Ecomorphological Variation in *Trithemis* (Odonata, Libellulidae) Dragonfly Wings Reconsidered

Norman MacLeod (✉ nmacleod@nju.edu.cn)

Nanjing University <https://orcid.org/0000-0002-7169-5036>

Benjamin Price

Natural History Museum

Zachary Stevens

Cardiff University

Research article

Keywords: morphology, ecology, geometric morphometrics, machine learning, convolution neural network

Posted Date: November 11th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-962784/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The phylogenetic ecology and wing ecomorphology of the Afro-Asian dragonfly genus *Trithemis* have been investigated previously. Curiously, results reported for the forewing and hindwing shape variation in the latter were, in some ways, at odds with expectations given the mapping of landscape and water-body preferences over the *Trithemis* cladogram. To confirm these results we conducted a wing-shape investigation of 27 *Trithemis* species that employed a robust statistical test for phylogenetic covariation, more comprehensive representation of *Trithemis* wing morphology and a wider range of morphometric data-analysis procedures. Contrary to results published previously, statistical comparisons of forewing and hindwing mean shapes with the *Trithemis* cladogram revealed no statistically significant pattern of phylogenetic covariation. Moreover, landmark-based and image-based geometric morphometric analysis results, as well as embedded image-contrast deep learning analysis results, all demonstrated that both wings exhibit substantial convergent wing-shape similarities among, and differences between, species that inhabit open and forested landscapes and species that hunt over temporary/standing or running water bodies. Geometric morphometric data and data-analysis methods yielded the worst performance in identifying wing shape distinctions between *Trithemis* habitat guilds and the direct analysis of wing images using an embedded, image-contrast, convolution (deep learning) neural network delivered the best performance. Bootstrap and jackknife tests confirmed that our results are not artifacts of overtrained discriminant systems or the “curse of dimensionality”. In addition to our conclusions pertaining to *Trithemis* ecomorphology, the discrepancy between the previous investigation’s results and ours appears to reflect decisions made with regard to the manner in which complex morphological structures are sampled and analyzed. Naturally, results and interpretations of patterns in morphometric data pertain only to the data collected, not necessarily to other aspects of the structures from which those data were collected. For samples of morphologically similar taxa, landmark-based sampling strategies may be effective provided a sufficient number of landmark points distributed across all structures of potential interest exist. However, in a large number of instances analysis of full digital images of the structures under consideration may prove to be a more robust and effective sampling strategy, especially when coupled with analysis via machine learning procedures.

Introduction

The interplay between the effects of phylogenetic relations among species, and the role of the environment in shaping the range of morphologies we observe in nature, have been subjects of perennial interest to evolutionary biologists. That both factors have affected biodiversity in the distant past, as well as in the present day, is beyond question. But appreciating the extent to which either factor has exerted dominance over the other – whether the forms we observe in nature are the result of phylogeny’s “pull” or the environment’s “push” – is an issue that can never be resolved fully. It must be considered on a case-by-case, and a feature-by-feature, basis.

Prior to the advent of phylogenetic systematics and the revolution wrought by the introduction of molecular data, the environment was considered to have the upper hand in this contest. Even into the

distant reaches of antiquity, whenever a species that exhibited a novel combination of morphological characteristics was discovered, the first question most naturalists asked was what environmental problem the structure was used to solve; what was its purpose?

In their influential 1979 essay, Stephen Jay Gould and Richard Lewontin [3] christened this point-of-view the “adaptationist programme”; the idea that the environment, through the agency of natural selection, optimized all aspects of a species’ morphology, physiology, behavior, etc. continuously and universally for the conditions present in the environment(s) at the time those species existed. These authors went on to review several outlandish examples of how previous generations of evolutionary biologists had tended to constrain their interpretations of comparative morphology to fit the precepts of this adaptationist programme, often in the absence of any supporting data. As an alternative, Gould and Lewontin offered a model that reinforced the role of phylogenetic history in the explanatory narrative. Of course, species must meet the challenges posed by their local habitats successfully. But did this really, or always, mean that every aspect of each species’ biology needed to fulfill some specific functional role(s) at all times throughout its evolutionary history? Or should species morphologies be regarded as constituting integrated Bauplänes, constrained by phylogeny and phyletic inheritance, in which some features might be either neutrally adaptive or mildly non-adaptive in terms of developmental efficiency, but maintained nonetheless by virtue of their phylogenetic heritage (as a character combination that did serve an adaptive purpose in some distant ancestor) or as by-products of a larger genetic/developmental/epigenetic system that serves a larger adaptive purpose? In raising these alternative explanations, Gould and Lewontin did not resolve the question of how best to interpret morphological structures. They merely extended the range of potential interpretations, and so the range and types of evidence that might be brought to bear on this question.

Today, we have much better conceptual and analytic tools that can be used to address this issue, along with much better ways of estimating degrees of phylogenetic relations within groups of species. In particular, Joe Felsenstein and later colleague’s work on the biological comparative method [4–9] have addressed many long-standing statistical difficulties in attempting to analyze patterns of similarity and difference across groups of taxa that are embedded within a network of ancestry and descent such that their morphological manifestations cannot be assumed to be independent of one another completely. While work remains to be done in this area (e.g., [10]), many outstanding problems that have complicated, and in many cases, compromised, the research of previous generations have been overcome.

In addition, new tools have become available for undertaking quantitative analyses of morphological structures. The geometric morphometric (GM) approach, which has done so much to encourage the quantitative analysis of morphological data and place these on a firm geometric footing, is now over 30 years old [11–16]. As such, it can no longer be described as a “new” approach to morphological analysis [17,18]. While GM remains a very valuable set of tools, procedures and standards for testing certain types of morphological hypotheses, its core principle — that patterns of morphological variation be described via reference to sparse sets of landmark and/or semilandmark locations usually selected at the outset of an investigation — requires that the features of greatest morphological interest be known at the outset of

investigation. The geometries of these features must also be capable of being represented adequately by a small number of two-dimensional (2D) or three-dimensional (3D) point coordinates, and these point-coordinate locations should, ideally, be (1) distributed more-or-less evenly across the morphologies or structures in question, (2) able to be located unambiguously at positions that correspond to true topological homologies and (3) able to be located on every specimen in the sample. So long as general morphological comparisons are being made across a set of well-preserved and morphologically similar species, this approach works well. However, as the taxonomic scope and/or spatial detail of a morphological investigation increases, the ability of samples to meet these rather stringent requirements often decreases, resulting in a concomitant decrease in the power, and so the appropriateness, of GM-style analyses.

Over the last two decades a completely novel approach to the discovery of patterns of morphological variation has become increasingly popular in the form of machine learning (ML) algorithms. While this approach has its conceptual origins in regression-based procedures that would be familiar to any GM practitioner (e.g., linear regression, principal component analysis or PCA, linear discriminant analysis), its core algorithms have been incorporated into such complex data-analysis system designs that their regression-based origins have become obscured to many casual users. Nevertheless, it is this complexity that gives ML algorithms their extraordinary power; a power that has provoked both amazement and, in certain areas, no small amount of concern. But despite the broad and convincing demonstration that ML-based approaches can sense and identify structured patterns of variation in a very wide range of data types, and in a very wide range of data-analysis contexts, systematic biologists have been slow to avail themselves of these new tools.

As pattern identifiers par excellence, ML algorithms are particularly well suited to deliver the sorts of analyses GM-style approaches struggle to provide. One prominent example of this is the exploratory search for morphological differences between a priori-defined taxonomic groups, especially in those (common) instances where there is little consensus among experts as to which sets of morphological features carry the greatest proportion of group-diagnostic signal(s). This is the problem so-called “deep learning”, or convolution (artificial) neural networks (CNNs), were designed originally to address [19–24]. Their level of performance in this area — as estimated from standardized, human expert-validated image-challenge datasets — is what has been responsible for the recent renaissance of interest in neural network designs [25].

Insect wing morphology has long been recognized as having great potential for exploring and illustrating the advantages of quantitative morphological analysis in both taxonomic and ecological contexts. Some of the first publications illustrating the use of GM methods took insect wing morphology as their subject (e.g., [26–28]). Insect wings are complex structures for which an extensive descriptive nomenclature has been developed [28,29]. In particular, the intersections or nodes of insect wing veins represent classic, Type 1 landmarks [13], and many wing vein and vein intersections can be located unambiguously across a surprisingly broad range of species. This observation lends support to the widely-held belief among entomologists that wings constitute biological homologues for insects as a whole [30].

Morphometric, GM-style approaches to the analysis of variation in insect-wing morphology have been used to identify species (e.g., [25,26,32]), populations (e.g., [33]) and even sexes [34]. While much remains to be learned about the relation between wing shape and aerodynamic function [35], good mechanistic and observational evidence indicates that insect wings of different shapes and internal structural arrangements are associated with different aerial capabilities [37]. This suggests that, like the wings of birds [37,38] and bats [39], insect wing form might reflect aspects of a species' preferred environment, a suggestion for which there is already a limited amount of positive evidence (see [35]).

Unfortunately, the morphological complexities of insect wings also pose a number of challenges for quantitative morphological analysis. The number of vein domains (sensu [40]; see also [41]) contained in many insect wings makes it difficult to decide how to characterize wing morphology along with the level of detail required to resolve particular morphology-related hypotheses. It is, of course, always tempting to employ as much data as can be collected in efforts to resolve outstanding controversies and/or develop comprehensive summaries of morphological trends among taxonomically and/or ecologically diverse data sets. Nevertheless, the well-known "curse of dimensionality" often renders such datasets – in which the number of variables greatly exceeds the number of samples or specimens available – difficult to analyze (see [42]), especially when the task is to achieve reliable between-groups discrimination ([43], but see [44,45] and, more recently, [46,47]). Related to this question are the perennial concerns of whether it is better to focus analysis on the locations of landmark configurations that represent aspects of the wing's internal morphology or on the geometry of the wing outline. In addition, there is the question what to do about patterns of coloration that are an intrinsic part of the morphologies of many insect wings. While color-banding, striping or spotting may have no effect on flight performance, these aspects of wing morphology may have significant species-characterization/identification, ecological and/or behavioral roles that may influence aspects of species' distribution.

The difficulties raised by these considerations are compounded in many studies by the simple fact that, at the outset of an investigation, many researchers have little idea which aspect(s) of the morphologies under study are best suited to resolving particular questions concerning the association between morphological variation and particular functional role(s), or the lack thereof. Yet, the outcome of any hypothesis test involving the quantitative representations of morphology is dependent entirely on the appropriateness of decisions taken regarding which aspects of the morphologies under study to collect and analyze. If a poor choice is made, and a negative result obtained, is it appropriate to conclude that the structure or character complex in question does not exhibit the pattern of variation predicted by the hypothesis test? Or could it be the case that the data extracted from the structure or character complex in question does not happen to exhibit the pattern of variation predicted by the hypothesis test whereas other data that could be extracted from the same specimen set and/or same structure, might?

In 2013 Outomuro et al. [2] published a study that compared wing shape and habitat variation in dragonflies belonging to the Afro-Asian genus *Trithemis*. Species assigned to this genus occur in a variety of ecological habitats with some preferring forests and others open country [1]. Similarly, some *Trithemis* species prefer to hunt in the vicinity of permanent running streams whereas others are found typically

around temporary or standing pools [1]. This study also compared sexual dimorphism in wing shape across the species included in the sample. These authors found (1.) no significant association between wing shape and water body type after correcting for (but not testing the significance of) phylogenetic covariation, (2.) a contrast between forewing and hindwing shape in terms of the ability of these structures to reflect landscape type, and (3.) a distinct difference among males and females belonging to the same species. Accordingly, it was agreed that “natural and sexual selection are acting partially independently on [*Trithemis*] forewings and hindwings and with differences between the sexes, despite evidence for phenotypic correlation of wing shape between males and females” ([1], p. 1866).

We find nothing wrong or amiss with any of these conclusions. Nevertheless, we wonder whether this represents the whole of the story embodied by *Trithemis* wing morphologies. Outomuro et al. [2] employed a decidedly non-standard manner of representing both forewing and hindwing shapes ([2], Fig. 1), mandated, it seems, by having collected these data from specimens in which the outlines of the forewings and hindwings overlapped. This may have prevented precise location of landmarks along the trailing, proximal edges of the forewing. In addition the location of forewing landmarks was quite uneven with the morphology of the trailing edge being sampled more intensively than that of the leading edge and the wing tip being sampled much more intensively than the proximal wing margin. As for the hindwing, again, the morphology of the trailing edge was sampled much more intensively than that of the leading edge and the wing tip more intensively than the wing's basal curve. Use of sparse sets of landmarks distributed unevenly across the wing would result in the characterization of only portions of each wings' overall geometry and the differentially weighting of variation in those parts of the wings sampled more intensively relative to those sampled using fewer landmarks.

In addition to this, Outomuro et al. [2] appear to have used an odd scheme of angles to locate peripheral landmarks in the anal region of the hindwing's trailing edge, with landmarks located at 10°, 27°, 54°, 73°, 81° and 90° from a chord drawn between the wing's anterior thoracic articulation and the intersection between the nodus and the wing periphery (at least, as shown in [2], Fig. 1). Other than this figure, no definitions of either the forewing or the hindwing landmarks were provided though, in all cases other than those noted above, the landmarks appear to have been placed at the intersections between major wing veins (e.g., costa, subcosta, first radius), or pterostigma (see [2], Fig. 1), and the wing periphery. Standard GM practice would be to represent each wing's outline shape using a set of (initially) equally spaced semilandmark points from a common starting landmark in order to ensure consistent resolution across the dataset and across all forms comprising the sample.

Given these aspects of the Outomuro et al. [2] analysis, in addition to the fact that no internal landmarks were employed in the characterization of wing morphology, we feel important aspects of the relation between *Trithemis* wing form and landscape and water body type guilds may not have been included in this previous analysis. Again, we stress that we are not challenging the results obtained by the Outomuro research team and presented in their 2013 report. We are quite certain those are correct for the data they chose to collect, for the specimens they chose to collect those data from, and given the manner in which they chose to analyze those data. But there are many ways to characterize complex morphologies and

results obtained under one characterization scheme cannot necessarily be assumed to cover all such schemes. Accordingly, we chose to determine (1) whether we could reproduce the findings of Outomuro et al. [2] in terms of the relation between wing shape and both landscape and water body preference using different form-characterization standards and different data-analysis methods in an effort to determine whether there was anything more this dragonfly genus had to tell us about the relation between wing morphology, ecology and natural selection. This comparison was also undertaken to (2) clarify the range of options available to quantitative morphologists interested in confronting similar problems in other taxonomic groups and (3) to provide a relative assessment of the power of different data types and data-analysis methods when used in similar ecomorphological contexts.

