Reassessment of morphological species delimitations in the *Cyperus margaritaceus-niveus* complex using morphometrics

Martin Xanthos[1], Simon J. Mayo[1], Isabel Larridon[1,2]

1   Royal Botanic Gardens, Kew, Richmond, UK

2   Ghent University, Department of Biology, Systematic and Evolutionary Botany Lab, Gent, Belgium

Corresponding author: Martin Xanthos (m.xanthos@kew.org)

Abstract

Background and aims – The *Cyperus margaritaceus-niveus* complex is a group of ten tropical species from sub-Saharan Africa and Madagascar: *C. karlschumannii*, *C. kibweanus*, *C. ledermannii*, *C. margaritaceus*, *C. niveus*, *C. nduru*, *C. obtusiflorus*, *C. somaliensis*, *C. sphaerocephalus*, and *C. tisserantii*. They are characterised by a capitate head of white-yellow spikelets and modified culm bases and recent molecular analysis puts them in a distinct clade. The group lacks a modern taxonomic revision, and the taxa described in the Flora treatments of the past 50 years differ considerably in their circumscription. In this study, morphometric analyses are used to test species limits to establish more stable morphological delimitations of the taxa.

Material and methods – An examination of 15 morphological characters on 489 herbarium specimens was carried out and the data was analysed using Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) with cross-validation, and Classification and Regression Tree (CART) analysis. *Cyperus kibweanus* was not further considered due to lack of material.

Key results – Both PCA and LDA showed varying degrees of overlap in the nine remaining taxa, with no single group clearly separating in multivariate space. However, cross-validation clearly showed *C. margaritaceus* as a distinct entity despite its overwhelming presence in the PCA. Both LDA and CART failed to separate *C. niveus* as a distinct group as its specimens were dispersed among the other groups. Differing results were obtained for other taxa depending on the type of analysis. *Cyperus margaritaceus*, *C. nduru*, and *C. sphaerocephalus* were divided into two groups by CART but re-examination of the specimens does not definitively support the idea that these infraspecific groups represent separate taxa.

Conclusions – The results show that eight morphospecies are recognised by LDA and six morphospecies by CART. Characters used to separate the taxa in Flora treatments scored high loadings in the analysis showing their high taxonomic utility value. The methods used can be applied to resolving other complexes in the Cyperaceae.

Keywords

Cyperaceae, morphometrics, species complex, taxonomy

INTRODUCTION

The *Cyperus margaritaceus-niveus* complex is here understood to comprise a group of ten species (*C. karlschumannii* C.B.Clarke, *C. kibweanus* Duvign., *C. ledermannii* (Kük.) S.S.Hooper, *C. margaritaceus* Vahl, *C. niveus* Retz., *C. nduru* Cherm., *C. obtusiflorus* Vahl, *C. somaliensis* C.B.Clarke, *C. sphaerocephalus* Vahl, and *C.* *tisserantii* Cherm.) distributed throughout sub-Saharan Africa and Madagascar. The group has been considered a species complex since Kükenthal (1936), who treated these taxa as varieties under *C. margaritaceus*, except for *C. niveus* and *C. obtusiflorus*, which he considered to be separate species, and *C. kibweanus*, which was first described in 1963 (Duvigneaud and Denaeyer-de Smet 1963). Other varieties of *C. margaritaceus* were also

Table 1. Comparison of the morphological characters used in classical Floras to delineate taxa in the *Cyperus margaritaceus-niveus* complex.

Flora of West Tropical Africa (Hooper and Napper 1972)	Sedges and Rushes of East Africa (Haines and Lye 1983)	Flora of Somalia (Lye 1995)	Flora of Ethiopia and Eritrea (Lye 1997)	Flora of Tropical East Africa (Hoenselaar et al. 2010)	Flora Zambesiaca (Browning et al. 2020)
Rhizomes vs modified stem bases					Rhizomes vs modified stem bases
					Basal sheath texture
Culm length	Culm length				Culm length
Leaf width	Leaf length and width				Leaf width
Length of involucral bracts	Length of involucral bracts			Length and width of involucral bracts	Length of involucral bracts
No. of involucral bracts					
Spikelet length and width				Spikelet length and width	
	No. of spikelets per head				No. of spikelets per head
Confluent vs discrete spikelets					
Glume length		Glume length	Glume length		
Glume shape	Glume apex shape and texture				
				Length of anthers and filaments	
	Nutlet length and width			Nutlet width	
		Nutlet surface	Nutlet surface	Nutlet surface	

recognised by Kükenthal (1936), but these have since been reduced to synonymy of other taxa in this group. A recent molecular phylogenetic study about resolving the relationships in the C$_4$ *Cyperus* clade has shown that some members of the *Cyperus margaritaceus-niveus* complex form a monophyletic group separate from and sister to most of the remaining C$_4$ *Cyperus* species (Larridon et al. 2020). Although three taxa – *C. kibweanus*, *C. sphaerocephalus*, and *C. somaliensis* – were not included in that phylogenetic study, the results suggested that the morphological combination of a capitate inflorescence of multiflowered spikelets with white distichously arranged glumes, sometimes with streaks of green or yellow, with modified culm bases is unique to this species complex.

Current species delimitations for the taxa in this complex are based on classical morphotaxonomy. Despite a considerable overlap of the characters used to delimit the taxa across the regional African Floras (Table 1), Flora treatments have not reached a consensus on the species limits in this complex. This can be explained by the fact that different authors placed different taxonomic emphasis on a particular character, i.e. whether it should be used at specific, subspecific, or varietal rank. The characteristic white glumes are not used as a character per se in any of the Floras, although the taxa are placed in the 'capitate species'

group of *Cyperus* within the key along with other species with cream, yellow, or brown glumes. As a result of this lack of consensus, the Flora treatments published over the past 50 years differ considerably in their circumscription of the taxa (Table 2). In Flora of West Tropical Africa, Hooper and Napper (1972) recognised five taxa as distinct species, excluding *C. kibweanus*, *C. obtusiflorus*, *C. niveus*, *C. somaliensis*, and *C. sphaerocephalus* as these occurred outside the region covered by the flora. Haines and Lye (1983) only recognised *C. margaritaceus* and *C. niveus* as separate species based on small differences in stem base size and degree of spikelet compression. The remaining taxa, excluding *C. kibweanus*, *C. karlschumannii*, and *C. somaliensis*, were reduced to varieties of these two species, while *C. obtusiflorus* was considered to be a synonym of *C. niveus* var. *leucocephalus* (Kunth) Fosberg. This followed Fosberg (1977) who commented on the few differences between African specimens of *C. niveus* and *C. obtusiflorus*, and, in his account for Flora of Aldabra, he recognized three varieties of *C. niveus*, sinking *C. obtusiflorus* into *C. niveus* var. *leucocephalus*; *niveus* being the older name. Later, both Flora of Tropical East Africa (Hoenselaar et al. 2010) and Flora Zambesiaca (Browning et al. 2020) broadly followed the concepts of Haines and Lye (1983) separating *C. margaritaceus* and *C. niveus*

Table 2. Comparison of the taxonomy of the complex and its status in relevant literature.

	Flora of West Tropical Africa (Hooper and Napper 1972)	Sedges and Rushes of East Africa (Haines and Lye 1983)	Flora of Somalia (Lye 1995)	Flora of Ethiopia and Eritrea (Lye 1997)	Flora of Tropical East Africa (Hoenselaar et al. 2010)	Flora Zambesiaca (Browning et al. 2020)
Cyperus karlschumannii	*C. karlschumannii*	Not present	Not present	Not present	Species of doubtful occurrence	Not present
Cyperus ledermannii	*C. ledermannii*	Synonym of *C. niveus* var. *ledermannii*	Not present	Not present	Synonym of *C. niveus* var. *leucocephalus*	Synonym of *C. niveus* var. *leucocephalus*
Cyperus margaritaceus	*C. margaritaceus*	*C. margaritaceus*	Not present	Not present	*C. margaritaceus*	*C. margaritaceus*
Cyperus niveus	Not present	Typical variety not recorded	*C. niveus*	*C. niveus*	*C. niveus*	*C. niveus*
Cyperus nduru	*C. nduru*	Synonym of *C. margaritaceus* var. *nduru*	Not present	Not present	*C. nduru*	*C. nduru*
Cyperus obtusiflorus	Not present	Synonym of *C. niveus* var. *leucocephalus*	Synonym of *C. niveus* var. *leucocephalus*	Synonym of *C. niveus* var. *leucocephalus*	Synonym of *C. niveus* var. *leucocephalus*	Synonym of *C. niveus* var. *leucocephalus*
Cyperus somaliensis	Not present	Not present	*C. somaliensis*	Not present	Not present	Not present
Cyperus sphaerocephalus	Not present	*C. niveus* var. *flavissimus*	*C. niveus* var. *flavissimus*	Not present	*C. flavissimus*	*C. flavissimus*
Cyperus tisserantii	*C. tisserantii*	Synonym of *C. margaritaceus* var. *tisserantii*	Not present	Synonym of *C. niveus* var. *tisserantii*	Synonym of *C. niveus* var. *tisserantii*	Synonym of *C. niveus* var. *tisserantii*

but elevating *C. nduru* to species level, transferring *C. tisserantii* to a variety of *C. niveus* – not *C. margaritaceus* cf. Haines and Lye (1983) – and reducing *C. ledermannii* to synonymy under *C. niveus* var. *leucocephalus*. In the notes accompanying the species account of *C. niveus* written for Flora of Tropical East Africa, Hoenselaar et al. (2010) mentioned that presence and absence of rhizomes was used as an additional character in their key to separate *C. margaritaceus* from *C. niveus* but noted it is possible that they could represent one species. Lye (1997), in his account for the Flora of Ethiopia and Eritrea, only includes *C. niveus*, but recognises two varieties: var. *leucocephalus* and var. *tisserantii*. Haines and Lye (1983) considered *C. sphaerocephalus* to be no more than a variety of *C. niveus* placing it under *C. niveus* var. *flavissimus*; a concept followed by Lye (1995). Hoenselaar et al. (2010) and Browning et al. (2020) elevated the variety to species on account of its distinctive yellow inflorescence. *Cyperus somaliensis* was first described by Clarke (1895) from among many collections brought to him from present-day Somalia. Due to its endemicity, it has not been treated in other African Floras and thus lacks a morphological comparison with the other taxa, though Lye (1995) compared it with *C. niveus* in Flora of Somalia, separating the two taxa only on glume length. *Cyperus kibweanus* was first described by Duvigneaud

(Duvigneaud and Denaeyer-de Smet 1963) from a single collection in present day Democratic Republic of the Congo and it is not present in any of the aforementioned Floras.

