taxonomic groups of *Linum suffruticosum s.l.* (Figs 3–6, Supporting Information Figs S1–3) with significant differences being found in genome size among most taxonomic entities and cytotypes (2C – $F_{14,\ 1695.25}$ = 4521.5, $P < 0.001$; 1C*x* – $F_{14,\ 1695.25}$ = 336.54, $P < 0.001$; Fig. 6).

Diploids

Two diploid taxa were detected, *L. salsoloides* and *L. suffruticosum* var. *milletii* (Fig. 3). *Linum salsoloides* was observed in central Spain, and was homogeneously diploid, with the lowest values and the smallest variation in genome size (Fig. 3A, Supporting Information, Table S3). Because we were unable to grow seedlings from

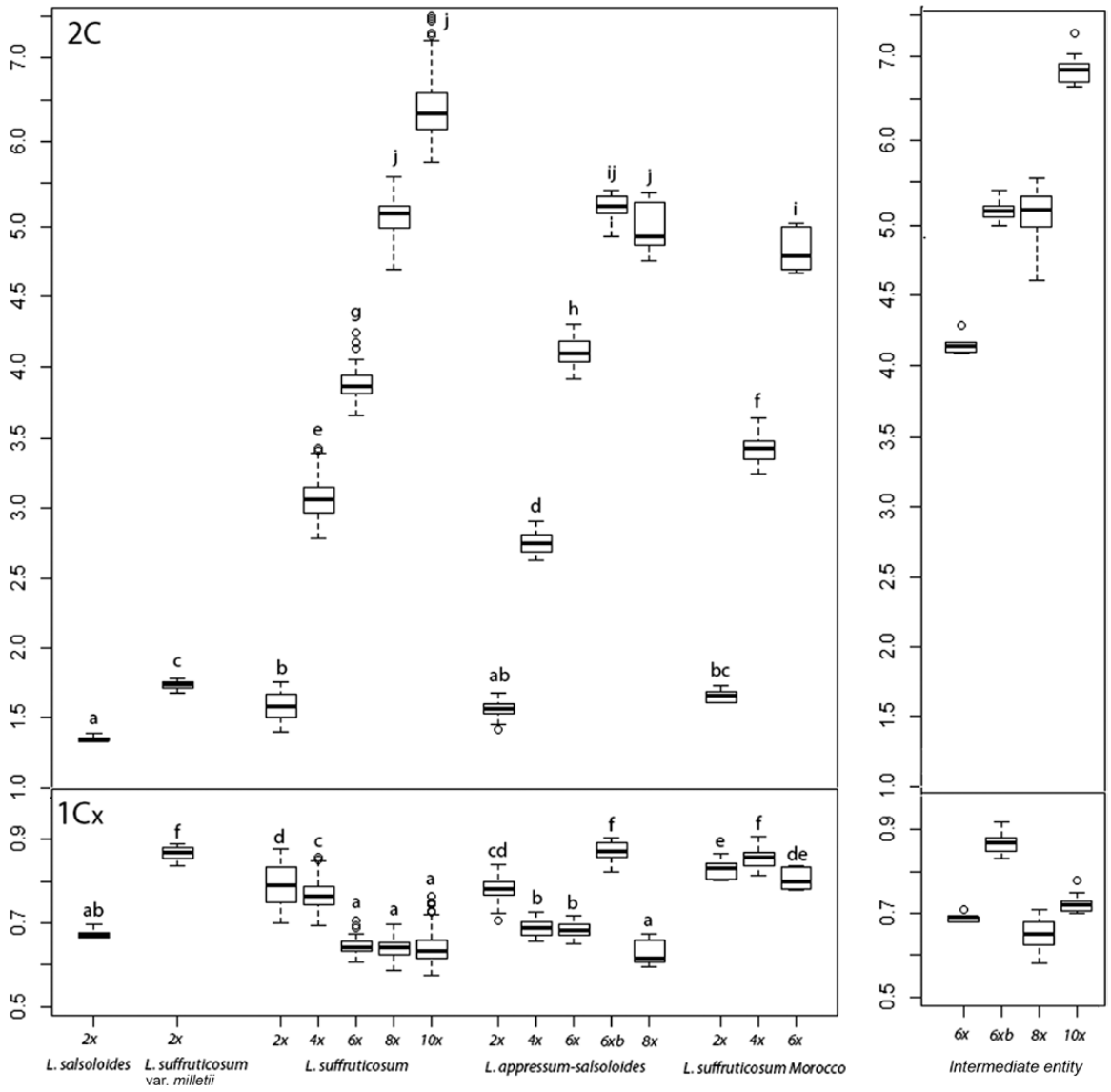

Figure 6. Holoploid and monoploid genome size range and mean (black line) of diploid, tetraploid, hexaploid, octoploid and decaploid populations of each taxonomic entity. Abbreviations: 2C, holoploid genome size; 1C*x*, monoploid genome size; 2*x*, diploid; 4*x*, tetraploid; 6*x,* 6*x*b, hexaploid; 8*x*, octoploid; 10*x*, decaploid. Outliers are also given as white circles. Different letters correspond to statistically significant differences at $P < 0.05$.

this species, information on the number of chromosomes is available only from the literature, with $2n = 18$ chromosomes (Supporting Information, Table S1). *Linum suffruticosum* var. *milletii* grows in the eastern Pyrenees and was homogeneously diploid (Figs 1B, 3B). In contrast to *L. salsoloides*, this taxon had the largest genome size values among diploids, also having 18 chromosomes (Figs 3, 6, Supporting Information, Table S3).