Methods

Materials

Trithemis is a large genus (> 40 species) of mainly African dragonflies referred to commonly as “dropwings” owing to their habit of holding their wings at a negative angle to their bodies, rather than horizontally, when at rest. In this investigation 27 *Trithemis* species (Table 1) were used, all of which were sourced from the insect collections of The Natural History Museum (London). Images of both forewings and hindwings from males and females of each species were collected using a low-magnification, digital SLR-based photo-microscopy system supplied by the museum. Plates 1 and 2 are composed of the photographs of the wings of each species’ forewings and hindwings, respectively. All images were taken from mounted specimens and a universal stage was employed to correct the orientation of each specimen prior to imaging so that the wing surface was normal to the optic axis of the microscope.

During photography every effort was made to block out the specimen label, which was impaled on the mounting pin beneath each specimen, by placing a white card over the label prior to image capture. In the case of some specimens, this operation could not be performed without risking damage to the specimen. Photomicrographs of these specimens were collected and used in the collection of GM data where imperfections in the image were not relevant to the collection of landmark or semilandmark data. However, owing to the writing and specimen label border artifacts that were included in the images of the specimens for which the label could not be blocked, these specimens were excluded from the image dataset, both for the linear discriminant and ML analyses of wing images. Preferred landscape and water body assignments for these species were made according to the ecological information provided by [1] updated for a few species by this report’s second author.

Methods

Image processing

All forewing and hindwing images were segmented from dorsal view, whole-specimen images, mounted in a variably sized image frame against a flat white background, converted from 8-bit RGB color to 8-bit greyscale format and adjusted for consistent average brightness and contrast values. In all cases the two

pairs of wings present on each individual were inspected and the best preserved/imaged forewing and hindwing set selected to represent the specimen. In those cases where the best preserved/imaged wing was collected from the body's right side the wing image was mirrored to the left-side orientation to render the wing dataset comparable in pose across all species. Once these processing and pose-standardization procedures had been carried out the processed wing images were written out to separate image files in the non-compressed TIFF file format to form an archive of *Trithemis* forewing and hindwing images. Plates 1 and 2 were assembled from these archive images. Copies of the original and processed images are also available in the Supplementary Material that accompanies this article.

Geometric morphometric analysis

In order to compare our *Trithemis* ecomorphological wing shape results to those of Outomuro et al. [2] a GM-style morphometric analysis was carried out on a combined landmark-semilandmark dataset that included a set of internal vein-node landmarks as well as peripheral outline landmarks and semilandmarks. Figure 2 illustrates the positions of these landmark and semilandmark point locations on a representative set of *Trithemis annulata* wings. One advantage of working with a group whose species exhibit such similar forewing and hindwing morphologies is that the same landmark and semilandmark points could be located on both the forewings and hindwings of every specimen in the dataset. This ensured comparable geometric coverage thus facilitating comparisons across wing types as well as across habitat groups. Use of internal landmark, peripheral outline landmark and peripheral outline semilandmark point locations to characterize wing morphologies also ensured accurate and consistent representation of localized morphological similarities and differences across species. The inclusion of internal landmarks added additional information to the analysis and assisted in making the collection of geometric information across the wing forms as even as possible.

In all, 13 landmarks located at the origins, intersections or peripheral termini of major wing veins (see S3 for landmark definitions), and 25 semilandmarks located in five different peripheral outline zones defined by landmarks 1, 6, 13, 20, 25 and 31, were used to represent wing form in the GM-style analysis. The number and location of peripheral semilandmarks were determined using the extended eigenshape protocol of MacLeod [48,49]; see also [50]) which allows the sample itself to determine how many equally-spaced semilandmarks were required to represent the geometry of outline zone peripheries to a consistent level of geometric accuracy across all specimens. Under the sampling scheme selected for this investigation all zone peripheries were represented to an accuracy of greater than 95 percent. This procedure usually (1) results in a reduction in the number of semilandmark points required to represent form outlines accurately (thus minimizing dataset dimensionality), (2) ensures consistent geometric converge of all portions of the peripheral outline, (3) allows more "shape rich" aspects of the outline to have a greater influence in data-analysis results and (4) minimizes the artificial inflation of form/shape similarity estimates that results from oversampling geometrically simple peripheral outline regions.

Following collection of these data, forewing and hindwing landmark-semilandmark configurations were aligned and scaled using the generalized least-squares Procrustes procedure [51]. The aligned shape coordinates were then used to produce the species-specific mean shape configurations that were

employed in the test for phylogenetic covariation in wing shapes, against the *Trithemis* ultrametric tree provided by Damm et al. ([1]; Fig. 3). The multivariate generalization (K_{mult}) of the K statistic (see [53]) described by Adams [54] was employed to test for phylogenetic covariation in the wing-shape datasets.

In order to determine whether the shapes of *Trithemis* forewings and/or hindwings, as represented by these landmark-semilandmark shape configurations, exhibit consistent and statistically significant differences in shape between species found typically in contrasting landscape and water body habitats, the dimensionality of the landmark-semilandmark shape-coordinate data was first reduced by subjecting the Procrustes-aligned point-coordinate data to a covariance-based PCA. Component scores on the set of eigenvectors sufficient to account for 95 percent of the pooled-sample shape variation were retained and submitted to secondary canonical variates analysis (CVA) using the landscape and water body habitat assignments (successively) as grouping variables. Projections of the PC configuration-shape scores onto the single linear discriminant vector resulting from this analysis enabled visualization of the degree to which shape distinctions existed among our *Trithemis* forewing and hindwing shape configurations as represented by our landmark-semilandmark data. A number of recent authors in various natural history, ML, and archeological fields have employed a combined PCA-CVA approach similar that used in this investigation to facilitate the analysis of group separations in a linear multivariate context (e.g., [44,45,55–57]). Recently Rohlf [58] reviewed data-analysis strategies for coping with high-dimensional data in group-discrimination contexts and identified this PCA-CVA technique as one that can possibly circumvent the “curse of dimensionality” issue.

A detailed geometric interpretation of the between-groups shape distinctions discovered by this procedure was created by back-projecting the CV scores into the PCA space and then back-projecting those coordinate positions into the space of the original shape variables (see [59] for a description of this technique). The last step of this procedure involved testing the statistical significance of the observed difference in mean vector orientations for the landscape and water body groups using a bootstrapped version of the standard Hotelling’s T^2 test [60–62].

Direct analysis of images

In order to compare the GM-style analysis of wing morphology as represented by a sparse set of landmarks and semilandmarks with a mathematically equivalent direct analysis of wing images, subsets of these same forewing ($n = 217$) and hindwing ($n = 227$) images that did not include labels in the image frame were processed to standardize their frame sizes, image sizes, orientations, and pixel color scales in order to render their images geometrically comparable. This processing operation also involved a reduction in the overall sizes of the image frames in order to reduce pixel redundancies, boost the images’ geometric information content, and perform an initial reduction in the image datasets’ dimensionality. After processing, the forewing images all occupied the central region of a 200 x 56 pixel image frame and the hindwing images a 200 x 81 pixel image frame. For both image sets the background pixels (outside the wing peripheries) were set to white.

Despite the high level of resolution reduction entailed by this procedure, all taxonomically critical aspects of wing morphology remained clearly visible on the processed images including the forms of the wing outline, all major wing veins and the size, location and intensity of the colored areas (e.g., distal pterostigma, see Fig. 1; proximal darkly pigmented hindwing regions of *T. tropicana* and *T. kirbyi*, the lightly pigmented regions of *T. annulata* and *T. bredoi*, see Plate 2). Once the greyscale pixel brightness values had been exported and reformatted into a data matrix these were submitted to the same PCA-CVA-based data-analysis procedure employed in the GM-style analysis to facilitate direct comparison with the landmark-semilandmark morphology-characterization results.

Machine learning analysis

In order to determine whether morphological distinctions between habitat categories could be improved and/or clarified by adopting a non-linear style of discriminant analysis, a “deep learning” convolution neural network (CNN) was employed to analyze the image datasets directly. This CNN architecture was based on the LeNet-5 system [22,24], which is arguably, the CNN that sparked initial interest in “deep learning” using convolution-based, multi-layer artificial neural networks. The LeNet-5 architecture achieved 98.5% accuracy when tested on the 10,000 test images included in the 70,000-image Modified Nation Institute of Standards and Technology (MNIST) image database (see <http://yann.lecun.com/exdb/mnist/>) after being trained on the remaining 60,000 28 x 28 pixel digital images.

All CNNs consist of an input layer that receives the information to be processed (in our case images) and an output layer that makes the final allocation of the processed data into one of a number of categories or classes. Between these a variable number of connected or “hidden” layers exist that process the data by (1) accepting the information from the input or previous layers, (2) evaluating this information for patterns that are consistent with those established by a previously identified training set that have been allocated to their appropriate categories (in our investigation, sex) and (3) passing the processed data on to the next layer. This layered design is used to overcome the problem of full connectivity which is impractical to apply to large images, but can be applied successfully to small images [44,63,64]. For our analysis we adopted the standard LeNet default of autoencoding, or “stepping down”, the input image resolutions to, in our case, a set of 40 x 40, 8-bit grayscale pixel values as an initial processing step.

Although LeNet-5 is but one of several advanced, gradient-descent CNN architectures for image-based automated identification applications (see <https://resources.wolframcloud.com/NeuralNetRepository/>), it remains one of the most efficient, best understood, and most flexible of the CNN architectures available currently. The LeNet architecture also has advantages over more elaborate CNN designs in that their complexity requires (optimally) that they be trained with large numbers of example images in order to avoid the ‘curse of dimensionality’ problem [42]. The LeNet architecture is relative simple – containing only eight processing layers – and so better suited to the analysis of small training sets, especially if only a limited number of group differences are of interest. The overall structure of the LeNet-5 architecture employed in this investigation is listed in Table 2.

One of the most severe limitations of CNN training in many taxonomic and systematic contexts is sample size. Owing to the number of inter-layer weights whose values must be calculated recursively, CNNs are usually trained on datasets whose sizes are vast by systematic-research standards. A training set such as ours, though relatively large by biometric and morphometric standards, would be considered far too small for CNN training by most data scientists. This problem can be circumvented, however, by opting for training as an embedded, distance-based, image-contrast learning system in which the aim is not to learn the characteristics of a priori-defined groups themselves but, rather, implicit patterns of similarities and differences between pairs of images, quantified either by image distances or angular separations, that either do, or do not, belong to the same training group (Fig. 4). Recent applications of this strategy have focused on systems for describing differences between image pairs drawn from large datasets using text-based descriptors [65,66] as well as image-based analyses [34,67–69].

In the analysis of small-to-modestly sized samples, there are many advantages to this approach, insofar as all, or most, pairwise comparisons between images in a dataset can be employed in CNN training. Thus, despite the fact that our samples contained images of only 217 (forewings) and 227 (hindwings) individuals, 46,872 (forewings) and 51,302 (hindwings) informative pairwise comparisons can be drawn from them, including large numbers of both within-group and between-group pairs. By focusing CNN training on differences among images of the same group, and between images of different groups, training can proceed more efficiently, and more comprehensively, than would be possible otherwise.

In order to visualize the trained *Trithemis* feature space the t-distributed stochastic neighbor embedding (t-SNE) algorithm [70,71], was employed to summarize the pattern of wing-morphology similarities and differences in a reduced-dimensional feature space. Other dimensionality reduction procedures and algorithms are available (e.g., PCA, see [70]; UMAP, see [72,73]). However, the t-SNE approach has become a standard dimensionality-reduction technique in many ML contexts and is often now preferred over many longer-established approaches (e.g., PCA, Linear discriminant analysis, multidimensional scaling). Owing to its sensitivity, care must be taken when interpreting t-SNE results, as it is well known that apparent clustering can result, even in cases where there is no structure (e.g., when applied to data derived artificially from a single statistical distribution, [74]). To avoid this issue multiple t-SNE analyses were preformed using a graded sequence of perplexity and iteration settings.

Traditionally, the performance of discriminant functions are tested by evaluating the statistical significance of mean vector separations and by using the trained discrimination system to place members of an independent “validation” set, whose true class identification is known, into inferred groups or classes. With respect to the latter, test class identification accuracies are typically tabulated in a “confusion matrix”.

A 1000-iteration bootstrap variant of Hotelling’s T^2 test [56–62] was used to obtain non-parametric estimates of the statistical significance of the training-set group separations. Stability of the *Trithemis* discriminant space was tested using a leave-one-out jackknifed or cross-validation strategy [75] which was applied to a randomly selected subset of 25 training-set specimens drawn from the full image

datasets. For this procedure the sample size of 25 was selected to balance the need to base the stability/accuracy test on a representative sample of *Trithemis* wing images against the time required to train the embedded image LetNet-5 CNN system on a GPU-enabled computer workstation (c. 1 hour per cross-tabulation iteration).

Results

Phylogenetic signal analyses

The degree to which patterns of shape variation in the *Trithemis* species covary with *Trithemis* phylogenetic relations was examined using the K_{mult} test [54]. This statistic tests the null hypothesis that the degree of morphological shape similarity existing among a set of species constitutes a direct reflection of the structure of phylogenetic relations existing between species as represented by a time-calibrated tree under the assumption of morphological evolution conforming to a Brownian-motion model with an expected value of 0.0 and a variance (σ^2) proportional to the elapsed time since speciation from a common ancestor. As noted by Adams [54], the random, or Brownian, expectation for GM data is derived by calculating the ratio between the observed square of the distance-based deviation of each species' mean shape-coordinate configuration from the phylogenetic mean (= mean square error observed) and the expected square of the distance-based deviation of each species' mean shape-coordinate configuration from the phylogenetic mean along the phylogenetic tree (= mean square error expected). This ratio is then scaled by a representation of the phylogenetic covariance matrix. Values less than, or greater than, 95 percent of the expected distribution of K_{mult} values for a sample reflect patterns of shape variation that were either less than, or greater than, expected under a random phylogeny model (= no phylogenetic signal) respectively. In order to avoid the need to make unsubstantiated assumptions with regard to interactions among variables the expected forms of the K_{mult} distribution conforming to the null hypothesis were estimated via a bootstrap strategy, by permuting the tips of the *Trithemis* tree randomly 1000 times and calculating the expected K_{mult} value for each permuted tree.

