

Insect Morphometry is Reproducible Under Average Investigation Standards

Sandor Csosz¹, Bernhard Seifert², Istvan Mikó³, Brendon Boudinot⁴, Marek Borowiec⁵, Brian Fisher⁶, Matthew Prebus⁵, Jayanthi Puniamoorthy⁷, Jean Claude Rakotonirina⁸, Nicole Rasoamanana⁸, Roland Schultz², Carolyn Trietsch⁹, Jonah Ulmer¹⁰, and Zoltán Elek¹¹

¹MTA

²Senckenberg Museum of Natural History Görlitz

³University of New Hampshire Department of Biological Sciences

⁴University of California Davis Department of Entomology and Nematology

⁵University of Idaho

⁶California Academy of Sciences

⁷National University of Singapore

⁸Madagascar Biodiversity Center

⁹Pennsylvania State University University Park

¹⁰Staatliches Museum für Naturkunde Stuttgart

¹¹MTA-ELTE-MTM, Ecology Research Group

August 28, 2020

Abstract

Morphometric research is being applied to a growing number and variety of organisms. Discoveries achieved via morphometric approaches are often considered highly transferable, in contrast to the tacit and idiosyncratic interpretation of discrete character states. The reliability of morphometric workflows in insect systematics has never been a subject of focused research, but such studies are sorely needed. In this paper, we assess the reproducibility of morphometric studies of ants where the mode of data collection is a shared routine. We compared datasets generated by eleven independent gaugers, i.e. collaborators, who measured 21 continuous morphometric traits on the same pool of individuals according to the same protocol. The gaugers possessed a wide range of morphometric skills, had varying expertise among insect groups, and differed in their facility with measuring equipment. We used Intraclass correlation coefficients (ICC) to calculate repeatability and reproducibility values (i.e., intra-, and inter-gauger agreements), and we performed a multivariate Permutational Multivariate Analysis of Variance (PERMANOVA) using the Morosita index of dissimilarity with 9999 iterations. The calculated average measure of intraclass correlation coefficients of different gaugers ranged from $R = 0.784$ to $R = 0.9897$ and a significant correlation was found between the repeatability and the morphometric skills of gaugers ($p = 0.016$). There was no significant association with the magnification of the equipment in the case of these rather small ants. The inter-gauger agreement, i.e. the reproducibility, varied between $R=0.872$ and $R=0.471$ (mean $R=0.690$), but all gaugers arrived at the same two-species conclusion. A PERMANOVA test revealed no significant gauger effect on species identity ($R^2 = 0.69$, $p=0.58$). Our findings show that morphometric studies are reproducible when observers follow the standard protocol; hence, morphometric findings are widely transferable, and will remain a valuable data source for alpha taxonomy.

Keywords

measurement error, repeatability, morphology, entomology, taxonomy, species delimitation

Introduction

The phenotype of organisms varies continuously during development and through evolutionary time. Continuous morphological variation is captured for numerous purposes in the life sciences via the practice of morphometry: the measurement of the size and shape of anatomical forms. Morphometry has yielded novel findings in evolution (Esquerré et al., 2020) and has been used to assess fluctuating asymmetry (Palmer, 1993; Klingenberg, 2015), ontogeny (Csősz & Majoros, 2009; Shingleton et al., 2007), ecomorphism (Mahendiran et al. 2018; Tomiya & Meachen, 2018; Anderson et al., 2019), and in human clinical practice (Bartlett & Frost, 2008). Among other applications, morphometric data are also key for alpha taxonomy, the discipline of formally differentiating and describing species and higher taxa. This is exemplified by the development of phenetics in the twentieth century (Michener & Sokal, 1957; Sokal & Sneath, 1963) and by numerous modern studies in other frameworks, such as for plants (Savriama, 2018; Chuanromanee, Cohen & Ryan, 2019), animals (Villemant, Simbolotti & Kenis, 2007; Inäbnit, 2019), and other organisms (Fodor et al., 2015; McMullin et al., 2018). Continuous data are also valuable, for modeling evolutionary histories (e.g., Parins-Fukuchi, 2017, 2020). Thus, the morphometric approach constitutes a fundamental and crucial practice for the study of phenotypes in biodiversity research.

Morphology is traditionally considered to comprise both continuous and discrete traits (Artistotle, 350; Thompson, 1917; Rensch, 1947; Remane, 1952). Discrete states were established as the basic comparative units in animal alpha taxonomy from its formalization (Linnaeus 1758), and have become a key means of scoring data for phylogenetic analysis, particularly after Hennig (1950, 1966). The reproducibility of scoring discrete states is an issue, however, as qualitative perception of phenotype not only requires specific training and considerable experience but can also be plagued by arbitrariness (Bond & Beamer, 2006), meaning that variation may simply come from individual (mis-)interpretation. The qualitative approach commonly uses verbal species descriptions that are often subjective or difficult to articulate. Therefore, information transfer, if at all reliable, is based on one-to-one knowledge sharing mechanisms, and requires logically-structured linguistic hierarchies such as the Hymenoptera Anatomy Ontology (Yoder, Mikó, Seltmann, Bertone & Deans, 2010).

In contrast to this relatively idiosyncratic approach, morphometry is considered transferable. It converts variation of shape, size of anatomical traits, and number and arrangement of anatomical elements into numerical values, allowing for the dissemination of reproducible, phenotype-based knowledge. Today, an increasing number of morphology-based insect alpha-taxonomists use morphometric data and provide numeric keys to species (Steiner Schlick-Steiner & Moder, 2006; Csősz Heinze & Mikó 2015; Seifert, 2018). If observers arrive at the same conclusion by measuring traits according to the same protocol, findings are believed to be reliable and transferable. If one can measure a trait, anyone else should be able to reproduce it.

However, measurements come with error. Agreement among different observers and within a single observer's measurements is affected by a number of sources, such as the skills of the observer (if human input is required), the precision and accuracy of the equipment, clear interpretation and appropriate understanding of the character recording protocol, and other parameters. All of the uncertainty factors mentioned above are common in practice, and the fact that it is impossible to control every source of measurement variation challenges morphometry-based research (Wolak, Fairbairn & Paulsen, 2012). Understanding of the degree to which measurement errors may affect the transferability of findings is urgently needed. During the last few decades, reproducibility issues have been studied in vertebrate systematics (e.g., Oxnard, 1983, Corruccini, 1988; Yezerinac, Loughheed & Handford, 1992; Helm and Albrecht, 2000; Takacs Vital, Ferincz & Staszny, 2016; Fox, Veneracion & Blois, 2020), clinical research (e.g., Bland & Altman, 1986; Ridgway et al., 2008; Phexell et al., 2019), social science (e.g., Salganik et al. 2020), molecular phylogeny and genetic clustering

(e.g., Huelsenbeck, 1998; Jones et al., 1998; DeBiasse & Ryan, 2019), and morphometric data generally (Andrew et al., 2015). However, to date, reproducibility assessments of morphometric data in entomology are extremely limited (Mutanen & Pretorius, 2007; Johnson et al., 2013).

In order to address the question “to what extent is insect morphometry reproducible?”, we compiled a broad database of morphometric data and performed robust statistical analyses. We used ants, a group in which the application of morphometric data has a long tradition (e.g., Brown, 1943; Brian & Brian, 1949), as a model organism. Morphometry has been employed widely in recent myrmecological studies (e.g., Ward, 1999; Baroni Urbani, 1998; Seifert, 1992, 2003, 2019; Csősz Heinze & Mikó, 2015; Wagner et al., 2017) as the primary method of interpreting anatomical forms and their variation. Eleven participants of diverse levels of skill and expertise, working with different taxonomic routines over three continents and six countries, were asked to perform repeated measurements on the same set of ant specimens, according to the same measurement protocol, with their own equipment. The wide range of morphometric skills and the quality of microscopes used provided us with an overview of the level of reproducibility of morphometric interpretation as it works in daily practice. Our findings are a first step in exploring the reproducibility of morphometric data across entomology.