Polyploids

Two polyploid series were observed, *L. suffruticosum* and *L. appressum-salsoloides*, with high variability in genome size and chromosome numbers. *Linum suffruticosum* was the taxon with the highest levels of variability (including 2*x*, 4*x*, 6*x*, 8*x*, 10*x* and a few 3*x* and aneuploids), and genome size estimates enabled us to unambiguously assign ploidy to all individuals analysed (Fig. 4, Supporting Information, Table S3). Diploids with $2n = 16$ and 18 chromosomes and 2C and 1C*x* genome size values intermediate to the two diploid taxa were found growing in mountain regions in southern Spain and in central Spanish Pyrenees (Fig. 1B). Tetraploid *L. suffruticosum* with 2n = 32 or 36 chromosomes and double the genome size of the diploid (Figs 4C, 6, Supporting Information, Table S3)

was found mostly in southern Spain in a parapatric distribution with diploids (Fig. 1B). The highest ploidy cytotypes were found mostly in northern Spain, with an increase in ploidy from west to east (Fig. 1B). Hexaploids were mostly found in central Spain with variable chromosome numbers ($2n$ = 48, 54, Figs 4D, 6, Supporting Information, Table S3), octoploids were found from Valencia to Zaragoza all with $2n$ = 72 (Figs 4E, 6, Supporting Information, Table S3) and decaploids with $2n$ = 90 were mostly found close to the coast in the north-eastern Iberian Peninsula, although a few populations were also found in mountain regions of La Rioja and Teruel (Figs 4F, 6, Supporting Information, Table S3). The increase in genome size among cytotypes is not proportional. As a result, the genome size of octoploids and decaploids did not differ statistically, and 1C*x* values of *L. suffruticosum* in Spain decreased with increased ploidy (Fig. 6). In Morocco, *L. suffruticosum* was rarer, but diploids, tetraploids and hexaploids were detected (Fig. 1B); diploids showed genome sizes similar to those found in diploid *L. suffruticosum* in Europe, and tetraploids and hexaploids showed genome sizes two and three times larger than diploids, respectively, showing less evidence of genome downsizing (Fig. 6, Supporting Information, Tables S3, S5).

Linum appressum-salsoloides, with populations distributed in northern Spain, France and Italy, includes all cytotypes except decaploids (Fig. 1B). Diploid populations with $2n$ = 16 and 18 (Fig. 5A, Supporting Information, Table S3) were mainly found in the Pyrenees, France and Italy, with only a few populations being found in northern and central Spain (Fig. 1B). In contrast, polyploids were concentrated in northern Spain, overall at higher latitudes than *L. suffruticosum* polyploids (Fig. 1B). The 2C values of diploids are similar to those found in diploid *L. suffruticosum* (Fig. 6), but such specimens are morphologically distinct. Tetraploids were found from central to northern Spain and have $2n$ = 36 (occasionally $2n$ = 38) (Figs 1B, 5B, Supporting Information, Table S3). Among hexaploids, two distinct groups based on genome size were found (Figs 5C, D, 6, Supporting Information, Table S3). One of the groups was found in north-western Spain with a genome size falling within the hexaploid range of values (2C = 4.24 ± 0.35 pg) and $2n$ = 54 (Fig. 5C, Supporting Information, Table S3). The other group was found in the western Pyrenees and had significantly higher genome sizes falling within the range of octoploid estimates (2C = 5.23 ± 0.15 pg; Fig. 5D), with homogeneous chromosome counts of $2n$ = 54 (Fig. 5D, Supporting Information, Table S3). Consequently, the latter group of hexaploids had distinctive 1C*x* values (Fig. 6), being an exception to the pattern of genome downsizing with increasing ploidy in this taxon. Finally, octoploids with $2n$ = 72 were found from the western Pyrenees to Soria (Figs 1B, 5E, Supporting Information, Table S3).

Intermediate entities

Specimens classified as intermediate (nine populations) also showed variability in chromosome numbers and genome size estimates, although such populations comprise mostly higher ploidies (hexaploids, octoploids and decaploids; Fig. 6, Supporting Information, Tables S3, S4). As in *L. appressum-salsoloides*, we observed hexaploids with distinct genome sizes (2C = 4.15 ± 0.07 pg and 2C = 5.22 ± 0.15 pg, Supporting Information, Fig. S3), both with $2n$ = 54 chromosomes (Supporting Information, Tables S3, S4). The intermediate populations were detected in regions of contact among cytotypes and taxa (Fig. 1B).

DISCUSSION

Cytogenetic diversity in *Linum suffruticosum s.l.*

Linum L. comprises > 200 species and is particularly diverse in the Mediterranean region (McDill *et al.*, 2009; Ruiz-Martín *et al.*, 2018). Economically relevant groups have been studied in detail, including their genetic and cytogenetic diversity (e.g. *L. usitatissimum* L. from section *Dasylinum* (Plach.) Juz, and the group of *L. perenne* L.; Ockendon, 1968; Chennaveeraiah & Joshi 1983; Bolsheva *et al.*, 2015), whereas other sections, highly diverse and with complex reproductive features, have received less attention and their diversity remains largely unknown [e.g. *Linum* section *Linopsis* (Rchb.) Engelm., Nicholls, 1985a; McDill *et al.*, 2009; Muravenko *et al.*, 2010]. *Linum suffruticosum s.l.* has been described as a polyploid group, but our study reveals the occurrence of wider cytogenetic variability in the complex than previously thought. Here, we observed that variation occurs at three levels, namely chromosome number, ploidy and genome size. First, we observed two basic chromosome numbers (n = 8 and 9; with the former being reported here for the first time in *L. suffruticosum s.l.*) and consequently, we observed different chromosome numbers within the same cytotype in diploids, tetraploids and hexaploids. This suggests the occurrence of chromosomal rearrangements, such as chromosome loss or gain in chromosome pairing, which may have played a role in promoting the cytogenetic diversity of the group. Processes of chromosome fusion, translocations and/or inversions have been described in other species of the genus (Muravenko *et al.*, 2010; Bolsheva *et al.*,

