New Phytologist

Research

Methods

Quantification of complex modular architecture in plants

Catherine Reeb[1], Jaap Kaandorp[2], Fredrik Jansson[2], Nicolas Puillandre[1], Jean-Yves Dubuisson[1], Raphaël Cornette[3], Florian Jabbour[1], Yoan Coudert[4], Jairo Patiño[5,6], Jean-François Flot[7] and Alain Vanderpoorten[8]

[1]Institut de Systématique, Évolution, Biodiversité (ISYEB – UMR7205 – Sorbonne Universités MNHN, CNRS, EPHE) Muséum national d'Histoire Naturelle, 57 rue Cuvier CP 50, 75005 Paris, France; [2]Computational Science Lab, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, the Netherlands; [3]Équipe Évolution et Développement des Variations Phénotypiques (ISYEB – UMR7205 – MNHN, CNRS, Sorbonne Universités EPHE) Muséum national d'Histoire Naturelle, Sorbonne Universités, 57 rue Cuvier CP 50, 75005 Paris, France; [4]Laboratoire Reproduction et Développement des Plantes, Ecole Normale Supérieure de Lyon, CNRS, INRA, Université Claude Bernard Lyon 1, 46 Allée d'Italie, 69007 Lyon, France; [5]Island Ecology and Evolution Research Group, Instituto de Productos Naturales β Agrobiología (IPNA-CSIC), La Laguna, Tenerife, Spain; [6]Department of Environmental Science, Policy and Management, University of California, Berkeley, CA 94720, USA; [7]Evolutionary Biology & Ecology, Université Libre de Bruxelles, Avenue F.D. Roosevelt 50, C.P. 160/12, 1050 Brussels, Belgium; [8]Institute of Botany, University of Liège, B22 Sart Tilman, 4000 Liège, Belgium

Author for correspondence:
Catherine Reeb
Tel: +33 140773200
Email: catherine.reeb@mnhn.fr

Received: *16 November 2017*
Accepted: *7 January 2018*

New Phytologist (2018) **218:** 859–872
doi: 10.1111/nph.15045

Key words: integrative taxonomy, modular architecture, morphogeometry, software, species delimitation analysis.

Summary

- Morphometrics, the assignment of quantities to biological shapes, is a powerful tool to address taxonomic, evolutionary, functional and developmental questions. We propose a novel method for shape quantification of complex modular architecture in thalloid plants, whose extremely reduced morphologies, combined with the lack of a formal framework for thallus description, have long rendered taxonomic and evolutionary studies extremely challenging.
- Using graph theory, thalli are described as hierarchical series of nodes and edges, allowing for accurate, homologous and repeatable measurements of widths, lengths and angles.
- The computer program MorphoSnake was developed to extract the skeleton and contours of a thallus and automatically acquire, at each level of organization, width, length, angle and sinuosity measurements.
- Through the quantification of leaf architecture in *Hymenophyllum* ferns (Polypodiopsida) and a fully worked example of integrative taxonomy in the taxonomically challenging thalloid liverwort genus *Riccardia*, we show that MorphoSnake is applicable to all ramified plants. This new possibility of acquiring large numbers of quantitative traits in plants with complex modular architectures opens new perspectives of applications, from the development of rapid species identification tools to evolutionary analyses of adaptive plasticity.

Introduction

Morphometrics, the assignment of quantities to biological shapes, is a powerful tool to address taxonomic, evolutionary, functional and developmental questions (Biot *et al.*, 2016; Coudert *et al.*, 2017). Morphometric analysis has undergone a dramatic renaissance in recent years, embracing a range of novel computational and imaging techniques to provide new approaches to phenotypic characterization (Stanton & Reeb, 2016) assisted by the development of a wide array of image analysis softwares (Lobet, 2017).

In particular, morphometrics appears as a promising tool in the fast-developing field of integrative taxonomy (Puillandre *et al.*, 2009). As a result of the development of molecular techniques for examining biological diversity in the context of large-scale DNA barcoding projects and of statistical techniques allowing fast and objective species delimitation (Fontaneto *et al.*, 2015), the speed of discovery of new 'molecular' species has substantially increased (Monaghan *et al.*, 2009; Fujita *et al.*, 2012). Although such species are increasingly mentioned in the recent literature, they may not necessarily differ in 'traditional' taxonomic characters and are almost never formally described, typified and named by their discoverers because many taxonomists are reluctant to describe species based solely on molecular characters (Pante *et al.*, 2015). As a result, the magnitude of the 'taxonomic impediment' – that is, the time-lag between species recognition and description – increases (Oliver & Lee, 2010), calling for the development of methods for the fast acquisition of new morphological characters.

Morphometric studies involve the quantitative comparison of homologous organs such as leaves, internodes, phytomeres and growth units (Esteve-Altava, 2017). Homology assessment is, however, an issue in modular organisms, which are characterized by the indefinite growth of repeated, similar units, called modules

(Harper *et al.*, 1986), making it difficult to compare modules of the same order among specimens. For root systems, the Root System Markup Language was developed to provide a standard format of root phenotypic characterization (Lobet *et al.*, 2015). In thalloid organisms, such a standardized framework is still lacking, so that descriptions by different authors may use different terminologies and may not draw correct homologous comparisons among different portions of the thallus. In a discussion on the difficulty to define botanical structures such as 'rhizome' or 'ramicaule', Tomlinson (1987) already suggested that it is impossible to describe ramification patterns and processes solely 'by eye' or by using *a priori* definitions; a sound understanding of the ramification ontogenesis and of its function is necessary; mathematical formalism entails universal abstraction and thus makes it possible to name the parts of a branched object without ambiguity. Tomlinson (1987) and Mishler & De Luna (1991) therefore stressed the need to develop a standardized terminology for an accurate designation of the different component parts of branched structures.

Here, we propose that a formal description of thalli, using a standardized terminology based on a mathematical background to unambiguously identify homologous modules among specimens, is a necessary prerequisite to any morphometric analysis. We then use this standardized descriptive framework to present a new tool aimed at automatically measuring a series of homologous quantitative traits through the development of a new computer program, MorphoSnake. We present the main features of this software and illustrate its use through two examples of application in thalloid liverworts and ferns.

Description

Thallus description

We applied concepts developed for the description of branched marine organisms (Kaandorp & Kübler, 2001; Kruszyński *et al.*, 2007). A thallus (Fig. 1a) can be represented by its skeleton (Fig. 1b) and its contour (Fig. 1c). The skeleton is the median axis of the shape and can be depicted as lines connecting the centers of maximum discs reaching the contour (Fig. 1b). The formalization of thallus terminology is conducted as an application of graph theory (Chatrand, 1985; Diestel, 2010). A thallus skeleton constitutes nodes v (= vertex; a set of vertices = V) interconnected by edges e (set = E) (Fig. 1c). There is only one path between two nodes with no fused branches. One can define the oldest part of the thallus, born from the development of the diaspore, and the youngest parts at the apex of terminal segments, near the embryonic apical cell. The graph is now oriented as a rooted tree. Each thallus node can be characterized by the number of neighbour nodes and assigned to coordinates in a 2D system $d_g(v_i)$. Three node types are distinguished (Fig. 1d):

Terminal nodes Connected to only one neighbour. The root is the oldest point of the thallus and the only terminal node far from an apical cell; terminal nodes $d_g(v_i) = 1$.

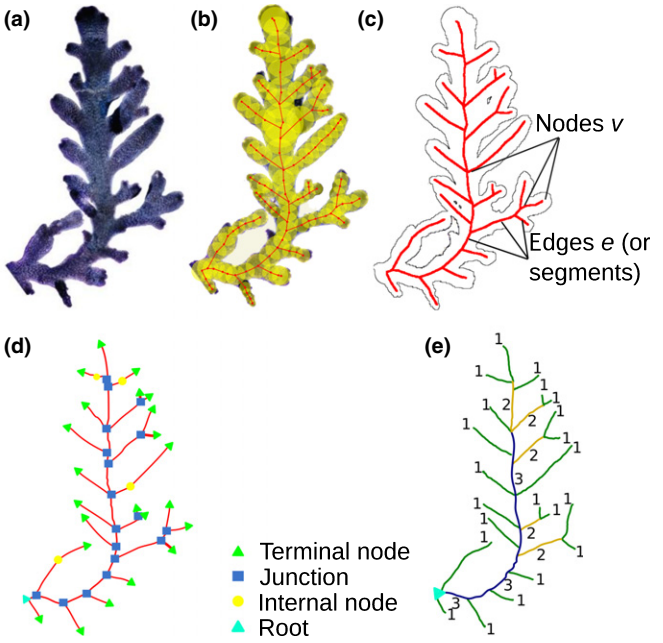

Fig. 1 Construction of the contour and Strahler ordering of the skeleton. (a) Original thallus of the thalloid liverwort *Riccardia longispica*. (b) Construction of the skeleton resulting from the connection of the centers of discs reaching the contour. (c) Contour and skeleton. (d) Junctions, terminal nodes and edges. (e) Strahler ordering of the segments.

Junction nodes Connected to more than two neighbours. They represent the meeting points of the skeleton, the ramification points; junction $d_g(v_i) > 1$.

Internal nodes Connected to two neighbours; internal nodes $d_g(v_i) = 2$.

Edge is the portion between nodes v_0 and v_n such as $d_g(v_{0,n+1}) \neq 2$ and $d_g(v_{1\ldots n-1}) = 2$.

The thallus as a hierarchical structure

Much work on branched structures was conducted in the middle of the 20th Century, with particular attention to the hierarchy of river networks, as originally proposed by Horton (1945) and Strahler (1952), and subsequently developed by Zanardo *et al.* (2013), among others. By analogy with a hierarchy of tributaries, the Strahler number is a numerical measure of branching complexity and the Strahler order corresponds to a given hierarchical level. By definition, the terminal node of a thallus and the corresponding edge are of level 1 ($\omega = 1$). The hierarchy of the following nodes v and edges e is built in a centripetal way: the order of the daughter branch after a junction is calculated by taking the maximum order of the two parent edges and adding a δ. If two parent branches are of the same order, then $\delta = 1$; if they are of different orders, then $\delta = 0$ (Figs 1e, 2c).

