Caddisfly Species Identification via Wing Morphometrics and Potential Applications

Colette Christensen

ABSTRACT

Species identification is an important aspect of many biological studies, and something that remains incredibly challenging in entomology. Researchers strive to discover new methods of identification, using more technological resources in recent years. In this study, I determined if geometric morphometrics of caddisfly wing venation could identify differences between species in the family Hydropsychidae and sexes. To examine wing venation with the geometric morphometrics, I mounted fore and hind wings on glass slides and identified 45 landmarks on each individual. I analyzed differences between individuals via Generalized Procrustes Analysis and Principal Component Analysis. The tests revealed variation largely in the central crossveins; PCA scores showed at least 50% variation between wing landmarks, enough to distinguish species into three distinct clusters. Ideally, this data can be used in future applications of machine learning identification.

KEYWORDS

Geometric morphometrics, wing venation, hydropsychidae, machine learning

INTRODUCTION

Morphological characters have traditionally been used to identify species for applications such as biomonitoring and conservation; this approach continues with modern identification tools available to the public. For example, the website iNaturalist uses a photo of a sample to identify organisms in the environment. (iNaturalist.org). A similar application, Leafsnap, allows users to take photos of leaves to identify tree species (Kumar et al. 2012). Although these types of sites can be useful, especially for those with little scientific background, these identification methods are limited to organisms with highly distinguishing characteristics, or with larger libraries of images.

Other approaches, such as DNA barcoding, require cells of an organism to amplify a fragment of DNA to create a species "barcode" (Zhou et al. 2016); it is especially useful for building phylogenetic trees. Barcoding has a number of applications, such as unifying species catalogs within a database (Zhou et al 2016) or for invasive species identification (Armstrong & Ball 2005). Barcoding can quicken identification in laboratories; however, it is not necessarily an ideal method. Looking at only the genetic aspects of individuals hinders researchers' ability to look at emerging patterns, and therefore sacrifices the benefit of a holistic perspective into species' morphology and dynamics (Will & Rubinoff 2004). Because it requires a laboratory, accessibility to this type of identification is limited, restricting this field of science to a select few countries. Overall, each identification tool has its own set of complications, limiting its effectiveness.

In the field of entomology, identifying an ideal identification method remains a challenge. Insect species identification typically relies on morphology of the male genitalia. However, differences in genitalia can be incredibly subtle, and requires specialist knowledge, specimen preparation, a high level of magnification, and large portions of time to identify specimens. Additionally, females cannot always be confidently identified via genitalia, as they do not have distinct genitalic structures like males, and keys to females are not often developed. These factors can make it difficult to narrow specimens down to the species level.

The use of computer-assisted analysis may provide the precision needed for identification. In this case, geometric morphometrics can be used to identify caddisfly species. Geometric morphometrics uses landmark coordinates to analyze shape and differences between specimen. This type of analysis can be used to find specific morphological differences, rather than just differences in size or orientation. Geometric morphometrics have already been used in the field of

entomology, particularly in various beetle species; even wing venation variance among beetle genera have been explored (Su et al. 2015). However, due to the more complex structures and wing venation patterns of Trichoptera, geometric morphometrics has not been as readily applied to caddisflies, especially not to the species level.

In this study, I used geometric morphometrics to analyze variation in caddisfly wing venation, as wing size, shape, and venation patterns can be powerful indicators in species identification. My objectives are to determine whether venation (1) varies significantly within and among sexes of the same species, and (2) varies between species. I anticipate that both the different species and sexes will have distinct characteristics.

METHODS

Study subject

My study subjects were two species of caddisflies in the family Hydropsychidae. I did not identify the specimens by species beforehand, but rather sorted them by visible genitalic differences. The two male species, two morphologically different species by genitalia with unknown differences in wing venation, were designated as Species A and B depending on genitalia (Figure 1). Females were all grouped together without sorting by species, and designated as X; prior collection data confirmed that at least two distinct species were likely within the females group. These species tend to emerge around June, when the intermittent stream is dry and the number of adults in the environment is larger.



Figure 1. Genitalia of two male species. Male specimens were organized as Species A, or Species B, depending on their genitalic structures.