2015). Second, we observed new ploidies in the group, with triploids, hexaploids and decaploids being described here for the first time. Although multiple ploidies have been reported in some species of *Linum* (e.g. Nilsson & Lassen, 1971; Rogers *et al.*, 1972; Chennaveeraiah & Joshi, 1983), none encompasses the level of variability observed here (including five dominant cytotypes). This large variation in ploidy is probably the highest known for this genus, for which ploidy has been reported in *c*. 25% of the taxa with available data, usually only with diploid and tetraploid levels (Ruiz-Martín *et al.*, 2018). A recent review on mixed-ploidy species revealed that most of the well-studied polyploid complexes harbour two (77%) or three ploidies (14%), and more rarely additional ploidies (9%) (Kolář *et al.*, 2017). This supports the idea that whole genome duplications (alone or with hybridization events) are one of the key mechanisms in the diversification of *L. suffruticosum s.l.* Finally, we observed variation in genome sizes within ploidies and, consequently, variation in 1C*x* values (e.g. for hexaploids and diploids or between *L. suffruticosum* individuals from Spain and Morocco), supporting different evolutionary histories (discussed below).

Cytogeographical patterns

Our large-scale sampling revealed complex geographical patterns of *L. suffruticosum s.l.* cytotypes. The different ploidies were distributed parapatrically, with several contact zones among cytotypes and with mixed-ploidy populations being rarely found. Polyploids were found in the Iberian Peninsula and northern Africa with the remaining areas of the species distribution in Europe being characterized by homogeneously diploid populations only. In northern Africa, we found diploid, tetraploid and hexaploid populations, with the species being less abundant there than in Europe. In the Iberian Peninsula, although cytotype distribution was complex, cytotypes had a spatially structured distribution. Diploid populations were more abundant in southern Spain and in the Pyrenees; this wide distribution of diploids may reflect a complex history of movements and lineage sorting across the range since the early Pleistocene, when the group originated (Ruiz-Martín *et al.*, 2018) and explain the lack of Mantel correlation between geographical distance and genome size. In contrast, tetraploid populations were scarce, being concentrated in the south where a clear contact zone of diploid–tetraploid populations was found. The majority of the higher polyploids were found in central and northern Spain, with spatial segregation and some contact zones. Hexaploid populations are distributed in central and north-eastern Spain, but some populations can also be found in the Pyrenees. The majority of octoploid and decaploid populations are found in lower and arid zones in eastern Spain, with the latter being found near the coast. In general, the more recent polyploid populations had a narrower range. Thus the Mantel correlation of the genome size is significant and positive, except for the scarcer decaploid populations. Spatial segregation has been shown in several polyploid complexes (Husband & Schemske, 1998; Balao *et al.*, 2009; Kolář *et al.*, 2009; Sonnleitner *et al.*, 2010; Castro *et al.*, 2012) and has been proposed as one of the most effective barriers for successful polyploid establishment (Levin, 2002; Li *et al.*, 2004; Baack & Stanton, 2005). The capacity to disperse and colonize new niches escaping competition with the progenitor individuals increases the probability of establishment by reducing the minority cytotype disadvantage (Levin, 1975; Ramsey, 2011; Hao *et al.*, 2013). Polyploidization has been shown to have consequences in the ability of polyploids to grow in habitats that differ from their progenitors, enabling polyploids to expand to new areas (Levin, 1975; Buggs & Pannell, 2007; Ramsey, 2011; Hao *et al.*, 2013). In *L. suffruticosum s.l.* the current distribution patterns may be associated with niche differentiation among cytotypes that promotes spatial segregation and consequently reproductive isolation, enabling the establishment and maintenance of polyploid lineages. A strong association between the spatial distribution of cytotypes and their environmental requirements has been explored using niche modelling tools in several polyploid complexes (Glennon *et al.*, 2014; Thompson *et al.*, 2014; Visger *et al.*, 2016; Muñoz-Pajares *et al.*, 2018). Further studies to model niche preferences in *L. suffruticosum s.l.* are already being developed in order to disentangle the ecological requirements that might explain polyploid success and the current distribution patterns.

Although rare, mixed-ploidy populations were also detected (15%), usually with one of the cytotypes in low frequency in the populations. These populations were mostly found at zones of contact between different ploidies, but were also observed in areas dominated by a cytotype. Contact zones are frequent in most polyploid complexes and enable cytotype interactions; however, mixed-ploidy populations are considered a transitory stage and are expected to be rare because positive frequency-dependent selection will exclude the cytotype in minority (Levin, 1975). In most mixed-ploidy populations of *L. suffruticosum s.l.* the lower ploidy was rarer. This may reflect dynamic contact zones where: (1) new polyploids are formed or disperse to existing localities; (2) between-ploidy hybridization occurs or (3) a decline in the frequency of a once dominant cytotype is ongoing. For example, in diploid regions, diploid populations with a few triploids suggest the production of unreduced gametes and emergence of new cytotypes, but in contact zones with

tetraploids or hexaploids it may reflect inter-cytotype hybridization; in contrast, mixed-ploidy populations such as tetraploid–hexaploid, hexaploid–octoploid and octoploid–decaploid where the lower ploidy occurs in low frequency suggest a successful expansion of higher ploidies over lower ones. To sum up, the detection of mixed-ploidy populations in *L. suffruticosum s.l.* clearly supports the idea that cytotype interactions at contact zones exist and that such populations are dynamic. The levels of interaction will influence the genetic structure and diversity at contact zones. Future studies at mixed-ploidy populations involving controlled pollinations and plant fitness assessment will provide insights into the dynamics of this polyploid complex at contact zones.