The maximum order of ramification is obtained after applying the hierarchy to the entire thallus (Fig. 1e).

The terms describing a thallus can thus be redefined as follows (Fig. 2):

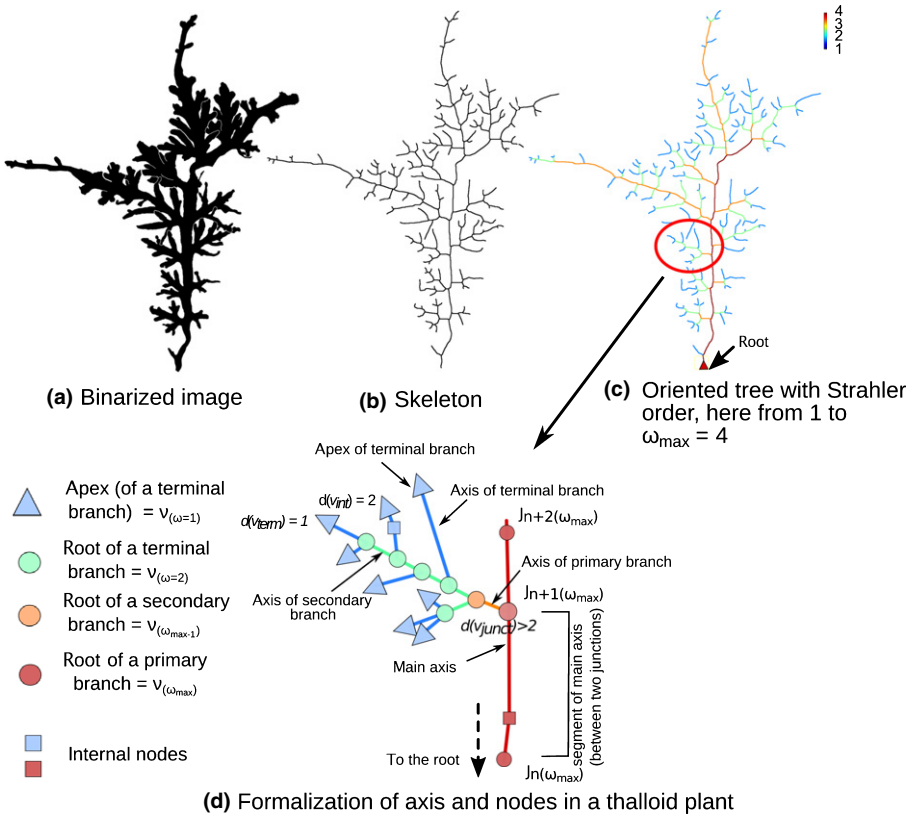

Fig. 2 MorphoSnake workflow. (a) Binarized image of a thallus of the thalloid liverwort *Riccardia multifida* obtained by ImageJ. (b) Skeleton obtained by the 'Skeletonize' function in ImageJ. (c) Strahler ordering heatmap obtained by the Strahler plugin in ImageJ. (d) Formalized thallus description.

An *axis of order* ω_n is a set of edges of order ω_n
$\text{axis}(\omega_n) = \sum_1^i(e)$ with $\omega(e) = n$
$\text{axis}(\omega_n) = \sum_1^i(e)$ with $\omega(e) = n$

The *main axis* is of order ω_{max} and contains the root of the thallus main axis, ω_{max} : main axis $(\omega_{max}) = \sum_1^i(e)$ with $\omega(e) = \omega_{max}$

A *branch of order* ω_n is a sub-tree of the thallus, rooted by a junction v of order ω_{n+1} and its axis contains the root.
A *primary branch* is a branch rooted by a junction v with $\omega(v) = \omega_{max}$
A *secondary branch* is a branch rooted by a junction v with $\omega(v) = \omega_{max-1}$ and a *branch of order n* is a branch rooted by a junction v with $\omega(v) = \omega_{max-(n-1)}$
A *terminal branch* is a branch of order 1 rooted by a node v_{term} with $\omega(v_{term}) = 2$.

Thallus measurements

Following an approach proposed for corals and sponges (Kaandorp, 1999; Kaandorp & Kübler, 2001), three categories of measurement are recognized: widths, lengths and angles (Fig. 3).

Widths These can be measured at junctions (IW) and apical points (TW) and are characterized by a maximum (IW_{max}) and mean (IW_{mean}) value at each segment. Width is measured by node diameter or segment width. Node diameter (IW, TW) is defined as the diameter of the largest possible disc centered on the node, which is fully inside the object. Thus, node diameter is twice the distance from the node to the closest background pixel. Segment widths are defined in a similar way: at each pixel in the skeleton, width is defined as twice the distance to the closest background pixel. For branches, maximum, minimum and average widths are calculated over all skeleton pixels in the branch.

Lengths This can be measured between two junctions (TL, IL) and two apices (ApDist). These distances are calculated as Euclidean distances or following the skeleton.

Angles These are measured between axes; terminal angles (TA) are distinguished from internal angles (IA).

Based on these measurements, TL : TW, IL : IW ratios, which provide information on shape variation, as well as a sinuosity index (Tcurv, Icurv), which measures the ratio between the total length along the skeleton and the Euclidean distance between two nodes, can be computed.

Automated thallus morphometrical measurements

A visualization of the skeleton (Figs 1c, 2b) and a quantitative description of its hierarchization (Figs 1e, 2c) can be obtained from a binarized image (Figs 1a, 2a) using the software ImageJ (Schindelin *et al.*, 2012; Schneider *et al.*, 2012) and the Strahler plugin (Ferreira, 2016) (Supporting Information Methods S1). Building on Branchometer (Konglerd *et al.*, 2017), which allowed automatic size measurements, but was not designed to take into account the hierarchical Strahler order, we developed

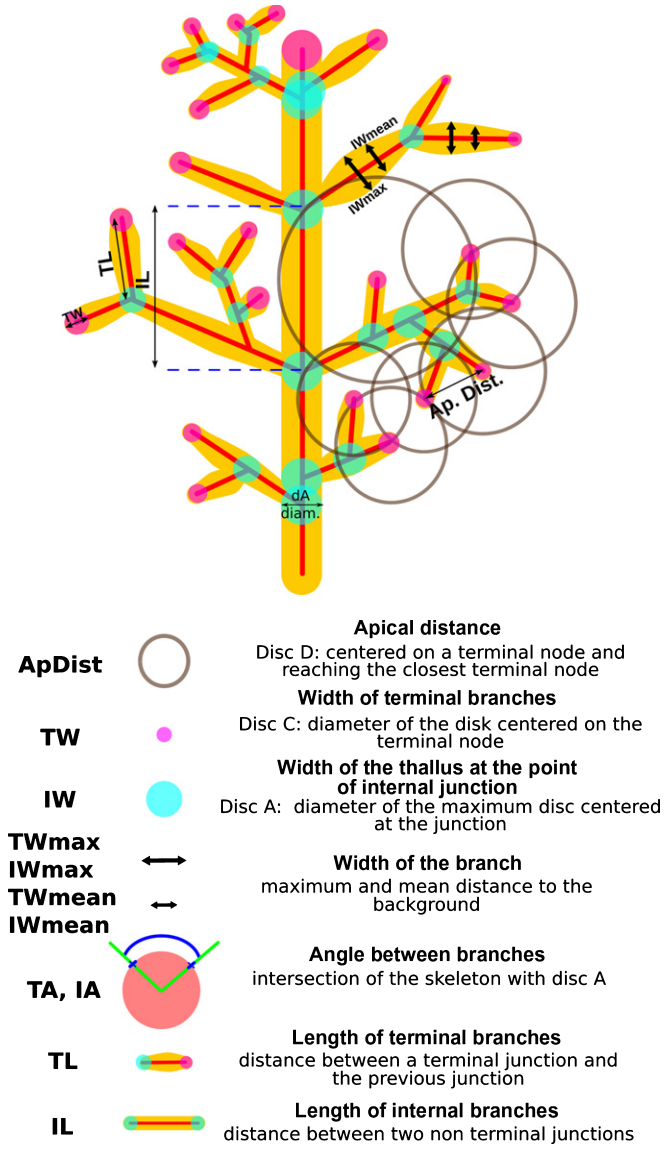

Fig. 3 Quantitative thallus shape variables available in MorphoSnake.

the new program MorphoSnake. MorphoSnake was written in the modular Python language to automatically acquire series of quantitative traits at each level of the organization of the thallus, a crucial feature for subsequent homologous comparisons that differentiates MorphoSnake from Branchometer. Similar morphometric programs were developed for the analysis of root systems (e.g. Lobet *et al.*, 2011), but could not be directly applied to thalloid plants for two main reasons. First, although previous programs made it possible also to automatically identify hierarchical branching patterns, they required the root branch to be first defined. In thalloid plants, very complex branching patterns often make it impossible to readily identify the root branch. In MorphoSnake, therefore, the hierarchy of the branches is built in a centripetal way (see above).

We illustrate the range of application of MorphoSnake using two examples: a fully worked example of integrative taxonomy, from the molecular delimitation of species in the taxonomically challenging thalloid liverwort genus *Riccardia* to the morphometrical characterization of their inter- and intraspecific variation; and the quantification of leaf architecture in *Hymenophyllum* ferns (Polypodiopsida) in the context of interspecific discrimination.