Study Site

My study location was Lower Curry Canyon at Mount Diablo. Mount Diablo is located in Contra Costa County in Northern California. Patina Mendez (University of California, Berkeley) and her undergraduate researchers collected approximately 13 different caddisfly families over the span of four years from four different sites, along one kilometer of Curry Creek, an intermittent stream. For this project, I worked exclusively with species of Hydropsychidae. The team used UV blacklight ethanol pan traps to collect insects monthly, which they then sorted by site, family and sex (Gladkov et al. 2018).

Data Collection - Wing Mounts and Landmarks

To prepare the caddisfly samples for the geometric morphometric analysis, I first created wing mounts for each specimen. I clipped the right fore and hind wings off of 15 individuals from Species A, B and X, and positioned them on glass square microscope coverslides for imaging and analysis (Blahnik & Holzenthal 2004). To mount the wings, I spread the wings out with forceps and clipped each wing at the base, taking precaution to not tear the wings. After both wings were clipped, I placed them onto a glass coverslip with a water droplet on top. Using small paintbrushes, I properly positioned the wings in the center of the slip, and unfolded the anal region of the hind

wing. After positioning, I removed excess water with a kimwipe, not touching the wings. Once done, I placed another glass coverslip on the first one, covering the wings. I then placed another folded kimwipe on top, and weighed the stack down with pennies to remove the rest of the water. After the coverslips dried overnight, I glued the corners of the glass coverslips together using Gelva, then glued the slips onto a glass slide. I organized the samples with an alpha-numerical code (A1, A2, B1, B2, etc.) to mark specimen number, sex, and species. Once the mounts were complete, I scanned the wings on an Epson Perfection C600 Photo Scanner at 2400 dpi, determined the most appropriate resolution for insect collection voucher imaging (Mendez 2018), to digitize them for analysis. Mounts were scanned in batches of 10, with a 1 cm scale bar for comparison.

To further prepare the mounts and begin the analysis, I edited the images and placed landmarks. I used Photoshop to crop the images and to digitally correct poorly mounted specimens with torn wings. Any wing mounts too torn or with missing pieces were not used in the analysis. Using ImageJ (https://www.indiana.edu/~g562/Handouts/Collecting%20Landmarks.pdf), I selected 45 different landmarks (Figure 2) at visible central crossveins and wing margin vertices on each fore and hind wing. Once the landmarks were established and coordinates collected, I converted the data tables into a .tps file to be analyzed in RStudio.

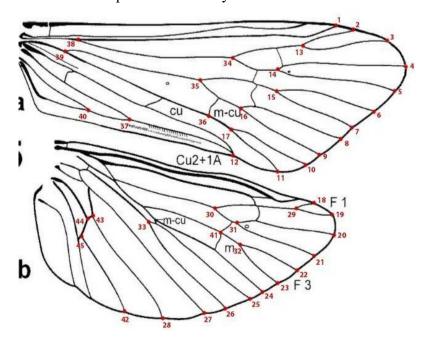


Figure 2. Location of the 45 set landmarks on the fore and hind wing. Although this model was based off a species not used in the experiment, *H. askalophos* (Schefter 2005), I selected landmarks that exist at the family level to make this model generalizable across the group.

Generalized Procrustes Analysis

To visualize the general location of variation on the wings, I ran a Generalized Procrustes Analysis (GPA). GPA scales and aligns landmarks from the different samples, and superimposes them the separate images for comparison (Sherratt 2014); this procedure removes differences due to wing size and orientation on the slide. Using RStudio, I ran this procedure, analyzing the variation for the landmarks of the two male species, as well as the males and the females combined in a full dataset (Adams et al. 2019).

Principal Component Analysis

To identify and condense critical points of variation onto one diagram, I conducted a Principal Component Analysis (PCA) on RStudio. PCA reduces the landmarks down into key components of variation, and highlights the largest differences between samples using warps. The PCA scores of each individual (PCA1, PCA2, etc.) are then graphed to visualize the relationships between species. Eigenvalues are used to calculate the variance explained from each PCA axis. I ran a PCA for the two male groups, and a second PCA with the full dataset (Sp A, Sp B, Sp X).