Genome size and chromosome number as informative characters

By analysing the ploidies and genome sizes in the group extensively, our study reveals cryptic diversity that has not been taken into consideration in previous studies, but that constitutes a strong reproductive barrier. Some of the specimens are easily distinguishable morphologically, such *as L. suffruticosum* var. *millettii* and *L. salsoloides*, and are here distinguished also by different 2C and 1C*x* genome size values, the former taxon having the highest genome sizes of diploids and the latter the lowest values. Most of the ploidy variability was found in the Iberian Peninsula, and polyploids are not restricted to *L. suffruticosum* as previously described in the literature (Supporting Information, Table S1), but also occur in *L. appressum-salsoloides*. The latter can be found as a diploid throughout France and Italy, but in central and northern Spain it comprises diploid and polyploid populations. In contrast, *L. suffruticosum* appears to be more constrained to central and southern Spain, and its cytotypes have a clear geographical structure. Within *L. suffruticosum*, we also found different 1C*x* values for Spanish and Moroccan populations, with genome downsizing being observed in the former, but not the latter region. This may reflect different evolutionary trajectories, as observed in several other plant groups (e.g. Hohmann *et al.*, 2014; Mandák *et al.*, 2016; Krahulcová *et al.*, 2017), especially across the Strait of Gibraltar (Rodríguez-Sánchez *et al.*, 2008) and supports the need for an extensive review of the group in its entire distribution range, as well as dated phylogenetic and phylogeographic studies.

Additionally, another source of complexity lies in contact zones (e.g. between *L. suffruticosum* and *L. appressum-salsoloides*) where populations of higher ploidies and intermediate characteristics are found (here treated as intermediate entities). Hybridization and introgression processes have been suggested to be involved in creating the variability found in *Linum* in the Iberian Peninsula and, in some cases, to generate populations of individuals with contiguous characters among closely related taxa (Martínez-Labarga & Garmendia, 2015; Ruiz-Martín, 2017), making the taxonomy of this group even more difficult. Previous works revealed the importance of cytogenetic traits for taxonomic and relationship evaluations in complex plant groups (e.g. Murray, 2005; Hohmann *et al.*, 2014; Habibi *et al.*, 2018; Prančl *et al.*, 2018) and in this genus (Nicholls,1985c; McDill *et al.*, 2009; Muravenko *et al.*, 2010; Bolsheva *et al.*, 2015; Talebi *et al.*, 2015). The latest review for the *Flora Iberica* segregates groups that are hardly distinguishable morphologically in the field (Martínez-Labarga & Garmendia, 2015), but our results also demonstrate that previous treatments (e.g. Ockendon & Walters, 1968; López González, 1979) do not accommodate all the diversity found in natural populations. All this cytogenetic information is useful as a tool to define geographical units that are hardly distinguishable morphologically, although it cannot be used alone as taxonomic character. Thus, in combination with morphological characterizations and dated phylogenetic relationships based on molecular data, our results can be a helpful tool for clarifying the taxonomy of *L. suffruticosum s.l.* in future studies.

Can cytogenetic data provide insights into the origin of cytotypes of *Linum suffruticosum s.l.*?

The *L. suffruticosum s.l.* group is monophyletic, but internal phylogenetic relationships are still unclear (Ruiz-Martín *et al.*, 2018). Multiple origins of polyploids from the same and/or different progenitors and rapid genomic changes immediately after polyploid formation may have contributed to the diversity in *L. suffruticosum s.l.* Our results support, at least for some of the cytotypes, origins and evolutionary pathways. First, geographical overlap and high cytogenetic diversity detected in natural populations suggest that unreduced gamete formation and hybridization events seem to be frequent in this complex and might be involved in recurrent auto- and/or allopolyploid formation and in gene flow among cytotypes. This agrees with molecular analyses of the group (Ruiz-Martín, 2017) and with the occurrence of morphologically intermediate individuals (e.g. Martínez-Labarga & Garmendia, 2015). Second, differences in monoploid genome size for different geographical areas have been described in several plant groups (Balao *et al.*, 2009; Kolář *et al.*, 2009) and reflect dissimilar evolutionary relationships among polyploids (e.g. Hohmann *et al.*, 2014; Mandák *et al.*, 2016; Krahulcová *et al.*, 2017). Indeed, the differences in

2C and 1C*x* genome sizes observed for populations from different geographical areas support distinct origins, e.g. that detected in hexaploid individuals of *L. suffruticosum s.l.* Additionally, differences in 1C*x* genome sizes between Spanish and Moroccan populations of *L. suffruticosum* support different evolutionary histories, in Morocco probably involving autopolyploidy, whereas northern Spain would have been a melting pot in which different taxa occur and allopolyploidy and/or multiple origins were probably involved. Finally, *L. suffruticosum s.l.* exhibits a rare three-dimensional reciprocal heterostyly with associated heteromorphic self-incompatibility (Armbruster *et al.*, 2006). Our observations throughout the geographical range suggest that there is a constant presence of this 3D heterostyly in all populations of the group. Moreover, most of the examined populations showed a 1:1 ratio of style morphs (A. Afonso and J. Arroyo, field observations), suggesting that this complex breeding system with heteromorphic self-incompatibility is maintained (Barrett, 2002) across the range, irrespective of ploidy. Breeding systems are hypothesized to change along polyploid complexes, particularly when hybridization is involved (Naiki & Nagamasu, 2004; Guggisberg *et al.*, 2006), whereas in outcrossing plants, polyploids tend to be formed mostly through autopolyploid events (Ramsey & Schemske, 1998). However, it is difficult to distinguish between the different processes without appropriate molecular markers, and further phylogenetic and phylogeographic studies are needed to confirm these hypotheses.

CONCLUSIONS

This study revealed complex geographical distribution patterns of cytotypes of *L. suffruticosum s.l.* The large-scale screening showed an outstanding cytogenetic diversity, with triploids, hexaploids and decaploids being described here for the first time. The different ploidies were distributed parapatrically and thus have a geographical structure and several contact zones. Genome size and/or chromosome counts might be useful tools for identifying individuals of *L. suffruticosum s.l.* However, the complexity and morphological variability of the group requires additional taxonomic studies accounting with the diversity found here. In addition, the origin of the polyploids is not easy to disentangle. The geographical overlap and high cytogenetic diversity detected here suggest multiple origins of the polyploids from the same and/or different progenitors. Future phylogenetic and phylogeographic studies coupled with niche modelling analyses are needed to understand the relationships among *L. suffruticosum* entities and to disentangle the ecological requirements that might explain the success of polyploids and their current distribution patterns.