Terminology [Textbox 1.]

A number of terms (e.g. “accuracy”, “precision”, “reliability”, “repeatability”, and “reproducibility”) commonly used in association with repeatability studies are defined differently in the literature. To increase the fluency of scientific discourse, we propose to adopt the standard terminology of the National Institute for Standards and Technology (NIST, Taylor & Kuyatt, 2001) of the USA and terms proposed by (Bartlett & Frost, 2008) in biological systematics:

- * Accuracy describes the average closeness of the measurement(s) to the value of the measurand (= subject or quantity to be measured) (Fig. 1). Accuracy is affected by systematic and random error. We follow the terminology proposed by the NIST in using the phrase “the value of the measurand” instead of the often-applied “true value of the measurand” (or “a true value”) (Taylor & Kuyatt, 2001).
- * Precision refers to the closeness of the measurements between pairs of measurements made on the same measurand and applying the same protocol. Precise measurements are tightly clustered, but are not necessarily accurate, i.e. close to the value of the measurand (Fig. 1). Precision is affected by random error.
- * Reliability refers to the amount of measurement error that occurs between observed measurements compared to the inherent amount of variability that occurs between measurands (Bartlett & Frost, 2008).
- * Repeatability refers to the degree of agreement between repeat measurements made on the same measurand under the same conditions, i.e. made by the same observer, using the same microscope, following the same measurement protocol (Taylor & Kuyatt, 2001). Repeatability can be assessed via intra-class correlation (ICC, see Lessells & Boag, 1987).
- * Reproducibility refers to the degree of agreement between measurements made on the same measurand under changing conditions, such as changing principle, method of measurement, observer, instrument, etc. (Taylor & Kuyatt, 2001).

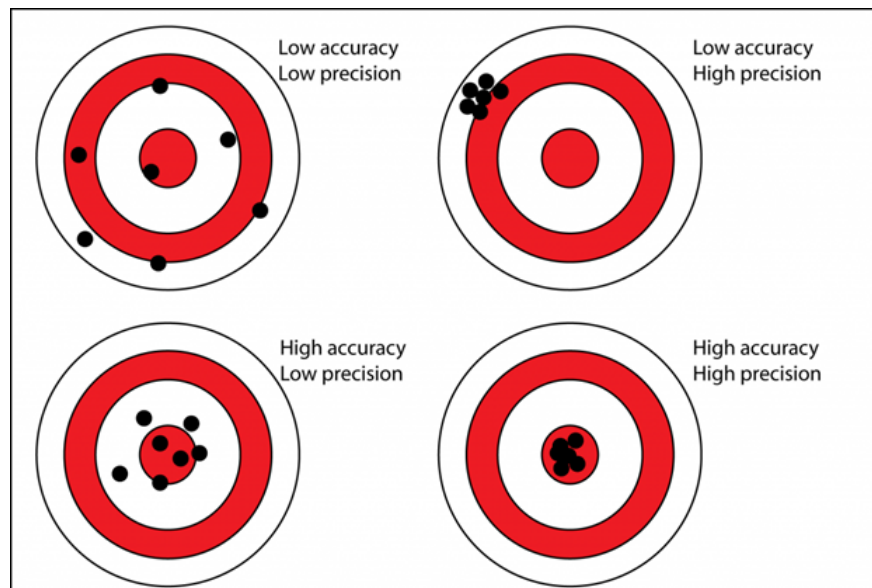


Fig. 1. Precision versus accuracy.

Sources of errors [Textbox 2.]

Recognized sources of error in morphometry include three broad classes of observational errors:

1. Random errors, which occur irregularly and hence are unpredictable. Such errors arise in three different ways: random oscillations of the apparatus, mechanical vibrations, and minor positional changes of the subject at every single measurement. This type of error results in dissimilar outcomes, which can be detected by replicated observations. Random error primarily affects precision.
2. Systematic errors, which can be subdivided into a) observational error, which arises from an individual's bias, unclear description of measuring procedures, lack of proper setting of the equipment, or false data recording due to parallax errors (Seifert, 2002); b) instrumental error caused by factors such as imperfect calibration, etc., and c) environmental error that can be ascribed to the effects of the external conditions on the measurements, e.g., temperature, illumination, etc. Systematic errors primarily influence a measurement's accuracy, but these sources are predictable.
3. Gross errors, arising from false readings, mistakes in recording data by an observer (e.g., reading or recording 88 instead of 38), or mistakenly set magnification. This type of error seriously affects both precision and accuracy. This source of error can be eliminated by careful reading or recording. This type of error can also be recognized post hoc via comparing the repeated measurements in a pairwise matrix scatterplot (Baur & Leuenberger, 2011).

Materials and Methods

The research objects –

As an ideal stress-test basis for evaluating repeatability of morphometric studies in insect systematic research, we selected ten specimens each of a cryptic species pair, *Nesomyrmex devius* (Csősz & Fisher, 2016) and *N. hirtellus* (Csősz & Fisher, 2016), for a total of twenty ant specimens. Every trait under observation shows overlapping ranges (Seifert 2009); thus, these species can be classified in multivariate fashion only. Today,

cryptic species pairs are considered the most difficult cases and pose extraordinary challenges to systematic biology.

The material is deposited in the California Academy of Sciences, San Francisco, California, U.S.A. The full list of material morphometrically examined in this work is listed in Supplementary Table S1 (available on Dryad at <https://doi.org/10.5061/dryad.q83bk3jfq>). Because two specimens suffered a certain degree of damage during the projects due to consecutive postal shipments, making the subsequent gaugers unable to measure them, final analyses were done on only 18 individuals. The ant specimens used in this study comply with the regulations for export and exchange of research samples outlined in the Convention on Biological Diversity and the Convention on International Trade in Endangered Species of Wild Fauna and Flora. For field work conducted in Madagascar, permits to research, collect, and export ants were obtained from the Ministry of Environment and Forest as part of an ongoing collaboration between the California Academy of Sciences and the Ministry of Environment and Forest, Madagascar National Parks and Parc Botanique et Zoologique de Tsimbazaza (Approval Numbers: N° 0142N/EA03/MG02, N° 340N-EV10/MG04, N° 69 du 07/04/06, N° 065N-EA05/MG11, N° 047N-EA05/MG11, N° 083N-A03/MG05, N° 206 MINENVEF/SG/DGEF/DPB/SCBLF, N° 0324N/EA12/MG03, N° 100 1fEF/SG/DGEF/DADF/SCBF, N° 0379N/EA11/MG02, N° 200N/EA05/MG02). Authorization for export was provided by the Director of Natural Resources.

Gaugers –

We addressed the question of whether or not the morphometric measurements performed by eleven gaugers (“measurers”) could be considered repeatable based on statistical thresholds. Eleven volunteers from three continents and six countries, who all have different levels of taxonomic training and skill, were asked to perform a pair of measurements on the same set of ant specimens with their own equipment. Eight of the volunteers are myrmecologists and three are non-myrmecologists (two are wasp specialists and one is a dipterologist). The wide range of the observers’ morphometric skills and the different levels of laboratory facilities and equipment, especially the types of microscopes used, provided an overview of morphometric reproducibility as it works in daily practice. Data belonging to gaugers appear anonymously in this paper, but in order to provide the most important information regarding their skills and their equipment’s quality, gaugers are coded in triad format as follows: expertise in field, estimated total number of specimens measured in their career, and the maximum magnification of the microscope used in the present study separated by underscores (e.g. MYRM_9000_100x).