Results of the K_{mult} tests for the *Trithemis* forewing and hindwing landmark-semilandmark datasets are shown in Figure 5. In both cases comparison of observed K_{mult} values to those generated as a result of randomizing the phylogeny failed to indicate that these wing shapes contained a statistically significant phylogenetic covariation signal. Given the number of studies demonstrating significant patterns of phylogenetic covariation in a host of morphological, ecological, and behavioral variables, this result might strike some as unexpected. However, most of this extensive literature involves the analysis of single or pairs of metric traits or categorical variables. Owing to the lack of generalized tests for the extent of phylogenetic signals in morphometric data it is presently unknown whether this result would be regarded as common, or unusual. What can be said without doubt, however, is that these GM-based characterizations of *Trithemis* forewing and hindwing morphologies – which are the most comprehensive and detailed collected to date – failed to yield any evidence supporting the presence of phylogenetically structured shape variation.

An analysis of the Procrustes PCA spaces created as a by-product of the phylogenetic signal test both supports and explains the nature of this test result. Figure 6 shows the *Trithemis* cladogram (with hypothetical common ancestors) projected into the linear subspace created by the first two Procrustes principal components for the forewing and hindwing datasets. As expected, hypothetical ancestral shapes (= internal tree nodes) are clustered in the centers of these spaces showing closer correspondence to the phylogenetic mean shape configuration than is typical of the terminal taxa. Projected positions of the terminal branches (= *Trithemis* species) form a “halo” of relatively more extreme shapes surrounding the hypothetical *Trithemis* ancestors. However, the tree branches connecting internal nodes with each other, and with the terminal *Trithemis* species, exhibit a high degree of crossover. This result indicates that species from very different parts of the *Trithemis* phylogeny exhibit similar forewing and hindwing shapes, and that closely related species often exhibit quite different and distinct forewing and hindwing shapes. Thus, there is little evidence in these graphic data for any substantial degree of phylogenetic covariance in the structure of *Trithemis* forewing or hindwing shapes; a morphometric result that supports the results obtained from the K_{mult} statistical phylogenetic signal tests.

In addition to the phylogenetic results displayed in Figure 6, the lack of obvious habitat-based clustering of *Trithemis* species in this (or any other) Procrustes PCA subspace suggests there is little evidence supporting the idea that *Trithemis* wing shapes reflect either landscape or water body habitat preferences. Nevertheless, these analytic results constitute a weak test of this ecomorphological hypothesis insofar as the shape ordinations that result from any PCA are optimized to reflect the directions of maximum variance in the pooled sample, without taking account of any differences that may, or may not, exist between subsidiary groupings. The correct interpretation of these results with respect to testing hypotheses of eco-group distinction is that, if any group-level shape distinctions are present within these data, they are not well aligned with the directions of pooled-sample shape variance. Owing to the nature of the optimizations present in PCA ordination spaces, this somewhat subtle interpretational constraint applies not only to visual inspections of the ordination-space patterns, but also to any statistical test based on PCA score data, no matter how many components are included in such tests. Accordingly, neither inspection of PCA ordination plots such as these, nor analysis of PCA score data, constitutes a sufficient basis on which to conclude that group-level shape distinctions either do, or do not, exist within these *Trithemis* wing morphologies.

Geometric morphometric (GM)-style analyses of landmark-semilandmark datasets

In order to address the issues of whether habitat-based, group-level, wing-shape distinctions do exist within these GM-style data a CVA was performed on the Procrustes principal components scores of wing-shape configurations projected into the set of Procrustes principal components (= shape-covariance eigenvectors) that together accounted for 95 percent of the observed shape variance for both the forewing (13 components) and hindwing (14 components) datasets. Results of these analyses, in the form of histograms of the projected wing-shape positions on the single linear discriminant axis, are presented in Figure 7.

Although it is clear that, for this GM-style representation of *Trithemis* wing shape variation, broad zones of overlap exist between species that prefer open and forested habitats, and among species that prefer to hunt over running, as opposed to temporary or standing, water, a clear distinction does exist between the wing morphologies found typically among these two opposing sets of ecological habitats. Tests of this distinction using a bootstrapped version of Hotelling's multivariate extension of the two-sample t -test — to avoid interpretative constraints imposed by any failure of these data to meet the assumptions of the parametric T^2 test — demonstrate that the shape differences shown in Figure 6 are significant statistically at well beyond the standard $p = 0.95$ confidence level. Since this result cannot be interpreted as an epiphenomenon of phylogenetic covariation (see above), the most reasonable alternative interpretation is that these forewing and hindwing shape differences result from morphological convergence on airfoil designs suited for aerial hunting in these different environments. The fact that these habitat-related shape differences do not account for the major directions of wing shape variation present within the sample suggests that a variety of *Trithemis* forewing and hindwing shapes are viable functionally. But based on these results there is a strong indication that, despite the wide range of wing shape variations present, there does exist a subtle, but definite difference between forest and open habitat-dwelling, and between running and temporary/standing water-hunting, *Trithemis* species' wing morphologies, at least for the species included in our investigation.

Which aspects of *Trithemis* wing morphology are responsible for these habitat-level distinctions? Figure 8 summarizes the wing-geometry changes represented by the landscape and water-body CVA results for *Trithemis* forewing and hindwing landmark configurations. Open landscape species, and species that hunt over standing water bodies, exhibit slightly broader forewings relative to forest-dwelling and running water-hunting counterparts. In both cases this narrowing is achieved via lateral, outboard migration of all costal landmarks and semilandmarks with the displacement magnitude reaching an acme in the middle of the interval between the nodus (landmark 6) and the costal terminus (landmark 13) with these displacement vectors incorporating a subordinate posterior orientation in the vicinity of the wing apex. Displacements of descending nodus vein - radius + media vein vertex (landmark 32), R_3 bifurcation vertex (landmark 33) and distal IR_3 (landmark 34) also follow this general displacement pattern with the latter exhibiting a contrary subordinate anterior orientation. In contrast, landmarks and semilandmarks along the posterior wing margins exhibit inboard-anterior displacement vectors whose magnitudes culminate in the vicinity of the R_3 vein terminus (landmark 20). These displacements have the effect not only of making the wings of forested landscape and running-water species narrower than those of open-dwelling and temporary/standing water-species, but also more uniform in width. In addition, the proximal posterior wing margins of open landscape and temporary/standing water species are more gently curved than forest-dwelling and running water-species which exhibit a more sharply angled anal proximal area or "corner".

With regard to hindwing morphology, the landmark-semilandmark displacement trends are similar to those of the forewing in general, but with a few interesting differences. The hindwings of *Trithemis* species that inhabit forested landscapes and hunt over running waters are slightly narrower and have

more pronounced anal area “corners” compared to species that inhabit open landscapes and hunt over temporary/standing waters. As with the forewings, this shape transformation is accomplished via lateral outboard migration of the costal landmarks and semilandmarks (1–13), lateral inboard and anterior-ward migration of the distal and medial posterior margin landmarks and semilandmarks (14–24), and a pronounced lateral outboard and posterior-ward migration of the posterior-margin landmarks and semilandmarks located in the wing’s anal area (25–29). Internal hindwing landmarks also mimic the displacements seen in their analogous forewing landmarks, for the most part. Nevertheless, differences exist in the orientation of hindwing displacement vectors at the extreme proximal and extreme distal ends of the wing. Whereas landmarks and semilandmarks documenting the *Trithemis* forewing apex (12-14) exhibit a subordinate posterior-ward direction, in the hindwing these topologically homologous landmark points exhibited a distinct-but-subordinate anterior-ward direction. Similarly, whereas the forewing anterior wing-attachment landmarks (1 and 31), along with the closely associated semilandmark (30) displayed a dominant posterior-inboard migration in concert with the anal-area semilandmarks, in the hindwing these landmarks display a pronounced anterior-ward migration in opposition to the anal-area semilandmarks. This transformation further accentuates the proximal width of *Trithemis* hindwings in those species that inhabit forested landscapes and hunt over running waters.

The consistency of the CVA results displayed in figures 7 and 8 raise the question of whether these two habitat-based distinctions in wing shape represent separate, or conjugate, aspects of selection on these dragonfly species. This determination cannot be made from the frequency histograms, but instead requires a regression analysis to determine the level of similarity between landscape and water-body linear discriminant analysis scores for both forewing and hindwing analyses. Results of these tests are presented in Figure 9.

The scatterplots shown in Figure 9 document an abundant degree of variation in the comparison of wing-shape scores in landscape and water body-optimized linear discriminant vectors. But these data also document a well-established and convincing linear trend in these data for both *Trithemis* forewings and hindwings. The null hypothesis for these regression tests is that a strong linear relation does not exist between the projections of wing shape in discriminant spaces optimized for landscape and water-body aspects of the *Trithemis* habitat. In evolutionary terms, this hypothesis amounts to a statement that selection on *Trithemis* wing shape operated in a different manner in its attempt to modify forewing and hindwing shapes for optimal function in forested, as opposed to open, landscapes and above running, as opposed to temporary/standing, waters. Accordingly, it would seem that, despite the fact that some differences are present in the species assigned to different landscape and water body character states, these differences are insufficient to accept the null hypothesis that, for this dataset, the distinctions that exist between the forewing and hindwing forms of species inhabiting forested landscapes and running water bodies, or between species inhabiting open landscapes and temporary/standing water bodies reflect random patterns of shape variation.

Failure to identify a clear distinction between these two habitat categories is, perhaps not surprising in that it may be a by-product of the *Trithemis* samples included in this investigation. All samples came

from museum collections where, it can be assumed, priority was given to collecting specimens that represent the species in question, not necessarily the range of habitats in which members of that species might occur. Species-specific sample sizes are also modest in terms of most eco-morphological investigations (see Table 1). Nevertheless, it is undeniably the case that just under 75 percent of our *Trithemis* species which occur typically in “forested” habitats have also been assigned by Damm et al. [1] to the “running” water-body category and those occurring typically in “open” habitats to the “temporary/standing” water-body category. To the extent that these assignments are correct and accurate, our finding that no wing shape distinction exists between these habitat-pair combinations may be a fully justified overall reflection of *Trithemis* biology. Of course, this result is also tied to the manner in which our landmark-semilandmark data represent *Trithemis* wing morphology. As such, our advice is to regard this result as an accurate description of our sample, but an interpretation of *Trithemis* eco-morphology that is provisional, pending more focused ecomorphological investigation with larger intra-specific samples collected with this specific issue in mind.

The last issue to cover in the presentation of our results from the GM-style analysis of the *Trithemis* landmark-semilandmark dataset is the overall quality of the discriminant partition obtained. Table 3 summarizes the performance of these landmark-semilandmark configuration datasets for the purpose of discriminating between landscape and water-body habitat groups. While previous results did demonstrate that a GM-style approach to *Trithemis* wing-shape characterization was sufficient to document the existence of subtle, but consistent habitat-based shape differences, these data were, on the whole, unable to separate different species based on their landscape or water body habitat preferences with high degrees of accuracy. The degrees of apparent shape overlap between forest and open landscape species, and between running and standing water species, are simply too great to rely on either forewing or hindwing shape alone as an accurate diagnostic indicator of these environmental preferences. Of course, the fact that identification accuracies as high as 76 to 83 percent were able to be delivered by the leave-one-out jackknife analyses is, itself, quite noteworthy. We doubt accuracies this high could be delivered routinely by taxonomic experts for species identifications who had access only to whole-wing images, much less configuration plots of 38 mathematical point locations. Still, overall Matthews Correlation Coefficient [76–78] accuracies of less than 0.75, plus the time and technology required to collect even a single set of landmark-semilandmark coordinates on which to base an identification, suggests that GM-style approaches to ecomorphological analyses, while better than multivariate morphometric analysis or, worse still, nothing at all, may represent, at best, a suboptimal solution to the ecomorphological group-discrimination from wing-shape problem in *Trithemis*.

Geometric morphometric (GM)-style analyses of image datasets

One of the inherent limitations of the classic GM approach to form analysis is that it is based on the collection of – and representation of morphologies by – relatively sparse sets of topologically homologous landmark and semilandmark point locations across all specimens comprising the sample. In our *Trithemis* analysis, this approach restricted us to the use of intersections between major wing

veins, and of these veins with the wing periphery, to represent wing form. Accordingly, the results presented in the previous section pertain only to those aspects of *Trithemis* wing shape represented by these shape-coordinate configurations. But *Trithemis* wings are, obviously, composed of far more, and far more complex, structural elements than those represented by our GM-style form-representation system. In order to determine whether a more comprehensive and holistic representation of wing morphology might improve our ability to detect, document, and model differences in *Trithemis* habitat guilds, we conducted a parallel GM-style data analysis in which wing images – as represented by grey-scale brightness values for all pixels in a standardized 200 x 56-pixel (forewings) and 200 x 81-pixel (hindwings) image frame – were substituted for the shape-coordinate datafiles. The fact that the same data-analysis procedures were employed in both sets of analyses renders their results comparable directly despite the slightly reduced dataset sizes being employed for the wing-image analyses. Perhaps even more importantly, use of segmented whole-wing images avoids the need to select any aspect of the wing morphology for investigation at the outset of an analysis as well as ensuring that those aspects of the morphology that cannot be represented by single-coordinate, or single-pixel, locations (e.g., color or shading patterns) participate in the analysis along with those aspects that can.

Figure 10 summarizes results obtained for the PCA-CVA analysis of whole wing images for those *Trithemis* specimens in which the images did not include representations of the specimen labels. Comparison of these habitat-contrast shape distributions with those obtained from the landmark-semilandmark dataset, shows that, despite the minor differences in the sample composition between the two analyses (a difference that would favor lower discrimination power), use of whole-wing images resulted in a marked improvement in between-group separations for each wing type and for each habitat contrast.

