SCIENTIFIC REPORTS

OPEN

Leaf epidermis images for robust identification of plants

Núbia Rosa da Silva[1,2], Marcos William da Silva Oliveira[1,2], Humberto Antunes de Almeida Filho[2], Luiz Felipe Souza Pinheiro[3], Davi Rodrigo Rossatto[4], Rosana Marta Kolb[3] & Odemir Martinez Bruno[1,2]

Received: 30 September 2015
Accepted: 19 April 2016
Published: 24 May 2016

This paper proposes a methodology for plant analysis and identification based on extracting texture features from microscopic images of leaf epidermis. All the experiments were carried out using 32 plant species with 309 epidermal samples captured by an optical microscope coupled to a digital camera. The results of the computational methods using texture features were compared to the conventional approach, where quantitative measurements of stomatal traits (density, length and width) were manually obtained. Epidermis image classification using texture has achieved a success rate of over 96%, while success rate was around 60% for quantitative measurements taken manually. Furthermore, we verified the robustness of our method accounting for natural phenotypic plasticity of stomata, analysing samples from the same species grown in different environments. Texture methods were robust even when considering phenotypic plasticity of stomatal traits with a decrease of 20% in the success rate, as quantitative measurements proved to be fully sensitive with a decrease of 77%. Results from the comparison between the computational approach and the conventional quantitative measurements lead us to discover how computational systems are advantageous and promising in terms of solving problems related to Botany, such as species identification.

Green plants (**Viridiplantae**) are among the most important living beings in the natural world. They are multicellular photosynthetic eukaryotic organisms forming a clade that includes flowering plants, conifers and other gymnosperms, ferns, clubmosses, hornworts, liverworts, mosses and green algae[1], providing most of the world's photosynthetically fixed carbon[2], and are the basis of all life on earth. According to O.W. Archibold[3], 11 major types can be recognized forming the majority of earth ecosystems: tropical forests, tropical savannas, arid regions (deserts), Mediterranean ecosystems, temperate forest ecosystems, temperate grasslands, coniferous forests, tundra (both polar and high mountain), terrestrial wetlands, freshwater ecosystems and coastal/marine systems. In South America, forests and savannas predominate[4], and in the Brazilian territory, the Cerrado is included as one of the most important tropical savannas[5]. Its geographical reach accounts for 22% of the country's land area (extending marginally into Paraguay and Bolivia), and is the second most important ecosystem in South America, after the Amazon Forest. The World Wide Fund for Nature called it the biologically richest savanna in the world, with about 12,000 plant species, 44% of which are endemic, according to[6,7]. Additionally, the Cerrado has been considered a hot-spot, suffering from severe deforestation and agricultural use[8].

Due to its incredible plant diversity, which is reflected in a great diversity of morphological structures[9,10], the Cerrado ecosystem presents itself as a great opportunity to test methods and approaches to provide new tools for plant identification. To date, the main approach used to identify any plant species is by using morphological traits[11]. This approach is widely based on morphological traits of reproductive organs that are not always found in the plant, such as flowers and fruits[12]. Alternatively, in recent years, some computational approaches have been proposed to identify species based on images of leaves that are often available throughout the year in tropical and subtropical regions of the world[13–18]. Such methods are able to differentiate the species based on leaf image properties, where texture is the main analyzed feature.

[1]Institute of Mathematics and Computer Science, University of São Paulo, USP, Avenida Trabalhador são-carlense, 400, 13566-590 São Carlos, São Paulo, Brazil. [2]Scientific Computing Group, São Carlos Institute of Physics, University of São Paulo, PO Box 369, 13560-970, São Carlos, SP, Brazil. [3]Department of Biological Sciences, Faculty of Sciences and Languages, Univ Estadual Paulista, UNESP. Av. Dom Antônio, 2100, 19806-900, Assis, São Paulo, Brazil. [4]Department of Applied Biology, Faculty of Agriculture and Veterinary Sciences, Univ Estadual Paulista, UNESP, Via de Acesso Prof. Paulo Donatto Castellane S/N. 14884-900, Jaboticabal, São Paulo, Brazil. Correspondence and requests for materials should be addressed to O.M.B. (email: bruno@ifsc.usp.br)

Generally, texture is associated to the feel of different materials to human touch. Texture image analysis is based on visual interpretation of this feeling[19]. By this fact, this descriptor indicates smoothness, coarseness and regularity in images[20]. In computational analysis of plant images, assessing texture of leaf surface is related to different characteristics of the plant, e.g. presence and type of trichomas, stomata types, etc., producing different patterns that can be identified. The application of such methods has been used in leaf cross-sections (analyzing internal structures) or on the leaf surface (where subsamples of the entire scanned leaf were analyzed)[13–18,21].

Another possibility to be explored in computational methods is analysing the leaf epidermal surface by its dissociation[22]. The dissociation process is normally used to infer structural patterns such as size, position and density of stomata, as well the distribution and shape of epidermal cells and characteristics of the cuticle, such as striation patterns[22,23]. These traits provide important information for plant identification and can even provide important taxonomic characteristics for phylogeny assembly[24–27]. When analysing the epidermis surface, the anatomical procedures are relatively simple compared to the preparation of transversal cuts[28], as the process to obtain the leaf surface is done in less than 12 hours, and many samples can be processed at the same time[29]. Furthermore, the full leaf is not needed to identify the plant. Nevertheless, few studies have assessed the morphometry, anatomical structure and texture of leaf epidermis images with identification purposes. In[30], dissociated samples of leaf epidermis were used, however only one species was analyzed.

The aim of this paper is to present an innovative use of microscopic images of leaf epidermis: by extracting texture features and classifying the species by analysing these characteristics. Computational methods are advantageous because they can provide many useful features to identify species[31]. The feature extraction process was carried out using three different methods: Fourier descriptors, corrosion-inspired texture analysis and local binary patterns and the classification was done using two classifiers: *k*-Nearest Neighbors and linear discriminant analysis.