Statistical Analysis

To confirm the degree of difference between the three sample groups, I tested the data gathered from the the GPA and PCA with analysis of variances (ANOVA). Using pairwise comparisons, I confirmed whether the groups were significantly different from one another. Specifically, I ran Randomized Residuals in a Permutation Procedure (RRPP) with 1000 permutations (Collyer and Adams 2019, Collyer and Adams 2018). The statistical model used the value α =0.05.

Species Associations

To confirm the accuracy of the PCA groupings, I re-examined the genitalia of the caddisflies. I compared genitalia again to see what specific species each male was, as there was a

strong likelihood of another additional species, or a misclassified one. In addition, I also identified the females by species, to see if structures potentially matched the structures of the paired male from the PCA. Although the PCA confirmed that at least two species existed in the female group, other cryptic species may have also been mixed in, which I have not been able to corroborate until after the morphometric analysis.

RESULTS

Generalized Procrustes Analysis

Geometric morphometric analysis revealed slight differences between male species wing venation. In the Generalized Procrustes Analysis (GPA), the most variation resided in the crossveins in the center of the wing for both the fore and hind wings of the male samples (n=30) (Figure 3). Specifically, forewing landmarks 12-17, and 36-40 were the most variable. The hind wing landmarks 18, 28, 30-33, and 41-42 displayed the most variability. The outermost vertices where the veins meet the wing margin on the fore and hind wings were less variable.

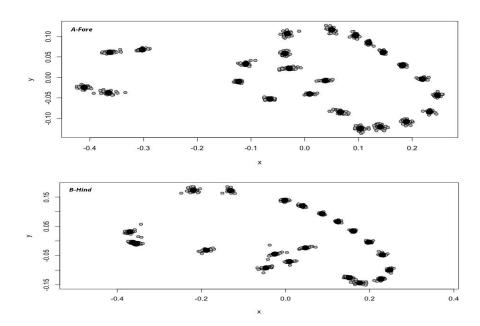


Figure 3. Procrustes analysis of (a) male fore and (b) male hind wings. These GPA graphs display variation between samples at distinct landmarks. Note that both images are flipped on the horizontal axis, where the top of the image is the bottom wing margin. Numbers for landmarks are available in Figure 2.

Much like the males, the GPA for the full dataset (males and females of both species) samples (n=45) highlighted middle crossveins as more variable than the outer vertices (Figure 4). The full dataset was similar in variability as the male dataset, with more variability displayed in the forewing landmarks 12-17, 36-40, and hind wing landmarks 18, 28, 30-33, 41-42. Most notably, however, the landmarks 1-4, and 8-11 on the outer wing margin of the forewings were also more variable. Overall, however, the landmarks did not vary widely between the samples from the same genus.

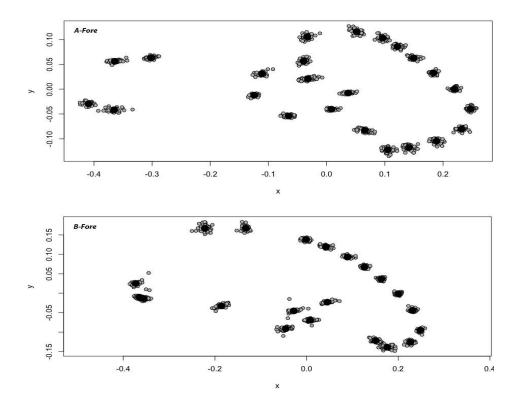


Figure 4. Procrustes analysis of full dataset for the (a) fore and (b) hind wings. These GPA graphs display variation between samples at distinct landmarks. Note that both images are flipped on the horizontal axis, where the top of the image is the bottom wing margin.

Principal Component Analysis

Similar to the GPA, Principal Component Analysis highlighted distinct variation between the two male caddisfly species. Species A and Species B clustered in two visibly different groups (Figure 5), which revealed some variation between their venation. Some of the individuals were outliers, most notably specimens 16, 20, 22 and 30 on the fore wing, and 19 on the hind wing.