ACKNOWLEDGEMENTS

We thank Mohammed Ater of the University Abdelmalek Essaadi of Morocco for providing the permits to collect the plants (92/2018; 132/2018). We are also thankful to Sara Lopes, Daniela Tavares, Alejandra de Castro, Albano Figueiredo, António Afonso and Montserrat Martínez Ortega and collaborators from the University of Salamanca for assistance in field collections, to Mariana Castro, Daniela Tavares and Lucie Mota for assistance in flow cytometric analyses and to Zoila Díaz for assistance with chromosome counts. This research was financed by POPH/FSE funds by the Fundação para a Ciência e Tecnologia with a doctoral grant to AA (SFRH/ BD/108451/2015) and a starting grant and exploratory project to SC (IF/01267/2013), by Project RENATURE financed by the Programa Operacional Regional do Centro (Centro2020 - CENTRO-01-0145-EDER-000007), a full graduate fellowship to EOV funded by the Consejo de Ciencia y Tecnologia-Mexico (VU-363401), and research grants to JA by Ministerio de Economía y Competitividad and Ministério de Ciencia e Innovación (CGL2013-45037-P and PGC2018 099608-B-100). The authors are also extremely grateful to the Associate Editor and to two anonymous reviewers for all the constructive and helpful comments, all of which enabled to significantly improve the work.

AUTHORS' CONTRIBUTIONS

AA, SC, JL and JA designed the experimental approach. AA, JA, EOV and JL made field collections. AA, SC and JL performed the flow cytometric analyses. AA identified all the specimens and performed chromosome counts. AA, with the collaboration of all co-authors, analysed the data, discussed the results and wrote the manuscript.

REFERENCES

Armbruster WS, Pérez-Barrales R, Arroyo J, Edwards ME, Vargas P. 2006. Three-dimensional reciprocity of floral morphs in wild flax (*Linum suffruticosum*): a new twist on heterostyly. *New Phytologist* **171:** 581–590.

Baack EJ, **Stanton ML. 2005.** Ecological factors influencing tetraploid establishment in snow buttercups (*Ranunculus adoneus*): niche differentiation and tetraploid establishment. *Evolution* **59:** 1936–1944.

Balao F, **Casimiro-Soriguer R**, **Talavera M**, **Herrera J**, **Talavera S. 2009.** Distribution and diversity of cytotypes in *Dianthus broteri* as evidenced by genome size variations. *Annals of Botany* **104:** 965–973.

Barrett SCH. 2002. Sexual interference of the floral kind. *Heredity* **88:** 154–159.

Blanc G, **Wolfe KH. 2004.** Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *The Plant Cell* **16:** 1667–1678.

Bolker BM, **Brooks ME**, **Clark CJ**, **Geange SW**, **Poulsen JR**, **Stevens MHH**, **White JSS. 2009.** Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution* **24:** 127–135.

Bolsheva NL, **Zelenin AV**, **Nosova IV**, **Amosova AV**, **Samatadze TE**, **Yurkevich OY**, **Melnikova NV**, **Zelenina DA**, **Volkov AA**, **Muravenko OV. 2015.** The diversity of karyotypes and genomes within section *Syllinum* of the genus *Linum* (Linaceae) revealed by molecular cytogenetic markers and RAPD analysis. *PLoS One* **10:** e0122015.

Brochmann C, **Brysting AK**, **Alsos IG**, **Borgen L**, **Grundt HH**, **Scheen AC**, **Elven R. 2004.** Polyploidy in Arctic plants. *Biological Journal of the Linnean Society* **82:** 521–536.

Buggs RJA, **Pannell JR. 2007.** Ecological differentiation and diploid superiority across a moving ploidy contact zone. *Evolution* **61:** 125–140.

Castro M, **Castro S**, **Figueiredo A**, **Husband B**, **Loureiro J. 2018.** Complex cytogeographical patterns reveal a dynamic tetraploid-octoploid contact zone. *AoB Plants* **10:** 1–18.

Castro M, **Castro S**, **Loureiro J. 2012.** Genome size variation and incidence of polyploidy in Scrophulariaceae *sensu lato* from the Iberian Peninsula. *AoB Plants* **2012:** pls037.

Castro M, **Loureiro J**, **Serrano M**, **Tavares D**, **Husband BC**, **Siopa C**, **Castro S. 2019.** Mosaic distribution of cytotypes in a mixed-ploidy plant species, *Jasione montana*: nested environmental niches but low geographical overlap. *Botanical Journal of the Linnean Society* **190:** 51–66.

Castro S, **Loureiro J**, **Procházka T**, **Münzbergová Z. 2012.** Cytotype distribution at a diploid-hexaploid contact zone in *Aster amellus* (Asteraceae). *Annals of Botany* **110:** 1047–1055.

Chelaifa H, **Monnier A**, **Ainouche M. 2010.** Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species *Spartina* × *townsendii* and *Spartina anglica* (Poaceae). *New Phytologist* **186:** 161–174.

Chennaveeraiah MS, **Joshi KK. 1983.** Karyotypes in cultivated and wild species of *Linum*. *Cytologia* **48:** 833–841.

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

Doležel J, **Sgorbati S**, **Lucretti S. 1992.** Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. *Physiologia Plantarum* **85:** 625–631.

Doyle JJ, **Doyle JL**, **Rauscher JT**, **Brown AHD. 2004.** Evolution of the perennial soybean polyploid complex (*Glycine* subgenus *Glycine*): a study of contrasts. *Biological Journal of the Linnean Society* **82:** 583–597.

Dray S, **Dufour A. 2007.** The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* **22:** 1–20.

Emberger L, **Maire R. 1941**. *Catalogue des plantes du Maroc*, **Vol. IV**. Algiers: Minerva.