Materials and Methods

Intra- and interspecific morphological characterization in the simple thalloid liverwort genus *Riccardia*

We compared the usefulness of quantitative morphometrical characters (hereafter called 'quantitative characters') acquired by MorphoSnake to that of traditional categorical morphological characters (hereafter categorical characters) for the characterization of infra- and interspecific variation. Our analysis focused on simple thalloid liverworts of the genus *Riccardia* from Africa, a taxonomically very challenging case. Indeed, *Riccardia* is the largest member of the taxonomically little-understood family of simple thalloid liverworts Aneuraceae, and African species are particularly poorly known. Each thallus is variously branched (Fig. S1a) and homogeneous in cross-section (Fig. S1b), thus offering few characters with substantial overlap and high levels of plasticity (Preußing *et al.*, 2010). We therefore defined species molecularly before partitioning the variance of each of the categorical and quantitative characters within and among species and determining whether combinations of the two kinds of characters vary depending on taxonomic identity at the species and generic levels, and habitat conditions within and among species.

Specimen sampling and molecular protocols In total, 183 collections from the 10 *Riccardia* species currently reported from sub-Saharan Africa (Reeb & Bardat, 2014) were sampled for molecular species delimitation (Table S1). Collections of Aneura (11), the sister genus to *Riccardia* (Preußing *et al.*, 2010), and six collections of the newly segregated genus *Afroriccardia* (Rabeau *et al.*, 2017) were selected as outgroups.

Before DNA extraction, specimens were cleaned in distillate water under the dissecting scope to remove contaminants such as epiphyllous liverworts or moss debris. Three to four thalli of 1–2 cm long were transferred into 2-ml round-bottom microcentrifuge tubes and dried in an incubator at 60–70°C for 3–4 h. Two tungsten beads (2 mm) and one volume of pure silica sand were added to each sample to facilitate the grinding of the tissues using a TissueLyser (Qiagen). DNA extraction was performed using the DNA Mini Kit (Qiagen) following the manufacturer's instructions. For old herbarium specimens, a preliminary step was applied, during which 400 µl AP1 lysis buffer, 30 µl CTAB (cetyltrimethylammonium bromide) buffer and 30 µl proteinase K were added to each specimen. Each tube was then placed for 20–24 h in a thermocycler at 42°C before 460 µl of chloroform isoamyl alcohol (96 : 4) were added. The tubes were gently mixed by inversion and centrifuged for 15 min at 13 000 ***g***.

Preliminary investigation revealed that sequence variation at *rbcL*, which has been identified, along with *matK*, as a suitable chloroplastic barcode for plants (CBOL Plant Working Group,

2009), was insufficient in *Riccardia*. Therefore, we combined *matK* with a set of three other markers, namely the chloroplast *trnL-F* and *psbA-trnH* regions and the nuclear *ITS2* region based on their amplification success and variability at the species level. These loci were amplified with the primers and PCR conditions detailed in Table S2. PCR reactions took place in a total volume of 20 µl of PCR Buffer with 1 mM MgCl2, 0.26 mM of dNTP mix, 0.25 mM of each primer (0.5 mM for diluted DNA solution), 1 µl DMSO (dimethyl sulfoxide), 1 µl BSA (bovine serum albumine), and 0.06 mM of QbioTaq (Quiagen) DNA polymerase. PCR products were sequenced by Genoscope (Lille, France) or Eurofins (https://www.eurofins.fr/). Forward and reverse sequences were assembled and edited using Geneious 7.1 (http://www.geneious.com, Kearse *et al.*, 2012).

Morphological and morphometrical measurements Categorical and quantitative characters were scored on a subset of 143 thalli from the molecular dataset and from 11 additional outgroups of the sibling genera *Afroriccardia* (seven collections of *A. comosa*) and *Aneura* (four collections of *A. pseudopinguis*). Morphological characters included 25 multistate characters, including 19 unordered (categorical) and six ordered characters, which have been used for species differentiation in African *Riccardia* (Table S3a). For the purpose of multivariate analyses (see below), the unordered characters were transformed into additive binary characters, resulting in a total of 35 characters (Table S3a). In agreement with previous taxonomic treatments of the genus, our analysis focused on vegetative characters because gametangia and sporophytes, which are seldom produced, were not available on each specimen.

Specimens were first moistened and inspected visually to remove debris, identify damaged branches that are excluded from the analysis, and deal with overlapping branches, which present a challenge for the automated description of morphological features (Unger *et al.*, 2016). Unger *et al.* (2016) circumvented this issue by implementing a semi-automatic approach, wherein the user first exemplarily marks a few points on the leaf and on the background. Here, overlapping branches were manually moved during specimen processing so that they all fitted in 2D. Although fully automatic routines have been proposed (Corney *et al.*, 2012), our approach allowed efficient and flexible processing while dealing with overlapping and damaged branches.

For morphometrical analysis, each thallus was photographed using a Nikon Coolpix 6000 attached to a binocular microscope. Foreground thalli were extracted using the Simple Interactive Object Extraction (Friedland *et al.*, 2006) implemented in Gimp 2.8 to generate black-and-white images. Five to 15 thalli per collection were employed, and the observations were averaged across specimens for subsequent analyses. In order to compare levels of variation at higher taxonomic levels, we also included specimens of the sibling genera *Afroriccardia* (seven collections of *A. comosa*) and *Aneura* (four collections of *A. pseudopinguis*). Using MorphoSnake, we scored 17 quantitative characters, including average node diameter, length of the branch along the skeleton (branch length), Euclidian, minimal distance between the proximal and distal ends of the branch (branch length – e), angle from parent branch, measured at the node's disk (alpha), sinuosity, and maximum branch width for internal and terminal branches, apical distance for terminal branches, maximum order of the thallus and total number of branches. The habitat (substrate type) of each specimen was scored based upon label information and assigned to one of three categories, including wood, rock and soil.

Species delimitation analyses The various species delimitation analyses available to date can be classified into three groups distance-based, tree-based and allele-sharing based approaches (Flot, 2015). Based upon the fact that the different species delimitation analyses make different assumptions that are rarely met in a particular system, Carstens *et al.* (2013) advocated the use of several techniques followed by an assessment of the congruence among them. Here, we implemented one method of each of the three groups mentioned above, including the allele-sharing based approach Haplowebs (Flot *et al.*, 2010), the distance-based approach ABGD (Automatic BarCode Gap Discovery) (Puillandre *et al.*, 2012), and the tree-based approach GMYC (General Mixed Yule Coalescent) (Monaghan *et al.*, 2009; Zhang *et al.*, 2013).

The ABGD approach requires two user input values: P, the maximum intraspecific distance, and X, a proxy for the minimum gap width. Following Puillandre *et al.* (2012), we kept the default value of $W = 1.5$ but tested a range of *P*-values between 0.0001 and 0.01. We reported the results associated with the lowest and highest *P*-values.

For GMYC, the best-fit substitution model was assessed for each locus with jModelTest2. For each locus, we ran three independent evolutionary models (uncorrelated lognormal relaxed clock, constant coalescent, Yule) to obtain ultrametric trees using Beast v.2.4.4 (Drummond & Rambaut, 2007; Bouckaert *et al.*, 2014) on Cipres (Cyberinfrastructure for Phylogenetic Research) (http://www.phylo.org/sub_sections/portal/). We ran three independent chains of 100 million generations each, sampling every 5000 generations. We plotted the likelihood values of each run to make sure that each chain had reach stationarity and that all three chains had converged, and further checked that the effective sample size of each parameter was sufficient (> 200) using Tracer v.1.6 (Rambaut *et al.*, 2014). The best-fit of the three tested evolutionary models, as assessed using the Akaike Information Criterion through Markov chain Monte Carlo (Raftery *et al.*, 2007) implemented by Tracer v.1.6, was an uncorrelated relaxed clock for *matK* and *ITS2* and a constant coalescent model for *trnLF* and *pbsbA-trnH*. For each marker, we combined the three runs for the best-fit model with LogCombiner 2.4.5 (Bouckaert *et al.*, 2014) in order to generate the maximum clade credibility tree. A single-threshold GMYC model (Fujisawa & Barraclough, 2013) was applied to the maximum clade credibility tree of Beast using the Splits package (Ezard *et al.*, 2009).

For Haplowebs, forward and reverse *ITS2* sequences were assembled and the contigs visually examined with Sequencher 4 as detailed in Fontaneto *et al.* (2015). Homozygote consensus sequences were obtained directly from the contigs. For heterozygotes, haplotypes were constructed manually when the

chromatograms exhibited double peaks at a single position. When double peaks were observed at several positions, haplotypes were reconstructed using CHAMPURU (Flot, 2007) when they had different lengths (resulting in very large numbers of double peaks; Flot *et al.*, 2006) and using SEQPHASE (Flot, 2010) otherwise. Haplotype sequences were aligned using MEGA6 and GENEIOUS 7.1. Haplowebs were drawn from a RAXML tree of haplotypes using CIPRES RAXML-HPC2 on XSEDE (https://www.phylo.org/portal2/) using rapid bootstrapping with random seed. Specimens of species that have recently diverged typically share haplotypes and are considered to belong to the same 'field for recombination' (Doyle, 1995). By contrast, specimens from long-isolated species do not share any haplotypes (Fontaneto *et al.*, 2015).

Following recent studies (Castro-Romero *et al.*, 2016; Montagna *et al.*, 2016; Song *et al.*, 2017), outgroup species were included in the species delimitation analyses for two main reasons. First, although related species tend to be unduly lumped by the GMYC model when distantly related species are included (https://francoismichonneau.net/gmyc-tutorial/), this was clearly not the case here, as more species were recognized with the GMYC criterion than in the two other analyses (Fig. 5). Second, Talavera *et al.* (2013) explicitly recommended that outgroups should be included in GMYC analyses to maximize the number of species included, as the GMYC model poorly performs when the number of species is low (Dellicour & Flot, 2015).