The morphometric character recording protocol –

Gaugers were asked to measure 21 continuous morphometric characters in each specimen twice in order to collect data for testing both intra-gauger error, equivalent to repeatability, and inter-gauger error rate, equivalent to reproducibility. Every gauger was provided the same measurement protocol, including visual and verbatim trait definitions to follow (Fig. 2 and Table 1). The protocol was assembled based on an existing set of characters used in published papers (Seifert, 2006, 2018; Csősz & Fisher, 2016; Schlick-Steiner et al., 2006; Wagner et al., 2017). In the current work, we addressed the question of to what extent random and systematic errors affect the rate of reproducibility. Therefore, all gaugers were encouraged to eliminate extraordinary differences due to gross error (occurring due to misreading, mistyping or erroneously set magnification) by comparing the values of the repeated observations.

Table 1. Verbatim trait definitions for morphometric character recording.

Trait abbreviation: description	Character definition
CL: Head capsule length	Maximum cephalic length in median line. The head must be carefully tilted t
CW: Width of head including eyes	Maximum width of the head including compound eyes.
CWb: Width of head capsule	Maximum width of head capsule posterior of the eyes.

Trait abbreviation: description	Character definition
EL: Eye length	Maximum diameter of compound eye. All structurally visible ommatidia, pi
FRS: Frontal carinae width	Distance of the frontal carinae immediately caudal of the posterior intersection
ML: Mesosoma length	Measured from caudalmost point of propodeal lobe to transition point between
STPL: Propodeal spine tip erection	Maximum distance from the center of the propodeal stigma to the margin of
MW: mesosoma width	In workers: maximum width the pronotum excluding the pronotal spines.
NOH: Maximum height of the petiolar node	Measured from the uppermost point of the petiolar node perpendicular to a r
NOL: Length of the petiolar node	In lateral view NOL is measured orthogonally from the reference line fitted to
PoOC: Postocular distance	Use a cross-scaled ocular micrometer and adjust the head to the measuring p
PEH: Maximum petiole height	Measured perpendicular to a ventral reference line defined as follows: the cho
PEL: Petiolar length	Diagonal petiolar length in lateral view; measured from the tip of subpetiolar
PEW: Petiole width	Maximum width of petiole in dorsal view. Nodal spines - if any - are not con
PPH: Postpetiole height	Maximum height of the postpetiole in lateral view. Measured perpendicularly
PPL: Postpetiole length	The longest anatomical line that is perpendicular to the posterior margin of t
PPW: Postpetiole width	Postpetiole width. Maximum width of postpetiole in dorsal view.
SL: Scape length	Maximum length of the scape excluding the neck of articulatory condyle.
SPST: Spine length	Distance between the centre of propodeal stigma and spine tip. The stigma c
SPBA: Spine base width	The smallest distance of the lateral margins of the spines at their base. This
SPTI: Propodeal spine tip distance	Distance of propodeal spine tips in dorsal view; if spine tips are rounded or t

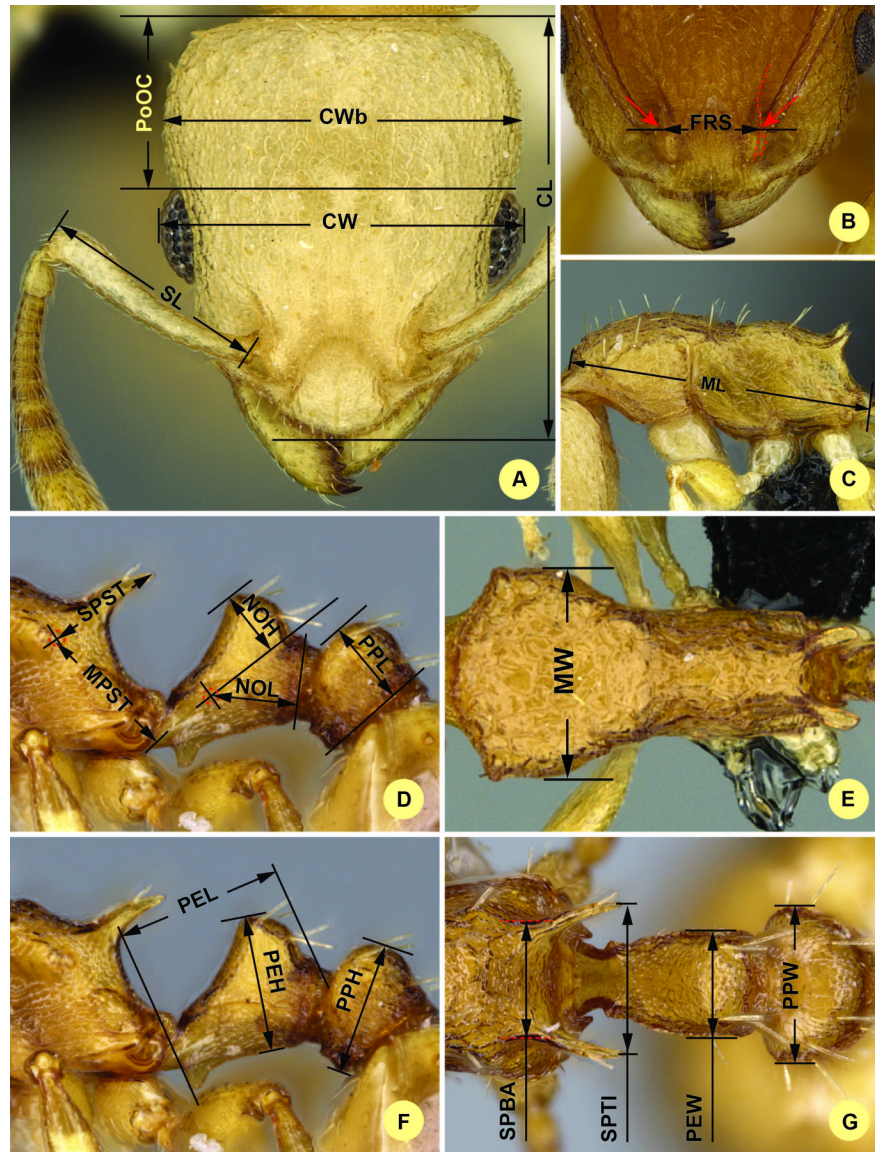


Fig. 2. Illustrations for morphometric characters.

Data analysis –

Distribution patterns of objects (i.e. specimens represented by 21 characters measured by the eleven different gaugers) were displayed in a scatterplot via Principal Component Analysis (PCA; Venables & Ripley, 2002) using a standardization to zero mean and the variance unit (Legendre & Gallagher, 2001). A Permutational Multivariate Analysis of Variance (PERMANOVA) was performed using the Morosita index of dissimilarity with 9999 iterations (Anderson, 2001).