Given the differing taxonomic opinions outlined above, in this study, we use the currently accepted species, as presented in the relevant Floras, as our baseline taxa, which function as our hypotheses. Morphometrics is then used as an algorithmic process to test the results of the traditional taxonomic procedure and thereby test the robustness of the authors' original circumscriptions from the Flora accounts. As such, *C. margaritaceus* is the most widespread species, occurring in west tropical Africa, central Africa, east Africa, and southern Africa but not in Madagascar. *Cyperus karlschumannii*, *C. ledermannii*, and *C. tisserantii* are native to west tropical Africa, although *C. tisserantii* extends south into the Democratic Republic of the Congo. *Cyperus niveus* and *C. nduru* – the latter characterised by bottle-like plant bases and hardened basal leaf sheaths (Fig. 1E, F) – have overlapping distributions; the former mostly found in east Africa and the latter mainly in west tropical Africa and southern tropical Africa. *Cyperus obtusiflorus* occurs from Sudan and Ethiopia down to South Africa and is widespread in Madagascar. *Cyperus somaliensis* is endemic to Somalia, while *C. sphaerocephalus* extends from tropical east
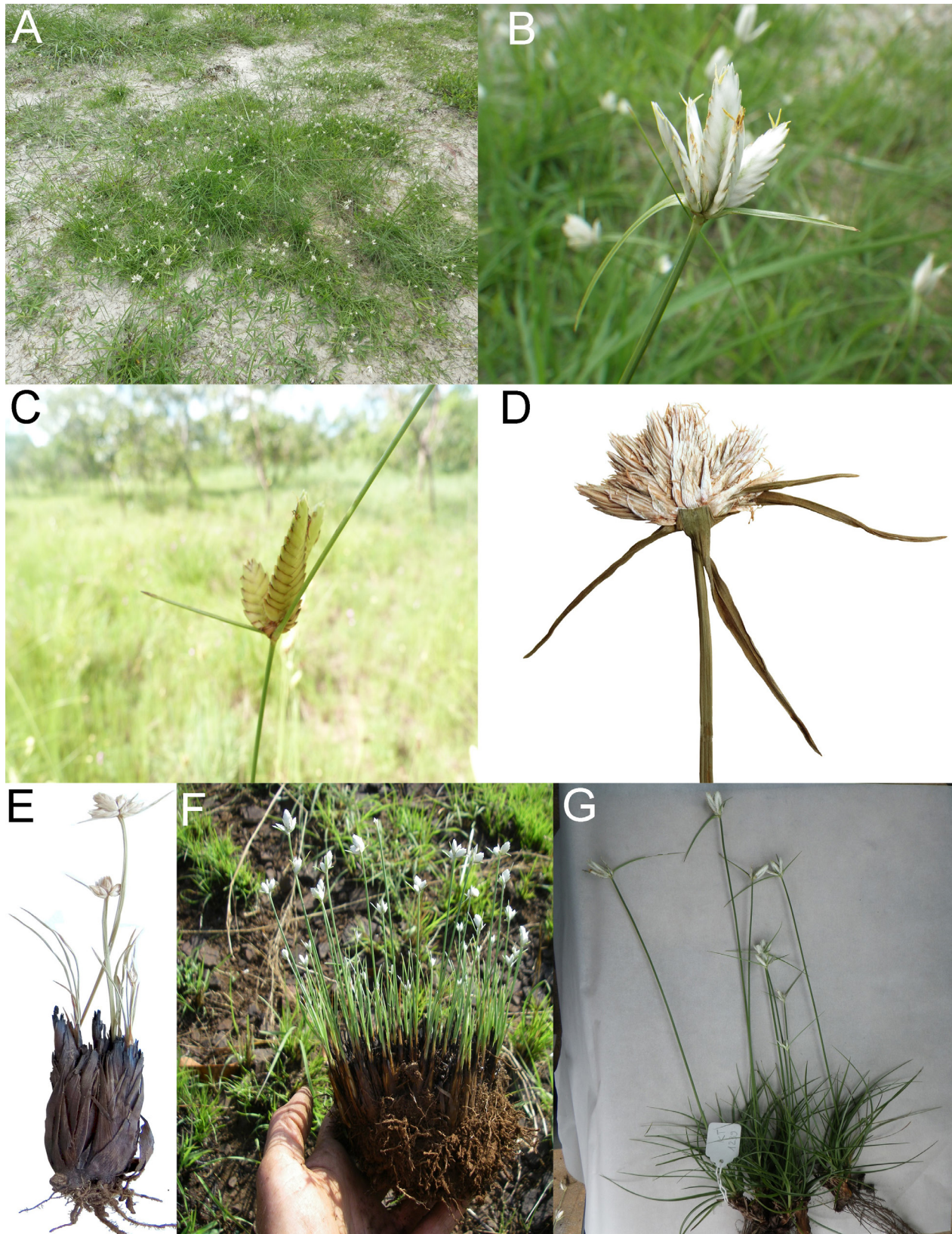
Figure 1. Representatives of the *Cyperus margaritaceus-niveus* complex. **A**. Habitat of *Cyperus margaritaceus.* **B**. *Cyperus margaritaceus* inflorescence. **C**. *Cyperus karlschumannii*. **D**. *Cyperus ledermannii*. **E–F**. *Cyperus nduru* showing the elongated hardened stem bases. **G**. Unmounted herbarium specimen of *C. margaritaceus.* Photos A, B, C, F, G by Xander van der Burgt; D, E by Jane Browning.

Africa to South Africa. Figure 2 shows the locations of the herbarium specimens used in this study.

Morphometrics has been shown to be a useful tool to investigate species limits in closely related taxa (Marhold 2011), either used alone (e.g. Atkinson and Codling 1986; Brysting and Elven 2000; Rivero-Guerra 2011) or in conjunction with other approaches such as molecular phylogenetics (e.g. Perný et al. 2005; Barrett and Freudenstein 2009; Kučera et al. 2010). Within Cyperaceae, a number of species complexes have been assessed using morphometrics (e.g. Rosen 2006; Naczi and Moyer 2016; Di Natale et al. 2020), with a particular focus on *Carex* L. (e.g. Naczi et al. 1998; Smith and Waterway 2008; Míguez Rios 2017), which is the most species-rich genus in the family, as well as one of the largest angiosperm genera. Comparatively, few morphometric studies have focussed on *Cyperus* L. (e.g. Carter and Bryson 2000; Lowe 2018) although such studies have resulted in species new to science being described (Gardner et al. 2014; Gray and Stott 2017).

In this paper, we use multivariate analyses as a tool to support hypotheses concerning the taxa involved in the *Cyperus margaritaceus-niveus* complex. In the analyses, we include most of the characters used in the relevant Flora treatments (Table 1) to test their utility for assessing species limits. Unlike the Flora accounts based on partial subsets of the taxa of the complex, all species were included in the present study and across the entire geographical range, except for *C. kibweanus* due to a lack of sufficient number of specimens. Three different types of analysis were carried out: Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Classification and Regression Tree (CART) analysis. PCA and LDA used only the quantitative characters present in the data set, whereas CART used the full data set, i.e. including the qualitative characters.