The congruence between each of the three species delimitation techniques applied to each locus was assessed by building a conspecificity matrix (Debortoli *et al.*, 2016; Johnston *et al.*, 2017) using HEATMAP3 (Zhao *et al.*, 2014). This matrix was obtained by computing, for each pair of individuals, a conspecificity score equal to the number of loci/methods supporting the hypothesis of their conspecificity minus the number of loci/methods suggesting that they belong to different species, then reordering the rows and columns to maximize the scores along the diagonal.

Infra- and interspecific morphological and morphometrical variation Following the assignment of each specimen to one of the molecularly identified species, we first used one-way ANOVA to partition the variance of each categorical and quantitative character within and among species through the computation of R^2, the ratio of the sum of squares for the model divided by the sum of squares for the corrected total, which measures how much variation in the dependent variable can be accounted for by the model. An ANOVA was subsequently used to test the hypothesis that, on average, the infraspecific variance of quantitative characters among collections was higher than that of categorical ones.

We then determined whether variation in combinations of categorical and quantitative characters globally varied depending on taxonomy at the species and genus levels and habitat conditions. We first summarized the information from each of the categorical and quantitative character matrices by means of a principal component analysis (PCA) based on the correlation matrix among characters. We then used ANOVAs to determine whether there were significant differences in the PCA scores for the categorical and quantitative characters, respectively, depending on the factor 'substrate type'. We performed this analysis both with the complete sampling of specimens across all *Riccardia* species and among conspecific specimens within *R. longispica*, which was the species for which the largest number of specimens ($n = 69$) was available.

We finally employed a linear discriminant analysis (LDA) to determine whether the categorical and quantitative characters significantly vary among *Riccardia* species and between *Riccardia*, *Afroriccardia* and *Aneura*. LDA is designed for the analysis of variables that are normally distributed within each category (here the molecular species), is sensitive to multicollinearity, and is designed to work with matrices including more observations in the category with the lowest sampling size than variables (Press & Wilson, 1978; Pohar *et al.*, 2004; Liong & Foo, 2013). To reduce the number of variables, solve the problem of multicollinearity, and generate a set of continuous variables to avoid working on a large number of categorical characters that do not meet the criterion of normality in the case of the morphological data, we employed the PCA axes, which summarize the information included in the raw matrix, and are orthogonal, as variables in the LDA. The normality of the PCA scores of each species was not rejected according to the Shapiro–Wilk W test. To avoid including species with lower number of specimens than the number of PCA axes included in the analysis (see below), we removed singletons and species with fewer than five specimens from the analyses. The optimum number of PCA axes to be included in the model was selected by stepwise variable selection with a probability level to enter and stay in the model of 0.05. The performance of the model, that is, its ability to assign a specimen to a particular species based on its categorical and quantitative traits, was subsequently assessed by a leave-one-out cross-validation procedure. We finally employed canonical discriminant analysis, which constructs linear combinations of morphological characters that maximize the differentiation among molecularly defined species, to graphically visualize the taxonomic performance of the morphological characters investigated.

All of the analyses were run with SAS 9.1.

Quantification of leaf architecture and interspecific differentiation in *Hymenophyllum* ferns

We tested the application of MORPHOSNAKE to other branched organisms and took the leaf ramification pattern of the fern genus *Hymenophyllum* as an example. We focused on closely related and morphologically similar species of subgenus *Mecodium* from the Comoros archipelago and the western Indian Ocean, including *H. capense* Schrad., *H. inaequale* (Poir.) Desv. and *H. kuhnii* C.Chr. (Saïd *et al.*, 2017), and also included for comparison representatives of *H. ivohibense* Tardieu (subgenus *Sphaerocionum*). Images from 18 herbarium specimens (Table S4) were acquired and treated by GIMP as specified in the Morphological and morphometrical measurements section above. Enlarged indusia from sporangia were erased in order to keep only vegetative branches. MORPHOSNAKE measurements were acquired for two to five leaves per specimen (Table S5).

Results

Software availability

MorphoSnake is available on GitHub (https://github.com/fjansson/MorphoSnake), along with a detailed user manual. MorphoSnake skeletonizes thallus images using the Zhang-Suen skeletonization method (Zhang & Suen, 1984), as implemented in the Scikit image Python library (http://scikit-image.org/). This thinning algorithm removes all the contour points (pixels) except those belonging to the skeleton. In order to preserve end-points and pixel connectivity, the algorithm uses two sub-iterations, which delete specific boundaries points under certain conditions (Zhang & Suen, 1984). From the skeleton, a graph is constructed, with vertices at the branching points and edges between vertices. In the course of skeletonization, spurious branches and loops may be generated. These can be removed manually by deleting nodes or vertices from the graph. Finally, to distinguish between developed and undeveloped axes (or, in other words, at which stage of growth a notch becomes a branch of order 1, an issue that is similar to the presence of dormant buds in angiosperms), branches < 5px are not taken in account.

Measurements are calculated for three sets of branches according to their hierarchical position: all branches; internal branches, not containing a terminal node, preceded by the letter 'I' in the data (e.g. IW, IL); and terminal branches, preceded by the letter 'T' in the data (e.g. TW, TL).

For each thallus, two files containing (respectively) the detailed measurements of each branch ('f_branches.txt') and junctions ('f_nodes.txt') are created. These files include detailed information on the maximum Strahler order, the number of branches by order and the total number of branches. A graphical representation of each thallus analyzed with junctions and skeleton is provided by MorphoSnake and saved automaticaly ('-graph.pkl'). The entire picture, or a focus on a specific area, can be saved in '.png' format (Figs 4, 8). In the graph, an index is assigned to each edge and node according to the Strahler hierarchy. After each MorphoSnake session, a file ('results.txt') is generated, containing the means of the measurements from internal and terminal branches of each thallus.

Infra- and interspecific morphological characterization in the simple thalloid liverwort genus *Riccardia*

The results of the molecular species delimitation analyses that were performed in the thalloid liverwort genus *Riccardia* based on sequence variation at the *ITS*, *matK*, *psbA* and *trnL* (57.8%, 57.8%, 54.2% and 50.3% of variable sites in the ingroup, respectively) before morphometrical characterization of the molecular species are summarized in Fig. 5. With ABGD, between 12 and 19 species were identified depending on the marker used and the *P*-value selected (Fig. 5). With Haplowebs, 17 fields of recombination (Figs S2, 5), considered as distinct species, were considered. Between 14 and 21 species were identified by GMYC depending on the locus investigated (see Fig. 5 for *trnL-F*, and Fig. S3a and b for *psbA-trnH* and *matK*, respectively). According to the conspecificity matrix across loci and species delimitation techniques, 14 *Riccardia* species were retained, including three singletons (Fig. 6), confirming the status of eight species while revealing the existence of at least six undescribed species. The taxonomical consequences of our analyses will be presented elsewhere, but our results are consistent with previous studies in *Plagiochila*, wherein Renner *et al.* (2017) estimated that real diversity in Australasian species is 29% higher than currently recognized and 36% of currently accepted and previously untested

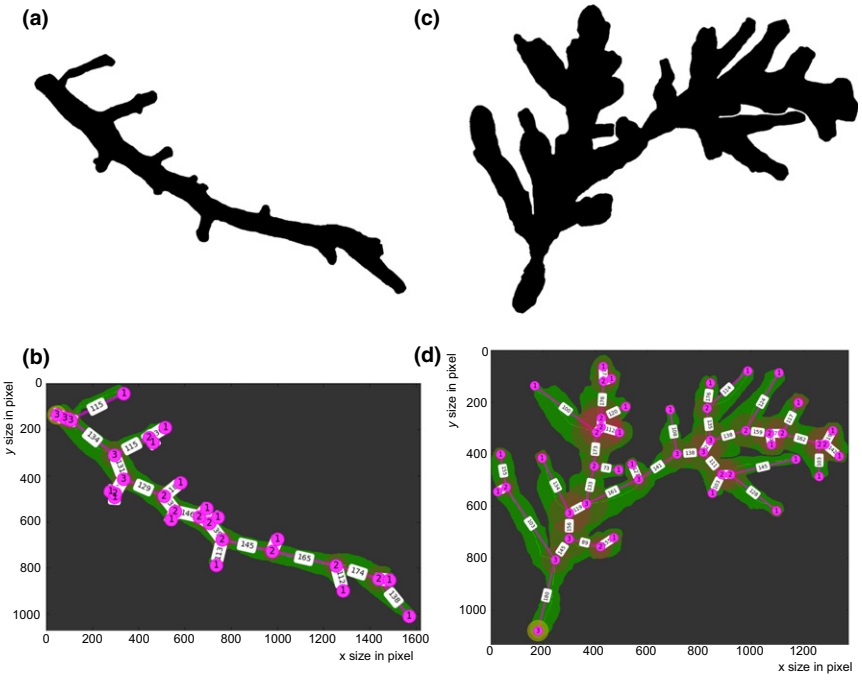

Fig. 4 Example of MorphoSnake input binarized pictures: (a) *Riccardia fastigiata*, (c) *Riccardia saccatiflora*. MorphoSnake output: (b) *R. fastigiata*, (d) *R. saccatiflora*. The numbers in the white squares are the angle measurements between axes along the skeleton. The numbers in the pink disks represent the Strahler ordering of the branch.