Some might be tempted to interpret this improvement as consistent with the well-known tendency for high-dimensional datasets to yield artificially large apparent group distinctions when subjected to linear-discriminant analysis owing to the sparse distribution of data points in high-dimensional mathematical spaces [43,58]. If such an interpretation of our results was the correct this should be revealed by a bootstrap analysis of between-group separation relative to within-group dispersion via any of a number of statistical tests. However, when this experiment was carried out using the well-established Hotelling's T^2 test, for each of the four comparisons shown in Figure 10, observed values of the T^2 statistic fell well beyond the ranges of T^2 -value distributions obtained from 1000 random permutations of the data (see Supplementary Material). Based on these results it seems clear that the between-group separations shown in Figure 10 cannot be interpreted as mere artifacts of variable number-sample size interactions, but rather reflect clear and consistent differences in the shapes of *Trithemis* wings for species that preferentially inhabit different landscape and water-body habitats. These results support interpretations offered for the previous GM-style landmark-semilandmark datasets and imply that (1) the distinctions present between the wing morphologies of these different habitat groups include aspects not represented in the landmark-semilandmark dataset and perhaps aspects that cannot be represented under GM-style morphology sampling conventions and (2) reliance on GM-style morphology sampling conventions

resulted in a substantial underrepresentation of the actual degree of difference between a priori-defined *Trithemis* habitat groups in terms of wing-form differences.

What aspects of *Trithemis* wing morphology might be responsible for the greater between habitat-groups separations shown in Figure 10? Figure 11 illustrates color-coded comparisons of shape models for the end-member coordinate positions along the different habitat-group linear discriminant axes for both *Trithemis* wing image sets. These shape models are consistent with the landmark-semilandmark displacement patterns shown in Figure 8 for *Trithemis* forewings and hindwings but, owing to their more complete coverage of wing morphology, they contain much additional information about habitat-structured wing-shape differences.

With regard to the forewings, the dark blue line most evident in the lower right and upper left of the forewing shape models (but present around the entire wing) marks the periphery of the typical open landscape and temporary/standing water habitat wing forms. The solid white areas in the regions adjacent to this periphery signal that, in typical forested landscape and running water habitats the distal portions of *Trithemis* forewings have adopted a slightly longer and narrower form with the the proximal posterior margin exhibiting a pronounced more posterior expansion along with a more prominently expressed anal-area “corner”. The forewing tip also appears less rounded (more asymmetrical) in typical forest-dwelling and running water habitats species.

These modifications in forewing periphery shape are consistent with the forewing landmark-semilandmark displacements shown in Figure 8, the overall effect of these displacements is much easier to see in the image-based comparison of linear discriminant-axes models. Moreover, the arrangement of both the internal major wing veins, and the polygonal cells formed by the minor wing veins, also exhibit consistently well-structured displacements that reflect wing-margin alterations. In other words, these results suggest no major reorganization of the internal *Trithemis* forewing structure is part of the distinction either between open or forested landscape-dwelling species or between running or temporary/standing water body-dwelling species. This too is consistent with results of the landmark-semilandmark forewing dataset analysis. But whereas that analysis only sampled seven discrete internal wing landmark locations, these image-based data facilitate a remarkably detailed examination of internal structures across the entire forewing at a level of morphological resolution that simply could not be matched by landmark-semilandmark sampling schemes without engaging a very laborious and time-consuming data-collection program.

Interestingly, variations in the forewing peripheral outline, along with the apparently passive general displacement of major wing textural elements, are *not* the most prominent, or the most important, morphological distinctions that characterizing these different species cohorts and serve to distinguish them from one another. As can be seen clearly in both forewing model comparisons, the largest distinctions between these habitat-guild pairs occurs in the extreme proximal portion of the forewings, close to the attachment between the wing and the body. The most prominent of these differences involves the posterior wing attachment which is composed of a complex array of morphological

structures (e.g., distal plate, proximal plate, axillary plates I-III). The reason this region plays such a prominent role in between-habitat group distinction appears to be linked to the pronounced changes in the form of the extreme proximal posterior forewing outline which force this complex of elements to migrate inboard and posteriorly to accommodate and support the proximally wider forewing typical of forest-dwelling *Trithemis* species.

Resolution in this area of the wing morphology was lost in the case of the landmark-semilandmark dataset because the posterior wing attachment was represented by only a single landmark. This landmark did exhibit a large displacement relative to surrounding landmarks in the extreme proximal region of the wing (see Fig. 8), but the overall morphological complexity of this attachment was, underestimated by the decision to represent it with only a single landmark location. The desire to spread landmarks as evenly as possible over the morphological in question often results in difficult, and perhaps disadvantageous, decisions having to be made regarding how to represent complex morphologies under GM-style landmark-based sampling systems; decisions whose implications are difficult to appreciate when no alternative summaries of wing-shape variation modes are available. The higher, more comprehensive, and much easier to collect morphological resolutions available via the use of whole-form images as morphological variables facilitate the complete and comprehensive analysis of all available morphological structures. In addition, and as these results demonstrate, identification of highly localized (and so taxonomically informative) differences in comparative morphology can result from image-based analyses.

Along with the posterior wing-body attachment complex, our linear discriminant model-based comparison of between open and forested landscape-dwelling species, and between running and temporary/standing water body-dwelling species, also identified the proximate limbs of the costa, subcosta and radius + media veins as important sites of distinction between these habitat groups. In the effort to ensure even landmark-semilandmark placement across the entire forewing form, and because of difficulties in defining a landmark point to represent the form of a laterally extensive wing vein, these sites were also not represented by any landmarks in the previous GM-style analysis.

Much the same story told for the hindwing image dataset (Fig. 11), but with a rather obvious addition in hindsight. Relative to the typical hindwings of open landscape and temporary/standing water body-dwelling species, typical forested landscape and running water habitat species possess hindwings that are more distally elongate, with greater asymmetry at the wing apex as well as being distally wider. The angle of the peripheral hindwing is slightly greater in the latter groups, but not as much so as in the forewings. Overall, the hindwing textural elements appear to have responded (again) more-or-less passively to these changes in peripheral outline shape, though there is some suggestion that the pterostigma has shifted to a position slightly more proximal along the wing's anterior margin than would be expected as a result solely of the increase in distal hindwing angularity. Nevertheless, typical forested landscape and running water habitat species possess proximally wider hindwings.

Also, similar to the forewings, both the anterior and posterior wing attachments, as well as the proximal costa, subcosta and radius + media veins along with, in the hindwings, the proximal cubitus vein exhibit pronounced variations along the transition from open landscape and temporary/standing water-dwelling species to forested landscape and running water-dwelling species. Once again, as no landmarks were placed in these areas of the hindwing morphology, these aspects of hindwing variation were invisible to the previous landmark-semilandmark-based analysis.

By far the most prominent aspect of hindwing form difference, however, is the presence of the medium-to-dark pigmented region proximal to the body. This is characteristic – to varying degrees – of such temporary/standing water body-dwelling species and such open landscape-dwelling species as *T. annulata*, *T. arteriosa*, *T. aurora*, *T. kalula*, *T. kirbyi* and *T. monardi*. While it is true that *T. tropicana* – a forest/running water-dwelling species – exhibits the darkest proximal color bands of all species examined (see Plate 2), this is an atypical condition for the majority of forest and running water-dwelling species (e.g., *T. dichroa*, *T. dorsalis*, *T. ellenbeki*, *T. nuptalis*). Accordingly, wing pigmentation patterns represent an additional character that can assist in the correct evaluation of *Trithemis* wing-form differences that characterize different ecological habitats. Owing to the constraints imposed by standard landmark/semilandmark-based representation, this aspect of wing form cannot be included in any GM-style analysis, at least at a level of spatial resolution commensurate with other structural elements whereas in whole-structure image datasets pigmentation patterns can be accommodated readily.

The degree to which the linear discriminant analyses of *Trithemis* forewing and hindwing images were successful in identifying landscape and water-body habitat groups is summarized in the confusion matrices presented in Table 4. Comparing these with analogous results obtained from the GM-style landmark-semilandmark datasets (Table 3) shows how dramatic the difference can be between the amount of ecomorphologically important information represented by these different data types. For a post-hoc analysis of the training datasets, group-identification accuracies for the forewing landmark-semilandmark datasets ranged from 0.66 to 0.75. These accuracy values rose to between 0.92 to 0.95 for the image-based dataset. Similarly, the hindwing analysis group-identification accuracies ranged from 0.85 to 0.85 for the landmark-semilandmark datasets, but rose to values between 0.95 to 0.98 for the image-based dataset.

The fact that such high accuracy values can be achieved based on wing morphology alone seems both unexpected and remarkable. However, our enthusiasm for these results is tempered by the results achieved in the jackknife (leave one out) analysis of discriminant-function stability. While unexpectedly high and quite respectable accuracies were achieved by the image datasets under this constraint (ranging from 0.61 to 0.68 for forewings and 0.65 to 0.74 for hindwings), these accuracy values are comparable those achieved for the landmark-semilandmark dataset for the *Trithemis* forewings and inferior to those achieved by the landmark-semilandmark dataset for the hindwings. This result in no way compromises the statistical validity of the results achieved by the image-based linear discriminant analysis for the individuals analyzed as a part of this investigation since the bootstrap version of Hotelling's T^2 test takes the possibility of overtraining into consideration explicitly. Nevertheless, we would not advocate use of

these image-based forewing or hindwing discriminant functions as a basis for making any critical habitat-based identifications. This fall-off in performance stability was most likely caused by dramatically higher dimensionality of the image datasets and consequent need (ideally) to increase the image-dataset size dramatically in order to offset problems arising from sample size-dimensionality interactions. In addition, it may be the case that the somewhat poor stability results reported for both the landmark-semilandmark and image-based datasets arose not only, and not exclusively, from sample-size issues but, indeed, from the wisdom of assuming that geometric distinctions between these different habitat groups can be modeled accurately via the use of linear approaches to morphological data analysis.

Machine-learning (ML) analyses of image datasets

In order to address the issue of linear, as opposed to non-linear, modelling of ecomorphological group discrimination, one further set of analyses was undertaken for the *Trithemis* image datasets based on use of the deep-learning LetNet-5 CNN architecture trained using an embedded, group-contrast training strategy. In the case of the forewing analysis a total of 37,671 image contrasts, included both within and between habitat-group comparisons, were used to train the system to discriminate between the wing morphologies of landscape and water body group species. Training was organized into three rounds of 1767 image batches. Thus, in total, the LetNet-5 system was trained redundantly using a total of 113,088 image contrasts. Training history plots showed that both landscape and water body training rounds exhibited good convergence profiles with post-training error-loss values of 3.05×10^{-5} and 9.74×10^{-5} for these group analyses respectively. Similarly, the hindwing analysis employed a total of 42,320 image contrasts organized into three rounds of 1986 image batches. The hindwing training sequence, then employed a total of 127,104 image contrasts and achieved convergence in both cases with post-training error-loss values of 2.38×10^{-5} and 1.00×10^{-4} for the landscape and water-body group analyses respectively.

Figure 12 summarizes results of the LeNet-5 “deep-learning” analysis of wing-shape differences between landscape and water-body guilds for *Trithemis* forewings and hindwings. In all four cases this non-linear, ML approach to morphometric analysis was able to identify morphological features characteristic of these different ecological groups with 100 percent accuracy using image-contrast training sets. Indeed, the degree of between-groups separation achieved — which is indicative of the statistical confidence the algorithm has in its results — was such that within-group discriminant-score variation (of which there is a bit), has been obscured totally by selecting the same number of histogram bins used to illustrate the discriminant results of previous analyses.

As is always the case with the analysis of high-dimensional data, the results shown in Figure 12 might be regarded as suspect; the result of an overtrained learning system. Two arguments can be cited against this interpretation. The first involves the actual dimensionality of the data submitted to the LeNet-5 system for analysis. The first step in the LeNet-5 procedure was to process the raw images into 40 x 40-pixel thumbnail images. This operation reduced the original image sizes (200 x 56 pixels for the

forewings and 200 x 81 pixels for the hindwings) to a standard 1600 pixel-value datasets. Naturally, some morphological detail was lost during this interpolation process. While this dimensionality remains quite high by geometric morphometric standards, its effect was offset by employing an embedded ML approach, which focused not on the number of raw images in the datasets, but on the number of image contrasts inherent in the raw image sets. Thus, the effective sample sizes used to train the LeNet-5 system (113,088 and 127,104 pairwise comparisons for the forewing and hindwing datasets respectively) were almost two orders of magnitude greater than the dimensionality of the data being discriminated. Accordingly, there is little reason to suspect that overtraining might arise in these results as a consequence of the curse of dimensionality.

In a more empirical sense, however, empirical evidence of the stability, and appropriateness, of the results shown in Figure 12 can be generated using a standard leave-one-out jackknife strategy. Owing to the amount of time required to train the LeNet-5 system on a GPU-enabled platform (c. 1 hour per training round), it was considered impractical to submit the entire forewing and hindwing image datasets to the jackknife procedure. Instead, a series of 25 randomly chosen images were removed successively, the *Trithemis* forewing and hindwing discriminant systems retrained on the basis of the $n-1$ image sets, and the jackknife identification procedure carried out on each of these 25 sequestered images in order to estimate of the accuracy of the identifications and, hence, the overall stability of each discriminant system. This selection was carried out independently for each of the habitat group analysis and for each of the forewing and hindwing datasets to ensure each estimated stability result was independent of the subsample selected for jackknife sequestration.

Table 5 summarizes results of these four jackknife analyses for the *Trithemis* habitat datasets. In each case the LeNet-5 system, trained with the entire forewing and hindwing pairwise image-contrast datasets minus the sequestered specimen images successively, were able to identify the sequestered specimens with 100 percent accuracy overall. These results indicate that discriminant analyses of each system calculated from the *Trithemis* hindwing and forewing datasets, partitioned either by landscape or water-body habitat groupings, exhibit compelling stability with no evidence of any identification inaccuracies that could indicate substantial – or indeed any – overtraining. Further, these results not only agree with our previous morphometric results in suggesting consistent and stable morphological differences exist in the wing morphologies of *Trithemis* species that inhabit open and forested landscapes, and that hunt above temporary/standing or running waters, they imply that these differences are more substantial, more consistent and more stable than was indicated by either the standard GM-style linear analyses of landmark-semilandmark data, or the linear analysis of image pixel data. Since the species cohorts assigned to the landscape and water-body groups show some, but not total, identity, these results also appear to indicate that different sets of morphological features are responsible for these landscape and water-body ecomorphological discriminations.