This study aims to provide testable circumscriptions of taxonomic species (see Mayo 2022) based primarily on morphological data. The data sets (Supplementary material 1) and analysis scripts are available for future testing and improvement. Our objective is to contribute a computable morphological framework for future integrative species taxonomy in Cyperaceae (Dayrat 2005; Will et al. 2005; Pante et al. 2015).
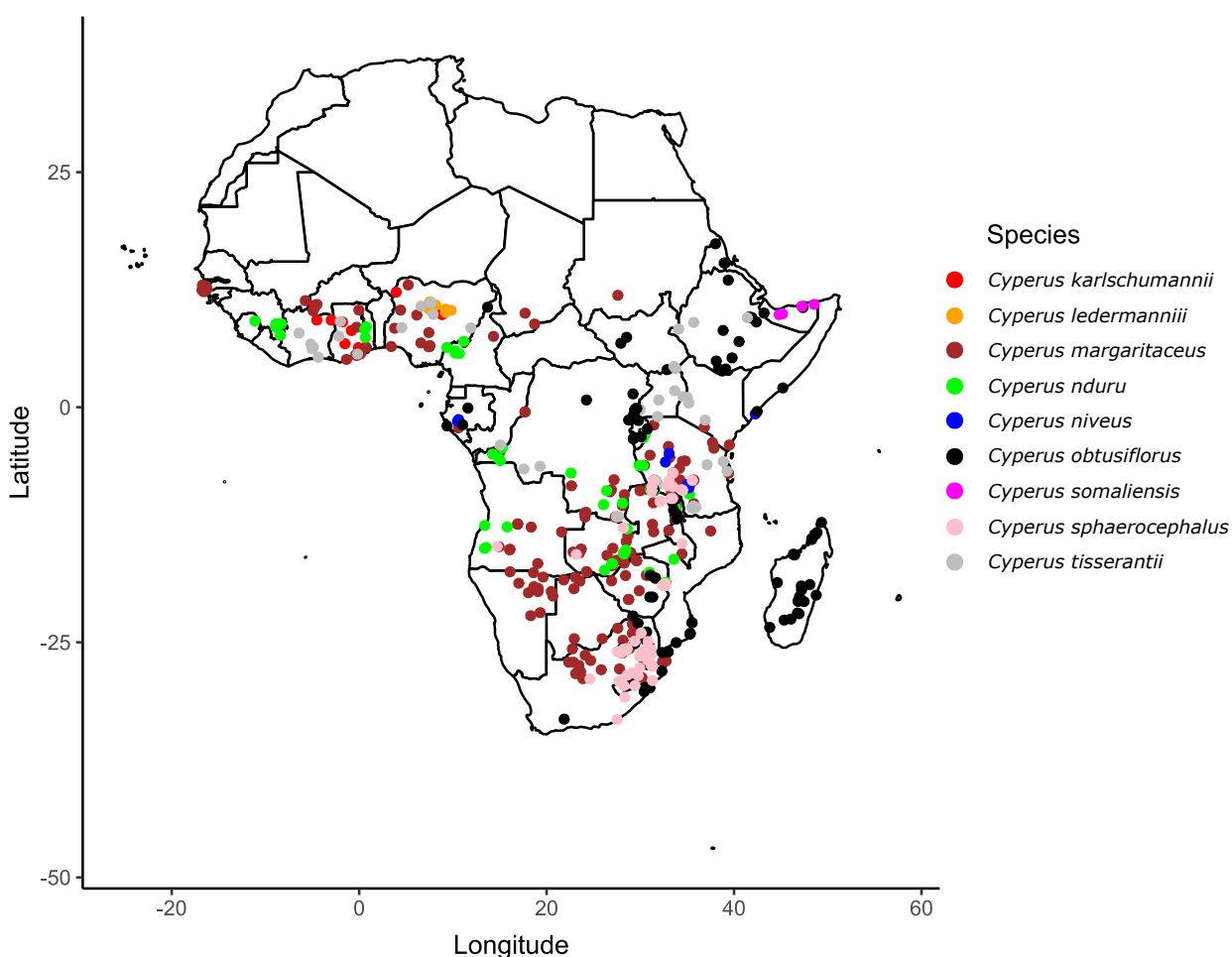
Figure 2. Distribution of herbarium specimens used in this study. These records include specimens not used in the analysis due to missing characters but with sufficient locality data. Map created in R Studio (R Studio Team 2021).

Table 3. List of characters measured and scored for morphometric analysis. [a]characters excluded from PCA and LDA. [b]characters excluded from all analyses due to multiple missing values.

Vegetative characters	
LFLEN	Leaf length (cm)
LFWID	Leaf width (mm)
CULLEN	Culm length (cm)
CULWID	Culm width (mm)
NUMBRAC	Number of involucral bracts
LENBRAC	Length of involucral bracts (cm)
BSHPERS[a]	Basal sheath persistency (flattened/fibrous)
BSHTEXT[a]	Basal sheath texture (papery/firm/hard)
BSHGLOSS[a]	Basal sheath surface (glossy/dull)
Floral characters	
NUMSPK	Number of spikelets per inflorescence
SPKLEN	Spikelet length (cm)
SPKWID	Spikelet width (mm)
GLLEN	Glume length (mm)
GLWID	Glume width (mm)
NUTLEN[b]	Nutlet length (mm)
NUTWID[b]	Nutlet width (mm)
NUMGL	Number of glumes per spikelet

MATERIAL AND METHODS

Sampling and measured characters

In this study, 480 specimens and 9 type images were examined from the herbaria B, C, K, LD, and P (Supplementary material 1). We define a specimen as a single herbarium sheet, which may include several plants. Specimens were named using a personalised system consisting of either one or two letters for the species name followed by two letters for its country and the collector number, e.g. a specimen of *C. margaritaceus* from Ghana with the collection number 357 was labelled as M.Gh.357.

Fifteen morphological characters (12 quantitative and 3 qualitative) were used for the analyses (Table 3). Measurements were taken using a ruler and a Leica S6E stereomicroscope with a graticule. Where possible, up to 10 measurements for each quantitative character were taken per specimen. The average value was calculated for each character per specimen and this average value was used in the character matrix.

PCA, LDA, and CART analysis

All analyses were carried out in R (R Core Team 2021) using R Studio v.1.3.1056 (RStudio Team 2021). The data set was imported as a matrix with the specimens in rows and the characters in columns. Missing values were examined using the summary function. The characters nutlet length and nutlet width, originally recorded for the study, were removed from the analyses because they included a large number of missing values (266 individuals out of 489). The other missing values within the matrix were imputed using the missForest package (Stekhoven and Bühlmann 2012) based on random forest algorithms, which accepts both continuous and discrete data and is non-parametric. A distance matrix was computed from the imputed data set (quantitative and qualitative characters) with the cluster package (Maechler et al. 2019), using Gower's distance coefficient (Legendre and Legendre 2012) in the daisy function.

The PCA was carried out on a matrix of the scaled quantitative variables of the imputed data set using the prcomp function from the stats package (loaded automatically with R). The number of significant principal components was computed using the evplot function (Borcard et al. 2011).

The LDA was carried out using the lda function from the package MASS v.7.3-49 (Venables and Ripley 2002; Ripley 2018). The analysis followed the approach of Legendre and Legendre (2012) and Borcard et al. (2011). Standardized data were used for computing the contribution of variables to the discriminant function axes whilst untransformed data was used for testing the allocation of individuals to species. The data sets were tested for homogeneity of covariance matrices using the betadisper function from the vegan package (Oksanen et al. 2018). Since the data usually violate the basic assumptions of the parametric LDA (multivariate normal distribution, homoscedastic group covariance matrices),

intergroup differences were compared using cross-validation based on the untransformed data set.

Classification and Regression Tree (CART) analysis is a non-parametric data-mining algorithm that can use qualitative and quantitative data sets, either singly or mixed, to classify individuals into pre-defined categories (classification trees) or predict the value of some quantitative trait of interest (regression trees); this procedure is known technically as binary recursive partitioning (Crawley 2013). In our study, the 489 specimens grouped into the hypothetical taxa (see Introduction) were classified into these species categories by the CART algorithm. The purpose of this was to evaluate how well the result of the algorithmic operation would match the original taxonomic determinations. Significant features of the CART analysis are that the results are clearly understandable in terms of the original characters (variables) used; there are no prior assumptions as regards probability distributions of the variables, and an a priori classification of the individuals is needed (i.e. it is a supervised classification method; Ripley 1996). This makes it a suitable method for testing conventional taxonomic determination of individuals into species.

The classification algorithm creates a dichotomous tree, similar to a key, beginning by dividing (partitioning) the complete set of individuals (the root node) into two daughter nodes, and then successively dividing these into subnodes and so on until some stopping criterion is reached. At every node, the algorithm searches each variable separately to find the variable value (threshold) that divides the node set of individuals into the two least heterogeneous subnodes (Foulkes 2009; Varmuza and Filzmoser 2009). The heterogeneity (or impurity) of a node is measured by the proportion of individuals correctly assigned to their original species, by comparing with the pre-determined classification of the taxonomist. The variable that produces the least impure daughter nodes then becomes the splitting criterion for that stage of the tree, and its threshold value is used to partition the individuals, e.g. in the first split of our tree, the left daughter node comprises the individuals meeting the criterion of glume width ≥ 1.95 mm and the right daughter node those with glume width < 1.95 mm. We used the rpart package (Therneau and Atkinson 2018) to carry out the analysis. For classification trees, the default measure for node impurity is the Gini index (Foulkes 2009: 162). The tree plot was made using the rpart.plot package (Milborrow 2020).

To avoid overfitting, the initial tree must be pruned back to the subtree that is considered optimally predictive (Foulkes 2009: 173). This is achieved by cost complexity pruning. The cost complexity of the tree is a measure that balances the error associated with the tree and the number of terminal nodes it has (i.e. the size of the tree). This is calculated using a complexity parameter α ≥ 0, which penalizes the tree (increases overall error) as tree size increases. By using the pruning process, a selection of subtrees is computed, using a cross-validation procedure, for a range of values of the complexity parameter. The overall error (tree impurity) of these optimal subtrees is also estimated by cross-validation, and the subtree that minimizes the cost complexity is selected as the best (Foulkes 2009: 174–175). The terminal nodes (leaves) are usually a mixture of individuals of the pre-determined species, and the algorithm automatically decides the name of each node by the species with majority representation. For both the LDA and CART analysis, the cross-validation assignments of the individuals of the hypothetical species to the resulting groups (or nodes) were visualized with stacked barplots. The assignments of each specimen were listed and exported to an Excel file, which made it possible to track the assignments of the type specimens.

RESULTS

Principal Component Analysis (PCA)

The purpose of the PCA was to show which characters contribute most to the overall variability of the data set, without considering differences between the species.