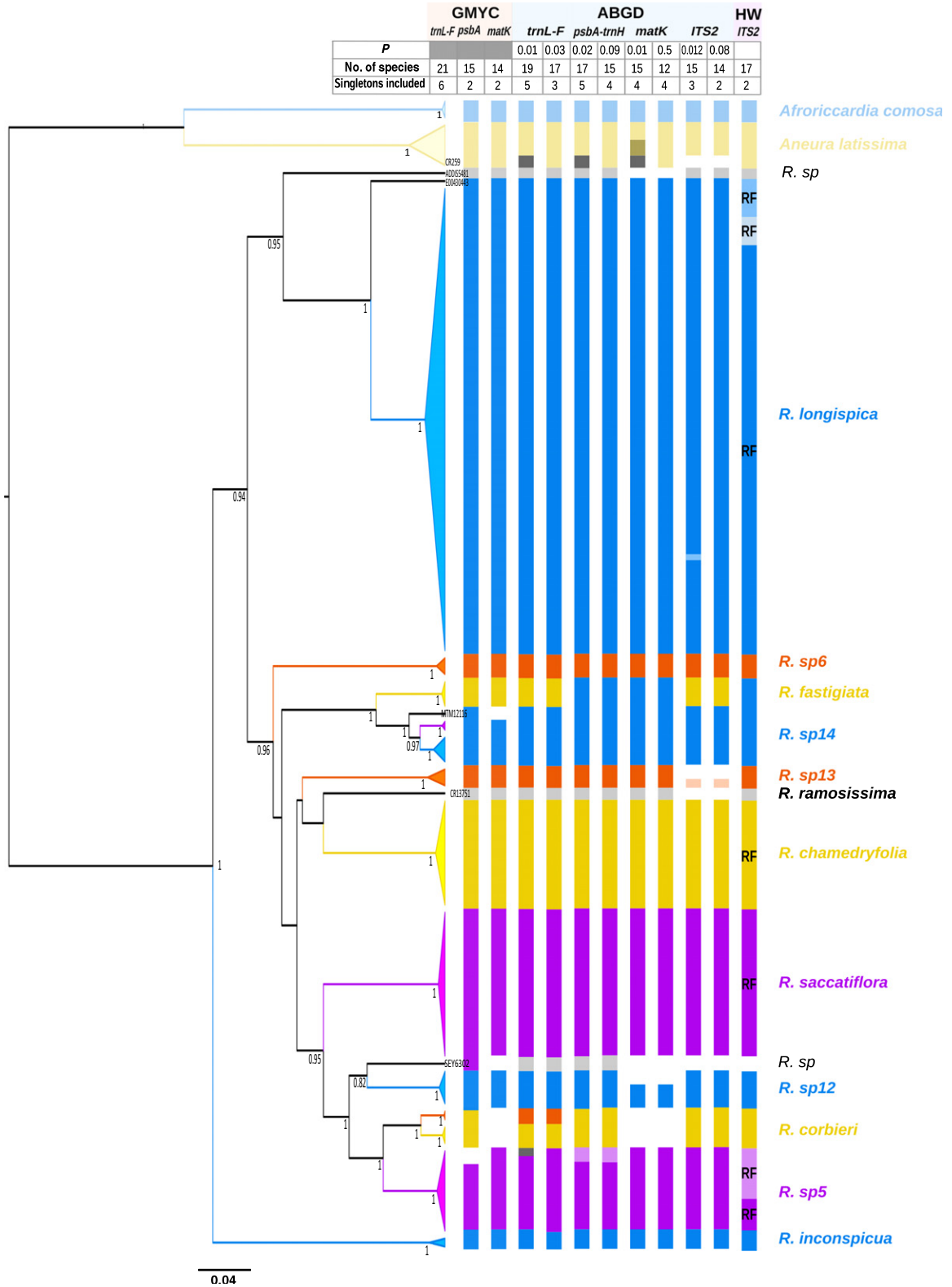

Fig. 5 Comparison of species delimitation analyses (General Mixed Yule Coalescent (GMYC), Automatic BarCode Gap Discovery (ABGD) and Haplowebs) based on sequence variation at *trnL-F*, *matK*, *psbA-trnH* and *ITS2* in African *Riccardia* on the *trnL-F* maximum credibility tree. Numbers below branches are the posterior probabilities > 0.7. Singletons are represented by grey boxes. *P*-values are the maximum intraspecific distance in ABGD. RF, field of recombinations recognized by the haplowebs analysis.

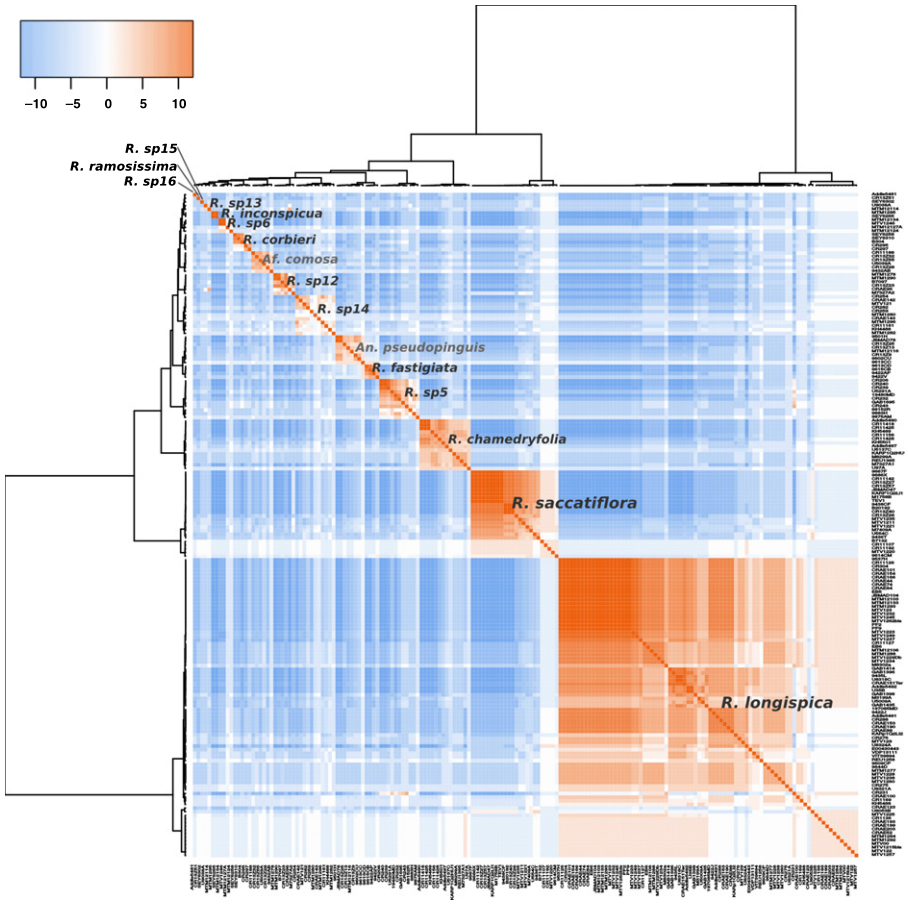

Fig. 6 Conspecificity matrix among species delimitation analyses (General Mixed Yule Coalescent (GMYC), Automatic BarCode Gap Discovery (ABGD) and Haplowebs) and loci (*trnL-F*, *matK*, *psbA-trnH* and *ITS2*) in African *Riccardia*. Colours represent the conspecificity score (number of independent markers and species delimitation methods supporting the hypothesis of the conspecificity between pairs of individuals) and range from orange (all methods and loci consider the pair as conspecific) to blue (this pair of specimens is never assigned to the same species by any locus and any method). The trees above and left from the conspecificity matrix are the results of the clustering of individuals based on their conspecificity score. *Af.*, *Afroriccardia*; *An.*, *Aneura*.

Table 1 *F*- and *P*-values of the ANOVA between the first three principal component analysis (PCA) axes of the categorical (ML) and quantitative (MM) characters and 'substrate type' (a) across species and (b) within *Riccardia longispica*

	%var	*F*	*P*
(a)			
MM1	0.47	1.0	0.37
MM2	0.15	9.30	0.0002
MM3	0.09	3.21	0.043
ML1	0.13	1.13	0.32
ML2	0.12	9.88	< 0.0001
ML3	0.09	6.49	0.0021
(b)			
MM1	0.50	0.02	0.98
MM2	0.16	4.57	0.014
MM3	0.10	0.60	0.55
ML1	0.20	0.53	0.59
ML2	0.18	0.79	0.45
ML3	0.10	0.99	0.37

%var, proportion of variance accounted for by each PCA axis.

species are para- or polyphyletic. Altogether, these results stress the necessity of molecular species delimitations combined with thorough morphological investigation in organisms with reduced morphologies like bryophytes.

Table 2 Performance of the categorical (ML) and quantitative (MM) morphological characters for the characterization of molecularly defined sub-Saharan African *Riccardia* species as measured by the correct classification rate (CCR) of a discriminant analysis after cross-validation

Species	CCR (%)
ML	
All	87.4
1. *R. sp5*	84.6
2. *R. longispica*	88.2
3. *R. fastigiata*	100
4. *R. sp6*	100
5. *R. saccatiflora*	80.0
6. *R. chamedryfolia*	78.6
7. *R. sp14*	87.5
8. *R. corbieri*	100
9. *R. sp12*	100
MM	
All	38.5
1. *R. sp5*	46.1
2. *R. longispica*	35.3
3. *R. fastigiata*	75.0
4. *R. sp6*	60.0
5. *R. saccatiflora*	35.0
6. *R. chamedryfolia*	35.7
7. *R. sp14*	0.0
8. *R. corbieri*	60.0
9. *R. sp12*	66.7

The percentage of variation among species (33.9 ± 17.3% (0.05–0.76) for the categorical and 28.1 ± 11.9% for the quantitative characters; Table S6) did not significantly differ between the categorical and quantitative characters (ANOVA $F = 1.54$, $P = 0.22$). There was a significant difference in the PCA scores depending on the factor 'substrate type' for both the categorical and quantitative characters with the full dataset including all the different *Riccardia* species investigated (Table 1a). When the analysis was conducted among conspecific accessions of *R. longispica*, variation in quantitative, but not in categorical characters, correlated with 'substrate type' (Table 1b).