Reliability depends on the magnitude of the error in the measurements to the inherent variability between subjects. These measures of variability can be expressed as standard deviations (SDs). Reliability is defined as a quadratic term of the measured values divided by the sum of the quadratic term of the measured plus the square standard deviation. It is formally described by Bartlett and Frost (2008) as

$$(SD \text{ of subject's true values})^2 = (SD \text{ subjects' true values})^2 + (SD \text{ measurement error})^2.$$

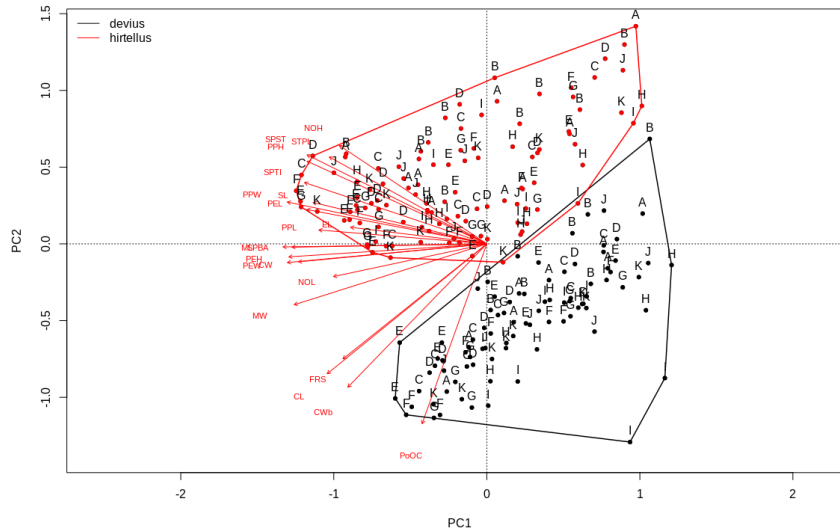
This measure of reliability is also known as intraclass correlation (ICC). If reliability is high, measuring error is small in comparison to the true differences between subjects, so that subjects can be relatively well distinguished (in terms of the quantity being measured) on the basis of the error-prone measurements (Bartlett & Frost, 2008).

To estimate the within-subject SD, we applied a one-way analysis of variance (ANOVA) to model the data containing the repeat measurements made on subjects. In addition, we also tested the effect of the gaugers' expertise and their equipment's performance on the accuracy of ICC estimation by using Spearman's rank correlation. The analyses were carried out in R 3.6.2 (R Core Team, 2019) by using the "Vegan" package (version 2.5-6, Oksanen et al., 2019) for PCA and PERMANOVA and "car package" (version 3.0-7, Fox & Weisberg 2019). Repeatability was calculated for each gauger respectively in order to assess whether the gauger's skills or equipment quality played major roles in measurement consistency.

Results

Agreement in classification between gaugers –

The classification of the 18 pairs of independent observations made by eleven gaugers was successful for the two taxa according to the cumulative PCA analysis that involved all gaugers' observations in the same analysis. Each gauger arrived at the two-species hypothesis with only two misidentified observations (less than 1%) out of the total 198 (Fig. 3a); a single misclassification appeared in two different gaugers respectively. The results of the PCA revealed that the species identity was responsible for the differences based on the morphological traits, and was not ascribed to gauger effect (Fig. 3b). The PCA results were based on the inertia 5489, and the variance explained by the 1st axis was 62.59%, while the variance for the 2nd axis was 10.32%; thus, the overall explained variance was 72.91%. These patterns were also revealed by the PERMANOVA performed using the Morosita index of dissimilarity with 9999 iteration, where the gaugers were shown to have no significant effect on the species identity based on the measured morphological traits ($R^2=0.69$, $p=0.58$).



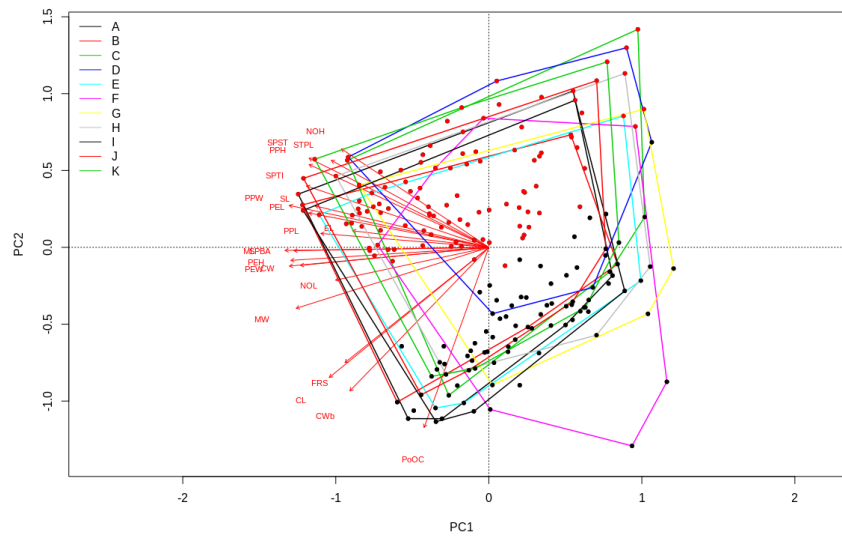


Fig. 3a,b. Ordination biplot for Principal Component Analysis based on (A) species identity and (B) the accuracy of the measurement.

Reproducibility (inter-gauger agreement) –

The Intraclass Correlation Coefficients (ICCs) indicated that the reproducibility of the examined 21 morphometric characters varied between $R=0.872$ and $R=0.471$ (mean $R=0.690$) when the inter-gauger agreement was considered across the 11 gaugers (Table 2). Reproducibility of 16 characters out of the total 21 were acceptable. Five morphometric traits (EL, FRS, NOL, PoOC, PPL) were found to be slightly reproducible, with intraclass correlation coefficient (R) scores between 0.471 to 0.526 (Table 2). These scores belong to physically smaller traits in the observed character pool, hence we examined to what extent absolute character size affects the reproducibility. The general linear model returned no significant correlation ($R = 0.3617$, $p = 0.1071$) between the trait size and ICC scores.

Table 2. Repeatability scores (R) calculated for each character.

character	R	lower bound	upper bound	n	average trait size (μm)
CL	0.8113	0.7050	0.9176	18	648.45
PoOC	0.5292	0.3511	0.7074	18	250.95
CW	0.8073	0.6993	0.9154	18	565.60
CWb	0.6278	0.4629	0.7927	18	540.93
FRS	0.5259	0.3476	0.7043	18	243.82
SL	0.6826	0.5306	0.8346	18	387.69
EL	0.5116	0.3324	0.6907	18	155.03
MW	0.7970	0.6846	0.9095	18	407.84
PEW	0.8434	0.7519	0.9349	18	264.86
PPW	0.7836	0.6655	0.9016	18	300.48
SPBA	0.8724	0.7955	0.9494	18	223.93
SPTI	0.7864	0.6695	0.9032	18	265.74
ML	0.7561	0.6275	0.8847	18	762.78
PEL	0.5951	0.4244	0.7658	18	330.85

character	R	lower bound	upper bound	n	average trait size (μm)
NOL	0.5087	0.3294	0.6880	18	180.08
STPL	0.6159	0.4487	0.7831	18	243.49
PEH	0.7283	0.5900	0.8666	18	249.68
NOH	0.6140	0.4465	0.7815	18	154.87
PPH	0.8169	0.7131	0.9207	18	228.55
SPST	0.8013	0.6907	0.9120	18	217.51
PPL	0.4709	0.2909	0.6508	18	185.64

Repeatability (intra-gauger agreement) –

A geometric mean of intraclass coefficients were calculated for every gauger in order to evaluate their personal performance in association with their skills and equipment quality. The calculated average measure of intraclass correlation coefficients of different gaugers ranged from $R = 0.7840$ to $R = 0.9897$ (Table 3). The non-parametric Spearman's Rank correlation revealed a significant correlation between ICC scores and gaugers' morphometric skills, represented by the estimated number of individuals measured in their personal career ($n = 11$, $R = 0.70$, $t = 2.94$, $p = 0.016$) and a non-significant association between the repeatability parameters and the maximum magnification of the microscope ($n = 11$, $R = 0.56$, $t = 2.03$, $p = 0.073$) applied by the gauger (Fig. 4).