Fennane M, **Tattou MI**, **Mathez J**, **Ouyahya A**, **El Oualidi J. 2007.** *Flore pratique du Maroc, manuel de détermination des plantes vasculaires. Angiospermae (Leguminosae-Lentibulariaceae)*. Rabat: Institut Scientifique de Rabat.

Fox J, **Weisberg S**, **Adler D**, **Bates D**, **Baud-Bovy G**, **S. E. 2005.** car: companion to applied regression. Available at: http://CRAN.R-project.org/package=car.

Galbraith DW, **Harkins KR**, **Maddox JM**, **Ayres NM**, **Sharma DP**, **Firoozabady E. 1983.** Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* **220:** 1049–1051.

Glennon KL, **Ritchie ME**, **Segraves KA. 2014.** Evidence for shared broad-scale climatic niches of diploid and polyploid plants. *Ecology Letters* **17:** 574–582.

Greilhuber J, **Doležel J**, **Lysák MA**, **Bennett MD. 2005.** The origin, evolution and proposed stabilization of the terms "genome size" and "C-value" to describe nuclear DNA contents. *Annals of Botany* **95:** 255–260.

Greilhuber J, **Temsch EM**, **Loureiro J. 2007.** Nuclear DNA content measurement. In: Doležel J, Greilhuber J, Suda J, eds. *Flow cytometry with plant cells: analysis of genes, chromosomes and genomes*. Weinheim: Wiley-VCH, 67–101.

Guggisberg A, **Mansion G**, **Kelso S**, **Conti E. 2006.** Evolution of biogeographic patterns, ploidy levels, and breeding systems in a diploid-polyploid species complex of *Primula*. *New Phytologist* **171:** 617–632.

Habibi F, **Vít P**, **Rahiminejad M**, **Mandák B. 2018.** Towards a better understanding of the *Chenopodium album* aggregate (Amaranthaceae) in the Middle East: a karyological, cytometric and morphometric investigation. *Journal of Systematics and Evolution* **56:** 231–242.

Hao GY, **Lucero ME**, **Sanderson SC**, **Zacharias EH**, **Holbrook NM. 2013.** Polyploidy enhances the occupation of heterogeneous environments through hydraulic related trade-offs in *Atriplex canescens* (Chenopodiaceae). *New Phytologist* **197:** 970–978.

Hastie TJ, **Pregibon D. 1992.** Generalized linear models. In: Chambers JM, Hastie TJ, eds. *Statistical models in S*. London, Boca Raton: Chapman & Hall/CRC, 195–246

Hohmann N, **Schmickl R**, **Chiang TY**, **Lučanová M**, **Kolář F**, **Marhold K**, **Koch MA. 2014.** Taming the wild: resolving the gene pools of non-model *Arabidopsis* lineages. *BMC Evolutionary Biology* **14:** 1–21.

Hothorn T, **Bretz F**, **Westfall P**, **Heiberger RM. 2017.** *Multcomp: simultaneous inference for general linear*

hypotheses. Available at: http://CRAN.Rproject.org/package=multcomp.

Husband BC, **Schemske DW. 1998.** Cytotype distribution at a diploid-tetraploid contact zone in *Chamerion* (*Epilobium*) *angustifolium* (Onagraceae). *American Journal of Botany* **85:** 1688–1694.

Jahandiez E, **Maire R. 1932.** *Catalogue des plantes du Maroc (Spermatophytes et Pteridophytes),* **Vol II**. Algiers: Minerva.

Kim S, **Rayburn AL**, **Parrish A**, **Lee DK. 2012.** Cytogeographic distribution and genome size variation in prairie cordgrass (*Spartina pectinata* Bosc ex Link). *Plant Molecular Biology Reporter* **30:** 1073–1079.

Kolář F, **Čertner M**, **Suda J**, **Schönswetter P**, **Husband BC. 2017.** Mixed-ploidy species: progress and opportunities in polyploid research. *Trends in Plant Science* **22:** 1041–1055.

Kolář F, **Tech M**, **Trávníek P**, **Rauchová J**, **Urfus T**, **Vít P**, **Kubesová M**, **Suda J. 2009.** Towards resolving the *Knautia arvensis* agg. (Dipsacaceae) puzzle: primary and secondary contact zones and ploidy segregation at landscape and microgeographic scales. *Annals of Botany* **103:** 963–974.

Krahulcová A, **Trávníček P**, **Krahulec F**, **Rejmánek M. 2017.** Small genomes and large seeds: chromosome numbers, genome size and seed mass in diploid *Aesculus* species (Sapindaceae). *Annals of Botany* **119:** 957–964.

Levin DA. 1975. Minority cytotype exclusion in local plant populations. *Taxon* **24:** 35–43.

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

Li BH, **Xu X-M**, **Ridout M. 2004.** Modelling the establishment and spread of autotetraploid plants in a spatially heterogeneous environment. *Journal of Evolutionary Biology* **17:** 562–573.

López González G. 1979. Algunas consideraciones sobre los linos del grupo *Linum tenuifolium* L. en España. *Memoires de la Société Botanique de Genéve* **1:** 99–109.

Loureiro J, **Rodriguez E**, **Dolezel J**, **Santos C. 2007.** Two new nuclear isolation buffers for plant DNA flow cytometry: a test with 37 species. *Annals of Botany* **100:** 875–888.

Madlung A. 2013. Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity* **110:** 99–104.

Mandák B, **Vít P**, **Krak K**, **Trávníček P**, **Drašnarová A**, **Hadincová V**, **Zákravský P**, **Jarolímová V**, **Bacles CFE**, **Douda J. 2016.** Flow cytometry, microsatellites and niche models reveal the origins and geographical structure of *Alnus glutinosa* populations in Europe. *Annals of Botany* **117:** 107–120.

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** (Suppl 1): 21–37.