Concerning the taxonomic partitioning of morphological variation among species, the 10 first PCA axes, explaining 69% of the total variance of the categorical characters, were retained by the stepwise variable selection of the LDA, with an overall correct classification rate of specimens after cross-validation of 87.4% (78.5–100%) across species (Table 2). With the quantitative characters, eight PCA axes (PCA1-6, 8, 10), cumulatively explaining 93% of the total variance, were retained by the stepwise variable selection of the LDA, with an overall correct classification rate of specimens after cross-validation of 38.5% (ranging between 0% for *R. sp14* and 75% for *R. fastigiata*) across species (Table 2). At the genus level, the quantitative characters allowed 99.5%, 71.4% and 100% of the *Riccardia*, *Afroriccardia* and *Aneura* specimens to be reassigned to the correct genus after cross-validation (Table 2). A graphical representation of the performance of the investigated characters to discriminate species and genera is given in Fig. 7(a–c).

Quantification of leaf architecture and interspecific differentiation in *Hymenophyllum*

An example of graphical representation of the skeleton and contours and a hierarchization of the nodes of the fronds of Afro-Malagasy *Hymenophyllum* species is presented in Fig. 8. Ten PCA axes resulting from the analysis of the MorphoSnake characters, accounting for 98% of the total variance, were selected by the stepwise discriminant analysis among species. After cross-validation of the model, 100% of the specimens were correctly assigned to the species they belong to. A graphical representation of the performance of the MorphoSnake characters to distinguish *Hymenophyllum* species is given in Fig. 7(d).

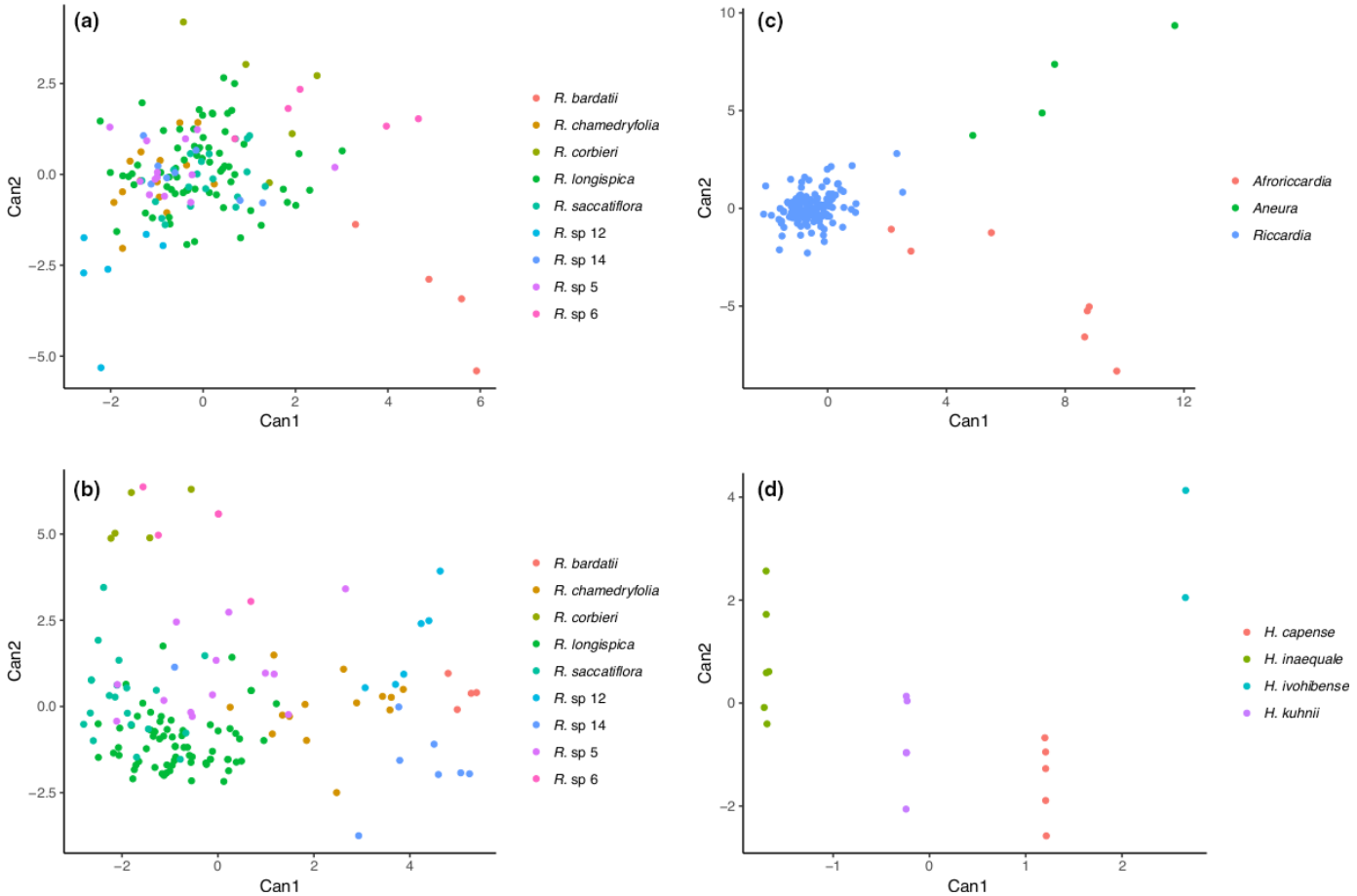

Fig. 7 Graphical representation of the performance of morphological characters to discriminate molecularly defined *Riccardia* species along the first two axes of a canonical discriminant analysis of morphological variation. (a) MorphoSnake characters among *Riccardia* species. (b) Traditional morphological characters among *Riccardia* species. (c) MorphoSnake characters among sibling genera (*Riccardia*, *Afroriccardia* and *Aneura*). (d) MorphoSnake characters among African *Hymenophyllum* species.

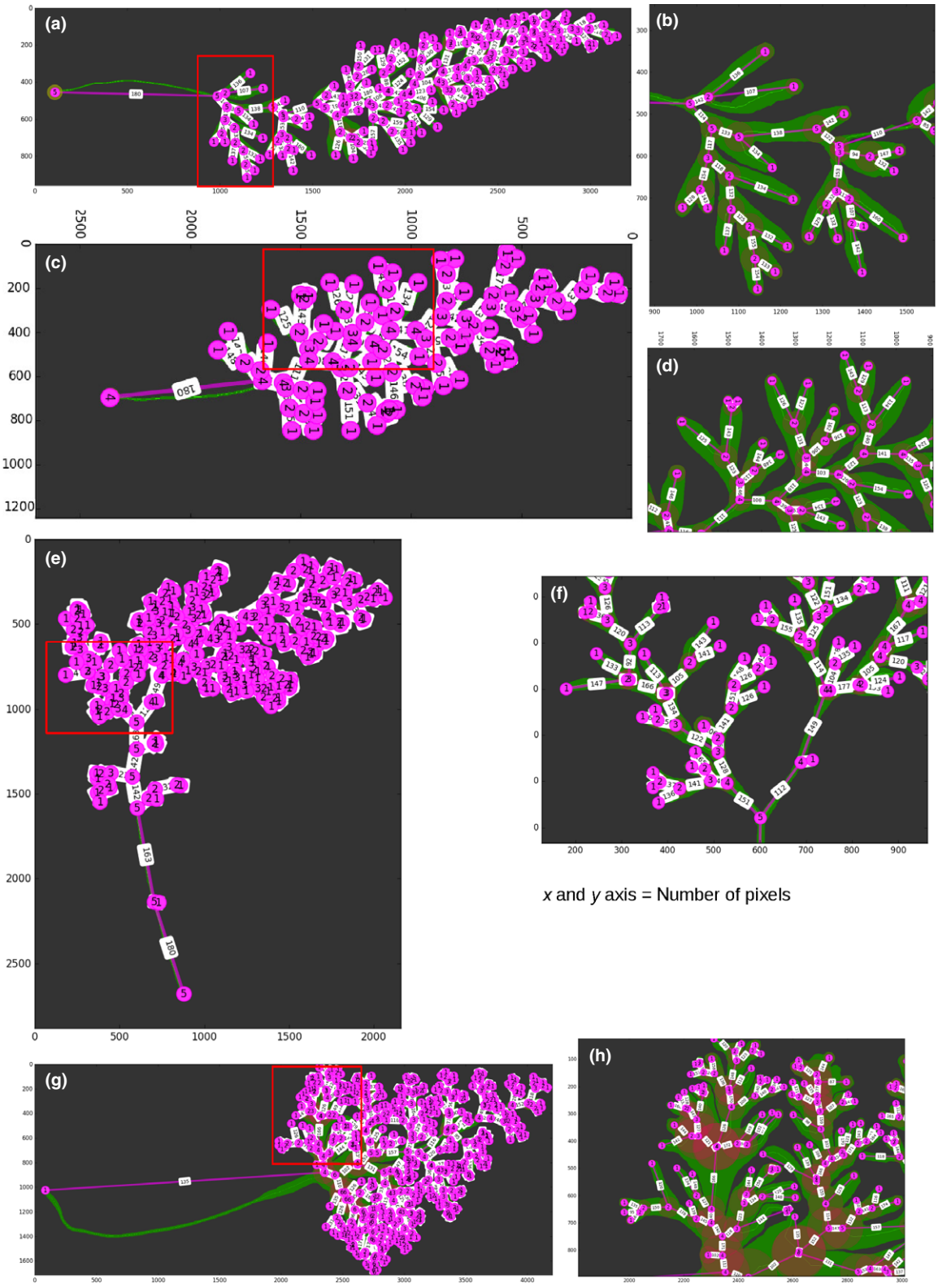

Fig. 8 MorphoSnake graphical representations of the contour and skeleton and Strahler hierarchization of the nodes in African *Hymenophyllum* (Polypodiopsida) species. The numbers in the white squares are the angle measurements between axes along the skeleton. The numbers in the pink disks represent the Strahler ordering of the branch. (b), (d), (f) and (h) correspond to focused pictures of the area within the red rectangles in (a), (c), (e) and (g), respectively. (a, b) *Hymenophyllum kuhnii* (P00312328). (c, d) *Hymenophyllum capense* (P00085205). (e, f) *Hymenophyllum inaequale* (P01305987). (g, h) *Hymenophyllum ivohibense* (P01320816).

