Towards more effective identification keys – a study of people identifying plant species characters: supporting material of statistics and results

A. Splitting participants into knowledge groups

We split participants into groups based on the self-assessment they provided in the first part of the questionnaire. These questions inquired about: (a) the number of species they were familiar with, (b) their experience with identification keys, (c) a self-assessment of plant knowledge, and (d) whether they are professionally involved with plants. We used a generalization of Gower's dissimilarity index to account for the mixture of ordinal and factor variables [2]. The resulting dissimilarity matrix was clustered using the partitioning around medoids algorithm (PAM) [1] and the average silhouette width was calculated to assess the quality of differing cluster numbers. We found a three-cluster solution to yield the highest value silhouette score of 0.63 and we therefore split participants into three different expertise groups based on their self-reported experience with plant identification.

B. Testing significance differences between expertise groups and plant organ related characters

A Kolmogorov-Smirnov test shows that our measured variables, i.e., answer correctness, self-assessed difficulty, self-assessed certainty, number of image views, and response time, are not normally distributed. Therefore, we compare groups in terms of these variables with non-parametric Mann-Whitney U tests (organ: leaf versus flower related characters) and Kruskal–Wallis tests (between three expertise groups).

C. Fitting a generalized linear mixed-effect model

Furthermore, we fit a generalized linear mixed-effect model (GLMM) with binomial error a priori utilizing all explanatory variables to model the probability of a character identification being conducted correctly. We consider the participants' identity and the species identity as random effects. We simplified the initial model using the Akaike information criterion (AIC) computed via R's MuMIn package [3]. We retain all models with ΔAIC<6 to be 95% sure that the most parsimonious models are maintained within the best supported model set [4]. Model averaging was used to calculate averaged parameter estimates and assess the relative importance (RI) of parameters using the natural averaging method [5]. Parameters within the resulting averaged model are considered significant if the p-value is < 0.05. The amount of variance explained by the fixed effects only and the combined fixed and random effects of the binomial GLMM models are calculated as marginal $R^2_{GLMM(m)}$ and conditional $R^2_{GLMM(c)}$ respectively following Nagakawa and Schlielzeth's method [6]. All analysis was performed with R version 4.0.2 [7].

The global model explained about 32% of the variation in the data ($R^2_{GLMM(c)} = 0.32$) while about 10% was explained by the fixed factors ($R^2_{GLMM(m)} = 0.10$). We produced a candidate model set consisting of all simplified versions of the global model and compared them based on their AIC. The top four models with $\Delta\text{AIC} < 6$ (cp. Tab. S2.1) were used to produce model averaged parameter estimates. "Difficulty", "certainty", "time", "organ", "number of character states" were all retained in each model within the candidate model set and had a relative importance (RI) of 1 in the final averaged model. Skill level and the number of viewed images were retained in 50% of candidate models with a RI of 0.75 and 0.29 respectively. The model-averaged parameter estimates highlight how correctness is strongly related to "certainty" followed by "number of character states" and "organ" with the steepest decline, while skill level and the number of pictures viewed had no statistically significant influence on correctness (Tab. S2.2).

	df	logLik	AIC ↓	ΔAIC	weight
Ability to correctly identify plant character states					
user group + difficulty + certainty + time + organ + # character states	9	-5347.91	10713.8	0	0.53
user group + difficulty + certainty + time + organ + # character states + # viewed images	10	-5347.79	10715.6	1.76	0.22
difficulty + certainty + time + organ + # character states	8	-5349.99	10716.0	2.15	0.18
difficulty + certainty + time + organ + # character states + # viewed images	9	-5349.94	10717.9	4.06	0.07
user group + certainty + time + organ + # character states	8	-5355.72	10727.4	13.60	0.00

Table S2.1. Results of top 5 models based on AIC (df = degrees of freedom, logLik = log-likelihood, AIC = Akaike information criterion, weight = Akaike weight)

	Parameter Estimate	Standard Error	Adjusted Standard Error	z-value	p-value
intercept	3.38	0.16	0.16	21.60	$< 0.001^{***}$
user group	0.06	0.05	0.05	1.23	0.22
difficulty	-0.19	0.05	0.05	3.88	$< 0.001^{***}$
# character states	-0.12	0.02	0.02	7.59	$< 0.001^{***}$
organ (leaf)	-0.29	0.06	0.06	4.70	$< 0.001^{***}$
certainty	-0.39	0.05	0.05	8.23	$< 0.001^{***}$
time	-0.01	0.00	0.00	6.13	$< 0.001^{***}$
# viewed images	0.00	0.01	0.01	0.21	0.84
marginal $R^2_{GLMM(m)}$ of fixed effects only = 0.10					
conditional $R^2_{GLMM(c)}$ of fixed and random effects = 0.32					

Table S2.2. Summary of results after model averaging for ability to correctly identify plant character states

D. Character identification and plant organs

Figure S2.1 provides an overview across all evaluated characters grouped by organ. For each character, a bar visualizes average identification correctness across all participants and their perceived difficulty encoded as bar color. Further columns refer to the number of character states and other metrics as known from the previous result tables.

E. Character state identification and species

Figure S2.2 shows identification correctness in relation to character state, rather than characters as shown in Figure S2.1. Furthermore the correctness of the character states is separated per species. Characters are again grouped into flower-related (upper part (a)) and leaf-related (lower part (b)) and shown from left to right. For each character CLxx and CFxx, different colors refer to results per character state and individual data points refer to the average identification correctness of this character state while identified from the same displayed species. The bubbles show the mean value independent of the species. The raw data of this plot is available in Table S2 of the supplementary material.