The PCA did not produce a very favourable dimensional reduction of the data since the first nine axes were needed to capture 95% of the total variance. However, only the first three principal components, which expressed 66% of total variance, were found to be significant according to the mean eigenvalue (Supplementary materials 2, 3), and only the first had a significant eigenvalue according to the "broken stick" model (MacArthur 1957). The hypothetical species show considerable overlap in the PCA ordination of the first two principal components with no single group clearly separating in the multivariate space (Fig. 3). This is primarily due to the overwhelming representation of *C. margaritaceus* – the largest sample in our study – contributing to much of the overlap with the other groups. *Cyperus ledermannii* and *C. somaliensis* are the only groups that show no overlap with *C. margaritaceus*. The other eight groups show some separation from each other in the multivariate space, notably *C. karlschumannii*, *C. ledermannii*, and *C. nduru*. *Cyperus obtusiflorus* separates from *C. karlschumannii* and *C. somaliensis* but overlaps to varying degrees with the remaining groups. The biplot of the first two principal components shows that the most influential characters on the first axis are leaf length, spikelet width, and culm length, whereas the most influential character on the second axis is the number of involucral bracts (Supplementary material 4).

Linear Discriminant Analysis (LDA)

The hypothetical species groups did not have homogeneous covariance matrices and so assessment of their relative distinctness was based mainly on the results of the cross-validation tests.

The LDA of the non-scaled data expressed 78.62% of the total variance in the first two discriminant axes.

When compared to the PCA, the ordination of these axes showed slight improvement in the separation of the nine taxa (Fig. 4). *Cyperus karlschumannii*, *C. ledermannii*, and *C. nduru* each form separate groups from each other in the multivariate space, while the remaining groups overlapped to varying degrees. *Cyperus margaritaceus* formed a more concentrated group with less general overlap in the multivariate space compared to the PCA.

The loadings determine which characters are the most influential in separating the species on the first and second discriminant axes (Supplementary materials 5, 6). The most important discriminating characters on the first axis (A) are glume width, leaf length, glume length, and number of spikelets; individuals with high values in the character combination glume width and leaf length are plotted on the left-hand side of the ordination, while those with high values of the combination glume length and number of spikelets are plotted on the right-hand side. On the second axis (B) individuals with high values for culm length and number of involucral bracts are plotted lower in the ordination and contrast with those with high values for glume length and number of spikelets are plotted higher up.

The LDA cross-validation was carried on the untransformed data set. This procedure gives a better guide to the distinctiveness of the species than the ordinations, because the algorithmic allocation of each individual to a group uses all the discriminant function axes, whereas the ordinations only show two-dimensional patterns. In Table 4, the rows show how many of the individuals of the original species were in fact assigned to each resulting group by the cross-validation algorithm. The columns of the table show the mixture of individuals from the original species in each resulting group; the name of each cross-validated group (column names) is that of the species with the largest number of individuals allocated to it. Figure 5 is a visual representation of the same data in which the bars represent the groups resulting from cross-validation and the coloured segments show the proportions of the different original species present
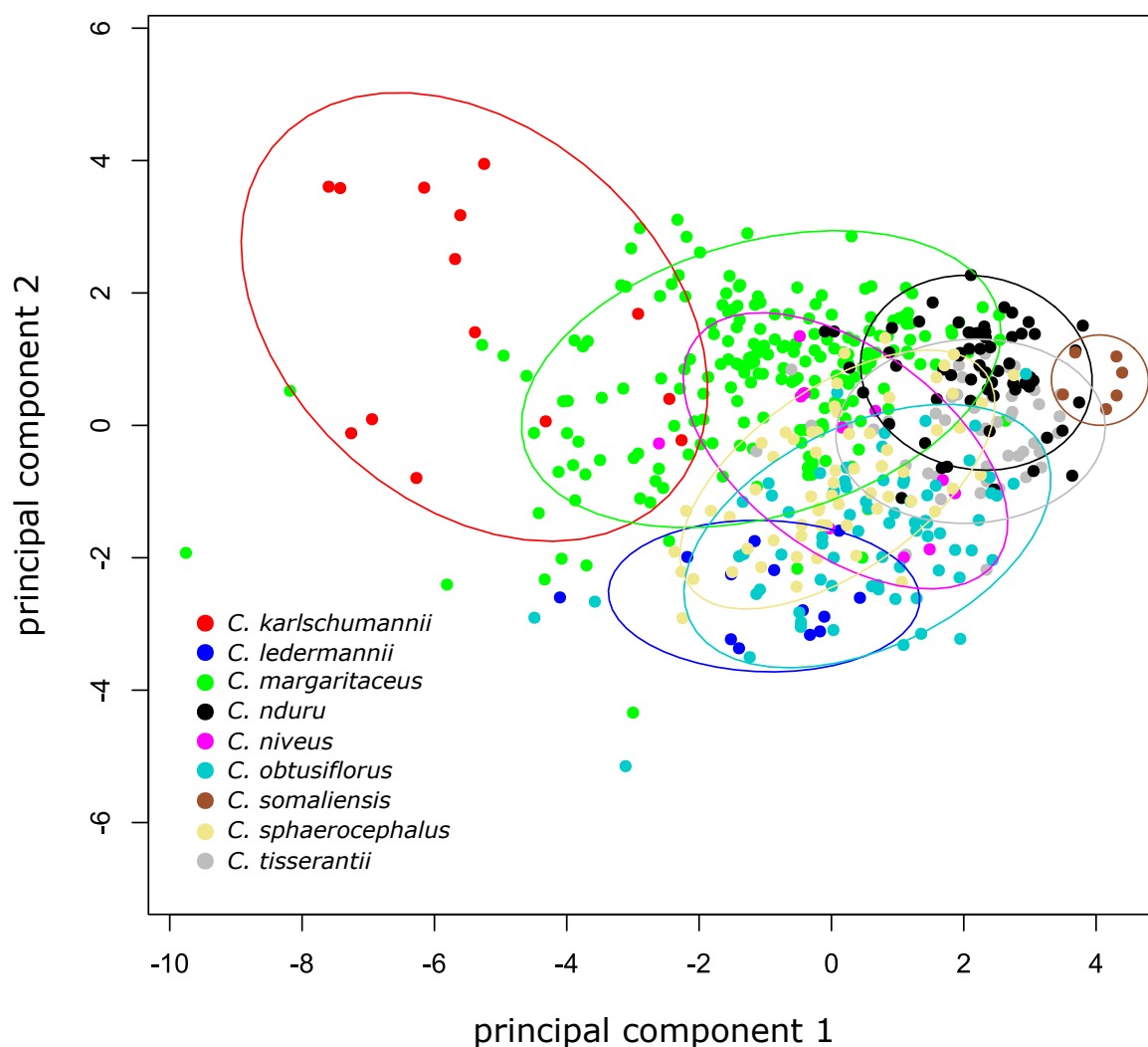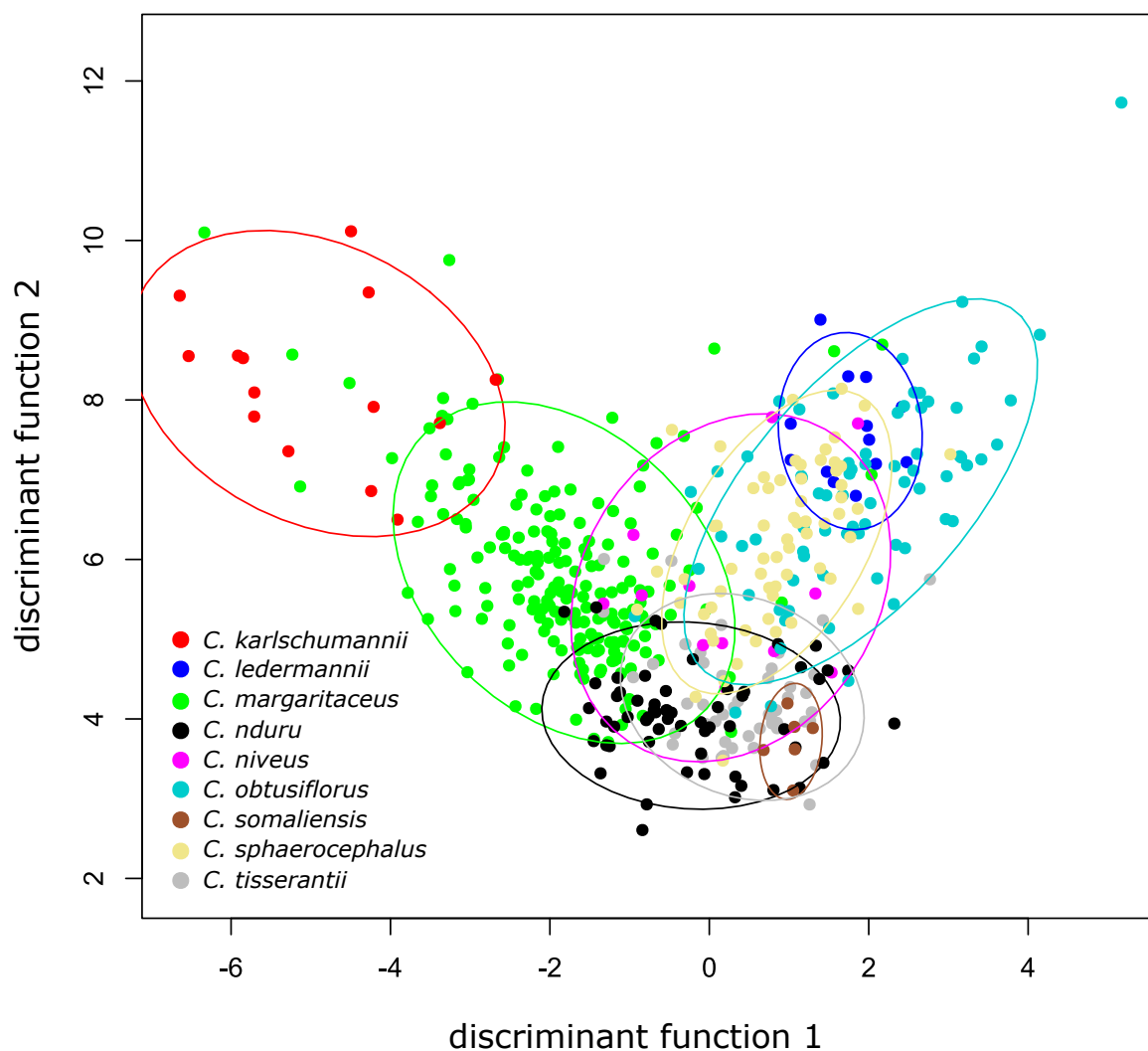