Table 3. Repeatability scores (R) calculated for gaugers.

no.	gauger code	ICC (R)	expertise	magnification	field
1	MYRM_9000_100x	0.93645	9000	100	myrmecologist
2	DIPT_0_100x	0.817519	0	100	dipterologist
3	MYRM_60000_360x	0.989691	60000	360	myrmecologist
4	MYRM_5000_288x	0.971863	5000	288	myrmecologist
5	MYRM_500_50x	0.831945	500	50	myrmecologist
6	MYRM_450_50x	0.903117	450	50	myrmecologist
7	MYRM_500_50x	0.816274	500	50	myrmecologist
8	WASP_1000_230x	0.884679	1000	230	parasitic wasps
9	WASP_0_230x	0.856756	0	230	parasitic wasps
10	MYRM_300_100x	0.784024	300	100	myrmecologist
11	MYRM_300_100x	0.87304	300	100	myrmecologist

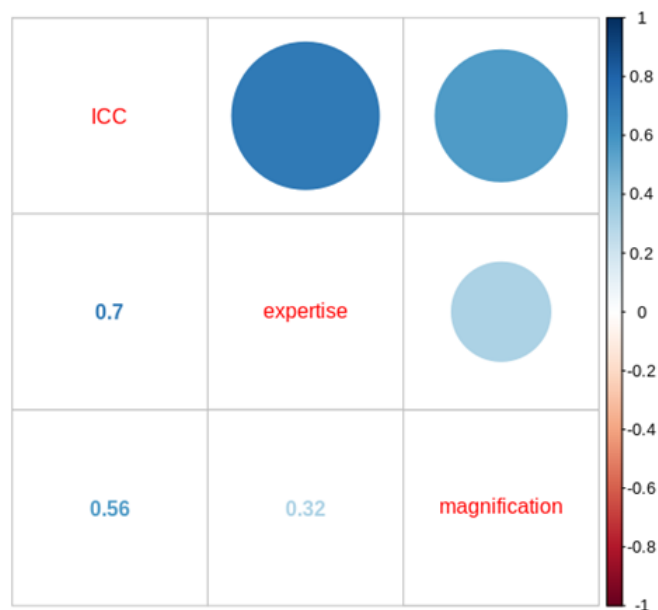


Fig. 4. The correlogram of the studied variables for testing repeatability by the Spearman rank correlation test.

Discussion

Morphometric characters proved reproducible in terms of inter-gauger agreement. The eleven gaugers successfully arrived at the same two-species conclusion despite a great variety of morphometric skills and microscopic equipment of differing quality. The PERMANOVA test revealed no significant gauger effect on the species identity ($R^2 = 0.69$, $p = 0.58$). The ratio of misidentifications on specimen level over all gaugers was only 1.0% within a total of 198 determinations. The non-parametric Spearman's Rank correlation revealed that gauger ICC scores and morphometric skills were significantly correlated, whereas repeatability parameters and maximum magnification used by the gauger were not significantly correlated. These results indicate that both observer experience (Fig. 4) and better optical resolution in microscopes reduces measurement error and increases repeatability (Table 3, Fig. 4).

In analyzing mean intra-gauger agreement character-wise, the mean ICC scores (R) varied between 0.471 in the least reproducible character and 0.872 in the most reproducible character. This rather low average reproducibility may have different causes. One of these may be the absolute physical size of a trait. Traits with smaller sizes tended to have lower ICC scores, but when we tested this with a generalized linear model (GLM) analysis there was no significant correlation between trait size and ICC score. This non-significance may be explained by the rather large minimum trait size (155 μm) in the *Nesomyrmex* test organisms where the given differences in resolution and magnification of the optical systems did not play a major role. The

situation might change dramatically if, for instance, 25- μ m long antennal segments of tiny *Plagiolepis* ants were to be measured. The solution of such a task requires measurement conditions as they were given in the gaugers MYRM_60000_360x and MYRM_5000_288x.

If mean trait size does not contribute much to the rather low ICC scores in the present study, these data are probably better explained by a combination of ten error sources as they were specified for stereomicroscopy by Seifert (2002). It is impossible to analyze which of these caused major disturbances in this study. All observers received verbal and picture-assisted character definitions (see Fig. 2 and Table 1) but were given no further advice or protocols on how to minimize stereomicroscopic measuring errors. Firstly, whether all observers avoided the parallax error is unknown. Secondly, whether all observers used an X-Y-Z-stage for spatial positioning of specimens (see Fig. 1 in Seifert, 2002) and which position stability this stage had are also unknown. In spatial positioning, it is important to place the two endpoints of a measurement in the same visual plane, which is more accurate the lower the depth of focus or the higher the magnification of the optical system. Thirdly, the performance and reliability (e.g., ratchet-step error) of the zoom microscopes used by gaugers in this study are unknown. Fourthly, it is unknown how the observers made their readings (by one tenth of a graduation mark, by entire graduation marks, by digital read-out systems, etc.). A fifth important error source is observer-specific, ambiguous translation of character definitions. These factors highlight the importance of presenting unambiguous character definitions and proposing accurate measurement procedures (see supplementary file SI4, the measuring protocol of the most advanced observer).

To conclude, besides the above-mentioned uncertainties that are common in regular practice in insect taxonomic research, morphometry has proven reproducible in our test setting. The best morphology, we believe, may be done through multi-modal means, such as combining multiple microscopic and morphometric methods (e.g., Richter et al., 2018; Sarnat et al., 2019; Hita-Garcia et al., 2019; Boudinot, 2019; Keklikoglou et al., 2019; Braga et al., 2019). Given the same size range of measured traits, the same range of observers' skill, and the same range of equipment, we expect the same reproducibility for other groups of arthropods, provided these have a similar exoskeleton stability and that specimens belong to a comparable developmental stage. Apart from this, we encourage research teams to replicate this study with taxa of different size classes, such as with tiny parasitic wasps and larger grasshoppers or crickets. The requirements for equipment will change, but we are keen to know if the basic conclusions prove comparable to our results with *Nesomyrmex* ants.

Acknowledgements

First, we would like to thank Michele Esposito, from CASC, for her enduring support with databasing, imaging processing, proofreading, and her overall support in the lab. The fieldwork that provided the specimens used in this study could not have been completed without the help of the Ministry of Environment and Forest, the Madagascar National Parks, and the Parc Botanique et Zoologique de Tsimbazaza, and the gracious support of the Malagasy people. The authors wish to thank Dr. Bethan Davies for her permission to use the graphics in Figure 1.

This study was supported by the National Science Foundation under Grant No. DEB-0072713, DEB-0344731, and DEB-0842395. Finally, SC was partly supported by the Schlinger Fellowship at the California Academy of Sciences.

The ant specimens used in this study comply with the regulations for export and exchange of research samples outlined in the Convention on Biological Diversity and the Convention on International Trade in Endangered Species of Wild Fauna and Flora. For field work conducted in Madagascar, permits to research, collect, and export ants were obtained from the Ministry of Environment and Forest as part of an ongoing collaboration between the California Academy of Sciences and the Ministry of Environment and Forest, Madagascar National Parks and Parc Botanique et Zoologique de Tsimbazaza (Approval Num-

bers: N° 0142N/EA03/MG02, N° 340N-EV10/MG04, N° 69 du 07/04/06, N° 065N-EA05/MG11, N° 047N-EA05/MG11, N° 083N-A03/MG05, N° 206 MINENVEF/SG/DGEF/DPB/SCBLF, N° 0324N/EA12/MG03, N° 100 l\feF/SG/DGEF/DADF/SCBF, N° 0379N/EA11/MG02, N° 200N/EA05/MG02). Authorization for export was provided by the Director of Natural Resources.