In order to show the importance of using computational methods to help identify plant species, the same species used in the previous experiments were classified using morphological characteristics. They were obtained manually by quantitative measurements using a light photomicroscope and the AxioVision microscope software from Zeiss. The analysed traits were the stomata density, the guard cell length and the stomatal complex width. A third set of experiments was conducted in order to verify if the texture based approaches and the conventional method using quantitative measurements are invariant to plasticity. Phenotypic plasticity is the morphological and physiological response of individuals to changes in the environment[32]. Factors such as climate change, land use change, invasiveness and resource limitation may cause morphological intraspecies differences, playing an important role when identifying species. To verify the robustness of plant identification methods to handle the variability of plasticity, we also analyzed the texture features of leaves from the same species growing in distinct environmental conditions.

Materials and Methods

Epidermal images acquirement system. Leaves were collected from representative woody species of the Cerrado (Table 1) in the IBGE Ecological Reserve ($15°56'42.04''S$, $47°52'43.74''W$) and in the Águas Emendadas Ecological Reserve ($15°33'56.84''S$, $47°36'04.22''W$), both in Brasília, Federal District, Brazil. Authorities from both ecological reserves issued permits to conduct scientific research on the samples. The abaxial surfaces of the epidermis were obtained from three or four leaves (one per individual) of each species. From each collected leaf, a sample of approximately 1 cm^2 was removed from the leaf's middle region, between the margin and midrib. Dissociation of the leaf epidermis was performed using a 1:1 solution of glacial acetic acid and hydrogen peroxide at 60 °C for 12 hours, or the time required to completely decouple the epidermis (modified from[29]). After this procedure, the abaxial surface of the epidermis was washed in distilled water, stained with safranin and mounted in glycerin. The images were captured at 20× objective lenses, using a *Zeiss Axio Scope A1* optical microscope, coupled to a digital camera (Zeiss AxioCam MRc model). Three images per leaf were used for the classification procedure. For the experiments using manual measurements, the stomata density, as well the guard cell length and the stomatal complex width, were obtained by using the AxioVision software from Zeiss. Stomatal counts were made in three fields (defined at random) per individual sampled at 10× objective lens. The size of the stomata was measured in at least three different fields, randomly set, totaling 90 stomata. In total, 309 microscopic images of epidermal leaf surfaces were obtained from 32 woody plants that are commonly found on our studied vegetation sites (Table 1).

Feature extraction for classification. Image texture can be defined as a function of the spatial variation in pixel intensities (gray level values). These textural properties can be extracted from the image and mapped on the form of a feature vector that will represent this image in a process called feature extraction. Afterwards, the texture patterns of each image need to be recognized to identify which texture pattern they belong to, which leads to a problem of texture classification. For this purpose, the texture patterns of all images under analysis are submitted to a classifier, which predicts the classes in each image considering T_1 images for training and T_2 images for testing.

Previous to feature extraction, some procedures can be used to enhance the image. For example, external interference could lead to obtaining dark, noisy or bright images, preventing capturing the original characteristics. Therefore, these procedures minimize the changes caused by these external interferences. The epidermis images obtained for this experiment underwent a staining procedure, which led to obtaining images with the original colour change of the leaves. For this reason, we converted the images to grayscale and then adopted a procedure for contrast enhancement to ensure there was no interference of the color samples in the feature extraction process. This procedure is described later on.

Three types of feature descriptors were used to describe each epidermis image, which are Fourier[33–35] with two approaches for feature representation: circular angular and circular, Corrosion-Inspired Texture Analysis[36]

Species (Family)
Baccharis linearifolia (Lam.) Pers. (Asteraceae)
Byrsonima laxiflora Griseb. (Malpighiaceae)
Calophyllum brasiliense Cambess. (Clusiaceae)
Campomanesia velutina (Cambess.) O. Berg (Myrtaceae)
Clusia criuva Cambess. (Clusiaceae)
Copaifera langsdorffii Desf. (Fabaceae)
Cupania vernalis Cambess. (Sapindaceae)
Duguetia furfuracea (A.St.-Hill.) Saff. (Annonaceae)
Eriotheca candolleana (K. Schum.) A. Robyns (Malvaceae)
Esenbeckia pumila Pohl. (Rutaceae)
Gaylussacia brasiliensis (Spreng.) Meisn. (Ericaceae)
Guapira noxia (Netto) Lundell (Nyctaginaceae)
Hymenaea stigonocarpa Mart. ex Hayne (Fabaceae)
Ilex affinis Gardner (Aquifoliaceae)
Matayba guianensis Aubl. (Sapindaceae)
Maytenus floribunda Reissek (Celastraceae)
Miconia cuspidata Naudin (Melastomataceae)
Miconia chamissois Naudin (Melastomataceae)
Myrsine guianensis (Aubl.) Kuntze (Myrsinaceae)
Myrsine ferruginea (Ruiz & Pav.) Spreng. (Myrsinaceae)
Ouratea hexasperma (A.St.-Hill.) Baill. (Ochnaceae)
Plenckia populnea Reissek (Celastraceae)
Pseudobombax longiflorum (Mart. & Zucc.) A.Robyns (Malvaceae)
Roupala montana Aubl. (Proteaceae)
Rourea induta Planch. (Connaraceae)
Salacia crassifolia (Mart. ex Schult.) G. Don (Celastraceae)
Symplocos mosenii Brand (Symplocaceae)
Symplocos nitens (Pohl) Benth. (Symplocaceae)
Handroanthus impetiginosus (Mart. ex DC.) Mattos (Bignoniaceae)
Tapirira guianensis Aubl. (Anacardiaceae)
Virola sebifera Aubl. (Myristicaceae)
Xylopia sericea A.St.-Hil. (Annonaceae)

Table 1. List of species from which we obtained the leaf epidermis images.

and Local Binary Pattern[37]. All these methods are described immediately after the description of the enhancement procedure. We carried out the classification using *k*-Nearest Neighbors and linear discriminant analysis classifiers.