Discussion

Despite intense interest in the results produced by advanced, deep-learning ML methods, these systems remain quite difficult to query in order to understand exactly what aspects of the patterns they are used to analyse (e.g., organismal morphologies) are being keyed upon to produce the (often amazing) group identification/discrimination results of which they are capable. The standard response most computer scientists and data analysts will give when asked about how advanced ML algorithms are able to provide the outstanding group identification results they clearly are capable of providing tends to be some variant, or combination, of “It’s complex”, “It doesn’t matter so long as the correct answer is produced reliably.”, and/or “You just have to trust the machine.”. In many instances, this range of responses is genuinely the best that can be provided, at least at present. Interpretation of ML results is a very active area of current computer-science research and much progress has been made. At the heart of this difficult issue are the number of components used to construct the ML system (e.g., the largest neural net systems at present are composed of literally tens of billions of artificial neurons each of which plays a role in all class assignment decisions, see [79]), the design complexity of the algorithms employed in advanced ML applications such as CNNs, and the non-linear mathematical spaces in which many of these systems operate.

Permutation feature importance (PFI) [80], saliency maps [81], local interpretable model-agnostic explanations (LIME) [82], Shapley values [83,84], scoped rules or anchors [85,86] and neural-backed decision trees (NBDT) [87] are among the current procedures available to evaluate the feature-based targets of inter-group discriminations of images. However, all of these approaches were designed originally to undertake evaluations of numerical datasets. When they are applied to large morphological datasets – and especially to image datasets – they test the contributions made by relatively large (and often somewhat randomly defined) segments of the data/images under consideration. While these approaches may be sufficient to identify the general regions in which group-diagnostic features are located, to date they have not been able to provide levels of spatial detail commensurate with those described routinely in taxonomic diagnoses. This is not to say that the ML algorithms themselves are incapable of basic discriminations at levels commensurate with (or even much finer than) those of taxonomic experts; only that, owing to the level of inter-pixel connectivity required to represent a diagnostic feature, coupled with spatial uncertainty as to where the boundaries of such features might lie, the relative sizes of the regions used by these and other algorithms in their attempts to identify group-diagnostic morphological or image features are, generally speaking, too large to be of much use to taxonomists (see [88,89] for examples). These approaches can be useful for ensuring group-diagnostic image features belong to parts of the image that pertain directly to the specimens being imaged (e.g., as opposed to some aspect of the background). But beyond this it appears we must await further developments in the field of ML interpretation before such algorithms can make a substantial contribution to revealing the morphological features they are sensing in order deliver their superior group-identification capabilities. One way to this interpretability issue might be addressed while we await further technical developments in this area is to employ a “middle range” strategy that bases exploratory morphometric investigations of the direct analysis of images using the linear, eigenanalysis-based

procedures that lay at the heart of multivariate analysis in general and GM in particular (e.g., PCA, canonical variates analysis or CVA, see [34,44,63,64].

Damm et al. [1] reconstructed ancestral character states for *Trithemis* landscape and water-body habitat characters using the stochastic analysis method of Bollback [90]. While it was not possible to infer all ancestral character states with complete certainty, these results were able to confirm that the transition from inhabiting open landscapes to forested landscapes, and from hunting above temporary/standing waters to flowing waters, arose multiple times in this clade. Our results, coupled with those of Damm et al. [1] suggest that, on each of those occasions, morphological changes in the phenotype were much more extensive than had been realized previously, involving not only those characteristics important for species identification, but also features – such as wing morphology – whose innate complexity had prevented their detailed ecomorphological analysis to date. In addition, our results suggest that, despite the independent origins of individual species (e.g., *T. persephone*, *T. purinata*) and sublineages (*T. grouti*-*T. stictica*-*T. nuptialis*-*T. aenea*-*T. aequalis*), common sets of morphological modifications characterize species that occupy both ancestral and derived ecological zones. Of course, it may also be the case that each of these derived ecological radiations also incorporates morphological variations unique to that species and/or to that radiation. The testing of this hypothesis must await more detailed analysis and, most likely, the collection of larger samples. Nonetheless, any species-specific or radiation-specific morphological traits in wing morphology were not of sufficient scope and/or consistency to obscure the common similarities within, and differences between, the wing forms that characterize these ecological guilds.

In 2013 Outomuro et al. [2] published an independent analysis of *Trithemis* ecomorphology that employed an approach to the analysis of wing morphology that differed from the ones we chose. That study concluded that (1) forewing and hindwing shape exhibited a statistically significant “phylogenetic structure”, (2) no significant association existed between wing shape and water body habitat, (3) male forewings and female hindwings differed with regard to the contrast between open and forest-dwelling species with the latter exhibiting characteristically broader wings, (4) hindwings of both sexes with greater coloration exhibited a broader base with this effect being more pronounced in males and (5) wing shape exhibited sexual dimorphism across all species considered. Obviously, the results obtained in our investigation contradict those of Outomuro et al. [2] in many respects even allowing for the different species compositions and foci of the two studies. Quite aside from the issue of what relations between wing morphology and ecological habitat actually characterize *Trithemis*, the different conclusions reached by these two analyses highlight the critical roles data collection and data-analysis strategy play in determining the results of all morphometric investigations.

The Outomuro et al. [2] study employed 32 species of which 25 (78%) were also included in our species list. Only single representatives of male and female forms for each species were employed and three of the species considered by Outomuro et al. [2] lacked females. In contrast, our dataset included multiple representatives of each species (see Table 1). The Outomuro et al. [2] investigation quantified wing morphology using 11 “biologically constant” forewing landmark points with a single semilandmark used

to quantify “wing curvature” along the posterior mid-wing margin. For the hindwings seven landmarks along the boundary outline were used to quantify wing form along the anterior margin (including two marking the position of the pterostigma) and distal half of the posterior margin. Owing to the lack of landmarks along the proximal half of the posterior wing margin, a somewhat arbitrary set of five semilandmarks were used. These were defined by an inter-landmark angular spacing of 9°, 18°, 36°, 63° and 90° from the chord joining the anterior wing attachment landmark to the nodus landmark (see [2], Fig. 1) In practice, this sampling scheme, which is reminiscent of that used in radial Fourier analysis, has the deficiency of the geometric resolution it offers being dependent on the length of the chord joining the specified semilandmarks to the wing nodus, which defined the angular vertex [91]. In any event, only these 11 (forewings) and 12 (hindwings) landmark/semilandmark points located along the wing periphery were used to quantify wing form. In contrast, our GM-style wing sampling scheme employed six landmarks (five of which were also employed by Outomuro et al. [2]), 25 semilandmarks (with constant inter-semilandmark spacing within landmark-defined wing-periphery regions) and seven landmarks located at topologically homologous positions in the wing interior. In addition to these data we also employed direct representation of the total wing morphology in the form of digital images. Accordingly, both our landmark-semilandmark and digital-image datasets incorporated more information related to wing morphology than the Outomuro et al. [2] study.

In terms of data analysis-strategy, Outomuro et al. [2] employed the method described by Klingenberg and Gidaszewski [92] to assess the strength of the phylogenetic signal on their wing landmark-semilandmark data whereas we employed the multivariate extension of the *K*-statistic method described by Adams ([54], see also [10]). The Klingenberg-Gidaszewski method fits phenotypic data to the tree using squared change parsimony and then obtains an estimate of the phylogenetic signal by summing the squared trait changes across all branches. Under this scenario the smaller the sum the greater the conformance with phylogeny.

However, as Adams [54] points out, this Klingenberg-Gidaszewski method (i.) relies on ancestral-state reconstruction which usually involves high levels of uncertainty [93] and (ii.) is unsuitable for use in evaluating phenotypic traits owing to the fact that geometric scaling is not taken into consideration and changes systematically as trait variation among species and/or with the number of traits increases. The Klingenberg-Gidaszewski method also incorporates a matrix inversion that limits its utility with datasets composed of a large number of phenotypical characteristics and a comparatively small number of species. Adams’ [54] multivariate *K*-statistic approach circumvents these limitations and is designed specifically for use with high-dimensional phenotypic data.

Finally, in terms of statistical testing, Outomuro et al. [2] relied on standard parametric tests whose accuracies rely on assumptions regarding the form of data distributions and equivalence of variances among variables, all of which are rarely met by morphometric data. In contrast, our study employed bootstrapping and jackknife variants of standard statistical and data-analysis tests to ensure the results of our hypothesis tests were robust to violations of distributional assumptions.

With regard to the test of phylogenetic signal strength in *Trithemis* forewing and hindwing morphometric data, our observed K_{mult} -statistic value fell well within the range of values expected from 1000 random permutations of the pruned Damm et al. [1] ultrametric tree, suggesting that our more completely realized representations of species-specific *Trithemis* wing morphology have a very low, and statistically non-significant, ratio of the range of phenotypic variation observed in both forewing and hindwing morphology and the range expected under a Brownian motion model. Indeed, for the *Trithemis* forewing dataset the observed K_{mult} value trends toward that which would be considered significant statistically for exhibiting a less than expected range of phenotypic covariation. These forewing and hindwing results are supported further by our calculation of the PC-based phylomorphospaces for both *Trithemis* species and their reconstructed ancestors based on the pruned Damm et al. [1] ultrametric tree. As illustrated by Adams [54], phylomorphospaces calculated from morphometric data that exhibit a strong covariance with phylogeny exhibit an organization in which closely-related species, along with their hypothetical ancestors, are grouped together in different regions of the Procrustes PC space with few tree branches that cross one another. Our *Trithemis* phylomorphospace results (Fig. 6) exhibit an overall pattern of morphospace distribution that is the opposite of this expectation, with closely related species projecting to positions in vastly different parts of the space and many crossed tree branches. Moreover, given the constraint (observed in our results) that estimated ancestral wing morphologies tend to occupy regions closer to the origin of the Procrustes PC space than the terminal taxa, the range of shape variation displayed by the former group is such that it is difficult for us to imagine any configuration of inferred hypothetical ancestral wing shapes that would be consistent with the expectations of a strongly supported phylogenetic covariation pattern. This ordination geometry indicates that our results, and our interpretations, are robust to the imprecision of ancestral node inferences, as noted by Losos [93].

The finding that many biological datasets do not exhibit significant patterns of phylogenetic covariation is well established – including for morphometric datasets – and can arise for many different reasons (see [93] and references therein). Notwithstanding the results reported by Outomuro et al. [1], *Trithemis* forewing and hindwing morphology appears to fall into this broad category. Our phylogenetic-signal results are also consistent with our finding of substantial ecomorphological covariation in *Trithemis* forewing and hindwing morphology as well as being somewhat inconsistent with the more limited findings in this area reported by Outomuro et al. [2] given the fact that the mappings of both landscape and water body characteristics of these species are distributed across the available *Trithemis* cladogram.

With regard to our finding of statistically significant wing shape differences for both *Trithemis* open versus forested landscape habitat guilds, and for temporary/standing versus running water body habitat guilds, the fact that our results differ from those reported by Outomuro et al. [2] may be the result of either a single-factor, or interactions between multiple-factor, differences in the two investigations. Owing to our failure to document a significant phylogenetic signal in either our *Trithemis* forewing or hindwing datasets, we did not follow Outomuro et al. [2] and subject any of our datasets to phylogenetic least-squares “correction”. This operation removes substantial amounts of information from the data and can only be justified when there is a clear data analysis-based concern with overall data independence. It may

be that Outomuro et al. [2] were misled in their interpretation of the degree of phylogenetic signal present in their data by their use of the Klingenberg-Gidaszewski phylogenetic signal test and so performed a data-standardization operation where none was actually required. Possibly this explains their, rather unusual, finding of significant differences among forest and open landscape-dwelling species for males, but not females. Sexual dimorphism was not a target of our study, but our dataset was, on the whole balanced in terms of the representation of male and female morphologies (44% females, 38% males, 14% uncertain) with all species being represented by individuals from both sexes. Accordingly, our findings of significant wing-morphology differences between landscape and water body groups imply that these differences pertain equally to both males and females, which is the more usual and expected pattern.

Alternatively, for our GM-style dataset, the difference between our habitat group findings and those of Outomuro et al. [2] may be a simple function of the degree to which, and the manner in which, we sampled the wing morphologies. In the Outomuro et al. [2] sampling scheme the spatially densest information was collected from the wing apex region and spatially sparsest from the anterior margin. Our sampling scheme achieved a much more even coverage of all parts of the peripheral outline. More importantly, the extended eigenshape sampling protocol we used to determine how the wing periphery should be sampled automatically places more semilandmark points in those regions that exhibit the greatest shape variation across the dataset as a whole. This had the effect of weighting our morphometric analysis toward those regions that exhibit the greatest amount of shape variation, thus ensuring appropriate advantage is taken of the information contained in those regions. No equivalent effort to focus the landmark-semilandmark data collected from *Trithemis* wings was used in the Outomuro et al. [2] study.

Our summary of the distribution of shape differences among the different habitat groups (Fig. 8) indicated that the wing periphery regions with the largest landmark/semilandmark displacements differed for the forewings and hindwings. In the case of the former the largest sampled-point displacements occurred mid-wing along the posterior, or trailing, margin, along the distal anterior margin, and along the proximal posterior margin, especially very close to the posterior wing attachment. While the same patterns are present in both the landscape and water body habitat-contrast results, displacements are much more pronounced for the landscape contrast. Yet, these regions were very weakly and unevenly sampled by the Outomuro et al. [2] wing morphology sampling scheme. Similarly, in the case of the hindwings, the largest displacements occurred along the proximal posterior margin, especially close to the point of maximum wing-periphery curvature (= the prominent proximate posterior “corner”), followed by the proximate anterior wing margin and the posterior mid-wing margin. As with the forewings, landmark-semilandmark displacements are much more pronounced for the landscape contrast. But again, these are areas where the Outomuro et al. [2] scheme obtained few, and unevenly distributed, samples of morphological variation.

At this point it is worth noting (again) that, in calling attention to the discrepancies between our findings and those of Outomuro et al. [2], and in attempting to offer explanations for those discrepancies, we are in no way leveling any criticisms at the authors of that study or at their interpretations. Indeed, we are

perfectly happy to state that the results and interpretations offered by Outomuro et al. [2] are correct, accurate, and fully justified *for the data they obtained and data-analysis results they produced*. Rather, the points we wish to make is that (1) representations of biological morphologies, especially those sampled by sparse sets of landmark-semilandmark points, cannot, should not, and must not be mistaken for the morphologies themselves and (2) different data-analysis procedures differ in the assumptions they make about the data that have been collected, the mathematical models applied to those data, and the power those models have to reveal patterns of similarities and differences within datasets; especially in the case where the point of the analysis is to compare groups defined a priori.