Fig. S2.1. Accuracies of answers given by the 484 participants in identifying plant characters. Bars indicate the percentage of correctly answered questions. Colors indicate mean difficulty per character rated by participants on a scale from 1 (easy) to 4 (difficult). The upper bars refer to leaf characters, while the lower refer to flower characters.

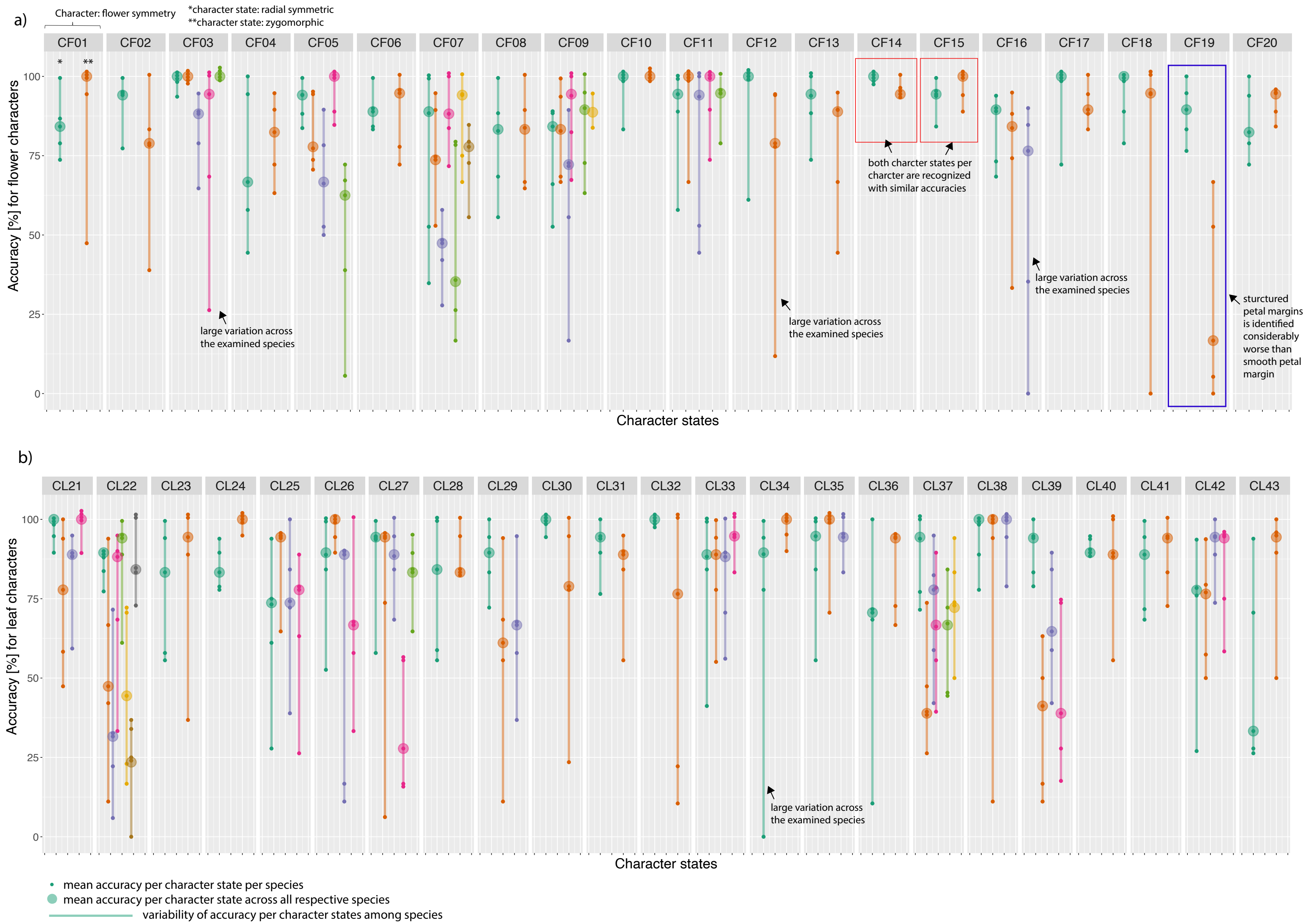

Fig. S2.2. Accuracies per character, character state, and species grouped into (a) flower-related and (b) leaf-related characters. Points show the mean accuracy per character state-species combination and lines visualize variability of accuracy among species. Bubbles show the overall mean value per character state across all respective species. Multiple points representing equal accuracies are drawn next to each other.

REFERENCES

1. Kaufman, L. and Rousseeuw, P. J. (1990) Finding Groups in Data. Wiley.
2. Pavoine, S., Vallet, J., Dufour, A.-B., Gachet, S. and Daniel, H. (2009) On the challenge of treating various types of variables: application for improving the measurement of functional diversity. Oikos, 118, 391–402.
3. Barton, K. (2019) MuMIn: multi-model inference. URL: http://r-forge.r-project.org/projects/mumin/. R package version 1.43.17.
4. Richards, S. A. (2008) Dealing with overdispersed count data in applied ecology. Journal of Applied Ecology, 45, 218–227.
5. Burnham, K. P. and Anderson, D. R. (2002) A practical information-theoretic approach. Model selection and multimodel inference, 2.
6. Nakagawa, S. and Schielzeth, H. (2013) A general and simple method for obtaining r2 from generalized linear mixed-effects models. Methods in Ecology and Evolution, 4, 133–142.
7. R Core Team (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.