Image Enhancement. In this procedure, histogram stretching was used to increase the image contrast. Let an image $f(x, y)$, the enhanced image $g(x, y)$ is obtained by

$$g(x, y) = \frac{f(x, y) - \min(f(x, y))}{\max(f(x, y)) - \min(f(x, y))} 2^{bpp}, \quad (1)$$

where bpp is the number of bits per pixel of the image $f(x, y)$. In our application, the image has 256 gray levels and $bpp = 8$. In the enhanced image $g(x, y)$, 1% of data is saturated at low and high intensities of the original image.

Fourier Descriptors. Fourier descriptors, proposed originally by Cosgriff in 1960[33], are a representation of a periodic signal through the coefficients of a finite combination of complex sinusoidal ordered by their frequencies. Coefficients of the sum of sines and cosines are obtained from the discrete Fourier transform (DFT). Since the image is at the frequency domain, complex values can be used as feature descriptors to represent an object or texture[34,35]. Lower frequency coefficients are shifted to extremities of the spectrum. Thus, a shift operation needs to be carried out, which moves the origin of the Fourier transform to the central coordinates at the frequency domain. Low frequency components describe the most relevant information of the behavior of a signal. On the other hand, high frequency components are related to abrupt changes and noise.

Fourier Circular-Angular. After the shift operation, $F(u, v)$ is partitioned into 64 sectors using eight circular rings and eight angles equally spaced over the image. Circular rings were calculated using radii equal to 3, 6, 9, 12,

15, 18, 21 and 24 pixels of distance. A total of 64 descriptors were given by the sum of the absolute values of each sector of the spectrum.

Fourier Circular. G circular rings using radii equal to 1, 2, …, $min(\lfloor M/2 \rfloor, \lfloor N/2 \rfloor)$, where the image has $M \times N$ pixel size, are calculated over $F(u, v)$ after the shift operation. For each circular ring, the sum of all the absolute values of the spectrum from the origin to the circular ring is performed obtaining G descriptors for each image. As the images obtained in this experiment have sizes of 2080×1540 pixels, 769 descriptors are generated per image.

Corrosion-Inspired Texture Analysis. Corrosion-Inspired Texture Analysis (CITA)[36] is a feature descriptor based on concepts from the process of metal pitting corrosion modeling and cellular automata. In this approach, a texture image is considered as a metal surface and thus represents the initial state of the cellular automaton. Each pixel in the image is considered as a cell of the automaton and, therefore 256 initial states are considered, which correspond to 256 gray levels of the image. The gray levels represent the depth of corrosion in a given surface, where 0 means no corrosion, and 255 means the greatest depth of corrosion. An update rule based on the analysis of the central pixel/cell and its neighborhood is set to update the state of each cell at time $t+1$.

Let an image be considered as a grid I of size $N \times M$. Each cell (i, j), where $i \in [1, N]$ and $j \in [1, M]$, is updated at time $t+1$ according to the level of depth calculated from the difference between the state of the central cell and the lowest state value within its neighborhood at time t.

$$d_{i,j} = s(I_{i,j}, t) - \min(\tilde{s}(N(c_{i,j}), t)), \tag{2}$$

where $s(I_{i,j}, t)$ is the state of the central cell $I_{i,j}$ and $\tilde{s}(N(c_{i,j}), t)$ is the set of states of the cells in the neighborhood of $I_{i,j}$.

Thus, the state of the cell $I_{i,j}$ at time $t+1$ is updated according to:

$$s(I_{i,j}, t+1) = \begin{cases} s(I_{i,j}, t) + C(d_{i,j}, \gamma) & \text{, if } 255 \geq d_{i,j} \geq \nu, \\ s(I_{i,j}, t) & \text{, if } d_{i,j} < \nu \text{ or } d_{i,j} > 255, \end{cases} \tag{3}$$

where ν is a surface roughness parameter, C is the level of corrosion to be applied, considering $d_{i,j}$ and the pitting power parameter $\gamma \in [0, 1]$. γ is the resistance of the metal to corrosion in a certain environmental condition.

$$C(d_{i,j}, \gamma) = \lfloor (255 - d_{i,j})\gamma \rfloor, \tag{4}$$

where $\lfloor \alpha \rfloor$ is the floor of α.

This process is performed T times. For each iteration, the cumulative mass of corroded metal of the CA-based model is used to construct the texture descriptor of size T for each image. In this context, 200 iterations were used to classify the image epidermis and the parameters were set up as $\nu = 2$ and $\gamma = 0.05$. With these parameters, 200 CITA descriptors were generated.

Local Binary Pattern. Local Binary Pattern (LBP)[37] is an operator of grayscale texture that characterizes the spatial structure of local binary patterns in the image in a circular neighborhood at any spatial resolution. This approach suggests the image consists of micro-patterns. Given a central pixel in the image, a number of patterns is computed by comparing its value with the pixels of its neighborhood:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \tag{5}$$

$$s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \tag{6}$$

where g_c is the gray level of the central pixel, g_p the gray level of the neighbor p, P is the number of neighbors and R is the radius of the neighborhood. Let an image of size $N \times M$ pixels, the LBP pattern of each pixel (i, j), $i \in [1, N]$, $j \in [1, M]$, is determined and a histogram of these patterns is calculated to represent the texture image. Consequently, a feature vector with 256 positions (the same number of gray levels) is generated for each image.

Experiments and Results

Texture analysis vs traditional methods. Following the procedure described in the Materials and Methods section, 300 samples of size 2080×1540 pixels distributed over 32 distinct woody species (see Table 1 for details) were obtained. Figures 1 and 2 show two samples of epidermis images of the same species per row. In the first figure, images with intra species similarity are shown, i.e., there is no significant variation of the images obtained from one individual to another. However, this is not the case for all species, as shown in Fig. 2, where wide variations in images of the same species (when considering different individuals) can be seen. This context makes the classification problem even more challenging.