Discussion

We developed a quantitative and reproducible approach to produce exhaustive quantitative measurements of plant architecture through the implementation of the new software MorphoSnake. Although the method was developed primarily for thalloid plants, we showed, with two examples of application in thalloid liverworts and ferns, that our approach has a broad range of application for shape quantification.

In African *Riccardia*, quantitative characters exhibited significant differences among species. However, traditional categorical morphological characters outperformed quantitative ones for the morphological characterization of molecularly defined species. *Riccardia* species are indeed assumed to exhibit high levels of plasticity (Preußing *et al.*, 2010) that would, at first sight, affect more quantitative traits of thallus shape and length measurements than categorical traits, rendering the latter more suitable for taxonomy. As Cope *et al.* (2012) indeed noticed in their analyses of leaf morphology, size is expected to be largely determined by the environment, whereas shape is more heritable. Yet, the percentage of infraspecific variance did not significantly differ between categorical and quantitative characters. We further showed that quantitative characters may be useful at higher taxonomic levels. Although MorphoSnake analyses exhibited a poor ability to discriminate *Riccardia* from *Afroriccardia*, a recently segregated genus characterized by sexual characters that were not taken into account in the present analyses, quantitative characters allowed distinguishing *Riccardia* from the sibling genus *Aneura*. MorphoSnake may further be applied to any branched organism, and we show in the fern genus *Hymenophyllum* that leaf shape quantification makes it readily possible to identify closely related species with an optimal performance. We therefore suggest that MorphoSnake may be used for taxonomic applications across a wide range of ramified organisms such as lichens, algae and ferns, and play a particularly crucial role in finding new taxonomic characters in organisms with reduced morphologies.

The procedure requires specimens to be prepared, which can be quite time-consuming in thalloid liverworts that need to be moistened and cleaned-up. The time required to prepare specimens and obtain suitable pictures may vary from one taxon to another, but our applications of MorphoSnake to ferns suggest that herbarium specimens can readily be used without any specific preparation. The availability of suitable images also is expected to increase in the context of the ongoing effort to digitize botanical collections (Unger *et al.*, 2016). The standardized protocol described here to acquire large numbers of quantitative traits in branched organisms from digitized pictures may further contribute to the development of rapid identification tools in the context of an increasing interest for automating the process of species characterization, and ultimately, identification (Cope *et al.*, 2012). To reach this goal, the MorphoSnake procedure will need to be fully automated to avoid any time-consuming manipulations that may further affect angle measurements. Furthermore, although actual measurements for randomly selected segments empirically confirmed MorphoSnake outputs, a thorough validation using artificial images generated by modelling tools would be desirable (Lobet, 2017).

Future improvements of MorphoSnake could include an analysis of morphological variation in three dimensions, wherein photographs taken at regular intervals along a *z*-axis would be stacked. The analysis presented here would be repeated at each level to take the 3D variation in branching patterns into account while readily solving the problem of overlapping branches.

Most importantly, although there was a significant partitioning of the variation in categorical characters depending on substrate type across species, pointing to substrate preferences among species, quantitative characters were the only ones to show significant correlations between infraspecific variation and substrate type. They may therefore reveal an important portion of infraspecific morphological variation linked to local habitat conditions to which traditional categorical characters do not respond. In fact, quantitative character analyses have already been used successfully to characterize patterns of intraspecific variation in correlation with ecological factors in marine organisms (Kaandorp *et al.*, 2003; Matsumoto, 2004). In this context, other potential uses of MorphoSnake include the morphological characterization of mutants in *Marchantia polymorpha*, which is used in molecular genetics as a model to study the evolution and diversity of regulatory systems in land plants (Ishizaki *et al.*, 2016), and of the development of branching patterns in mosses (Coudert *et al.*, 2017).

Acknowledgements

Many thanks are due to Robbert Gradstein for his comments on the manuscript, to Tim Vaughan for assistance with the GMYC analyses, to Valentin Faivre for acquiring images of *Riccardia*, and to Gwenaelle Saulnier for her assistance with the acquisition of the fern data. J.P. was funded by the Spanish 'Ministerio de Economía y Competitividad' through the Juan de la Cierva Program – Incorporation (IJCI-2014-19691) and Marie Sklodowska-Curie COFUND, Researchers' Night and Individual Fellowships Global (MSCA grant no. 747238, 'UNISLAND' Molecular data were acquired with the support of the Genoscope thanks to the BDV program 'Bibliothèque du vivant' (CNRS, MNHN, INRA and CEA) and LabEx ANR-10-LABX-0003-BCDIV with the support of ANR ANR-11-IDEX-0004-02 from the program 'Investissements d'avenir'. Sequences were acquired at the 'Service de Systématique Moléculaire' of the MNHN (UMS 2700 OMSI) and we thank Céline Bonillo and Josie Lamboudière for their kind help.

Author contributions

J.K. developed the original conceptual framework of the study; F.Jansson developed MorphoSnake; R.C. and J-Y.D. prepared the framework of the morphometric analysis; C.R. acquired the measurements with MorphoSnake and produced the molecular data; C.R. and J.P. performed the species delimitation analyses with the help of J-F.F.; N.P. and A.V. performed the statistical analyses; F. Jabbour, Y.C. and R.C. brought their expertise in

morphometry; and C.R. and A.V. wrote the paper with the assistance of F. Jabbour, Y.C. and J.K.

References

Biot E, Cortizo M, Burguet J, Kiss A, Oughou M, Maugarny-Calès A, Gonçalves B, Adroher B, Andrey P, Boudaoud A *et al.* 2016. Multiscale quantification of morphodynamics: MorphoLeaf software for 2D shape analysis. *Development* **143**: 3417–3428.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10**: e1003537.

CBOL Plant Working Group. 2009. A DNA barcode for plants. *Proceedings of the National Academy of Sciences, USA* **106**: 12794–12797.

Carstens B, Pelletier T, Reid N, Satler J. 2013. How to fail in species delimitation. *Molecular Ecology* **22**: 4369–4383.

Castro-Romero R, Montes MM, Martorelli SR, Sepulveda D, Tapia S, Martinez-Aquino A. 2016. Integrative taxonomy of *Peniculus, Metapeniculus*, and *Trifur* (Siphonostomatoida: Pennellidae), copepod parasites of marine fishes from Chile: species delimitation analyses using DNA barcoding and morphological evidence. *Systematics and Biodiversity* **14**: 466–483.

Chatrand G. 1985. *Introductory graph theory. Dover books on mathematics*. New York, NY, USA: Dover Publications.

Cope JS, Corney D, Clark JY, Remagnino P, Wilkin P. 2012. Plant species identification using digital morphometrics: a review. *Expert Systems with Applications* **39**: 7562–7573.

Corney D, Clark JY, Tang HL, Wilkin P. 2012. Automatic extraction of leaf characters from herbarium specimens. *Taxon* **61**: 231–244.

Coudert Y, Bell NE, Edelin C, Harrison CJ. 2017. Multiple innovations underpinned branching form diversification in mosses. *New Phytologist* **215**: 840–850.

Debortoli N, Li X, Eyres I, Fontaneto D, Hespeels B, Tang C, Flot JF, Van Doninck K. 2016. Genetic exchange among bdelloid rotifers is more likely due to horizontal gene transfer than to meiotic sex. *Current Biology* **26**: 723–732.

Dellicour S, Flot JF. 2015. Delimiting species-poor data sets using single molecular markers: a study of barcode gaps, haplowebs and GMYC. *Systematic Biology* **64**: 900–908.

Diestel R. 2010. *Graph theory. Graduate texts in mathematics collection, 4th edn.* New York, NY, USA: Springer.

Doyle J. 1995. The irrelevance of allele tree topologies for species delimitation and a non-topological alternative. *Systematic Botany* **20**: 574–588.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**: 214.

Esteve-Altava B. 2017. In search of morphological modules: a systematic review. *Biological Reviews* **92**: 1332–1347.

Ezard T, Fujisawa T, Barraclough T. 2009. *splits: SPecies' LImits by Threshold Statistics. R package version 1.0-11/r29.* [WWW document] URL http://r-forge.r-project.org/projects/splits/

Ferreira T. 2016. *hIPNAT: hIPNAT 1.0.1*. doi 10.5281/zenodo.49399.

Flot JF. 2007. CHAMPURU 1.0: a computer software for unraveling mixtures of two DNA sequences of unequal lengths. *Molecular Ecology Notes* **7**: 974–977.

Flot JF. 2010. SeqPHASE: a web tool for interconverting PHASE input/output files and FASTA sequence alignments. *Molecular Ecology Resources* **10**: 162–166.

Flot JF. 2015. Species delimitation's coming of age. *Systematic Biology* **64**: 897–899.

Flot JF, Couloux A, Tillier S. 2010. Haplowebs as a graphical tool for delimiting species: a revival of Doyle's "Field For Recombination" approach and its application to the coral genus *Pocillopora* in Clipperton. *BMC Evolutionary Biology* **10**: 372–384.

Flot JF, Tillier A, Samadi S, Tillier S. 2006. Phase determination from direct sequencing of length-variable DNA regions. *Molecular Ecology Notes* **6**: 627–630.

Fontaneto D, Flot JF, Cuong Q. 2015. Guidelines for DNA taxonomy, with a focus on the meiofauna. *Marine Biodiversity* **45**: 433–451.

Friedland G, Jantz K, Lentz T, Rojas R. 2006. *Extending the SIOX algorithm: alternative clustering methods, sub-pixel accurate object extraction from still images, and generic video segmentation*. Berlin, Germany: Free University of Berlin, Department of Computer Science.