Author contribution:

SC, BS, BLF and IM conceived the ideas and designed methodology; MB, MP, JP, JCR, NR, RS and JMU collected the data; ZE and SC analysed the data; SC, IM, BS, BEB, BLF, CT and ZE led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Data availability

Data available from the Dryad Digital Repository <https://doi.org/10.5061/dryad.q83bk3jfq>.

References

- Anderson, A. M., Friis, C., Gratto-Trevor, C. L., Morrison, R. G., Smith, P. A., & Nol, E. (2019). Consistent declines in wing lengths of Calidridine sandpipers suggest a rapid morphometric response to environmental change. *PloS One*, 14(4), e0213930. doi:10.1371/journal.pone.0213930.
- Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26, 32–46. doi:10.1111/j.1442-9993.2001.01070.pp.x.
- Andrew, R. L., Albert, A. Y. K., Renaut, S., Rennison, D. J., Bock, D. G., Vines, T. (2015). Assessing the reproducibility of discriminant function analyses. *PeerJ*, 3:e1137. doi:10.7717/peerj.1137
- Artistotle, 350. *Historia Animalium*, translated by Thompson, D.W., edited by Smith, J.A., Ross, W.D. London, Oxford University Press.
- Baroni Urbani, C. (1998). The number of castes in ants, where major is smaller than minor and queens wear the shield of the soldiers. *Insectes Sociaux*, 45, 315–333. doi:10.1007/s000400050091.
- Bartlett, J. W., Frost, C. (2008). Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound in Obstetrics & Gynecology*, 31, 466–475. doi:10.1002/uog.5256.
- Baur, H., Leuenberger, C. (2011). Analysis of ratios in multivariate morphometry. *Systematic Biology*, 60, 813–825. doi:10.1093/sysbio/syr061.
- Bland, J. M., Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 327, 307–310. doi:10.1016/S0140-6736(03)15270-1.
- Bond, J. E., Beamer, D. A. (2006). A morphometric analysis of mygalomorph spider carapace shape and its efficacy as a phylogenetic character (Araneae). *Invertebrate Systematics*, 20, 1–7. doi:10.1071/IS05041.
- Boudinot, B. E. (2019). Toward phylomics in entomology: Current systematic and evolutionary morphology. *Insect Systematics and Diversity*, 3(6), 1–4. doi:10.1093/isd/ixz019.
- Braga, J., Zimmer, V., Dumoncel, J., Samir, C., de Beer, F., Zanolli, C., Pinto, D., Rohlf, F. J., Grine, F. E. (2019). Efficacy of diffeomorphic surface matching and 3D geometric morphometrics for taxonomic discrimination of Early Pleistocene hominin mandibular molars. *Journal of Human Evolution*, 130, 21–35. doi:10.1016/j.jhevol.2019.01.009.

- Brian, M. V., Brian, A. D. (1949). Observations on the taxonomy of the ants *Myrmica rubra* L. and *M. laevinodis* Nylander (Hymenoptera: Formicidae). Transactions of the Entomological Society of London, 100(14), 393–409.
- Brown, W. L. Jr. (1943). A new metallic ant from the pine barrens of New Jersey. Entomological News, 54, 243–248.
- Chuanromanee, T. S., Cohen, J. I., Ryan, G. L. (2019). Morphological Analysis of Size and Shape (MASS): An integrative software program for morphometric analyses of leaves. Applications in Plant Science, 7(9):e11288. doi:10.1002/aps3.11288.
- Corrucini, R. S. (1988). Morphometric replicability using chords and cartesian coordinates of the same landmarks. Journal of Zoology, 215, 389–394. doi:10.1111/j.1469-7998.1988.tb02847.x.
- Csősz, S., Heinze, J., Mikó, I. (2015). Taxonomic synopsis of the Ponto-Mediterranean ants of *Temnothorax nylanderi* species-group. PLoS One, 10(11), e0140000. doi:10.1371/journal.pone.0140000.
- Csősz, S., Majoros, G. (2009). Ontogenetic origin of mermithogenic *Myrmica* phenotypes (Hymenoptera, Formicidae). Insectes Sociaux, 56, 70–76. doi:10.1007/s00040-008-1040-3.
- DeBiasse, M. B., Ryan, J. F. (2019). Phylotocol: Promoting transparency and overcoming bias in phylogenetics. Systematic Biology, 68, 672–678. doi:10.7287/peerj.preprints.26585v4.
- Esquerré, D., Donnellan, S., Brennan, I. G., Lemmon, A. R., Lemmon, E. M., Zaher, H., Grazziotin, G. G., Keogh, J. S. (2020). Phylogenomics, biogeography and morphometrics reveal rapid phenotypic evolution in pythons after crossing Wallace’s line. Systematic Biology, syaa024. doi:10.1093/sysbio/syaa024.
- Fodor, E., Hâruța, O., Milenković, I., Lyubenova, A., Tziros, G., Keča, N., Slavov, S., Diamandis, S., Kostov, K. (2015). Geometric morphometry of *Phytophthora plurivora sporangia*. Annals of Forest Research, 58, 275–294. doi:10.15287/afr.2015.411.
- Fox, J., Weisberg, S. (2019). *An R Companion to Applied Regression, Third edition*. Sage, Thousand Oaks CA. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.
- Fox, N. S., Veneracion, J. J., Blois, J. L. (2020). Are geometric morphometric analyses replicable? Evaluating landmark measurement error and its impact on extant and fossil *Microtus* classification. Ecology and Evolution, 10, 3260–3275. doi:10.1002/ece3.6063.
- Gotzek, D., Brady, S. G., Kallal, R. J., LaPolla, J. S. (2012). The importance of using multiple approaches for identifying emerging invasive species: the case of the raspberry crazy ant in the United States. PLoS One, 7(9), e45314. doi:10.1371/journal.pone.0045314.
- Helm, B., Albrecht, H. (2000). Human handedness causes directional asymmetry in avian winglength. *Animal Behaviour*, 60, 899–902. doi:10.1006/anbe.2000.1534c.
- Hennig, W. (1950). Grundzüge einer Theorie der phylogenetischen Systematik. Berlin, Deutscher Zentralverlag.
- Hennig, W. (1966). Phylogenetic Systematics, translated by Davis, D., Zangerl, R. Urbana, University of Illinois Press.
- Huelsenbeck J. P. (1998). Systematic bias in phylogenetic analysis: is the Strepsiptera problem solved? Systematic Biology, 47, 519–537.
- Hita-Garcia, F., Lieberman, Z., Audisio, T. L., Liu, C., Economo, E. P. (2019). Revision of the highly-specialized ant genus *Discothyrea* (Hymenoptera: Formicidae) in the Afrotropics with x-ray microtomography and 3D cybertaxonomy. Insect Systematics and Diversity, 3, 1–84.
- <http://www.antarcticglaciers.org/glacial-geology/dating-glacial-sediments-2/precision-and-accuracy-glacial-geology/> accessed 2020-03-05.

<https://www.nist.gov/pml/nist-technical-note-1297/nist-tn-1297-appendix-d1-terminology/> accessed 2020-03-05.

Inäbñit, T., Jochum, A., Kampschulte, M., Martels, G., Ruthensteiner, B., Slapnik, R., Nesselhauf, C., Neubert, E. (2019). An integrative taxonomic study reveals carychiid microsnails of the troglobitic genus *Zospeum* in the Eastern and Dinaric Alps (Gastropoda, Ellobioidea, Carychiinae). *Organisms Diversity and Evolution* 19:135–177.