Three texture descriptors were used to describe the images and subsequently perform the classification, which were Fourier, CITA and LBP descriptors. For the Fourier descriptors two approaches were used. The first one used only the Fourier circular descriptors, while the second one used the concatenated vectors of Fourier circular and Fourier circular-angular. The number of descriptors obtained varied from 200 to 833, therefore the Principal

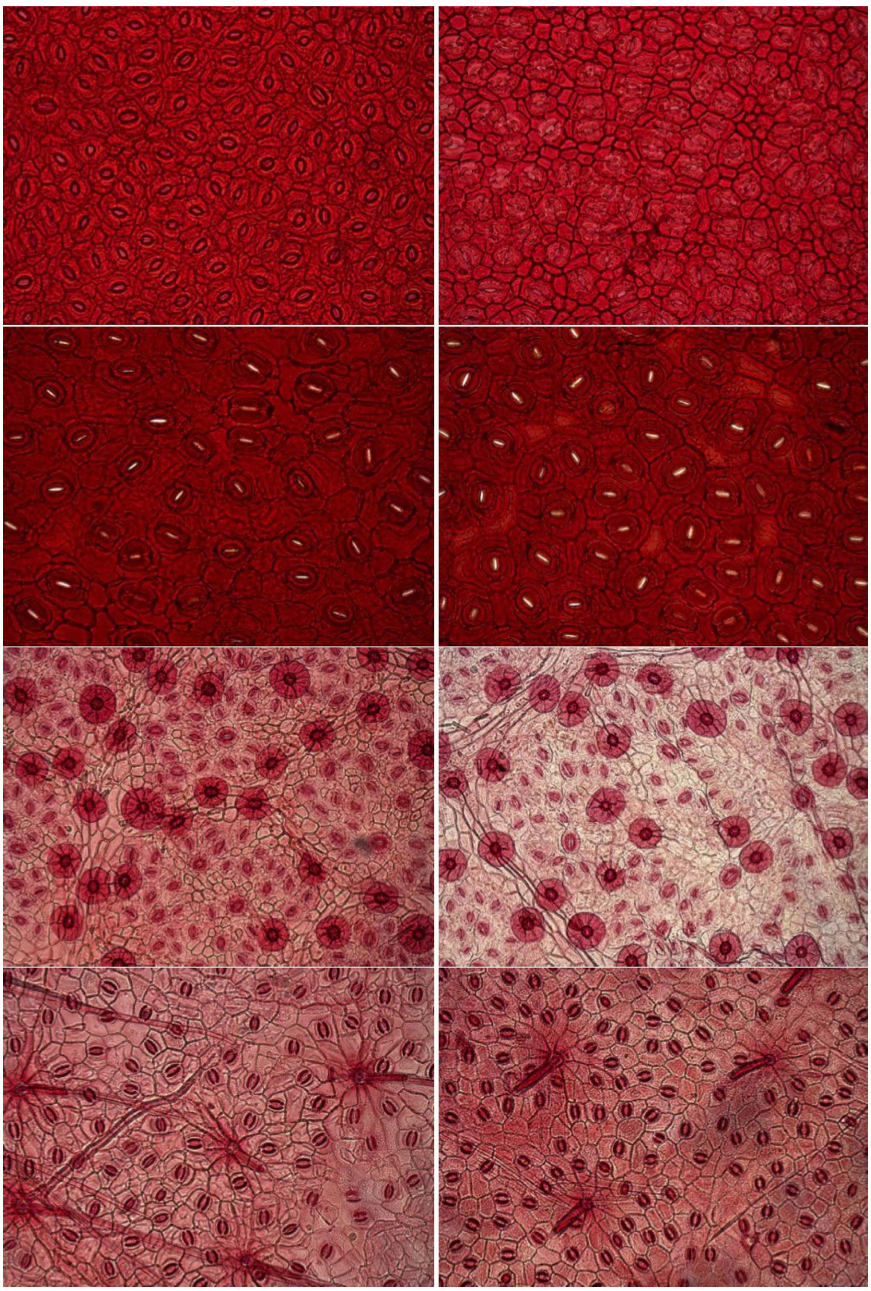

Figure 1. Samples of epidermal images with low variation in images of the same species (row). From top to bottom: *Ilex affinis*, *Myrsine guianensis*, *Handroanthus impetiginosus* and *Xylopia sericea*.

Component Analysis (PCA), a technique to reduce the dimensionality, was used for all the feature descriptors so that a smaller number of features could be used. PCA is an orthogonal linear transformation that converts a number of possibly correlated variables into a set of values of uncorrelated variables called principal components[38]. The first principal component has the greatest variance in the data, the second principal component and succeeding components have the highest variability in descending order.

A supervised classification was performed to assign each sample to a plant species. This process is called supervised because the species of the training set are known and the aim is to use a function to classify a new observation in one of the given species. A validation scheme termed stratified 6-fold cross-validation[39] was used. The set of samples was equally divided into six mutually exclusive subsets and samples of five subsets were used for training, and samples of one set were used for testing. This procedure was performed 6 times alternating the testing subset so that all samples were classified. In the end, the success rate for the data set was calculated. This procedure was executed 10 times to obtain the standard deviation.

To classify the samples, two classifiers were used: k-Nearest Neighbor (k-NN)[40,41], with $k = 1$ (experiments with $k = 1$, 3, 5 and 7 were carried out. For the analysed data, $k = 1$ obtained the best performance. k-NN for $k = 1$ is also called nearest neighbor algorithm), and Linear Discriminant Analysis (LDA)[42]. k-NN associates a class