Martínez-Labarga JM, **Garmendia FM. 2015.** *Linum* L. In: Castroviejo S, Muñoz Garmendia F, Navarro C, Quintanar A, Buira A, eds. *Flora Iberica. Vol. 9. Rhamnaceae-Polygalaceae.* Madrid: Real Jardín Botánico, CSIC, 172–259.

McDill J, **Repplinger M**, **Simpson BB**, **Kadereit JW. 2009.** The phylogeny of *Linum* and Linaceae subfamily Linoideae, with implications for their systematics, biogeography, and evolution of heterostyly. *Systematic Botany* **34:** 386–405.

Muñoz-Pajares AJ, **Perfectti F**, **Loureiro J**, **Abdelaziz M**, **Biella P**, **Castro M**, **Castro S**, **Gómez JM. 2018.** Niche differences may explain the geographic distribution of cytotypes in *Erysimum mediohispanicum*. *Plant Biology* **20** (Suppl 1): 139–147.

Muravenko OV, **Bolsheva NL**, **Yurkevich OY**, **Nosova IV**, **Rachinskaya OA**, **Samatadze TE**, **Zelenin AV. 2010.** Karyogenomics of species of the genus *Linum* L. *Genetika* **46:** 1339–1342.

Murray BG. 2005. When does intraspecific C-value variation become taxonomically significant? *Annals of Botany* **95:** 119–125.

Naiki A, **Nagamasu H. 2004.** Correlation between distyly and ploidy level in *Damnacanthus* (Rubiaceae). *American Journal of Botany* **91:** 664–671.

Nicholls MS. 1985a. The evolutionary breakdown of distyly in *Linum tenuifolium* (Linaceae). *Plant Systematics and Evolution* **150:** 291–301.

Nicholls MS. 1985b. Pollen flow, population composition, and the adaptive significance of distyly in *Linum tenuifolium* L. (Linaceae). *Biological Journal of the Linnean Society* **25:** 235–242.

Nicholls MS. 1985c. A systematic study of the *Linum tenuifolium* group (Linaceae). *Botanical Journal of the Linnean Society* **91:** 473–490.

Nicholls MS. 1986. Variation and evolution in *Linum tenuifolium* (Linaceae). *Plant Systematics and Evolution* **153:** 243–258.

Nilsson O, **Lassen P. 1971.** Chromosome numbers of vascular plants from Austria, Mallorca and Yugoslavia. *Botaniska Notiser* **124:** 270–276.

Ockendon DJ. 1968. Biosystematic studies in the *Linum perenne* group. *New Phytologist* **67:** 787–813.

Ockendon DJ, **Walters SM. 1968.** *Linum* L. In: Tutin TG, Heywood VH, Burges NA, Moore DM, Valentine DH, Walters SM, Webb DA, eds. *Flora Europaea*, **Vol. 2**. Cambridge: Cambridge University Press, 206–211.

Otto SP, **Whitton J. 2000.** Polyploid incidence and evolution. *Annual Review of Genetics* **34:** 401–437.

Prančl J, **Koutecky P**, **Travniček P**, **Jarolímová V**, **Lučanová M**, **Koutecká E**, **Kaplan Z. 2018.** Cytotype variation, cryptic diversity and hybridization in *Ranunculus* sect. *Batrachium* revealed by flow cytometry and chromosome numbers. *Preslia* **90:** 195–223.

Quézel P, **Santa S. 1962.** *Nouvelle flore de l'Algérie et des régions désertiques méridionales.* Paris: CNRS.

R Core Development Team. **2019.** *A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available at:http://www.R-project.org/.

Ramsey J. 2011. Polyploidy and ecological adaptation in wild yarrow. *Proceedings of the National Academy of Sciences of the United States of America* **108:** 7096–7101.

Ramsey J, **Schemske DW. 1998.** Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* **29:** 467–501.

Ramsey J, **Schemske DW. 2002.** Neopolyploidy in flowering plants. *Annual Review of Ecology and Systematics* **33:** 589–639.

Rodríguez-Sánchez F, **Pérez-Barrales R**, **Ojeda F**, **Vargas P**, **Arroyo J. 2008.** The Strait of Gibraltar as a melting pot for plant biodiversity. *Quaternary Science Reviews* **27:** 2100–2117.

Rogers CM, **Mildner R**, **Harris BD. 1972.** Some additional chromosome numbers in the Linaceae. *Brittonia* **24:** 313–316.

Ruiz-Martín J. 2017. *Evolución y función de la heterostilia en el género* Linum *(Linaceae).* Unpublished D. Bio. Thesis, Departamento de Biología Vegetal y Ecología, Universidad de Sevilla.

Ruiz-Martín J, **Santos-Gally R**, **Escudero M**, **Midgley J**, **Pérez-Barrales R**, **Arroyo J. 2018.** Style polymorphism in *Linum* (Linaceae): a case of Mediterranean parallel evolution? *Plant Biology* **20**(Suppl 1):100–111.

Sampson JF, **Byrne M. 2012.** Genetic diversity and multiple origins of polyploid *Atriplex nummularia* Lindl. (Chenopodiaceae). *Biological Journal of the Linnean Society* **105:** 218–230.

Segraves KA, **Thompson JN**, **Soltis PS**, **Soltis DE. 1999.** Multiple origins of polyploidy and the geographic structure of *Heuchera grossulariifolia*. *Molecular Ecology* **8:** 253–262.

Soltis DE, **Albert VA**, **Leebens-Mack J**, **Bell CD**, **Paterson AH**, **Zheng C**, **Sankoff D**, **Depamphilis CW**, **Wall PK**, **Soltis PS. 2009.** Polyploidy and angiosperm diversification. *American Journal of Botany* **96:** 336–348.

Soltis DE, **Buggs RJA**, **Doyle JJ**, **Soltis PS. 2010.** What we still don't know about polyploidy. *Taxon* **59:** 1387–1403.

Soltis DE, **Soltis PS. 1999.** Polyploidy: recurrent formation and genome evolution. *Trends in Ecology and Evolution* **14:** 348–352.