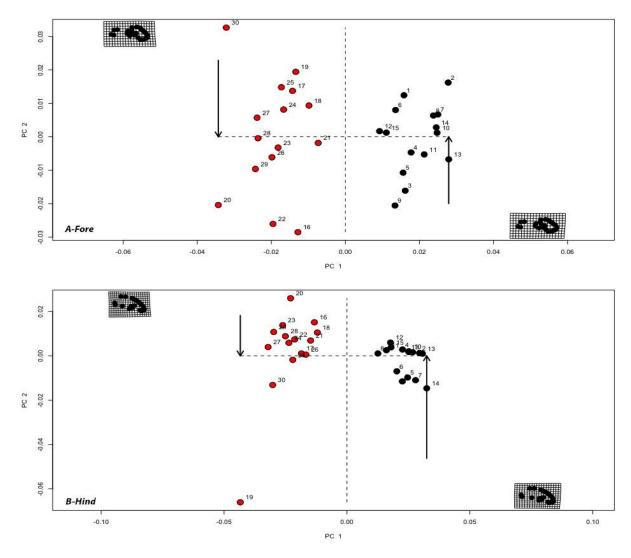


Figure 5. Principal component analysis of Species A (black) and Species B's (red). These PCA graphs display a separation between the two groups in both the (a) fore and (b) hind wings.

In the PCA of the SpA, SpB, and females (SpX), clustering also occurred as a result of differences in wing venation between the two species and females (Figure 6). In this analysis, the outliers were specimen 2 and 20 for the fore wing, and 17 and 19 for the hind wing.

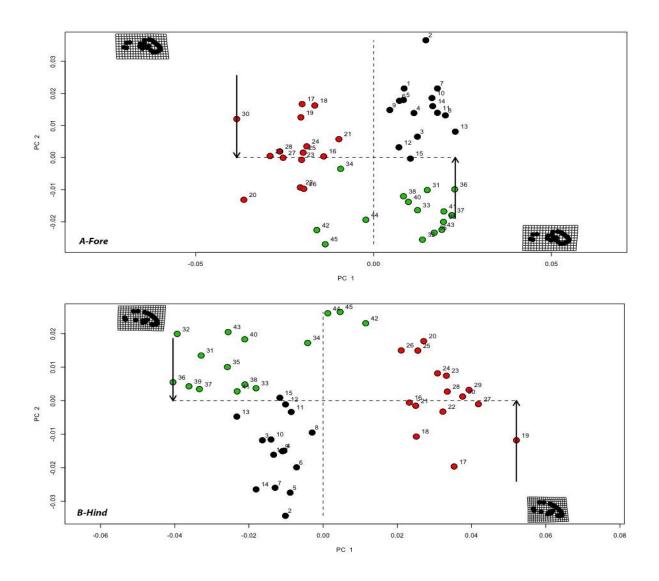


Figure 6. Principal component analysis of Species A (black), Species B (red), and unknown females (green). These PCA graphs display a separation between the three groups in both the (a) fore and (b) hind wings.

Further analysis revealed values for the combined PCA1 and PCA2 accounted for around half of the variation for both male (fore - 39.3%, hind - 57.6%) and full datasets (fore - 45.3%, hind - 57.7%) (Table 1). Hind wings for both datasets were more variable based on these values than the fore wings.

Table 1. Top two Principal Component Analysis scores. PCA1 and PCA2 accounted for most variation in the fore and hind wings of each sample group.

PCA1	PCA2	
34%	15.3%	
41.8%	15.8%	
26.4%	18.9%	
42.7%	15%	
	34% 41.8% 26.4%	34% 15.3% 41.8% 15.8% 26.4% 18.9%

Overall, the variation was not drastically different between the the male species and females, a conclusion confirmed by the warp diagrams (Figure 7). Much like the GPA, most of the variation was concentrated in the central crossveins. The slight distortion is apparent in only a small amount of general variation.

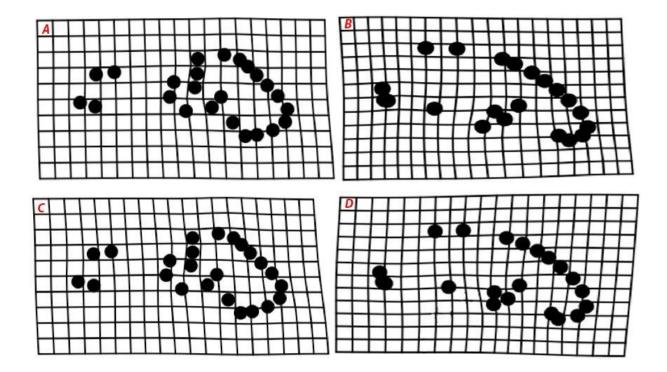


Figure 7. Warp diagrams of the (a) male fore wing, (b) male hind wing, (c) full dataset fore wing, and (d) full dataset hind wing. These warps displayed the landmarks where the variation is located.