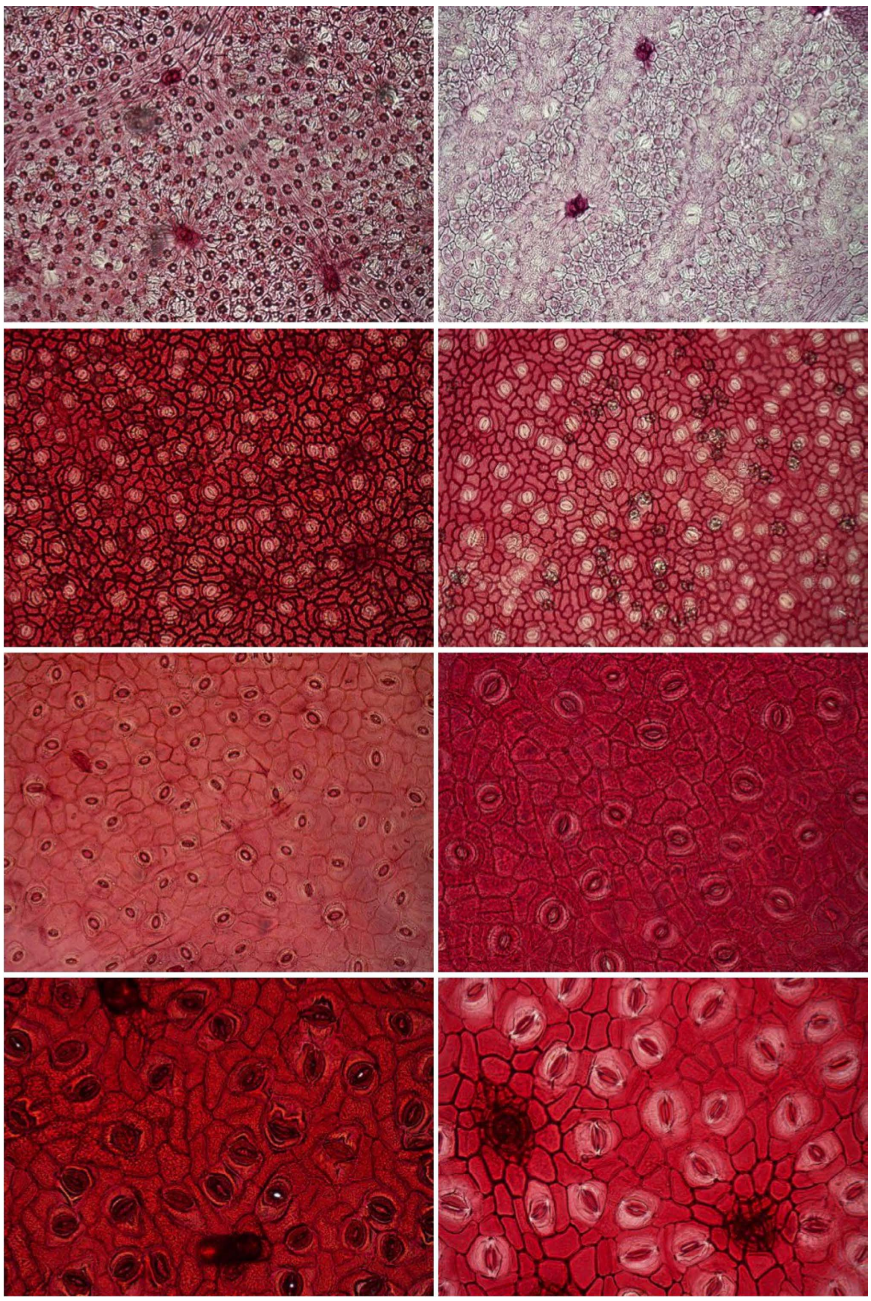

Figure 2. Samples of epidermal images with wide variations in images of the same species (row). From top to bottom: *Miconia cuspidata*, *Tapirira guianensis*, *Symplocos mosenii* and *Guapira noxia*.

to a sample according to the most repeated class of the surrounding neighborhood. LDA maximizes the ratio of between-class variance and within-class variance achieving the maximal separability. These procedures for classification were applied to the set of epidermal images and high probabilities of correct classification were obtained performing 96.6% of success rate using Fourier descriptors. All the results were compared in Tables 2 and 3 for *k*-NN and LDA, respectively.

Moving on further, the proposed method was compared to the traditional approach, where quantitative measurements were obtained manually from the epidermis images. The following features were considered: stomata density, the guard cell length and the stomatal complex width. The same 32 species and 300 samples used for the computational experiments were also used for the morphological experiment, as well the stratified 6-fold cross-validation scheme. The results using quantitative measurements are 61.33% with *k*-NN and 58.47% with LDA. For each of the texture methods, the three descriptors obtained manually (density, length and width) were added to the texture feature vector to verify the impact of using both features together for plant species recognition. It can be observed in Tables 2 and 3 that morphological characteristics may increase up to 10% success rate when combining both features.

	k-NN						
	300 images			309 images			
Feature	#	% (±std)	% *T. g.*	#	% (±std)	% *T. g. joint*	% *T. g. split*
Fourier Circular + Circular-Angular	20	96.00 (±0.05)	100	20	94.17 (±0.06)	80	33
Fourier Circular + Circular-Angular + Quantitative	23	98.67 (±0.03)	100	23	96.76 (±0.05)	80	44
Fourier Circular	19	95.00 (±0.06)	83	19	93.20 (±0.06)	67	33
Fourier Circular + Quantitative	22	97.33 (±0.04)	100	22	95.79 (±0.05)	80	44
CITA	19	74.33 (±0.12)	100	21	74.11 (±0.12)	80	44
CITA + Quantitative	22	84.67 (±0.09)	100	24	84.14 (±0.09)	80	44
LBP	29	70.00 (±0.13)	67	33	72.49 (±0.13)	80	78
LBP + Quantitative	32	80.67 (±0.11)	100	36	81.88 (±0.10)	87	56
Quantitative (Density + Length + Width)	3	61.33 (±0.15)	100	3	57.61 (±0.15)	33	44

Table 2. Classification accuracy of 32 plant species. The results are described by the number of PCA components (#), success rate and standard deviation (std) using Fourier, CITA and LBP feature descriptors and *k*-NN as classifiers. Moreover, the results are presented in two modes. The first one shows the success rate for 300 images, in which images of *Tapirira guianensis* only from the gallery forest are used. The second one, labeled as '309 images', shows the results including nine images of *Tapirira guianensis* from a marsh camp to draw a comparison of the classification rate with samples of the same species which grew in different environments. The success rate of identifying the *Tapirira guianensis* species using stratified 6-fold cross validation is shown in the columns '% *T. g.*' and '% *T. g. joint*', respectively, for 300 and 309 images and the column '% *T. g. split*' shows the result considering the nine images of *Tapirira guianensis* from the marsh camp as the testing set and the remaining 300 images as the training set.