Figure 3. Principal component analysis (PCA) ordination of the scores of the first two principal components, with 95% confidence ellipses shown for each species. Based on scaled data of 12 quantitative measured variables of specimens from the *Cyperus margaritaceus-niveus* complex. The colours represent the nine hypothetical species of the complex.

in each group. Essentially, each bar provides a visual impression of the consistency of the originally determined species: the more mixed the bars, the less consistent the species, whereas bars that are predominantly one colour show the species to be relatively more distinct. Thus, each bar receives the name of the species with majority representation.

The results showed that *C. margaritaceus* was the most consistent species, with 87.1% of the originally determined individuals correctly assigned to the resulting cross-validated group with the same name, and 86.6% of the individuals of this group belonging to the species (Table 4, Fig. 5). This result was surprising, given the larger sample in the analyses and geographical range of this species, the wide intra-specific variation shown in the two-dimensional ordinations (Figs 3, 4) and the fact that individuals of six species were assigned to its group. None of the individuals of *C. niveus* were recognized as this species by the cross-validation, with all its individuals being distributed among groups named as other species; of

the five individuals assigned to the cross-validated group named as *C. niveus*, two of them were originally named as *C. obtusiflorus*, two were named as *C. margaritaceus*, and one as *C. sphaerocephalus*. It is important to note that high purity of the cross-validated groups do not necessarily guarantee high % correct values, or vice versa. For example, although the cross-validated group named as *C. somaliensis* consists only of that species (% comp. = 100%), only 50% of the individuals originally named as *C. somaliensis* were assigned to that group; the remaining individuals were assigned to other cross-validated groups (Table 4). Insofar as allocation of the type specimens are concerned, five of the type specimens were assigned to their respective cross-validated groups (*C. margaritaceus*, *C. karlschumannii*, *C. ledermannii*, *C. nduru*, and *C. tisserantii*). For the remaining type specimens, each of them was assigned to different cross-validated groups; *C. niveus* assigned to *C. obtusiflorus*, *C. obtusiflorus* assigned to *C. sphaerocephalus*, *C. sphaerocephalus* assigned to *C. ledermannii*, and *C. somaliensis* assigned to *C. tisserantii*.

Figure 4. Linear discriminant analysis (LDA) ordination of the first two discriminant axes, with 95% confidence ellipses shown for each species. Based on scaled data of 12 quantitative measured variables of specimens from the *Cyperus margaritaceus-niveus* complex. The colours represent the nine hypothetical species of the complex.

Table 4. Cross validation results of LDA using non-standardised data. Rows show the original population memberships, while columns show the composition of the cross-validated populations. Numbers in **bold** show the number of specimens from the original taxa that are assigned to the cross-validated populations. Row sums show the total number of original individuals per taxon, and the percentage correct value shows the number of correctly assigned individuals to their original taxon. Column sums show the number of original individuals assigned to each cross-validated population, and the '% comp.' value is the percentage of individuals in the cross-validated population that belong to the original population with the same name.

	C. karlschumannii	*C. ledermannii*	*C. margaritaceus*	*C. nduru*	*C. niveus*	*C. obtusiflorus*	*C. somaliensis*	*C. sphaerocephalus*	*C. tisserantii*	SUM	% correct
C. karlschumannii	**9**	0	5	0	0	0	0	0	0	14	64.3
C. ledermannii	0	**9**	0	0	0	1	0	3	0	13	69.2
C. margaritaceus	9	3	**175**	5	2	3	0	2	2	201	87.1
C. nduru	0	0	5	**44**	0	0	0	1	10	60	73.3
C. niveus	0	0	4	1	**0**	2	0	1	4	12	0.0
C. obtusiflorus	0	0	4	2	2	**50**	0	10	4	72	69.4
C. somaliensis	0	0	0	1	0	0	**3**	0	2	6	50.0
C. sphaerocephalus	0	3	5	3	1	5	0	**45**	2	64	70.3
C. tisserantii	0	0	4	6	0	1	0	0	**36**	47	76.6
SUM	18	15	202	62	5	62	3	62	60		
% comp.	50	60	86.6	71.0	0	80.6	100	72.6	60		

Table 5. Cross validation results of CART using non-standardised data. Rows show the original population memberships, while columns show the composition of the cross-validated populations. Numbers in **bold** show the number of specimens from the original taxa that are assigned to the cross-validated populations. Row sums show the total number of original individuals per taxon, and the percentage correct value shows the number of correctly assigned individuals to their original taxon. Column sums show the number of original individuals assigned to each cross-validated population, and the '% comp.' value is the percentage of individuals in the cross-validated population which belong to the original population with the same name.

|	karl.crt	marg1.crt	marg2.crt	nduru1.crt	nduru2.crt	obtus.crt	sphaer1.crt	sphaer2.crt	tisser.crt	SUM	% correct
C. karlschumannii	**10**	4	0	0	0	0	0	0	0	14	71.4
C. ledermannii	0	0	0	0	0	8	2	2	1	13	0.0
C. margaritaceus	5	**175**	**5**	6	0	3	2	3	2	201	87.0
C. nduru	0	0	0	**36**	**18**	0	0	2	4	60	90
C. niveus	0	3	0	0	0	2	0	3	4	12	0.0
C. obtusiflorus	0	3	4	0	1	**47**	5	7	5	72	65.3
C. somaliensis	0	0	0	0	2	0	0	0	4	6	0.0
C. sphaerocephalus	0	8	0	0	1	5	**26**	**21**	3	64	73.4
C. tisserantii	0	9	0	3	2	1	0	3	**29**	47	61.7
SUM	15	202	9	45	24	66	36	41	52		
% comp.	66.7	86.6	55.6	80.0	75.0	71.2	72.2	51.2	55.8		

Classification and Regression Tree (CART) analysis

Three optimal subtrees were identified that had a relative error below the critical threshold (Supplementary material 7). Each of these trees resulted from one cycle of pruning and had 7, 9, and 11 leaves respectively. The objective of the complexity pruning function (cp) is to balance the prediction error of the tree with the number of terminal nodes (its complexity), which in this case means that the optimal tree had 7 leaves (i.e. the simplest tree with an error below the critical threshold). However, we chose to focus on the tree with nine leaves, since this addresses a taxonomic question of equal interest, that is, how well the algorithm sorted the individuals from the nine original species into nine reconstituted taxa. Our result suggests

further investigation is needed of the 7-leaf optimal classification produced by this CART analysis.