Soltis PS. 2005. Ancient and recent polyploidy in angiosperms. *New Phytologist* **166:** 1–5.

Soltis PS, **Doyle JJ**, **Soltis DE. 1992.** Molecular data and polyploid evolution in plants. In: Soltis PS, Soltis DE, Doyle JJ, eds. *Molecular systematics of plants.* Boston: Springer, 177–201.

Sonnleitner M, **Flatscher R**, **Escobar García P**, **Rauchová J**, **Suda J**, **Schneeweiss GM**, **Hülber K**, **Schönswetter P. 2010.** Distribution and habitat segregation on different spatial scales among diploid, tetraploid and hexaploid cytotypes of *Senecio carniolicus* (Asteraceae) in the eastern Alps. *Annals of Botany* **106:** 967–977.

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

Suda J, **Kron P**, **Husband BC**, **Trávníček P. 2007.** Flow cytometry and ploidy: applications in plant systematics, ecology and evolutionary biology In: Doležel J, Greilhuber J, Suda J, eds. *Flow cytometry with plant cells: analysis of genes, chromosomes and genomes*. Weinheim: Wiley-VCH, 103–130.

Talebi SM, **Sheidai M**, **Atri M**, **Sharifnia F**, **Noormohammadi Z. 2015.** Infraspecific morphological and genome size variations in *Linum glaucum* in Iran. *Biodiversitas* **16:** 69–78.

Thompson JD. **2005.** *Plant evolution in the Mediterranean*. Oxford: University Press.

Thompson KA, **Husband BC**, **Maherali H. 2014.** Climatic niche differences between diploid and tetraploid cytotypes of *Chamerion angustifolium* (Onagraceae). *American Journal of Botany* **101:** 1868–1875.

Valdés B, **Rejdali M**, **Kadmiri AA El**, **Jury SL**, **Montserrat JM. 2007.** *Catalogue des plantes vasculaires du nord du Maroc, incluant des clés d'identification / Checklist of vascular plants of N Morocco with identification keys.* Madrid: Consejo Superior de Investigaciones Cientificas.

Visger CJ, **Germain-Aubrey CC**, **Patel M**, **Sessa EB**, **Soltis PS**, **Soltis DE. 2016.** Niche divergence between diploid and autotetraploid *Tolmiea*. *American Journal of Botany* **103:** 1396–1406.

Wan J-N, **Guo Y-P**, **Rao G-Y. 2019.** Unraveling independent origins of two tetraploid *Achillea* species by amplicon sequencing. *Journal of Systematics and Evolution* doi: 10.1111/jse.12544.

Wood TE, **Takebayashi N**, **Barker MS**, **Mayrose I**, **Greenspoon PB**, **Rieseberg LH. 2009.** The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 13875–13879.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Chromosome number reported for *L. suffruticossum s.l.*

Table S2. Specimens and populations of *L. suffruticosum s.l.* analysed in this study. Information about the ploidy level of each population (Ploidy level), mean, standard deviation of the mean (SD), and minimum (min.) and maximum (max.) values of the holoploid (2C) and monoploid (1C*x*) genome size in picograms (pg) are given. The coefficient variation of the mean holoploid genome size (CV; calculated as the ratio of the standard deviation to the mean) in percentage (%), number of individuals with ploidy level estimations (*N* ploidy level), number of individuals with genome size estimates (*N* G.s.,), chromosome numbers (2*n*; individual chromosome counts are separated by a comma; '~' denotes approximate counts due to low availability of material), and number of individuals with chromosome counts (*N* number chro) are provided. Finally, information about country, locality,

coordinates, date of collection, collector, collector number and herbarium where specimen has been deposited are also given.

Table S3. Genome size and chromosome numbers among *L. suffruticosum* s.l. entities/taxa. Information about the mean, standard deviation of the mean (SD), and minimum and maximum (2C range) values of the holoploid (2C) and monoploid (1C*x*) genome size in picograms (pg) are given. The coefficient variation of the mean holoploid genome size (CV; calculated as the ratio of the standard deviation to the mean) in percentage (%), number of populations analysed (*N* pop.), number of individuals with genome size estimates (*N* G.s.), chromosome numbers (number chromosome, individual chromosome counts are separated by a comma; '~' denotes approximate counts due to low availability of material; rare chromosome counts are presented in parenthesis; * Denotes chromosome numbers reported in the literature; na indicates data not available), number of individuals with chromosome counts (*N* number of chromosome), and information about geographic distribution are also given. Different letters correspond to statistically significant differences at $P < 0.05$.

Figure S1. Genome size estimation and chromosome number of the intermediate entity. A, B, hexaploid; C, octoploid and D, decaploid populations. Abbreviations:6*x*, he*x*aploid; 8*x*, octoploid; 10*x*, decaploid; S.l., *Solanum lycopersicum*; Mean FL, mean relative fluorescence in picograms; DI, DNA index; CV(%), coefficient of variation of the peak in percent. Scale bar: 20 µm (black line).

Figure S2. Genome size estimation of *L. suffruticosum* from Morocco. A, diploid; B, tetraploid; C, hexaploid. abbreviations: 2*x*, diploid; 4*x*, tetraploid; 6*x*, hexaploid; S.l., *Solanum lycopersicum*; Mean FL, mean relative fluorescence in picograms; DI, DNA index; CV(%), coefficient of variation of the peak in percent.

Figure S3. Pure-ploidy populations sampled in Iberian Peninsula and Morocco with the respective ploidy levels (diploid: yellow; tetraploid: green; hexaploid with low genome size: dark blue; hexaploid with high genome size: light blue; octaploid: red; decaploid: pink) and with mixed-ploidy populations with the respective ploidy levels composition (diploid-triploid: yellow star; diploid-tetraploid: green, hexaploid-tetraploid: blue star, octoploid-hexaploid: purple star, decaploid-octoploid: red star). The base map was downloaded from https://www.diva-gis.org/gdata.