	LDA						
	300 images			309 images			
Feature	#	% (±std)	% *T. g.*	#	% (±std)	% *T. g. joint*	% *T. g. split*
Fourier Circular + Circular-Angular	45	96.60 (±1.27)	100	46	94.92 (±0.63)	89	67
Fourier Circular + Circular-Angular + Quantitative	48	97.43 (±0.39)	100	49	96.28 (±0.53)	89	67
Fourier Circular	31	95.83 (±0.63)	100	40	94.63 (±0.79)	92	67
Fourier Circular + Quantitative	34	97.43 (±0.49)	100	43	96.34 (±0.82)	92	67
CITA	64	78.13 (±1.35)	83	55	78.93 (±1.82)	86	56
CITA + Quantitative	67	86.87 (±1.31)	83	58	85.28 (±2.38)	86	44
LBP	87	82.83 (±1.39)	70	80	83.43 (±1.23)	83	89
LBP + Quantitative	90	91.10 (±1.44)	70	83	88.71 (±1.90)	83	100
Quantitative (Density + Length + Width)	3	58.47 (±1.02)	100	3	55.28 (±0.62)	36	0

Table 3. Classification accuracy of 32 plant species. The results are described by the number of PCA components (#), success rate and standard deviation (std) using Fourier, CITA and LBP feature descriptors and the LDA as classifiers. Moreover, the results are presented in two modes. The first one shows the success rate for 300 images, in which images of *Tapirira guianensis* only from the gallery forest are used. The second one, labeled as '309 images', shows the results including nine images of *Tapirira guianensis* from the marsh camp to draw a comparison of the classification rate with samples of the same species which grew in different environments. The success rate of identifying the *Tapirira guianensis* species using stratified 6-fold cross validation is shown in columns '% *T. g.*' and '% *T. g. joint*', respectively, for 300 and 309 images and the column '% *T. g. split*' shows the result considering the nine images of *Tapirira guianensis* from the marsh camp as the testing set and the remaining 300 images as the training set.

The effect of plasticity on identification. Plasticity is an important characteristic that makes it difficult to conventionally identify plant species because the individuals from the same species, which grow in different vegetation formations with different environmental factors, can present different morphological features[43]. The database used for the experiments contains 300 images from 32 species, including six samples of *Tapirira guianensis* species grown in a gallery forest environment. The results presented in Tables 2 and 3 at column '% *T. g.*' show the identification success rate of the *Tapirira guianensis* species grown in a gallery forest environment.

To verify if the approaches used in this work (the computational method and the traditional method) are invariant to plasticity, nine images of *Tapirira guianensis* species grown in a marsh camp environment were added to the database making a total of 309 images. For this new dataset, two approaches were considered. Firstly, images of *Tapirira guianensis* grown in a gallery forest and in marsh camp environment were placed together in the database and classified using a stratified 6-fold cross-validation scheme (Column '% *T. g. joint*' in Tables 2 and 3). Secondly, the nine samples of *Tapirira guianensis* species grown in a marsh camp environment were considered as the test set and the 300 images of the database were considered as the training set (Column '% *T. g. split*' in Tables 2 and 3). The first approach analyses the increase in the morphological variability, when samples of the

same species grown in another environment are added. The second one analyses the direct effect of plasticity on the identification methodology. In this case, for successful identification, methods must be able to observe characters that are less biased for the growing environment. Considering only samples of *Tapirira guianensis* from the gallery forest, the recognition rate was 100% using the traditional method. However, when adding samples obtained from *Tapirira guianensis* grown in a marsh camp, which increases the variability due to the plasticity ('309 images', column '% *T. g. joint*', Tables 2 and 3), the success rate using the traditional methodology fell sharply, while the texture descriptors remained robust, maintaining the high success rate of the *Tapirira guianensis* species identification.

Considering the experiment where the samples from the two grown environments were split, training with samples grown in a gallery forest and testing using samples grown in a marsh camp ('309 images', column '% *T. g. joint*', Tables 2 and 3), the results of the traditional approach showed a high decrease. It changed from 100% to 44% of success rate assessed by the *k*-NN classifier and to 0% estimated by the LDA classifier. On the other hand, in the case of the *k*-NN classifier, the texture descriptors had a high decrease, achieving a success rate less than 50%. Taking into account the LDA classifier, the results were better, showing 67% of success rate for most of the texture descriptors. The exception of this decrease was the LBP method, which had a high success rate for both classifiers, achieving 78% for *k*-NN and 89% when using the LDA, and 100% of success when combined with the traditional method using the LDA.

Discussion

Analysing morphological and anatomical traits has helped to discriminate species for a long time[23,43]. In this context, information about leaf epidermis features (distribution pattern of epidermal cells, types of trichomes and stomata, shape of guard cells and cuticular ornamentation) have been quite useful for taxonomical approaches[44–46]. However, these features could be limited to discriminating a large number of species. Considering this, computational analysis of the epidermis texture provides many suitable descriptors for plant discrimination, which shows it is a promising approach in this task.

In our study, four different approaches were used to obtain information about leaf epidermis to classify species, two of which were related to feature extraction based on Fourier transform. Both of these showed the best results. Furthermore, CITA and LBP feature descriptors were used, also proving to be effective to identify plant species. All the tests were carried out using both *k*-NN and LDA classifiers and similar results were obtained in most of the cases, showing the consistency of the results. Considering the plasticity experiment, columns '% *T. g. joint*' and '% *T. g. split*' in Tables 2 and 3, it can be observed that the LDA had a much better result than *k*-NN. The *k*-NN is a very simple classification method that proved to be very useful in most of the applications, while LDA is more sophisticated and uses linear transformations to explain the data better. Therefore, when the morphological variability between the training and testing sets are high, the result is also dependent on the adopted classifier, and the LDA is a better option.