As with the LDA cross-validation, the species with the largest number of individuals assigned to a terminal group provides the name of that group. This resulted in no terminal groups for *C. ledermannii*, *C. somaliensis*, and *C. niveus*, and two each for *C. margaritaceus*, *C. nduru*, and *C. sphaerocephalus* (Fig. 6). The results of the classification are expressed in the tree in Fig. 7, in which an estimate of the percentage success of determining individuals into each species given the data used, can be provided. *Cyperus margaritaceus* was split by number of spikelets, while *C. nduru* and *C. sphaerocephalus* were separated by glume width.
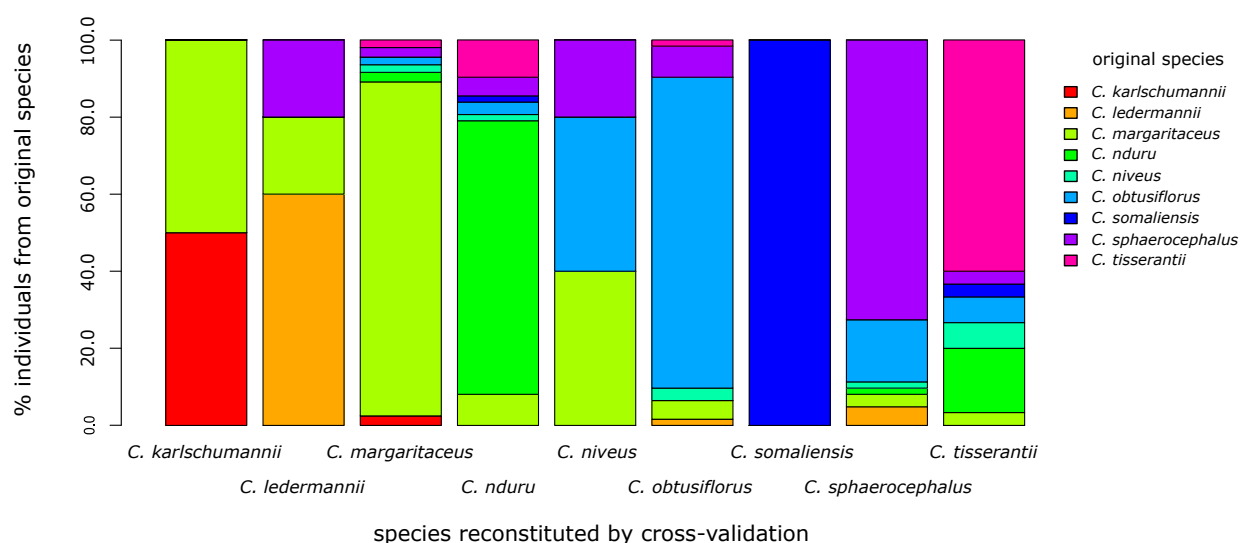
Figure 5. Bar plot representation of the cross-validation table for the LDA. Each bar represents the composition of one cross-validated group. The colours of species are shown in each bar as the proportion of originally determined individuals that make up that cross-validated group.
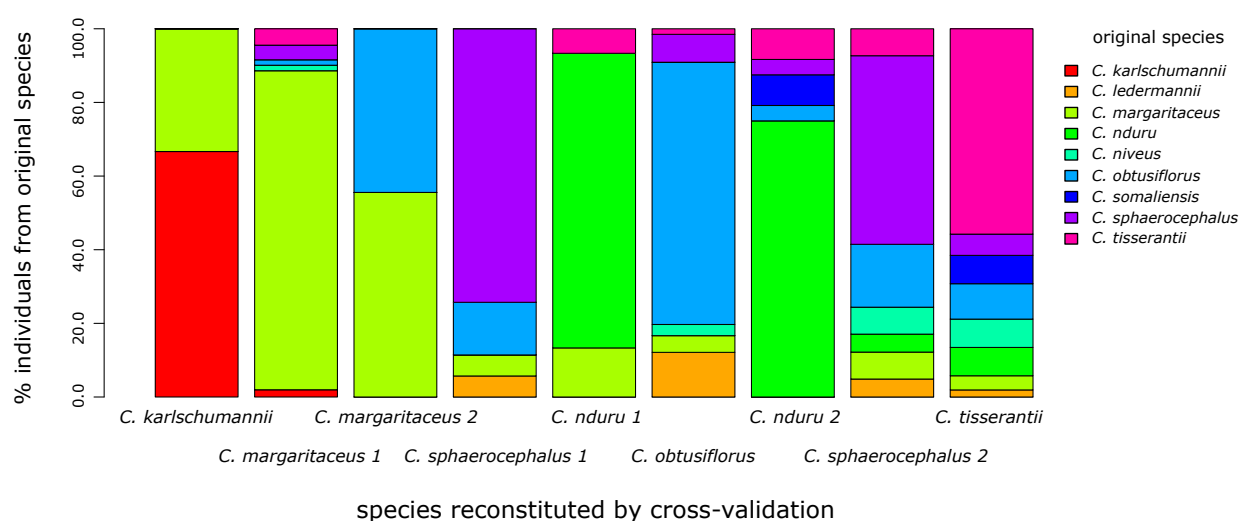
Figure 6. Bar plot representation of the cross-validation table for the CART analysis. Each bar represents the composition of one cross-validated group. The colours of species are shown in each bar as the proportion of originally determined individuals that make up that cross-validated group.

An overview of the assignments of individuals to the cross-validated groups resulting from the CART analysis are given in Table 5. The table is interpreted in the same way as the LDA cross-validation table in that the composition of the hypothetical species and the composition of the resulting "species" (i.e. the cross-validated groups that result from the analysis) are distinct. Depending on the distinctness of the original hypothetical sets of individuals (the species), the analysis will assign a greater or lesser proportion of them to the cross-validated group with the same name, and the remainder to other groups. Consequently, the cross-validated groups that result from the analysis are usually a mixture of individuals from the original species, as shown by the bars in Fig. 6. These groups (suffixed as .crt in Table 5) receive the name of the species with majority representation. Like the LDA cross-validation, high % correct values do not necessarily correlate with low impurity of the cross-validated groups, or vice versa.

Nevertheless, six of the hypothetical taxa that formed a cross-validated group bearing the same name had high percentages of correct assignments; *C. margaritaceus* and *C. nduru* with the highest percentages, 87% and 90% respectively. Of the three original taxa that failed to come out as separate groups, two of them had many of their individuals assigned to one group. *Cyperus ledermannii* had 62% of its individuals, including the type, assigned to the obtus.crt group. *Cyperus somaliensis* had 67% of its individuals, including the type, assigned to the tisser.crt group. In the case of *C. niveus*, not only did the original taxon fail to separate but none of its individuals received

majority representation in any cross-validated group. The type was assigned to the obtus.crt group. Only four of the type specimens were assigned to a cross-validated group bearing the same name: *C. margaritaceus*, *C. karlschumannii*, *C. nduru*, and *C. tisserantii*. For the remaining two taxa, although these come out as their own cross-validated groups, the types were assigned elsewhere. Thus, *C. obtusiflorus* was assigned to tisser.crt and *C. sphaerocephalus* was assigned to nduru2.crt.

DISCUSSION

Within the *Cyperus margaritaceus-niveus* complex, the analyses classify the individuals, originally assigned to nine species, into eight groups in the LDA and six groups in the CART analysis. The two results we wish to highlight are the failure of *C. niveus* to separate as a group in both analyses and the retention of *C. margaritaceus* as a distinct entity mixed together with a small number of individuals of other species. In the PCA, although the plots for *C. niveus* and *C. sphaerocephalus* are embedded within the overall scatter plot, *C. sphaerocephalus* is still separated by LDA and CART, while *C. niveus* is not separated by either analysis. In both the LDA and CART cross-validation, the 12 specimens of *C. niveus* were distributed among several taxa, although without consistency between the two analyses (Tables 4, 5). It is possible that characters not included in this study would help to separate this species as a cross-validated group, but our present results imply that *C. niveus* is not sufficiently distinct from the
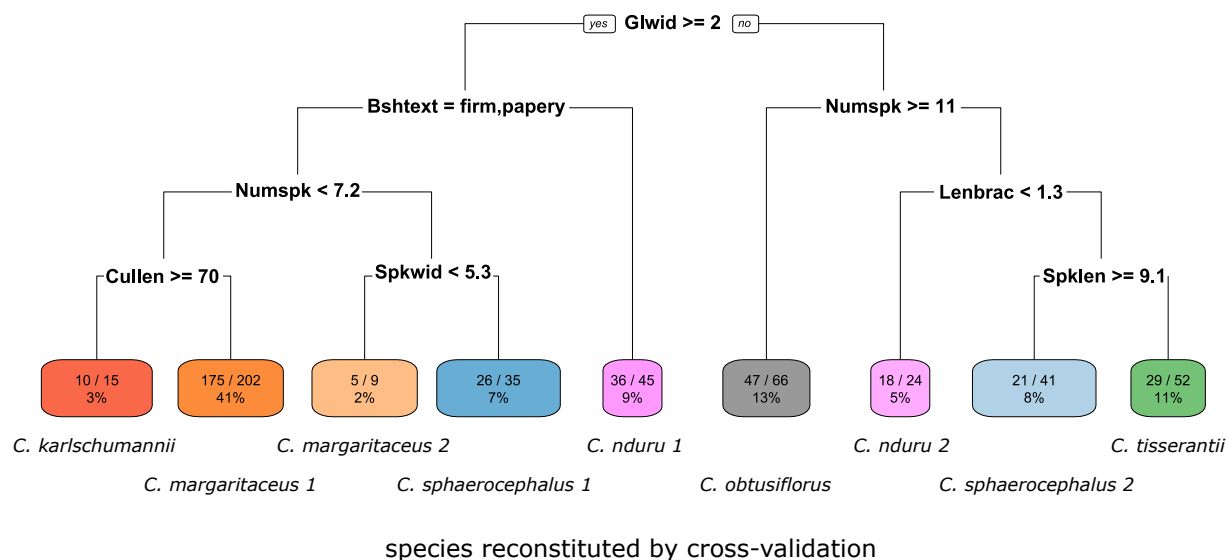
Figure 7. Classification tree after one cycle of cost complexity pruning. The terminal nodes are the result of assignment by cross-validation. The terminal cross-validated groups are named by the species with the largest number of individuals assigned to that group; the numbers separated by a slash represent on the left the number of individuals originally from the species of the leaf name, and on the right the total number of individuals assigned to that leaf by CART; the percentages represent the proportion of the total number of individuals of the study assigned to that leaf. Each node is marked by the character that provides the optimal binary split of the individuals at that node; the logical statement at each node (e.g. glume width at the root node) indicates that individuals for which the statement is true pass to the left-hand branch and those for which it is false to the right-hand branch.

other taxa in the complex. This corroborates the view of Fosberg (1977) who did not consider *C. niveus* as a separate species when compared to *C. obtusiflorus* even though he focussed on material from Aldabra. The taxonomic differences between the species according to Kükenthal (1936) were somewhat dismissed by Fosberg as "a weak difference indeed"; hence, his reduction of *C. obtusiflorus* into *C. niveus*, which subsequent regional Floras have followed. In our analysis the type specimen of *C. niveus* from India was also placed in the *C. obtusiflorus* group although Fosberg recognised the typical variety only of *C. niveus* as being confined to southern Asia. No Asian material was examined in this study, but clearly the taxonomic status of *C. niveus* requires further elucidation.