Johnson, L., Mantle, B.L., Gardner, J.L., Backwell, P.R.Y. (2013). Morphometric measurements of dragonfly wings: the accuracy of pinned, scanned and detached measurement methods. *ZooKeys*, 276:77–84.

Jones, C. J., Edwards, K. J., Castiglione, S., Winfield, M. O., Sala, F., Van der Wiel, C, ... Karp, A. (1998). Reproducibility testing of AFLPs by a network of European laboratories. pp. 191–192 in Karp A., Isaac, P.G. and Ingram D.S. (Eds) *Molecular tools for screening biodiversity*. London, New York, Melbourne, Madras, Chapman & Hall.

Keklikoglou, K., Faulwetter, S., Chatzinikolaou, E., Wils, P., Brecko, J., Kvaček, ... Arvanitidis, C. (2019). Micro-computed tomography for natural history specimens: a handbook for best practice protocols. *European Journal of Taxonomy*, 522:1–55.

Klingenberg, C. P. (2011). MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources*, 11:353–357.

Klingenberg, C. P. (2015). Analyzing fluctuating asymmetry with geometric morphometrics: concepts, methods, and applications. *Symmetry*, 7:843–934.

Legendre, P., Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129:271–280.

Lessells, C. M., Boag, P. T. (1987). Unrepeatable repeatabilities, a common mistake. *Auk*, 104:116–121.

Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, 50:913–925.

Linnaeus, C. (1758). *Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis*. Tomus I. Editio decima, reformata. Holmiae [= Stockholm], Laurentii Salvii.

Lösel, P., Heuveline, V. (2016). Enhancing a diffusion algorithm for 4D image segmentation using local information. *Proceedings of International Society for Optics and Photonics*, 9784:1–11, doi: 10.1117/12.2216202.

Mahendiran, M., Parthiban, M., Azeez, P. A., Nagarajan, R. (2018). In situ measurements of animal morphological features: A non-invasive method. *Methods in Ecology and Evolution*, 9:613–623.

McMullin, R. T., Maloles, J. R., Selva, S. B., Newmaster, S. G. (2018). A synopsis of *Chaenotheca* in North America, including a new species from southern Ontario, *C. selvae*, supported by morphometric analyses. *Botany*, 96:547–553.

Michener, C. D., Sokal, R. R. (1957). A quantitative approach to a problem in classification. *Evolution*, 11:130–162.

Miller, M. I., Priebe, C. E., Qiu, A., Fischl, B., Kolasny, A., Brown, T., ... Buckner R. L. (2009). Collaborative computational anatomy: an MRI morphometry study of the human brain via diffeomorphic metric mapping. *Human Brain Mapping*, 30:2132–2141.

Mutanen, M., Pretorius, E. (2007). Subjective visual evaluation vs. traditional and geometric morphometrics in species delimitation: a comparison of moth genitalia. *Systematic Entomology*, 32:371–386.

Nakagawa, S., Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological Reviews*, 85:935–956.

- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... Wagner, H. (2019). *vegan: Community Ecology Package*. R package version 2.5–6. <https://CRAN.R-project.org/package=vegan>
- Oxnard, C. E. (1983). Multivariate statistics in physical anthropology: testing and interpretation. *Zeitschrift für Morphologie und Anthropologie*, 73:237–278.
- Parins-Fukuchi, C. (2017). Use of continuous traits can improve morphological phylogenetics. *Systematic Biology*, 67:328–339.
- Parins-Fukuchi, C. (2020). Mosaic evolution, preadaptation, and the evolution of evolvability in apes. *Evolution*, 74:297–310.
- Phexell, E., Åkesson, A., Söderberg, M., Bolejko, A. (2019). Intra-and inter-rater reliability in a comparative study of cross-sectional and spiral computed tomography pelvimetry methods. *Acta Radiologica Open*, 8:2058460119855187.
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Randers, J. (2012). 2052 – A Global Forecast for the Next Forty Years. Chelsea Green Publishing, Vermont, 416 pp.
- Remane, A. (1952). Die grundlagen des natürlichen Systems, der vergleichenden Anatomie und der Phylogenetik. Akademische Verlagsgesellschaft Geest und Portig, Leipzig.
- Rensch, B. (1947). Neuere Probleme der Abstammungslehre die Transspezifische Evolution. Ferdinand Enke, Stuttgart. Pp. 407.
- Richter, A., Keller, R. A., Rosumet, F. B., Economo, E. P., Hita Garcia, F., Beutel, R. G. (2019). The cephalic anatomy of the ant species *Wasmannia affinis* (Formicidae, Hymenoptera, Insecta) and its evolutionary implications. *Arthropod Structure & Development*, 49:26–49.
- Ridgway, G. R., Henley, S. M. D., Rohrer, J. D., Scahill, R. I., Warren, J. D., Fox, N. C. (2008). Ten simple rules for reporting voxel-based morphometry studies. *NeuroImage*, 40:149–1435.
- Salganik, M. J. Lundberg, I., Kindel, A. T., Ahearn, C. E., Al-Ghoneim, K., Almaatouq, A., ... Datta, D. et al. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences*, 117:8398–8403.
- Sarnat, E. M., Hita-Garcia, F., Dudley, K., Liu C., Fischer, G., Economo, E. P. (2019). Ready Species One: exploring the use of augmented reality to enhance systematic biology with a revision of Fijian *Strumigenys* (Hymenoptera: Formicidae). *Insect Syst. Diversity* 3:1–43.
- Savriama, Y. (2018). A step-by-step guide for geometric morphometrics of floral symmetry. *Frontiers in Plant Science*, 9, 1433. doi:10.3389/fpls.2018.01433
- Schlick-Steiner, B. C., Seifert, B., Stauffer, C., Christian, E., Crozier, R. H., Steiner, F. M. (2007). Without morphology, cryptic species stay in taxonomic crypsis following discovery. *TREE*, 22:391–392.
- Schlick-Steiner, B. C., Steiner, F. M., Moder, K., Seifert, B., Sanetra, M., Dyreson, E., Stauffer, C., Christian, E. (2006). A multidisciplinary approach reveals cryptic diversity in Western Palearctic *Tetramorium* ants (Hymenoptera: Formicidae). *Molecular Phylogenetics and Evolution*, 40:259–273.
- Schlick-Steiner, B. C., Steiner, F. M., Seifert, B., Stauffer, C., Christian, E., Crozier R. H. (2010). Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual Review of Entomology*, 55:421–438.
- Seifert, B. (1992). A taxonomic revision of the Palaearctic members of the ant subgenus *Lasius* s.str. (Hymenoptera: Formicidae). -*Abh. Ber. Naturkundemus. Görlitz*. 66:1–67.