The ability of the Fourier descriptors to concentrate the low frequency components separately from the high frequency components enables us to analyse epidermal tissue (regular cells with coating function) separately from specialized cells (stomata) and epidermal appendices (trichomes). Thus, attributes of low and high frequency can be solely compared between samples and these comparisons are joined to enhance the separation of different species and decrease the distance between samples of the same species. Therefore, Fourier descriptors showed the best classification rate, achieving a very impressive success rate higher than 96%. CITA and LBP descriptors have similar methodologies for texture analysis considering the behavior of central pixel neighborhood for feature extraction. CITA considers regions with similar values as belonging to the same local surface and tries to maintain these regions throughout the iterations. These regions may be the epidermis tissue and regions within the structure of stomata. Furthermore, CITA attempts to erode regions with high abrupt differences so that structures on the epidermis can be analysed. Thus, CITA achieves a reasonable result in plant species classification. LBP accounts for patterns of neighbors by simply analyzing if they have a value higher or lower than the central pixel without considering the difference ratio. When considering the plasticity and a large variability from the training and testing sets, the LBP method obtained the best result. Among the texture descriptors evaluated, Fourier and LBP descriptors are the best options. The choice between them has to be made according to the variability of the samples used in the experiment.

Concerning the comparison between computational methodologies based on texture and the conventional approach using manual measurements, significant differences were observed. The methods based on texture features were better able to classify the species than just the quantitative data, which are more laborious to obtain. Texture methods are able to identify plant species even when the species presents plasticity in their quantitative measurements from the epidermis structures. These quantitative values can change according to the environment, however the texture information remains concise. An analysis using manual measurements considers only the shape and density of the stomata, whereas the texture analysis considers spatial orientation and geometric arrangement of the stomata, as well as their quantitative characteristics. The orientation of the stomata is an important attribute as in certain groups of species, it can be used as an important discriminating attribute, as in the case of monocotyledons and pine species. Furthermore, as the identification method based on texture uses the entire image of the epidermis, stomata are not segmented, thus epidermal tissue patterns and also other structures that might occur in some species, such as trichomes are also considered. The combination of all these features allows for a rich analysis, able to provide a strong and unique identity to different plant species.

The texture information of the epidermis when combined with the manual measurements of the stomata (density, length and width) enhance the information concerning species identification. When texture and morphological features are used together, the success rate may increase up to 10%. The main advantage of texture methods is that it is no longer necessary to segment the stomata. The stomata segmentation process is a difficult task, as the color and contrast of the stomata varies for each species and, in many cases, an automatic

segmentation methodology is not feasible. Considering that the texture method is automatic and can be analysed more quickly without manual intervention of the images while maintaining a good success rate, depending on the experiment, it can be used without combining it with the quantitative measurements.

Conclusions

The main method used to identify any plant species is based on the morphological analysis. Nevertheless, identifying plant species can become a difficult task for botanists and other researchers not specialized in plant taxonomy. In this context, computational approaches have became an important tool to help in the task of identifying plant species. Considering the efficiency of the texture descriptors to discriminate plant species by epidermis images and the simplicity of this approach, an innovative concept is proposed that can investigate common texture features extracted from the taxonomic hierarchical levels, which could help taxonomists to classify new unknown plant species using a pattern recognition system. These new discrimination strategies associated with extracting features of images of leaves can open up new possibilities for classification methods of modern taxonomy.