Cyperus margaritaceus shows considerable overlap with the other taxa in the PCA (Fig. 3) with 41% of the total sample study (Table 4) and the widest geographical range of the species in the complex. However, despite this apparent "clouding" of the data in the two-dimensional ordinations, the high percentage of correctly assigned individuals in the cross-validations of the LDA and CART analysis shows that *C. margaritaceus* is a robust entity, supporting its circumscription in the regional African Floras as a distinct species. Its delimitation as a separate species is further strengthened by a character not used in our analyses but highlighted by Hooper and Napper (1972): the presence of a rhizome consisting of a series of connected culm bases. While this character is also found in other species in the complex, in *C. margaritaceus* it is found in combination with discrete spikelets (as opposed to confluent spikelets) that are relatively few in number and moderate in length. The CART analysis separated the species into two groups based on the number of spikelets in an inflorescence; individuals in the group marg1.crt had less than 7.2 spikelets per inflorescence, while those in group marg2.crt, a considerably smaller group, had more spikelets. This character alone may not be sufficient for splitting *C. margaritaceus* into two species and furthermore, the number of original *C. margaritaceus* individuals in this group is tiny with no geographical correlation among them.

Cyperus nduru was treated as a variety of *C. margaritaceus* by Haines and Lye (1983) but as a separate species by Hoenselaar et al. (2010) and Browning et al. (2020). The LDA supports the assertion that *C. nduru* is a distinct entity, though the CART analysis separates them into two groups by glume width. All the specimens from west tropical Africa were assigned to the nduru1.crt group, while the specimens from D.R.Congo, east tropical and south tropical Africa were assigned to both the nduru1. crt and nduru2.crt groups. Our results support the notion that *C. nduru* is a distinct entity within the complex but that glume width alone may not be sufficient to recognise two separate taxa especially since the individual specimen assignments show no clear geographical distinction between the nduru.crt groups.

The taxonomic status of *C. tisserantii* differs amongst the regional Floras (Table 2). Hooper and Napper (1972) separated it from *C. nduru* based on the number of involucral bracts and their length in relation to the spikelets. Our analyses separate this species as a distinct group although the characters used by Hooper and Napper (1972) do not necessarily have high influence according to the LDA axes. Number of involucral bracts scored relatively high on both axes but length of involucral bracts scored very low for the first axis (Supplementary material 5). Nevertheless, the CART analysis shows that *C. tisserantii* is recognised as a distinct entity based on glume width, number of spikelets, length of bracts, and spikelet length. The length of the bracts was used by CART to separate this species from *C. tisserantii*, and although the PCA shows considerable overlap with other taxa, mainly *C. nduru*, the cross-validation analyses show that these two species – *C. tisserantii* and *C. nduru* – are two separate groups.

The analyses show strong support for *C. karlschumannii* as a distinct entity. This species can be easily distinguished from the other taxa in the complex by the considerably longer ovate spikelets, ranging from white to pale-yellow and with broader glumes.

The analyses differ as to the identity of *C. ledermannii*. While the LDA treats it as a separate entity, the CART analysis places most of the *C. ledermannii* individuals in the obtus.crt group. Hooper (1972) originally raised the status of *C. ledermannii* from variety to species commenting that "this taxon seems as worthy as others in the *C. obtusiflorus-margaritaceus* group of specific recognition", although there was no direct comparison made with *C. obtusiflorus*. *Cyperus ledermannii* is restricted to west tropical Africa though Hooper commented that its occurrence in east tropical Africa seemed probable. Haines and Lye (1983), who kept it as a variety of *C. niveus* following Kükenthal (1936), noted that, for east Africa, this taxon is only recorded from Tanzania. The type of *C. niveus* var. *ledermannii*, observed for this study on JSTOR, has a large of head of many discrete spikelets and in this respect is most similar to *C. obtusiflorus*. The CART analysis supports Hoenselaar et al. (2010) and Browning et al. (2020) who reduced *C. ledermannii* to synonymy under *C. niveus* var. *leucocephalus*. It is therefore possible that *C. ledermannii* does not represent a separate morphospecies.

Cyperus sphaerocephalus, known as the golden headed sedge in South Africa, was long considered as *C. obtusiflorus* var. *flavissimus*, until Hilliard and Burtt (1986) made morphological comparisons between it and the typical variety of *C. obtusiflorus*. They found that several quantitative characters (number of spikelets, spikelet length and width, glume length and width, involucral bract length, diameter of head), rather than inflorescence colour alone, separated the two varieties and hence elevated *C. obtusiflorus* var. *flavissimus* to species level as *C. sphaerocephalus*. Within South Africa, these two taxa have also been treated as separate species based on differing distributional patterns. Gordon-Gray (1995) observed how *C. sphaerocephalus* was mainly confined to Natal, while the "white-headed form" was predominantly

at "Coastland though Midlands, into Zululand and N Natal". Gordon-Gray (1995) stressed that further investigation should cover both entities across the whole of Africa. Since our methods take the whole range of these species into account, we conclude that our results support the notion of both species representing distinct entities, although the CART analysis separates the taxon further based on glume width. Individuals of *C. sphaerocephalus* make up the majority of both sphaer.crt groups but the individual assignments of this taxon for each group reveal a geographical correlation. The majority of individuals in the sphaer1.crt group are from South Africa, whereas those from sphaer2.crt are mostly from outside of South Africa. Further work is needed to determine whether they represent subspecies or varieties.

The entity of *C. somaliensis* also differs between the analyses. While the LDA distinguishes this species as a distinct entity, the CART analysis shows little consistency with regards to assigning individuals. This disparity may be due to the comparative lack of specimens used that belong to this taxon; only six specimens of *C. somaliensis* were used in the analyses. When Clarke (1895) originally described the species, no diagnosis was given though Clarke commented it was "near *C. leucocephalus*" a species from India, which has morphological similarities. In Flora of Somalia, Lye (1995) compared it with *C. niveus*, although *C. niveus* is not a distinct entity in any of the analyses and further work is needed to establish its relationship to other species in the complex.

The assignment of type specimens to different cross-validated groups across the analyses makes their classification difficult. Types may not be typical of the specimens assigned to their particular species, however, missing character values may alter their true position in the multispace and hence affect their assignment in the cross-validation.

For the complex as a whole, characters with the highest loadings (glume width, number of bracts, glume length, and number of spikelets per head) were also used by Haines and Lye (1983), Hoenselaar et al. (2010), and Browning et al. (2020) for delimiting some of these taxa at varietal rank. Similarly, these characters were used by Hooper and Napper (1972) to separate them at species level. Despite differences in taxonomic opinion (see Table 2), the fact that the characters used in these regional Floras also represent the highest loadings in the PCA indicate their taxonomic informativeness for delineating the taxa.

In summary, of the nine putative species studied, six morphospecies can be recognised based on our results: *C. karlschumannii*, *C. margaritaceus*, *C. nduru*, *C. obtusiflorus*, *C. sphaerocephalus*, and *C. tisserantii*. Each of the six morphospecies is distinguished by a combination of morphological characters and geographical distribution. The picture is more complex for *C. ledermannii* and *C. somaliensis*, given the differences between the LDA and CART analysis. Further work is desirable to elucidate the taxonomic affinities of these two taxa. None of the analyses considered *C. niveus* to be a separate entity

for Africa. We have also shown that the morphological characters used in classical taxonomic studies for this complex are of sufficient taxonomic value to delimit similar morphogroups as presented in the regional Floras. The analyses used in this study provide a transparent and repeatable methodology for justifying the recognition of morphology-based taxa within this complex, which also enables morphological characters, not included in this study (e.g. nutlet dimensions, anther length), to be added to our dataset to corroborate our results. We recommend the application of this methodology to resolve other known species complexes in the Cyperaceae.

DATA AVAILABILITY STATEMENT

ACKNOWLEDGEMENTS

REFERENCES

Atkinson MD, Codling AN (1986) A reliable method for distinguishing between *Betula pendula* and *B. pubescens*. Watsonia 16: 75–87.

Barrett CF, Freudenstein JV (2009) Patterns of morphological and plastid DNA variation in the *Corallorhiza striata* species complex (Orchidaceae). Systematic Botany 34: 496–504. https://doi.org/10.1600/036364409789271245

Borcard D, Gillet F, Legendre P (2011) Numerical Ecology with R. Second Edition. Springer, Cham, Switzerland, 1–435. https://doi.org/10.1007/978-3-319-71404-2

Browning J, Gordon-Gray KD, Lock M, Beentje, H, Vollesen K, Bauters, K, Archer C, Larridon I, Xanthos M, Vorster P, Bruhl J, Wilson K, Zhang X (2020) Cyperaceae. In: García MA, Timberlake JR (Eds) Flora Zambesiaca, vol. 14. Royal Botanic Gardens, Kew, 1–455.

Brysting AK, Elven R (2000) The *Cerastium alpinum-C. arcticum* complex (Caryophyllaceae): numerical analyses of morphological variation and a taxonomic revision of *C. arcticum* Lange s.l. Taxon 49: 189–216. https://doi.org/10.2307/1223835

Carter R, Bryson CT (2000) *Cyperus sanguinolentus* (Cyperaceae) new to the south eastern United States, and its relation to the supposed endemic *Cyperus louisianensis*. SIDA, contributions to Botany 19(2): 325–343. https://www.jstor.org/stable/41968941

Clarke CB (1895) CCCLXXV – Diagnoses Africanæ, VII. Bulletin of Miscellaneous Information, Kew 105: 229.