There are many ways to sample or represent any complex morphological structure. But once sampled, the results obtained from the analysis of those data pertain only to the samples that have been collected, not necessarily to the far more complex structure itself. Of course, all systematists and all morphometricians strive to obtain an adequate and accurate representations of the morphologies or structures they investigate. In some cases, and for some structures, this is straightforward. In others it is exceedingly difficult. If the question under examination is specific and tied intrinsically to an explicit aspect of the morphology in question (e.g., Are the forewings of male *Trithemis annulata* longer than those of females of the same species?) the data relevant to the hypothesis test can be obvious. But if the question under examination is non-specific and not tied intrinsically to any particular aspect of the morphology in question (e.g., Do the forewings of *Trithemis* species that inhabit forested landscapes differ in some way from those that inhabit open landscapes?) it often is difficult to know what to compare, what data to collect, and how to interpret the results of data-analysis procedures *in terms of the original question of interest*.

In attempting to address this more difficult type of question, Outomuro et al. [2] chose a wing morphology sampling scheme that achieved a representation of *Trithemis* forewing and hindwing morphology, but did so in quite an approximate manner. To some extent, their approach reflected, and was possibly encouraged by, conventions that have grown up around geometric morphometrics which prioritizes the representation of complex structures thorough the digitization of small sets of independently defined landmark points. Originally, GM even objected to the collection and use of boundary outline semilandmarks [14,94] though this prohibition has now been relaxed to some extent, largely for practical reasons (see [48]). But even given the belated acceptance of semilandmark points as useful means of sampling complex morphologies, few systematists would be comfortable with the proposition that the sampling scheme devised and employed by Outomuro et al. [2] was either an accurate, or entirely satisfactory, representation of a *Trithemis* dragonfly wing. That scheme quantifies some aspects of the wing morphology, but ignores the vast majority of the information available.

Any set of data can be collected from any insect wing, subjected to analysis and used to produce a result. That result is guaranteed to reflect patterns present in the data submitted for analysis. But can the interpretation of those results be extended beyond the data through which they were generated? The answer to this question must depend on the character of the result, the representativeness of the sample of individuals (drawn from a much larger population) and the fidelity with which those aspects of the

morphology that were sampled truly represents the morphology in question. In precisely the same way that a biased or unrepresentative sample drawn from a population of individuals may compromise the ability of numerical data analysis to tell us anything interesting or useful about a population of interest, irrespective of the fact that a data-analysis result will always be generated, a biased or unrepresentative sample of the morphologies in question may compromise the ability of numerical data analysis to tell us anything interesting or useful about patterns of morphological variation existing in a sample, irrespective of the fact that a morphological data-analysis result will always be generated. The simple act of collecting morphological data is insufficient, by itself, to ensure a reasonable and/or accurate answer to generalized questions about any set of morphologies has been, or can be, obtained. This prescription is especially pertinent when negative hypothesis-test results are obtained as they were in the case of the Outomuro et al. [2] study.

Outomuro et al. [2] found that (1) no significant association existed between wing shape and water body habitat for *Trithemis* species, (2) male, but not female, *Trithemis* forewings differed with regard to the contrast between open and forest-dwelling species, and (3) female, but not male, *Trithemis* hindwings differed with regard to the contrast between open and forest-dwelling species. These conclusions were presented as though they pertained to the forewing and hindwing morphologies themselves rather than to the shape-coordinate locations of 11 forewing landmarks and 7 hindwing landmarks augmented by 5 hindwing semilandmarks, both sets of which were confined exclusively to the wing peripheral margin. We contend that the differences between these results and the results obtained by our investigation are largely the result of differences in the way the two research teams *chose* to quantify *Trithemis* wing morphology, augmented by differences in sample composition, sample size, and the manner in which the data were analyzed. As they pertain to the data used to represent *Trithemis* wing morphology, there is no disagreement between these two sets of results. Both are correct summaries of patterns in the data collected by each research team. But, in terms of the larger question involving *Trithemis* ecomorphological wing-shape variation among binary parsings of landscape and water-body guild states, we believe our results are the more fully representative because they are based on more complete assessments of *Trithemis* forewing and hindwing morphology and because they produced broadly consistent, as well as progressively more fully realized results, under two different data-collection regimes and two different data-analysis strategies.

As a final discussion topic it is interesting to note the implications our study has for the practice of morphometrics. Nowadays it is common to read and hear reference made to the “morphometrics revolution”, which is to say the advent of geometric morphometrics which took place more than 30 years ago [11,12,13,14,16,95] (see also [17,18,96]). That revolution was actually a synthesis between three aspects of morphometric practice that had been pursued more-or-less separately until the mid-1980s: the representation of form through the use of sparse sets of topologically corresponding landmark-points (that served and the end-definitions of linear distances originally), the alignment of these geometric point-locations through use of a least-squares Procrustes fitting algorithm as sets of deviations from the mean configuration, and the representation of patterns of morphological variation via linear multivariate analysis. While advances in addition to these did figure in the development of geometric morphometrics

(e.g., centroid size, bending energy-based shape decomposition, graphic representation of shape deformation via use of thin-plate splines), and acknowledging that the GM synthesis has grown since its original formulation (e.g., admission of semilandmarks as useful morphology-sampling devices), these three core aspects are those most often used and referred to in GM investigations. This synthesis is powerful, enabling morphological analysis to be pursued quantitatively and at levels of detail, coherence and interpretability unprecedented by the formerly separate schools of morphometric practice. Owing to that power, the geometric morphometric synthesis has proven to be highly effective in addressing a wide range of problems in systematic and comparative morphology, as well as being quite popular among communities of biological, paleontological, systematics and evolutionary researchers. However, the geometric morphometric approach, like all data-analysis approaches, has its weaknesses as well as its strengths. Perhaps even more importantly the field of data analysis rarely remains static for long.

Over the last 20-25 years alternative – some might say a rival – to GM has appeared in the form of ML. Unlike GM, ML approaches were not developed by researchers whose primary interest was in the analysis of biological morphology. Nevertheless, one of the primary, and most popular, uses of ML approaches, as well as spurs to ongoing research into the general topic of ML, has been the ability of these algorithms to find previously unsuspected patterns in all sorts of data, but especially in morphological data.

In many ways, ML represents a natural complement to GM. Whereas GM was designed to operate on a specific type of morphological data (= configurations of landmark point locations), ML can be used to analyze any sort of morphological data, including configurations of landmark point locations. Thus, whereas the application of GM is limited to those situations in which forms can, reasonably, be represented by configurations of point coordinates, use of ML approaches opens the door to the consideration of a much wider range of morphological data and morphological problems. To date the overwhelming majority of GM analyses published in the biological, paleontological, systematic and evolutionary literature have been based on linear data-analysis models. However, ML approaches can be applied readily to situations in which the optimal models are non-linear, even if that is not known to the case at the outset of an investigation.

At present, ML models are inferior to their GM counterparts in terms of their ability to be queried and so used to identify which aspects of a set of morphologies are contributing disproportionately to overall sample variance, which are useful for group discrimination and/or the structure of variable correlations. To be sure, attempts have been made to improve the interpretability of ML models (see [97] for a review). But as we have outlined above, much more remains to be done in this area.

Inevitably some will claim that GM is their preferred option for generalized morphological data analysis, either because, despite its obvious limitations, they regard their study group(s) and research questions well-served by this approach and/or because they wish to retain a “geometric focus” in their analysis. In response we can only point out that all analyses of morphological data are “geometric” in character because morphology is composed entirely of the sizes and shapes of various parts, characters and characteristics as well as their spatial arrangements relative to other parts, characters and characteristics.

From the results we have presented above it is unquestionably clear that the strictly GM approach to the analysis of *Trithemis* wing-shape data was the one that performed least well in finding, summarizing, testing, and assembling sets of characteristics that could be used to answer the generalized questions of whether shape variance was distributed among *Trithemis* landscape and water-body ecological guilds in a continuous or disjunct manner. What is also clear is that this comparative finding is not an unusual or exceptional result [44,45,67–69,98–107].

Conclusions

In seeking to understand the structure of the natural world it is important to appreciate the roles played by phylogenetic inertia – that an absence of divergent selection pressure can fail to produce substantial morphological divergence (hence “cryptic” species) even after speciation has taken place – and adaptive environmental diversification – that can lead to substantial morphological diversification even among sister species – in creating and maintaining that structure. Prior to the publication of [4], phylogeny was thought to contribute little more than anecdotal historical information to comparative biology and less still to the quantitative analysis of modern organismal morphology. However, following what can only be described as Felsenstein’s seminal description of the problems inherent in taking a non-phylogenetic approach to comparative biology, the systematic, biological and evolutionary research communities were rapidly converted to the idea that “nothing in biology makes sense except in the light of phylogeny”.

Today, the pendulum has swung decidedly in favor of a phylogenetically informed approach to comparative biology, to such an extent that many comparative biologists [seem to] believe that a phylogeny is not only a necessary, but also a sufficient answer any evolutionary question [93]. But has comparative biology’s late 20th century course correction gone too far? Certainly, patterns of ancestry and descent are fundamental to the analysis of all biological data. But are there cases in which the structure of phylogenetic relations might provide little insight into understanding the morphological superstructure of the natural world; in which the demands of the environment may, indeed, have played the dominant role. Just as importantly, which are the best tools to use in determining whether the ranges of variation in complex biological structures manifested by one organismal group are the same as, or different from, those same structures as manifested by another group; whether there is any pattern that requires explanation.

In our investigation of *Trithemis* forewing and hindwing morphology we set out to address these twin concerns of comparative morphology in an ecomorphological context using the most sophisticated and powerful tools available currently. With respect to the issue whether a statistically significant pattern of phylogenetic covariation exists in our sample of 276 mixed male and female individuals representing 27 *Trithemis* species (including species from across Africa, Madagascar, and China), analysis of GM-style landmark-semilandmark data using the multivariate extension of the K_{mult} statistic failed to detect any significant pattern of covariation between species-specific forewing or hindwing shapes. This result contrasts with the result obtained previously by Outomuro et al. [2] who used a different phylogenetic

signal test, but is supported by the analytic superiority of the K_{mult} test and by PCA-based phylomorphospace results derived from the same dataset.

With respect to tests for ecomorphological differences in forewing and hindwing morphology between binary-parsed landscape and water body ecological guilds, linear discriminant analyses of GM-style landmark-semilandmark data, direct linear discriminant analyses of wing images, and ML (embedded CNN) analyses of wing images all detected statistically significant shape differences between both habitat-guild partitions for both wing complexes. The best between-groups separations were achieved by the ML (embedded CNN) analyses of image data, the worst by linear discriminant analyses of GM-style data. These results suggest that shape differences between both habitat groups are not focused solely on the wing outline and are distributed geometrically in a non-linear manner. Our ecomorphological results also contrast, to some extent, with those reported previously by Outomuro et al. [2] who used a different, though GM-based, morphometric data collection and data-analysis strategy. Nevertheless, we believe our results are more representative of the actual situation in *Trithemis* because (i.) a greater amount of wing-morphology information was included in our analyses and (ii.) we obtained consistent results from the analysis of radically different wing data sets and different data-analysis strategies. Our ecomorphological results are also consistent (iii.) with expectations of the mapping of these ecological habitat guilds onto the *Trithemis* phylogeny.

In *Trithemis*, radiation from the ancestral ecological conditions of open landscapes and temporary/standing water bodies, into forested landscapes and running water bodies, occurred frequently and at multiple locations across the phylogeny. Traits that evolve frequently and substantially within taxa are usually responding to the differing needs of life under different selective regimes [69]. This principle is not sex-specific, and would be expected to apply equally to males and females, as our results indicate in the case of *Trithemis*. Moreover, highly functional aspects of the phenotype, such as wings, are subject to mechanistic, physical principles of optimization that are similar within similar environments, but differ across different environments. This is a very well-established principle in the comparative morphology of bird wings [37,108], bat wings [39,109] and insect wings including members of the Odonata [73]. This principle is also consistent with our finding of little phylogenetic covariation in *Trithemis* wing-shape data.

In raising criticisms regarding the adaptationist paradigm, Gould and Lewontin [3] viewed phylogenetic ancestry as exerting a constraint on morphological change and proposed that this hypothesis be considered a possible alternative to the direct, adaptive modification of each aspect of a species' morphology, physiology, behavior, etc. to meet some environmentally mandated challenge. But as had been noted repeatedly by a number of researchers (e.g., [52,93,110]) phylogeny is not a constraint; rather it is a pattern. The fact that two closely related species might exhibit similar morphologies cannot necessarily be attributed to the closeness of their phylogenetic relation any more than the fact that morphological differences between distantly related species can be attributed necessarily to their phylogenetic distance. In both cases it is a trivial exercise to cite numerous counter examples. Species remain close to, or diverge widely from, one another morphologically because of the manner in which

they have met the challenges selection pressures have exerted upon them. These pressures need not cause every aspect of their morphologies to change and genetic, as well as mechanistic, linkages that constrain both the advent, and the character, of realizable options do exist. Phylogenies are indispensable for understanding the structure of the living world. But phylogenetic patterns of ancestry and descent, by themselves, cannot provide an adequate process-level explanation for any aspect of biological structure. Neither can associations between morphology and phylogeny be blithely regressed out of morphological datasets and dispensed with as though they were some uninteresting nuisance factor. Rather, such associations should be regarded as constituting a category, or mode, of variation existing within morphological datasets that demands its own set of process-level explanations; separate from – but perhaps linked to – explanations proposed for features not closely associated with phylogenetic patterns.

In addition to these considerations, it is important to note that the quality of any morphological analyses will depend critically on the tools used to discover patterns in biological data; patterns that can be compared to the pattern of phylogenetic ancestry and descent as well as to aspects of the natural environment. If mathematics can be regarded, as it is by many mathematicians, as the study of patterns in numbers [111], biology can be thought of as the search for patterns in the living world. Indeed, it is the existence of these patterns that provides the subject matter for biological study as well as providing the evidence that deterministic processes or factors responsible for these patterns exist. If such patterns did not exist – if everything in nature simply graded continuously and insensibly into everything else – it would not only be impossible to conduct any truly scientific biological investigation, it would be pointless.