Fujisawa T, Barraclough TG. 2013. Delimiting species using single-locus data and the generalized mixed yule coalescent approach: a revised method and evaluation on simulated data sets. *Systematic Biology* **62**: 707–724.

Fujita M, Leaché A, Burbrink F, McGuire J, Moritz C. 2012. Coalescent-based species delimitation in an integrative taxonomy. *Trends in Ecology and Evolution* **27**: 480–488.

Harper JB, Rosen B, White J. 1986. The growth and form of modular organisms. *Philosophical Transactions of the Royal Society B* **313**: 1–5.

Horton R. 1945. Erosional development of streams and their drainage basins, hydrophysical approach to quantitative morphology. *Geological Society of America Bulletin* **53**: 275–370.

Ishizaki K, Nishihama R, Yamato KT, Kohchi T. 2016. Molecular genetic tools and techniques for *Marchantia polymorpha* research. *Plant Cell Physiology* **57**: 262–270.

Johnston EC, Forsman ZH, Flot J-F, Schmidt-Roach S, Pinzón JH, Knapp ISS, Toonen RJ. 2017. A genomic glance through the fog of plasticity and diversification in *Pocillopora*. *Scientific Reports* **7**: 5991.

Kaandorp JA. 1999. Morphological analysis of growth forms of branching marine sessile organisms along environmental gradients. *Marine Biology* **134**: 295–306.

Kaandorp JA, Koopman EA, Sloot PMA, Bak RPM, Vermeij MJ, Lampmann LEH. 2003. Simulation and analysis of flow patterns around the scleractinian coral *Madracis mirabilis* (Duchassaing and Michelotti). *Philosophical Transactions of the Royal Society B* **358**: 1551–1157.

Kaandorp JA, Kübler J. 2001. *The algorithmic beauty of seaweeds, sponges and corals*. New York, NY, USA: Springer.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C *et al.* 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.

Konglerd P, Reeb C, Jansson F, Kaandorp JA. 2017. Quantitative morphological analysis of 2D images of complex-shaped branching biological growth forms: the example of branching thalli of liverworts. *BMC Research Notes* **10**: 103.

Kruszyński KJ, Kaandorp JA, van Liere RA. 2007. Computational method for quantifying morphological variation in scleractinian corals. *Coral Reefs* **26**: 831–840.

Liong CY, Foo SF. 2013. Comparison of linear discriminant analysis and logistic regression for data classification. *Proceedings of the 20th National Symposium on Mathematical Sciences* **1522**: 1159–1165.

Lobet G. 2017. Image analysis in plant sciences: publish then perish. *Trends in Plant Science* **22**: 559–566.

Lobet G, Pagès L, Draye X. 2011. A novel image analysis toolbox enabling quantitative analysis of root system architecture. *Plant Physiology* **157**: 29–39.

Lobet G, Pound MP, Diener J, Pradal C, Draye X, Godin C, Javaux M, Leitner D, Meunier F, Nacry P *et al.* 2015. Root System Markup Language: toward a unified root architecture description language. *Plant Physiology* **167**: 617–627.

Matsumoto A. 2004. Heterogeneous and compensatory growth in *Melithaea flabellifera* (Octocorallia: Melithaeidae) in Japan. *Hydrobiologia* **530**: 389–397.

Mishler B, De Luna E. 1991. The use of ontogenetic data in phylogenetic analysis. *Advances in Bryology* **4**: 123–167.

Monaghan M, Wild R, Elliot M, Fujisawa T, Balke M, Inward D, Lees D, Ranaivosolo R, Eggleton P, Barraclough T *et al.* 2009. Accelerated species inventory on Madagascar using coalescent-based models of species delineation. *Systematic Biology* **58**: 298–311.

Montagna M, Mereghetti V, Lencioni V, Rossaro B. 2016. Integrated taxonomy and DNA barcoding of Alpine midges (Diptera: Chironomidae). *PLoS ONE* **11**: e0149673.

Oliver P, Lee M. 2010. The botanical and zoological code impede biodiversity research by discouiraging publicatIon of unamed species. *Taxon* **59**: 1201–1205.

Pante E, Schoelinck C, Puillandre N. 2015. From integrative taxonomy to species description: one step beyond. *Systematic Biology* **64**: 152–160.

Pohar M, Blas M, Turk S. 2004. Comparison of logistic regression and linear discriminant analysis: a simulation study. *Metodološki zvezki* **1**: 143–161.

Press SJ, Wilson S. 1978. Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association* **73**: 699–705.

Preußing M, Olsson S, Schäfer-Verkimp A, Wickett N, Wicke D, Nebel M. 2010. New insights of the evolution of the liverwort family Aneuraceae (Metzgeriales, Marchantiophyta), with emphasis on the genus *Lobatiriccardia*. *Taxon* **59**: 1424–1240.

Puillandre N, Baylac M, Boisselier-Dubayle MC, Cruaud C, Samadi S. 2009. An integrative approach to species delimitation in *Benthomangelia* (Mollusca: Conoidea). *Biological Journal of the Linnean Society* **96**: 696–708.

Puillandre N, Lambert A, Brouillet S, Achaz G. 2012. ABGD Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology* **21**: 1864–1877.

Rabeau L, Gradstein R, Dubuisson JY, Nebel M, Quandt D, Reeb C. 2017. New insights into the phylogeny and relationships among the worldwide genus *Riccardia* (Aneuraceae, Marchantiophytina). *European Journal of Taxonomy* **273**: 1–6.

Raftery A, Newton M, Satagopan J, Krivitsky P. 2007. Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. In: Bernardo JM, Bayarri MJ, Berger JO, eds. *Bayesian statistics*. New York, NY, USA: Oxford University Press, 1–45.

Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014. *Tracer v1.6*. [WWW document] URL http://tree.bio.ed.ac.uk/software/tracer/

Reeb C, Bardat J. 2014. Studies on African types and related materials. *Cryptogamie Bryologie* **35**: 47–75.

Renner MAM, Heslewood MM, Patzak SDF, Schäfer-Verwimp A, Heinrichs J. 2017. By how much do we underestimate species diversity of liverworts using morphological evidence? An example from Australasian *Plagiochila* (Plagiochilaceae: Jungermanniopsida). *Molecular Phylogenetics and Evolution* **107**: 576–593.

Saïd AH, Hennequin S, Rouhan G, Dubuisson J-Y. 2017. Disentangling the diversity and taxonomy of Hymenophyllaceae (Hymenophyllales, polypodiidae) in the Comoros. *European Journal of Taxonomy* **313**: 1–53.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B *et al.* 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**: 676–682.

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**: 671–675.

Song S, Zhao J, Li C. 2017. Species delimitation and phylogenetic reconstruction of the Sinipercids (Perciformes: Sinipercidae) based on target enrichment of thousands of nuclear coding sequences. *Molecular Phylogenetics and Evolution* **111**: 44–55.

Stanton D, Reeb C. 2016. Morphogeometric approaches to non-vascular plants. *Frontiers in Plant Sciences* **7**: 916.

Strahler A. 1952. Hypsometria (area-altitude) analysis of erosional topography. *Bulletin of the Geological Society of America* **63**: 1167–1177.

Talavera G, Dincă V, Vila R. 2013. Factors affecting species delimitations with the GMYC model: insights from a butterfly survey. *Methods in Ecology and Evolution* **4**: 1101–1110.

Tomlinson P. 1987. Branching is a process, not a concept. *Taxon* **36**: 54–57.

Unger J, Merhof D, Renner S. 2016. Computer vision applied to herbarium specimens of German trees: testing the future utility of the millions of herbarium specimen images for automated identification. *BMC Evolutionary Biology* **16**: 248.

Zanardo S, Zaliapin I, Foulafoula-Giorgio E. 2013. Are American rivers Tokunaga self-similar? New results on fluvial network topology and its climatic dependence. *Journal of Geophysical Research* **118**: 1–18.

Zhang J, Kapli P, Pavlidis P, Stamatakis A. 2013. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* **29**: 2869–2876.

Zhang TY, Suen CY. 1984. A fast parallel algorithm for thinning digital patterns. *Communication of the ACM* **27**: 236–239.

Zhao S, Guo Y, Sheng Q, Shyr Y. 2014. Heatmap3: an improved heatmap package with more powerful and convenient features. *BMC Bioinformatics* **15**: P16.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 Branching patterns and cross-sections of the main-axis in mid-part in *Riccardia corbieri*, *R. saccatiflora* and *R. multifida*, respectively.

Fig. S2 Species delimitation analysis in sub-Saharan African Riccardia species based on the Haploweb analysis of the ITS2 region.

Fig. S3 Chronogram of the maximum clade credibility tree from the BEAST analysis in sub-Saharan African *Riccardia* species.

Table S1 Voucher information and GenBank accession numbers for African *Riccardia* specimens used for molecular species delimitation and MORPHOSNAKE analysis

Table S2 Primer and PCR conditions used to amplify and sequence the *matK*, *trnL-F*, *psbA-trnH* and *ITS2* regions in sub-Saharan African *Riccardia* specimens listed in Table S1

Table S3 Matrix of unordered (C) and ordered (N) categorical characters, and of quantitative morphological characters (average from measurements made on five to 15 specimens per collection) scored with MORPHOSNAKE on sub-Saharan African *Riccardia*, *Afroriccardia* and *Aneura* specimens (see Table S1 for voucher information)

Table S4 Voucher information for African *Hymenophyllum* specimens used for MORPHOSNAKE analysis

Table S5 Matrix of quantitative morphological characters (average from measurements made on two to six leaves per specimen) scored with MORPHOSNAKE on sub-Saharan African *Hymenophyllum* specimens (see Table S4 for voucher information)

Table S6 F statistics and associated *P*-values and r^2 of the ANOVA for unordered (C) and ordered (N) characters (see Table S3a for character labels) and quantitative morphological characters (see Table 3b for character labels) among molecularly defined *Riccardia* species

Methods S1 How to get skeletonized images with IMAGEJ.