Crawley MJ (2013) The R Book. Second Edition. Wiley, Chichester, 1–977. https://doi.org/10.1002/9781118448908

Dayrat B (2005) Towards integrative taxonomy. Botanical Journal of the Linnean Society 85: 407–415. https://doi.org/10.1111/j.1095-8312.2005.00503.x

Di Natale S, Lastrucci L, Hroudová Z, Viciani D (2020) A review of *Bolboschoenus* species (Cyperaceae) in Italy based on herbarium data. Plant 156(1): 261–270. https://doi.org/10.1080/11263504.2020.1852330

Duvigneaud P, Denaeyer-de Smet S (1963) Cuivre et végétation au Katanga. Bulletin de la Société Royale de Botanique de Belgique 96(2): 93–232. https://www.jstor.org/stable/20792417

Fosberg FR (1977) Miscellaneous notes on the flora of Aldabra and neighbouring islands: IV: a new *Bulbostylis* and observations on *Cyperus* (Cyperaceae). Kew Bulletin 31(4): 829–835. https://doi.org/10.2307/4109553

Foulkes AS (2009) Applied Statistical Genetics with R for Population-Based Association Studies. Springer, New York, 1–252. https://doi.org/10.1007/978-0-387-89554-3

Gardner LR, Weber O, Simpson DA (2014) *Cyperus beentjei*, a new species of Cyperaceae from Tropical East Africa delimited through morphometric analysis. Kew Bulletin 69: 9501. https://doi.org/10.1007/S12225-014-9501-5

Gordon-Gray KD (1995) Cyperaceae in Natal. Strelitzia 2: 1–218.

Gray A, Stott G (2017) *Cyperus stroudii* (CYPERACEAE), a new species from Ascension Island, South Atlantic Ocean. Phytotaxa 331(1): 65–74. https://doi.org/10.11646/phytotaxa.331.1.4

Haines RW, Lye KA (1983) The sedges and Rushes of East Africa. East Africa Natural History Society, Nairobi, 1–404.

Hilliard OM, Burtt BL (1986) Notes on some plants of Southern Africa chiefly from Natal: XIII. Notes from the Royal Botanic Garden Edinburgh 43(3): 345–405.

Hoenselaar K, Verdcourt B, Beentje HJ (2010) Cyperaceae. In: Beentje HJ (Ed.) Flora of Tropical East Africa. Royal Botanic Gardens, Kew, 1–466.

Hooper SS (1972) New taxa, names and combinations in Cyperaceae for the 'Flora of West Tropical Africa'. Kew Bulletin 26(3): 577–583. https://doi.org/10.2307/4120322

Hooper SS, Napper DM (1972) Cyperaceae. In: Hepper FN (Ed.) Flora of West Tropical Africa, vol. 2(3). Crown Agents, London, 278–349.

Kučera J, Marhold K, Lihová J (2010) *Cardamine maritima* group (Brassicaceae) in the amphi-Adriatic area: a hotspot of species diversity revealed by DNA sequences and morphological variation. Taxon 59: 148–164. https://www.jstor.org/stable/27757059

Kükenthal G (1936) Cyperaceae – Scirpoideae – Cypereae. In: Engler A. (Ed.) Das Pflanzenreich, vol. 101. Engelmann, Leipzig, 1–671.

Larridon I. Villaverde T. Zuntini AR, Pokorny L. Brewer, GE, Epitawalage N, Fairlie I, Hahn M, Kim J, Maguilla E, Maurin O, Xanthos M, Hipp AL, Forest F, Baker WJ (2020) Tackling rapid radiations with targeted sequencing. Frontiers in Plant Science 10: 1655. https://doi.org/10.3389/fpls.2019.01655

Legendre P, Legendre L (2012) Numerical Ecology. Third English Edition. Elsevier, Amsterdam, 1–990.

Lowe PD (2018) Studies in the Cyperaceae of Georgia: distribution of Georgia sedges, analysis of the *Cyperus squarrosus-granitophilus* complex & two new species. Master's Thesis, Valdosta State University, Georgia. https://hdl.handle.net/10428/3154

Lye KA (1995) Angiospermae (Hydrocharitaceae-Pandanaceae). In: Thulin M. (Ed.) Flora of Somalia, vol. 4. Royal Botanic Gardens, Kew, 98–166.

Lye KA (1997) Hydrocharitaceae to Arecaceae. In: Edwards S, Demissew S, Hedberg I (Eds) Flora of Ethiopia and Eritrea, vol. 6. The National Herbarium, Ethiopia, 391–511.

MacArthur RH (1957) On the relative abundance of bird species. Proceedings of the National Academy of Sciences 43: 293–295. https://doi.org/10.1073/pnas.43.3.293

Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K, Studer M, Roudier P, Gonzalez J, Kozlowski K, Schubert E, Murphy K (2019) cluster: "Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al. R package version 2.1.0. https://svn.r-project.org/R-packages/trunk/cluster/ [accessed 17.02.2023]

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 Verlag, Ruggell, Liechtenstein, 73–99.

Mayo SJ (2022) Plant taxonomic species and their role in the workflow of integrative species delimitation. Kew Bulletin 77: 1–26. https://doi.org/10.1007/S12225-022-10002-X

Míguez Ríos M (2017) Evolution of *Carex* section *Rhynchocystis* (Cyperaceae): phylogenetic, biogeographic and taxonomic approaches. PhD Thesis, Universidad Pablo de Olavide de Sevilla, Spain. https://hdl.handle.net/10433/4800

Milborrow S (2020) Plot 'rpart' Models: an enhanced version of 'plot.rpart'. R package version 3.0.9. https://cran.r-project.org/web/packages/rpart.plot [accessed 17.02.2023]

Naczi RFC, Moyer RD (2016) Revision of the *Rhynchospora glomerata* species complex, focusing on the taxonomic status of *R. leptocarpa* (Cyperaceae). Brittonia 69(1): 114–126. https://doi.org/10.1007/s12228-016-9452-2

Naczi RFC, Reznicek AA, Ford BA (1998) Morphological, geographical and ecological differentiation in the *Carex willdenowii* complex (Cyperaceae). American Journal of Botany 85(3): 434–447. https://doi.org/10.2307/2446335

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, Mcglinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2018) vegan: community ecology package. R package version 2.4-6. https://cran.r-project.org/web/packages/vegan [accessed 17.02.2023]

Pante E, Schoelinck C, Puillandre N (2015) From integrative taxonomy to species description: one step beyond. Systematic Biology 64(1): 152–160. https://doi.org/10.1093/sysbio/syu083

Perný M, Tribsch A, Stuessy TF, Marhold K (2005) Taxonomy and cytogeography of *Cardamine raphanifolia* and *C. gallaecica* (Brassicaceae) in the Iberian Peninsula. Plant Systematics and Evolution 254: 69–91. https://doi.org/10.1007/s00606-005-0317-5

R Core Team (2021) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ [accessed 17.02.2023]

Ripley BD (1996) Pattern Recognition and Neural Networks. Cambridge University Press, Cambridge, 1–403. https://doi.org/10.1017/CBO9780511812651

Ripley BD (2018) MASS: Support functions and datasets for Venables and Ripley's MASS. R package version 7.3-49. https://cran.r-project.org/web/packages/MASS [accessed 17.02.2023]

Rosen DJ (2006) A systematic study of select species complexes of *Eleocharis* subgenus *Limnochloa* (Cyperaceae). PhD Thesis, Texas A&M University, USA.

Rivero-Guerra AO (2011) Morphological variation within and between taxa of the *Santolina rosmarinifolia* L. (Asteraceae: Anthemideae) aggregate. Systematic Botany 36(1): 171–190. https://doi.org/10.1600/036364411X553261

R Studio Team (2021) RStudio: integrated development for R. PBC, Boston, USA. https://posit.co/products/open-source/rstudio/ [accessed 17.02.2023]

Smith TW, Waterway MJ (2008) Evaluating the taxonomic status of the globally rare *Carex roanensis* and allied species using morphology and amplified fragment length polymorphisms. Systematic Botany 33: 525–535. https://doi.org/10.1600/036364408785679824

Stekhoven DJ, Bühlmann P (2012) MissForest – non-parametric missing value imputation for mixed-type data. Bioinformatics 28(1): 112–118. https://doi.org/10.1093/bioinformatics/btr597

Therneau T, Atkinson B (2018) rpart: recursive partitioning and regression trees. R package version 4.1-13. https://cran.r-project.org/web/packages/rpart [accessed 17.02.2023]

Varmuza K, Filzmoser P (2009) Introduction to multivariate statistical analysis in chemometrics. CRC Press, Boca Raton, 1–336. https://doi.org/10.1201/9781420059496

Venables WN, Ripley BD (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York, 1–497. https://doi.org/10.1007/978-0-387-21706-2

Will KW, Mishler BD, Wheeler QD (2005) The perils of DNA barcoding and the need for integrative taxonomy. Systematic Biology 54(5): 844–851. https://doi.org/10.1080/10635150500354878

SUPPLEMENTARY MATERIALS

Supplementary material 1
Accessions used in the study and their morphological characters.
Link: https://doi.org/10.5091/plecevo.97453.suppl1

Supplementary material 2
Barplots representing eigenvalues of the PCA.
Link: https://doi.org/10.5091/plecevo.97453.suppl2

Supplementary material 3
Eigenvalues of the principal component analysis shown in Fig. 3.
Link: https://doi.org/10.5091/plecevo.97453.suppl3

Supplementary material 4
Biplot of the PCA.
Link: https://doi.org/10.5091/plecevo.97453.suppl4

Supplementary material 5
Character loadings of the first (A) and second (B) discriminant axes based on the scaled data of the Linear Discriminant Analysis.
Link: https://doi.org/10.5091/plecevo.97453.suppl5

Supplementary material 6
Loadings (eigenvectors) of the linear discriminant analysis shown in Fig. 4.
Link: https://doi.org/10.5091/plecevo.97453.suppl6

Supplementary material 7
Complexity pruning plot based on CART analysis.
Link: https://doi.org/10.5091/plecevo.97453.suppl7