Mathematical data analysis and statistics (the two are *not* synonymous) are tools that, when used properly, can be employed to discover patterns in data that can aid biologists in their attempts to understand the living world. They are not, substitutions for, or means through which careful reasoning by researchers with specialist knowledge and experience can be overruled. Rather they can, are, and should, be used to aid and support biological reasoning by extending the powers of human senses and perception, making patterns invisible to the unaided eye visible so they can be identified, discussed and interpreted. In the same way as new and progressively more powerful statistical tools are being made available to compare patterns within morphological datasets and between morphological data and other sources of information, new tools have recently been made available for discovering patterns in biological data; ML being perhaps the latest and most intriguing example. As we hope we have demonstrated above, the development of new and much more sophisticated ways of applying quantitative data-analysis procedures to the task of identifying patterns of variation in morphological data promises, at the very least, to invigorate, and perhaps also to revolutionize, the study of morphology.

Declarations

Acknowledgements

This project began as ZSs work-study project, under the supervision of NM, when ZS was an Sixth Form student at Uckfield College (<https://www.uckfield.college>). We gratefully acknowledge The Natural History Museum (London) for providing access to its *Trithemis* collections. All software used in this investigation was written by NM in the Wolfram *Mathematica*[™] scripting language and is available either as part of the Supplementary Materials associated with this contribution, or directly from the corresponding author.

Author's contributions

BP and NM conceptualized the study. NM selected the specimens for imaging. These were imaged by ZS and NM. NM developed the data-collection and data analysis strategies as well as writing custom software for the investigation and performing all data analyses. All authors collaborated closely on the interpretation of the results. NM wrote the initial draft of this manuscript, and both BP and ZS contributed equally to creation of the final text. All authors read and approved the manuscript's final version.

Funding

Not applicable.

Availability of data and materials

The image dataset supporting the conclusions of this article is included within the article (as a set of plates) and as an archive in the supplementary information. Additional files documenting intermediate data-analysis results, along with the Wolfram Mathematica[™] scripts for all data-processing and analysis software used in this investigation are also included in the Supplementary Information. The specimens analyzed are all part of the Life Science Collections of The Natural History Museum (London). These are available for study or loan by agreement with this institution.

Ethics approval and consent to participate

Not applicable.

Competing interests

All authors declare they have no competing or conflicting interests with regard to any aspect of this report.

Consent for publication

Not applicable.

Author details

¹School of Earth Science and Engineering, Nanjing University, Nanjing, China. ²Department of Life Sciences, The Natural History Museum, London, UK. ³Department of Geology, Cardiff University, Cardiff, UK

References

1. Damm S, Dijkstra KDB, Hadrys H. Red drifters and dark residents: The phylogeny and ecology of a Plio-Pleistocene dragonfly radiation reflects Africa's changing environment (Odonata, Libellulidae, *Trithemis*). *Mol Phylogenet Evol*. 2010;54:870–82. doi:10.1016/j.ympev.2009.12.006.
2. Outomuro D, Dijkstra KD, Johansson F. Habitat variation and wing coloration affect wing shape evolution in dragonflies. *J Evol Biol*. 2013;26:1866–74. doi:10.1111/jeb.12203.
3. Gould SJ, Lewontin RC. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc R Soc London Ser B*. 1979;205:581–98.
4. Felsenstein J. Phylogenies and quantitative characters. *Annu Rev Ecol Syst*. 1988;19:445–71.
5. Harvey PH, Pagel MD. The comparative method in evolutionary biology. Oxford: Oxford University Press; 1991.
6. Garland T, Harvey PH, Ives AR. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Syst Biol*. 1992;41:18–32.
7. Garland T, Midford PE, Ives AR. An introduction to phylogenetically based statistical methods, with a new method for confidence intervals on ancestral values. *Am Zool*. 1999;39:374–88.
8. Ives AR, Midford PE, Garland T. Within-species variation and measurement error in phylogenetic comparative methods. *Syst Biol*. 2007;56:252–70.
9. Harmon LJ. Phylogenetic comparative methods: learning from trees. Atlanta. Georgia: CreateSpace Independent Publishing Platform; 2018.
10. Adams DC, Collyer ML. Multivariate phylogenetic comparative methods: evaluations, comparisons, and recommendations. *Syst Biol*. 2018;67:14–31.
11. Kendall DG. Shape manifolds, procrustean metrics and complex projective spaces. *Bull London Math Soc*. 1984;16:81–121.
12. Kendall DG. Comment on "Size and shape spaces for landmark data in two Dimensions by Fred L. Bookstein). *Stat Sci*. 1986;1:222–6.
13. Bookstein FL. Size and shape spaces for landmark data in two dimensions. *Stat Sci*. 1986;1:181–242.

14. Bookstein FL. Morphometric tools for landmark data: geometry and biology. Cambridge: Cambridge University Press; 1991.
15. Bookstein FL, Rohlf FJ. Proceedings of the Michigan morphometrics workshop. 1990.
16. Goodall CR. Procrustes methods in the statistical analysis of shape. *J R Stat Soc Ser B*. 1991;53:285–339.
17. Adams DC, Rohlf FJ, Slice DE. Geometric morphometrics: ten years of progress following the ‘revolution.’. *Ital J Zool*. 2004;71:5–16. doi:10.1080/11250000409356545.
18. Adams DC, Rohlf FJ, Slice DE. A field comes of age: geometric morphometrics in the 21st century. *Hystrix*. 2013;24:13–20.
19. Fukushima K. Neural network model for a mechanism of pattern recognition unaffected by shift in position—Neocognitron. *Trans Inst Electron Commun Eng*. 1979;J62-A:658–65.
20. Fukushima K. Artificial vision by multi-layered neural networks: Neocognitron and its advances. *Neural Netw*. 2013;37:103–19.
21. LeCun Y, Simard P, Pearlmutter B. Automatic learning rate maximization by on-line estimation of the Hessian’s eigenvectors. In: Hanson S, Cowan J, Giles L, editors. *Advances in neural information processing systems*. Vol. 5. San Mateo: Morgan Kaufmann Publishers; 1993. pp. 156–63.
22. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE*. 1998;86:2278–323.
23. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw*. 2015;61:85–117. doi:10.1016/j.neunet.2014.09.003.
24. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–44.
25. Rawat W, Wang Z. Deep convolutional neural networks for image classification: a comprehensive review. *Neural Comput*. 2017;29:2352–449. https://www.mitpressjournals.org/doi/pdf/10.1162/neco_a_00990.
26. Rohlf FJ. Relative warp analysis and an example of its application to mosquito wings. In: Marcus LF, Bello E, García-Valdecasas A, editors. *Contributions to morphometrics*. Madrid: Museo Nacional de Ciencias Naturales 8. 1993:131–160.
27. Klingenberg CP, McIntyre GS. Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with Procrustes methods. *Evolution*. 1998;52:1363–75.
28. Klingenberg CP, McIntyre GS, Zaklan SD. Left-right asymmetry of fly wings and the evolution of body axes. *Proc R Soc B Biol Sci*. 1998;265:1255–9.
29. Comstock JH, Needham JG. The wings of insects. *Am Nat*. 1898;33:117–26.
30. Wootton RJ. Function, homology and terminology in insect wings. *Syst Entomol*. 1979;4:81–93.
31. Yang HP, Ma CS, Wen H, Zhan QB, Wang XL. A tool for developing an automatic insect identification system based on wing outlines. *Nat Sci Reports*. 2015;5:12786. doi:10.1038/srep12786.
32. Sontigun N, Sukontason KL, Zajac BK, Zehner R, Sukontason K, Wannasan A, et al. Wing morphometrics as a tool in species identification of forensically important blow flies of Thailand.

- Parasit Vectors. 2017;10:1–14. doi:10.1186/s13071-017-2163-z.
33. Hall MJR, MacLeod N, Wardhana AH. Use of wing morphometrics to identify populations of the Old World screwworm fly, *Chrysomya bezziana* (Diptera: Calliphoridae): a preliminary study of the utility of museum specimens. *Acta Trop*. 2014;138 Suppl:49–55. doi:10.1016/j.actatropica.2014.03.023.
34. MacLeod N, Hall MJR, Wardhana AH. Towards the automated identification of *Chrysomya* blow flies from wing images. *Med Vet Entomol*. 2018;32:323–33. doi:10.1111/mve.12302.
35. Wootton RJ. Functional morphology of insect wings. *Annu Rev Entomol*. 1992;37:113–40.
36. Blanke A. Analysis of modularity and integration suggests evolution of dragonfly wing venation mainly in response to functional demands. *J R Soc Interface*. 2018;15:20180277.
37. Altshuler DL, Bahlman JW, Dakin R, Gaede AH, Goller B, Lentink D, et al. The biophysics of bird flight: functional relationships integrate aerodynamics, morphology, kinematics, muscles, and sensors. *Can J Zool*. 2015;93:961–75.
38. Baliga B, Szabo I, Altshuler DL. Range of motion in the avian wing is strongly associated with flight behavior and body mass. *Sci Adv*. 2019;5:eaaw6670.
39. Norberg UM, Rayner JMV. Ecological morphology and flight in bats (Mammalia; Chiroptera): wing adaptations, flight performance, foraging strategy and echolocation. *Philos Trans R Soc London Ser B (Biological Sci)*. 2015;316:335–427.
40. Mengesha TE, Vallance RR, Barraja M, Mittal R. Parametric structural modeling of insect wings. *Bioinspiration Biomimetics*. 2009;4:1–15.
41. Salcedo MK, Hoffmann J, Donoughe S, Mahadevan L. Computational analysis of size, shape and structure of insect wings. *Biol Open*. 2019;8:bio040774.
42. Altman N, Krzywinski M. The curse(s) of dimensionality. *Nat Methods*. 2018;15:399–400.
43. Mitteröcker P, Bookstein FL. Linear discrimination, ordination, and the visualization of selection gradients in modern morphometrics. *Evol Biol*. 2011;38:100–14. doi:10.1007/s11692-011-9109-8.
44. MacLeod N. The direct analysis of digital images (eigenimage) with a comment on the use of discriminant analysis in morphometrics. In: Lestrel PE, editor. *Proceedings of the Third International Symposium on Biological Shape Analysis*. Singapore: World Scientific; 2015. p. 156–182.
45. MacLeod N. The quantitative assessment of archaeological artifact groups: beyond geometric morphometrics. *Quat Sci Rev*. 2018;201:319–48. doi:10.1016/J.QUASCIREV.2018.08.024.
46. Cardini A, O'Higgins P, Rohlf FJ. Seeing distinct groups where there are none: spurious patterns from between-group PCA. *Evol Biol*. 2019;46:303–16. doi:10.1007/s11692-019-09487-5.
47. Cardini A, Polly PD. Cross-validated between group PCA scatterplots: a solution to spurious group separation? *Evol Biol*. 2020;47:85–95. doi:10.1007/s11692-020-09494-x.
48. MacLeod N. Generalizing and extending the eigenshape method of shape visualization and analysis. *Paleobiology*. 1999;25:107–38.
49. MacLeod N. Going round the bend: eigenshape analysis I. *Palaeontol Assoc Newsl*. 2012;80:32–48.

50. Mayall P, Pilbrow V, Bitadze L. Migrating huns and modified heads: eigenshape analysis comparing intentionally modified crania from Hungary and Georgia in the migration period of Europe. PLoS One. 2017;12:e0171064.
51. Rohlf FJ, Slice D. Extensions of the Procrustes method for optimal superposition of landmarks. Syst Zool. 1990;39:40–59.
52. Nel A. Un nouvel Odonate fossile du Miocène de Bellver de Cerdana (Espagne) (Odonata, Libellulidae). Entomologica Gall. 1991;2:129–30.
53. Blomberg SP, Garland TJ, Ives AR. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution. 2003;57:717–45.
54. Adams DC. A generalized K statistic for estimating phylogenetic signal from shape and other high-dimensional multivariate data. Syst Biol. 2014;63:685–97.
55. Christenson AL, Read DW. Numerical taxonomy, r-mode factor analysis and archeological classification. Am Antiq. 1977;42:163–79.
56. Anderson MJ, Willis TJ. Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. Ecology. 2003;84:511–25.
57. Marrama G, Kriwet J. Principal component and discriminant analyses as powerful tools to support taxonomic identification and their use for functional and phylogenetic signal detection of isolated fossil shark teeth. PLoS One. 2017;12:e0188806. doi:10.1371/journal.pone.0188806.
58. Rohlf FJ. Why clusters and other patterns can seem to be found in analyses of high-dimensional data. Evol Biol. 2020;1–16. doi:10.1007/s11692-020-09518-6.
59. MacLeod N. Form & shape models. Palaeontological Association Newsletter. 2009;72:14–27.
60. Hotelling H. The generalization of Student's ratio. Ann Math Stat. 1931;2:360–78.
61. Hotelling H. The generalization of Student's ratio. Ann Math Stat. 1931;2:360–78.
62. Manly BFJ, Alberto JAN. Multivariate statistical methods: a primer, 4th Edition. Boca Raton, Florida: CRC Press; 2017.
63. Turk M, Pentland A. Eigenfaces for recognition. J Cogn Neurosci. 1991;3:71–86.
64. Turk M, Pentland A. Face recognition using eigenfaces. Proc IEEE Conf Comput Vis Pattern Recognit. 1991;5:586–91.
65. Jhamtani H, Berg-Kirkpatrick T. Learning to describe differences between pairs of similar images. ArXiv. 2018;1808.10584:4024–34.
66. Forbes M, Kaeser-Chen C, Sharma P, Belongie S. Neural naturalist: generating fine-grained image comparisons. ArXiv. 2019;1909.04101:708–17.
67. Hoyal Cuthill JF, Guttenberg N, Ledger S, Crowther R, Huertas B. Deep learning on butterfly phenotypes tests evolution's oldest mathematical model. Sci Adv. 2019;5:1–11.
68. MacLeod N, Kolska Horwitz L. Machine-learning strategies for testing patterns of morphological variation in small samples: sexual dimorphism in gray wolf (*Canis lupus*) crania. BMC Biol. 2020;18:1–26. <https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-020-00832-1>.