- Seifert, B. (2002). How to distinguish most similar insect species - improving the stereomicroscopic and mathematical evaluation of external characters by example of ants. *Journal of Applied Entomology*, 126:445–454.
- Seifert, B. (2003). The ant genus *Cardiocondyla* (Insecta: Hymenoptera: Formicidae) - a taxonomic revision of the *C. elegans* , *C. bulgarica* , *C. batesii* , *C. nuda* , *C. shuckardi* , *C. stambuloffii* , *C. wroughtonii* , *C. emeryi* , and *C. minutior* species groups. *Ann. Naturhist. Mus. Wien, B Bot. Zool.* 104:203–338.
- Seifert, B. (2009). Cryptic species in ants (Hymenoptera: Formicidae) revisited: we need a change in the alpha-taxonomic approach. *Myrmecological News*, 12:149–166.
- Shingleton, A. W., Frankino, W. A., Flatt, T., Nijhout, H. F., Emlen, D. J. (2007). Size and shape: the developmental regulation of static allometry in insects. *BioEssays*, 29(6):536–548.
- Sokal, R. R., Sneath, P. H. N. A. (1963). *Principles of Numerical Taxonomy*. San Francisco, W. H. Freeman.
- Steiner, F.,M., Schlick-Steiner, B.,C., Moder, K. (2006). Morphology-based cyber identification engine to identify ants of the *Tetramorium caespitum /impurum* complex (Hymenoptera: Formicidae). *Myrmecologische Nachrichten*, 8:175–180.
- Takacs, P., Vital, Z., Ferincz, Á., Staszny, Á. (2016). Repeatability, reproducibility, separative power and subjectivity of different fish morphometric analysis methods. *PLoS One*, 11:e0157890.
- Taylor, B. N, Kuyatt, C. E. (2001). Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results, [Online]. Available: <http://physics.nist.gov/TN1297> [2020, March 5]. National Institute of Standards and Technology, Gaithersburg, MD. Originally published as Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results, Barry N. Taylor and Chris E. Kuyatt, NIST Technical Note 1297 (1994 Edition).
- Thompson, D. W. (1917). *On Growth and Form*. Cambridge University Press, London.
- Tomiya, S., Meachen, J. A. (2018). Postcranial diversity and recent ecomorphic impoverishment of North American gray wolves. *Biological Letter*, 14:20170613.
- Venables, W. N., Ripley, B. D. (2002). *Modern Applied Statistics with S* . Springer-Verlag.
- Villemant, C., Simbolotti, G., Kenis, M. (2007). Discrimination of *Eubazus* (Hymenoptera, Braconidae) sibling species using geometric morphometrics analysis of wing venation. *Systematic Entomology*, 32:625–634.
- Yoder, M. J., Mikó, I., Seltmann, K. C., Bertone, M. A., Deans, A. R. (2010). A gross anatomy ontology for Hymenoptera. *PLoS ONE*, 5:e15991.
- Ward, P. S. (1999). Systematics, biogeography and host plant associations of the *Pseudomyrmex viduus* group (Hymenoptera: Formicidae), *Triplaris*-and *Tachigali*-inhabiting ants. *Zoological Journal of the Linnean Society*, 126:451–540.
- Wolak, M. E., Fairbairn, D. J., Paulsen, Y. R. (2012). Guidelines for estimating repeatability. *Methods in Ecology and Evolution*, 3:129–137.
- Wright, A. M. (2019). A systematist’s guide to estimating Bayesian phylogenies from morphological data. *Insect Systematics and Diversity*, 3:1–14.
- Yezerinac, S. M., Loughheed, S. C., Handford, P. (1992). Measurement error and morphometric studies: statistical power and observer experience. *Systematic Biology*, 41:471–482.

Table legends

Table 1.

Verbatim trait definitions for morphometric character recording. Abbreviations, definitions and descriptions of morphometric characters are given. This standard protocol was followed by each gauger.

Table 2.

Intraclass correlation coefficients (R), upper and lower bounds, number of cases (n) and average trait sizes are given for each observed characters. Descriptions for abbreviations of morphometric characters are as follows: CL: Head capsule length, CW: Width of head including eyes, CWb: Width of head capsule, FRS: Frontal carinae width ML: Mesosoma length; MW: mesosoma width; NOH: Maximum height of the petiolar node, NOL: Length of the petiolar node, PEH: Maximum petiole height, PEL: Petiolar length, PEW: Petiole width, PoOC: Postocular distance, PPH: Postpetiole height; PPL: Postpetiole length, PPW: Postpetiole width, SL: Scape length, SPBA: Spine base width, SPST: Spine length, SPTI: Propodeal spine tip distance, STPL: Propodeal spine tip erection.

Table 3.

Repeatability scores (R) calculated for gaugers. Gauger information in the table follows this format: experience in an insect group, estimated number of individuals measured in a career, maximum magnification of the microscope used, separated by underscores. ICC: intraclass correlation coefficient calculated from the repeated measurements. The gaugers are aligned according to the sequence of their contribution. Gauger alphabet codes in triad format: A: MYRM_9000_100x, B: DIPT_0_100x, C: MYRM_5000_288x, D: MYRM_60000_360x, E: MYRM_500_50x, F: MYRM_500_50x, G: MYRM_450_50x, H: WASP_1000_230x, I: WASP_0_230x, J: MYRM_300_100x, K: MYRM_300_100x.

Figure captions

Fig. 1.

Precision versus accuracy. The bullseye represents the value of the measurand. Accuracy is indicated by closeness to the bullseye—measurements closer to the bullseye are more accurate. Precise measurements are tightly clustered. Accurate and precise measurements are tightly clustered in the bullseye. Graphics produced and used with permission from Dr. Bethan Davies (antarcticglaciers.org).

Fig. 2.

Illustrations for morphometric characters. Head in dorsal view (a) with measurement lines for CL: Head capsule length, CW: Width of head including eyes, CWb: Width of head capsule, PoOC: Postocular distance and SL: Scape length; frontal region of the head dorsum (b) with measurement lines for FRS: Frontal carinae width (red accessory lines and arrows identify the torular lamella); lateral view of mesosoma (c) with measurement line for ML: Mesosoma length; lateral view of propodeum, petiole, and postpetiole (d) with measurement lines for STPL: Propodeal spine tip erection, NOH: Maximum height of the petiolar node, NOL: Length of the petiolar node, PPL: Postpetiole length, and SPST: Spine length; dorsal view of mesosoma (e) with measurement lines for MW: mesosoma width; lateral view of propodeum, petiole, and postpetiole (f) with measurement lines for PEH: Maximum petiole height, PEL: Petiolar length, and PPH: Postpetiole height; dorsal view of propodeum, petiole, and postpetiole (g) with measurement lines for SPBA: Spine base width, SPTI: Propodeal spine tip distance, PEW: Petiole width, and PPW: Postpetiole width. Detailed verbatim trait definitions for characters are given in Table 1.

Fig. 3a,b.

Ordination biplot for Principal Component Analysis based on (a) species identity and (b) the accuracy of the measurement. Black and red dots represent repeated observations on the same objects, while black dots represent *Nesomyrmex devius*, and red dots represent *N. hirtellus*. Convex hulls for spatial distribution of observations within morphospace represent (a) species and (b) gaugers. Descriptions for abbreviations of morphometric characters (red letters) are as follows: CL: Head capsule length, CW: Width of head including eyes, CWb: Width of head capsule, FRS: Frontal carinae width ML: Mesosoma length; MW: mesosoma width; NOH: Maximum height of the petiolar node, NOL: Length of the petiolar node, PEH: Maximum petiole height, PEL: Petiolar length, PEW: Petiole width, PoOC: Postocular distance, PPH: Postpetiole height; PPL: Postpetiole length, PPW: Postpetiole width, SL: Scape length, SPBA: Spine base width, SPST: Spine length,

SPTI: Propodeal spine tip distance, STPL: Propodeal spine tip erection. Gauger alphabet codes (B) in triad format: A: MYRM_9000_100x, B: DIPT_0_100x, C: MYRM_5000_288x, D: MYRM_60000_360x, E: MYRM_500_50x, F: MYRM_500_50x, G: MYRM_450_50x, H: WASP_1000_230x, I: WASP_0_230x, J: MYRM_300_100x, K: MYRM_300_100x. Compositional differences between treatments expressed as the results of the PERMANOVA (coefficient of determination, F and p values, details in the text).

Fig. 4.

The correlogram of the studied variables for testing repeatability by the Spearman rank correlation test. The size of each bubble is proportional to the estimated correlation value (r). The heat chart on the right shows the color correspondence for r values.