References

1. Keddy, P. A. *Plants and vegetation* (Cambridge University Press, 2006).
2. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281,** 237–240 (1998).
3. Archibold, O. W. *et al. Ecology of world vegetation*. (Chapman & Hall Ltd, 1995).
4. Ratter, J. A., Ribeiro, J. F. & Bridgewater, S. The Brazilian Cerrado vegetation and threats to its biodiversity. *Ann. Bot.* **80,** 223–230 (1997).
5. Lehmann, C. E. R. *et al.* Savanna vegetation-fire-climate relationships differ among continents. *Science* **343,** 548–552 (2014).
6. Mendonça, R. C. *et al.* Flora vascular do cerrado: checklist com 12.356 espécies. In Sano, S. M., Almeida, S. P. & Ribeiro, J. F. (eds) *Cerrado: ecologia e flora* 422–442 (Embrapa Informação e Tecnologia, 2008).
7. Walter, B. M. T. *Fitofisionomias do bioma Cerrado: síntese terminológica e relações florísticas.* Ph.D. thesis, Universidade de Brasília, Brasilia (2006).
8. Myers, N. The biodiversity challenge: Expanded hot-spots analysis. *Environmentalist* **10,** 243–256 (1990).
9. Gottsberger, G. & Silberbauer-Gottsberger, I. *Life in the Cerrado: Origin, structure, dynamics and plant use*. Life in the Cerrado (Reta Verlag, 2006).
10. Bieras, A. C. & das Graças Sajo, M. Leaf structure of the cerrado (Brazilian savanna) woody plants. *Trees-struct. Funct.* **23,** 451–471 (2009).
11. Bridson, D., Forman, L. & Royal Botanic Gardens, K. *The Herbarium Handbook* (Royal Botanic Gardens, 1998).
12. Stace, C. A. *Plant taxonomy and biosystematics* (Cambridge University Press, 1991).
13. Liu, J., Zhang, S. & Deng, S. A method of plant classification based on wavelet transforms and support vector machines. In Huang, D.-S., Jo, K.-H., Lee, H.-H., Kang, H.-J. & Bevilacqua, V. (eds) *Emerging Intelligent Computing Technology and Applications,* vol. 5754 of *Lecture Notes in Computer Science* 253–260 (Springer Berlin Heidelberg, 2009).
14. Casanova, D., de Mesquita Sa Junior, J. J. & Bruno, O. M. Plant leaf identification using gabor wavelets. *Int. J. Imaging Syst. Technol.* **19,** 236–243 (2009).
15. Backes, A. R., Casanova, D. & Bruno, O. M. Plant leaf identification based on volumetric fractal dimension. *Int. J. Pattern Recogn.* **23,** 1145–1160 (2009).
16. Rossatto, D., Casanova, D., Kolb, R. & Bruno, O. M. Fractal analysis of leaf-texture properties as a tool for taxonomic and identification purposes: a case study with species from neotropical Melastomataceae (Miconieae tribe). *Plant Syst. Evol.* **291,** 103–116 (2011).
17. SáJunior, J. J. d. M., Backes, A. R., Rossatto, D. R., Kolb, R. M. & Bruno, O. M. Measuring and analyzing color and texture information in anatomical leaf cross sections: an approach using computer vision to aid plant species identification. *Botany* **89,** 467–479 (2011).
18. Florindo, J. B. *et al. Brachiaria* species identification using imaging techniques based on fractal descriptors. *Comput. Electron. Agric.* **103,** 48–54 (2014).
19. Julesz, B. Visual pattern discrimination. *IEEE Trans. Inf. Theory* **8,** 84–92 (1962).
20. Gonzalez, R. & Woods, R. *Digital Image Processing* (Pearson/Prentice Hall, 2008), 3 edn.
21. da Silva, N. R. *et al.* Plant identification based on leaf midrib cross-section images using fractal descriptors. *PLOS ONE* **10,** e0130014 (2015).
22. Evert, R. & Eichhorn, S. *Esau's Plant Anatomy: Meristems, Cells, and Tissues of the Plant Body: Their Structure, Function, and Development* (Wiley, 2006).
23. Metcalfe, C. R. & Chalk, L. *Anatomy of the dicotyledons: volume 1. Systematic anatomy of leaf and stem, with a brief history of the subject* (Oxford, Clarendon Press, 1979).
24. Payne, W. W. Stomatal Patterns in Embryophytes: Their Evolution, Ontogeny and Interpretation. *Taxon* **28,** 117–132 (1979).
25. Keating, R. C. Leaf histology and its contribution to relationships in the Myrtales. *Ann. Missouri Bot. Gard.* **71,** 801–823 (1984).
26. Barthlott, W. *Applications of the scanning EM in taxonomy and functional morphology,* chap. Scanning eléctron microscopy of the epidermal surface in plants, 69–94 (Clarendon for the Systematics Association, 1990).
27. Cristina, M., Dinis, A. & Sales, F. Testing the reliability of anatomical and epidermical characters in grass taxonomy. *Microsc. Microanal.* **14,** 156–157 (2008).
28. Johansen, D. *Plant Microtechnique*. McGraw-Hill publications in the botanical sciences (McGraw-Hill, 1940).
29. Franklin, G. L. Preparation of thin sections of synthetic resins and wood-resin composites, and a new macerating method for wood [10]. *Nature* **155,** 51 (1945).
30. Ramos, E. & Fernández, D. S. Classification of leaf epidermis microphotographs using texture features. *Ecol. Inf.* **4,** 177–181 (2009).
31. Wang, H.-J., Zhang, G.-Q. & Qi, H.-N. Wood recognition using image texture features. *PLOS ONE* **8,** e76101 (2013).
32. Price, T. D., Qvarnström, A. & Irwin, D. E. The role of phenotypic plasticity in driving genetic evolution. *Proc. R. Soc. B* **270,** 1433–1440 (2003).
33. Cosgriff, R. L. Identification of shape. Ohio State University Research Foundation, Columbus, Rep. 820-11,. ASTIA AD 254 792 (1960).
34. Azencott, R., Wang, J.-P. & Younes, L. Texture classification using windowed Fourier filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **19,** 148–153 (1997).
35. Gonzalez, R. C. & Woods, R. E. *Digital Image Processing* (3rd Edition) (Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006).
36. da Silva, N. R., der Weeën, P. V., Baets, B. D. & Bruno, O. M. Improved texture image classification through the use of a corrosion-inspired cellular automaton. *Neurocomputing* **149, Part C**, 1560–1572 (2015).
37. Ojala, T., Pietikäinen, M. & Mäenpää, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **24,** 971–987 (2002).

38. Jolliffe, I. T. *Principal component analysis* (Springer, New York, 2002).
39. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning*. Springer Series in Statistics (Springer New York Inc., New York, NY, USA, 2001).
40. Aha, D. W., Kibler, D. & Albert, M. K. Instance-based learning algorithms. *Mach. Learn.* **6,** 37–66 (1991).
41. Amancio, D. R. *et al.* A systematic comparison of supervised classifiers. *PLOS ONE* **9,** e94137 (2014).
42. Fisher, R. A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics* **7,** 179–188 (1936).
43. Stuessy, T. *Plant Taxonomy: The Systematic Evaluation of Comparative Data* (Columbia University Press, 2009).
44. Yang, Z.-R. & Lin, Q. Comparative morphology of the leaf epidermis in *Schisandra* (Schisandraceae). *Bot. J. Linn. Soc.* **148,** 39–56 (2005).
45. Agbagwa, I. O. & Okoli, B. E. Leaf epidermal micromorphology in the systematics of *Abrus* (Papilionaceae) in parts of tropical West Africa. *Asian J. Plant Sci.* **5,** 41–49 (2006).
46. Edeoga, H. & Ikem, C. Comparative morphology of the leaf epidermis in three species of *Boerhavia* L. (Nyctaginaceae). *J. Econ. Taxon. Bot.* **19,** 197–205 (2001).

Acknowledgements

The authors gratefully thank the financial support of the São Paulo Research Foundation (FAPESP), with grant N^{o}s.: 2011/01523-1, 2011/23112-3 and 2011/21467-9, the National Council for Scientific and Technological Development (CNPq) with grant N^{o}s.: 307797/2014-7 and 484312/2013-8 and PROPE/UNESP (14/2012/Renove), and the Coordination for the Improvement of Higher Education Personnel (CAPES).

Author Contributions

O.M.B, R.M.K. and D.R.R. designed the study; L.F.S.P., D.R.R. and R.M.K. collected the samples, produced the microscope slides and acquired the images; N.R.S. and M.W.S.O. performed the image processing and analysis; N.R.S., M.W.S.O. and O.M.B. analyzed the data sets; N.R.S., M.W.S.O., H.A.A.F., D.R.R., R.M.K. and O.M.B. wrote the paper.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Silva, N.R. *et al.* Leaf epidermis images for robust identification of plants. *Sci. Rep.* **6**, 25994; doi: 10.1038/srep25994 (2016).