69. MacLeod N, Canty RJ, Polazek A. Morphology-based identification of *Bemisia tabaci* cryptic species puparia via embedded group-contrast convolution neural network analysis. Syst Biol. 2021.
70. van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res. 2008;9:2579–605.
71. van der Maaten L, Postma E, van den Herik J, University T. Dimensionality reduction: a comparative review. Tilburg: Tilburg University; 2009.
72. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2019;37:38–47.
73. Dorrity MW, Saunders LM, Queitsch C, Fields S, Trapnell C. Dimensionality reduction by UMAP to visualize physical and genetic interactions. Nat Commun. 2020;11:1–6. doi:10.1038/s41467-020-15351-4.
74. Wattenberg M, Viégas F, Johnson I. How to use t-SNE effectively. Distill. 2016;1–13.
75. Manly BFJ. Randomization, bootstrap and Monte Carlo methods in biology, third edition. Boca Ration, Louisiana: Chapman Hall/CRC; 2006.
76. Matthews BW. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim Biophys Acta - Protein Struct. 1975;405:442–51.
77. Jurman G, Riccadonna S, Furlanello C. A comparison of MCC and CEN error measures in multi-class prediction. PLoS One. 2012;7:e41882.
78. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genom. 2020;21:1–13.
79. Rosset C. Turing-NLG: A 17-billion-parameter language model by Microsoft. Microsoft Research. 2020;:1–10. <https://www.microsoft.com/en-us/research/blog/turing-nlg-a-17-billion-parameter-language-model-by-microsoft/>. Accessed 23 Nov 2020.
80. Breiman L. Random forests. Mach Learn. 2001;45:5–32.
81. Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: visualising image classification models and saliency maps. In: Benigo Y, LeCun Y, editors. 2nd International Conference on Learning Representations, ICLR 2014 - Workshop Track Proceedings. Banff, Canada: arXiv; 2014. p. 1–8. <https://www.robots.ox.ac.uk/~vgg/publications/2014/Simonyan14a/simonyan14a.pdf>.
82. Ribeiro MT, Singh S, Guestrin C. “Why should I trust you?” Explaining the predictions of any classifier. Proc ACM SIGKDD Int Conf Knowl Discov Data Min. 2016;August:1135–44.
83. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Guyon I, Luxburg U V, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. 31st Conference on Neural Information Processing Systems (NIPS 2017). Boston, Massachusetts: MIT Press; 2017, p. 10.
84. Ribeiro MT, Singh S, Guestrin C. Anchors: high-precision model-agnostic explanations. In: 32nd AAAI Conference on Artificial Intelligence, AAAI 2018. New Orleans, Louisiana: Association for the Advancement of Artificial Intelligence; 2018:1527–35.
85. Delaunay J, Galárraga L, Largouët C. Improving anchor-based explanations. In: Berendt B, de Vries A, editors. Proceedings of the 29th ACM International Conference on Information & Knowledge

- Management (CIKM '20). New York, New York: Association for Computing Machinery; 2020:3269–3272.
86. Wan A, Dunlap L, Ho D, Yin J, Lee S, Jin H, et al. NBDT: neural-backed decision trees. *arXiv*. 2020;1:1–15.
87. Arteaga C. Interpretable machine learning for image classification with LIME. *Toward Data Sci*. 2019;21 October:1–9. <https://towardsdatascience.com/interpretable-machine-learning-for-image-classification-with-lime-ea947e82ca13>.
88. Stewart M. Guide to interpretable machine learning. *Toward Data Sci*. 2020;2020 19 March:1–40. <https://towardsdatascience.com/guide-to-interpretable-machine-learning-d40e8a64b6cf>.
89. Bollback JP. SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics*. 2006;7:1–7.
90. MacLeod N. The center cannot hold I: Z-R Fourier analysis. *Palaeontol Assoc Newsl*. 2011;78:35–45.
91. Klingenberg CP, Gidaszewski NA. Testing and quantifying phylogenetic signals and homoplasy in morphometric data. *Syst Biol*. 2010;59:245–61. doi:10.1093/sysbio/syp106.
92. Losos JB. Seeing the forest for the trees: The limitations of phylogenies in comparative biology. *Am Nat*. 2011;177:709–27.
93. Zelditch ML, Fink WL, Swiderski DL. Morphometrics, homology, and phylogenetics: quantified characters as synapomorphies. *Syst Biol*. 1995;44:179–89.
94. Kendall DG, Barden D, Carne TK, Le H. Shape and shape theory. New York: Wiley; 1999.
95. Rohlf FJ, Marcus LF. A revolution in morphometrics. *Trends Ecol Evol*. 1993;8:129–32.
96. Molnar C. Interpretable machine learning: a guide for making black box models explainable. 2020:318. <https://christophm.github.io/interpretable-ml-book/>.
97. Appeltans W, Ahyong ST, Anderson G, Angel MV, Artois T, Bailly N, et al. The magnitude of global marine species diversity. *Curr Biol*. 2012;22:2189–202.
98. van de Lande LS, Papaioannou A, Dunaway DJ. Geometric morphometrics aided by machine learning in craniofacial surgery. *J Orthod*. 2019:1–3.
99. Courtenay LA, Yravedra J, Huguet R, Aramendi J, Maté-González M, González-Aguilera D, et al. Combining machine learning algorithms and geometric morphometrics: A study of carnivore tooth marks. *Palaeogeogr Palaeoclimatol Palaeoecol*. 2019;522(March):28–39. doi:10.1016/j.palaeo.2019.03.007.
100. Courtenay LA, Huguet R, González-Aguilera D, Yravedra J. A hybrid geometric morphometric deep learning approach for cut and trampling mark classification. *Appl Sci*. 2020:10.
101. MacLeod N. Automated taxon identification in systematics: theory, approaches, and applications. London: CRC Press, Taylor & Francis Group; 2007.
102. MacLeod N. Morphometric approaches to the delineation and analysis of taxonomic and phylogenetic characters. In: *BioSyst.EU 2017*. Gothenburg: University of Gothenburg; 2017. p. 92.

103. Van Bocxlaer B, Schultheiß R. Comparison of morphometric techniques for shapes with few homologous landmarks based on machine-learning approaches to biological discrimination. *Paleobiology*. 2010;36:497–515.
104. Criminisi A. Machine learning for medical images analysis. *Med Image Anal*. 2016;33:91–3. doi:10.1016/j.media.2016.06.002.
105. Favret C, Sieracki JM. Machine vision automated species identification scaled towards production levels. *Syst Entomol*. 2016;41:133–43.
106. Monson TA, Armitage DW, Hlusko LJ. Using machine learning to classify extant apes and interpret the dental morphology of the chimpanzee-human last common ancestor. *PaleoBios*. 2018;35:1–20.
107. Vincze O, Vágási CI, Pap PL, Palmer C, Møller AP. Wing morphology, flight type and migration distance predict accumulated fuel load in birds. *J Exp Biol*. 2019;222:4–10.
108. Marinello MM, Bernard E. Wing morphology of neotropical bats: a quantitative and qualitative analysis with implications for habitat use. *Can J Zool*. 2014;92:141–7.
109. Derrickson EM, Ricklefs RE. Taxon-dependent diversification of life-history traits and the perception of phylogenetic constraints. *Funct Ecol*. 1988;2:417–23. <http://library1.nida.ac.th/termpaper6/sd/2554/19755.pdf>.
110. Hardy GH. *A mathematician's apology*. Cambridge: Cambridge University Press; 1940.
111. 111. Hardy GH. *A mathematician's apology*. Cambridge, U.K. Cambridge University Press; 1940.

Tables

Due to technical limitations, tables are only available as a download in the Supplemental Files section.

Figures

Figure 1

Major morphological features of the *Trithemis* forewing (top) and hindwing (bottom). C = costa; N = nodus, Cs = subcosta; R+M = radius and media; R₁ = first radius; R₂ = second radius; R₃ = third radius; R₄ = fourth radius; MA = media anterior; IR₃ = intercalary vein behind R₃, Cu = cubitus; CuP = cubitus posterior, Ps = pterostigma. Note that, unlike other dragonfly genera, The major structures features of the *Trithemis* forewing and hindwing are quite similar (see also Plate 1), facilitating detailed direct comparisons between individual wings and across species within this genus.

Figure 2

Positions of the 38 landmarks (white) and semilandmarks (black) used to quantify *Trithemis* forewing (top) and hindwing (bottom) form. Landmarks are located at the origins, intersections or peripheral termini of major wing veins (see Fig. 1 and S3 for formal definitions). Peripheral landmarks 1, 6, 13, 20, 25 and 31 were used to subdivide the wing outlines into five zones with the number of evenly spaced semilandmarks interpolated within these zone being sufficient to represent the zone-based peripheral outline geometry to an accuracy of > 95% across all specimens in the sample. This form-sampling system ensures accurate and even representation of the wing outline characterization and improves inter-semilandmark correspondence across specimens.

Figure 3

Ultrametric *Trithemis* cladogram (modified from Damm et al.[1]) showing phylogenetic relations for all species included in this investigation. This tree was inferred from ND1 and 16S genetic sequences based on maximum likelihood branch lengths. Time calibration was based on the r8s sequence using a 10 mya estimate for the basal *Trithemis* node which itself was based on data from the fossil record [52]. Landscape characters also from [1]. Note the manner in which the distributions of landscape and water-body preferences are arranged as phylogenetic grades rather than phylogenetic clades.

Original Images

Difference Images

Figure 4

Example of embedded, paired comparison within *Trithemis* species' hindwings, including within group contrasts (vertical arrows), and between group contrasts (diagonal arrows). Upper Fig: original images two randomly selected specimens from the *T. annulata* image set (209-5-05 and 209-6-04) and the *T. tropicana* image set (214-4-04 and 214-4-07). Rather than simply comparing image vectors, and so being restricted to small sample sizes, the embedding image comparison protocol used in this investigation focused on pairwise comparisons across all images in the sample (e.g., six comparisons for this four-image example) with structure of group similarities and differences being represented by Euclidean distances between image vectors. Lower Fig: summary of all six comparisons possible for this four-image example in terms of difference images and image distance values (d) both for the full-resolution (680 x 277 pixel) and reduced resolution (40 x 40 pixel) image comparisons, the latter values in parentheses. Here, images that exhibit a higher degree of difference will appear darker than images that

exhibit a higher degree of similarity with particular regions and/or structures of distinction being represented as darker areas and/or structures represented by dark highlights. Note that, for both image resolutions, images belonging to the same group typically exhibit shorter image distances relative to those belonging to different groups. Under this approach to the representation of the image (= specimen) similarity structure CNN training focuses on constructing a set of convolution filters that maximize between-groups image distances and minimize within-groups image distances.

Figure 5

Bootstrapped phylogenetic signal distributions for the *Trithemis* forewing and hindwing landmark-semilandmark shape datasets. Note that, in both cases, the observed values of the K_{mult} statistic fall well into the range expected for randomized phylogenies. Accordingly, there are no grounds for concluding either forewing or hindwing shape distributions reflect phylogenetic covariation in *Trithemis* to a statistically significant extent.

Forewings

Hindwings

Landscape Habitat

Water Body Habitat

Figure 6

Phylomorphospace plots for the subspaces formed by the first two Procrustes principal components calculated from the *Trithemis* forewing and hindwing landmark-semilandmark datasets. Projected positions of the *Trithemis* (tree tip) species color coded to reflect their landscape and water body habitat preferences. Note the number of tree branches that cross one another in this space, indicating that the distribution of wing shapes is not being controlled, or reflecting phylogenetic structure to any substantial extent. These results support the findings of the statistical phylogenetic signal test (Kmult) shown in Fig. 1.

Figure 7

Frequency histograms for *Trithemis* forewing and hindwing shape distinctions for species preferring different landscape and water body habitats as assessed from the landmark-semilandmark coordinate datasets. Note that, despite the broad ranges of shape overlap, a clear distinction exists between the central tendencies of these habitat-based wing shape groups. Statistical comparison of the between-groups separation relative to within-groups dispersion using a bootstrapped version of Hotelling's T2 test indicate that all four comparisons are significant at well beyond the $p = 0.95$ confidence level.

Figure 8

Vector displacement models for *Trithemis* habitat-category contrasts calculated from the linear discriminant analysis of forewing and hindwing datasets. For illustrative purposes displacement-vector lengths have been exaggerated by a factor of 4. Note similarity in the geometric character of habitat-group displacement patterns across the landscape and water body linear discriminant results.

Figure 9

Standardized major axis linear regression analyses of linear discriminant scores calculated for landscape and water body habitat differences for *Trithemis* forewing and hindwing landmark-semilandmark datasets. Although there is a broad range of scatter in these data and the species assigned to these different habitat categories differ somewhat (see Fig 3), a statistically significant linear trend does exist

for both datasets. Accordingly, the null hypothesis that selection for wing shape differences operated in a different manner for landscape and water body aspects of the environment can be rejected.

Figure 10

Frequency histograms for *Trithemis* forewing and hindwing shape distinctions for species preferring different landscape and water body habitats as assessed from the wing-image datasets. Note the minimum extent of shape-range overlap and clear distinction exists between the central tendencies of these habitat-based wing shape differences. Statistical comparison of the between-groups separation relative to within-groups dispersion using a bootstrapped version of Hotelling's T² test indicate that all four comparisons are significant at well beyond the $p = 0.95$ confidence level.

Figure 11

Color-coded comparisons of *Trithemis* forewing and hindwing mean shapes illustrating which wing pixels changed their brightness values least (blue) and most (red) along the linear discriminant vectors separating landscape (left) and water-body (right) species. As with the previous landmark-semilandmark results, the transitions from forested to open landscape habitats and temporary-standing and running water-body habitats involve essentially identical shape transitions, despite very different species cohorts being placed into these habitat groups. Note that the forewing and hindwing areas exhibiting the greatest changes in pixel brightness values constitute either very small well specified regions (forewings) or large areas (hindwings) that were either not included in the landmark-semilandmark sampling scheme (forewings) or could not be so included owing to the fact that the wing character most consistent with between-group difference was a pigmentation region (hindwings).

Figure 12

Frequency histograms for *Trithemis* forewing and hindwing shape distinctions for species preferring different landscape and water body habitats as assessed from the wing-image datasets using the embedded LeNet-5 deep-learning algorithm. Note the complete, and quite marked, separation that was achieved for these habitat-based wing morphology differences in the case of both forewings and hindwings. Comparison of these results suggest with those of the linear-model analyses (Figs 7 and 10) suggests that the geometries of morphological differences between these habitat groups exhibit a strongly non-linear character.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table1.pdf](#)
- [Table2.pdf](#)
- [Table3.pdf](#)
- [Table4.pdf](#)
- [Table5.pdf](#)