Statistical Analysis

The statistical analysis revealed more in depth information about the wing venation variation between species and sexes. In the ANOVA results, the P values for all three pairwise relationships indicate that a significant difference exists not only between the two male species (Table 2), but between males and females as well (Table 3). The RRPP results also confirmed significance (Table 4).

Table 2. Pairwise results for fore wing. Distances between means, plus statistics.

	d	UCL (95%)	Z	Pr > d	
SpA:SpB	0.03859276	0.01764522	8.854829	0.001	
SpA:UnkF	0.03404098	0.01779681	7.35548	0.001	
SpB:UnkF	0.03778603	0.01798106	8.354132	0.001	

Table 3. Pairwise results for hind wing. Distances between means, plus statistics.

	d	UCL (95%)	Z	Pr > d	
SpA:SpB	0.04723585	0.02092305	8.530508	0.001	
SpA:UnkF	0.0303714	0.02144884	4.225457	0.002	
SpB:UnkF	0.05414844	0.02144884	9.942457	0.001	

Table 4. Random Residuals in a Permutation Procedure (RRPP). Results for the male fore wings, male hind wings, full fore wings and full hind wings.

	Df	SS	MS	Rsq	F	Z	Pr(>SS)
species (MF)	1	0.011171	0.0111707	0.30778	12.449	5.038	0.001
Residuals	28	0.025124	0.0008973	0.69222			
Total	29	0.036295					
species (MH)	1	0.016737	0.0167367	0.39055	17.943	4.9975	0.001
Residuals	28	0.026117	0.0009328	0.60945			

Total	29	0.042854					
species (F)	2	0.02038	0.0101899	0.35742	11.681	6.6349	0.001
Residuals	42	0.03664	0.0008724	0.64258			
Total	44	0.05702					
species (H)	2	0.030429	0.0152143	0.44942	17.142	5.841	0.001
Residuals	42	0.037277	0.0008876	0.55058			
Total	44	0.067706					

Species Associations

The species associations for the males confirmed that each individual was classified correctly as Species A or B. Through further analysis, I determined Species A to be *h. oslari*, and Species B to be *h. philo*. Although significant outliers did emerge on the PCA, (Figures 8, 9) further measurement and identification confirmed the individuals to be the same species. In the female classification, I identified at least two different species in the female specimens.

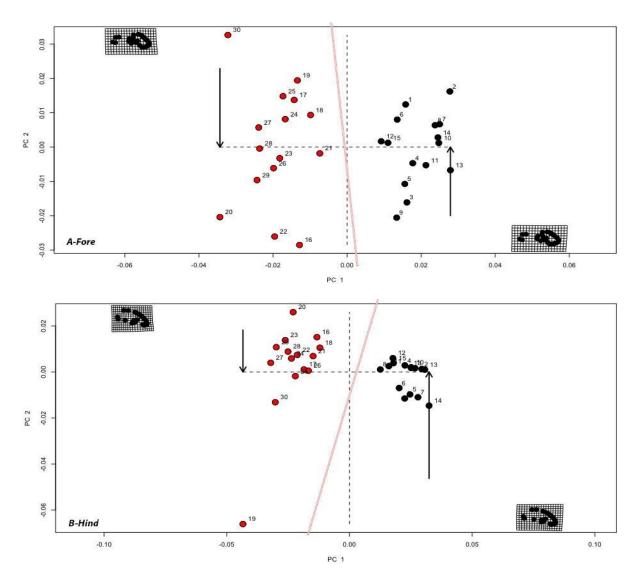


Figure 8. PCA of the male sample group with clustering. A distinguishing line separates the clusters of Species A and B, on the (a) fore and (b) hind wings.

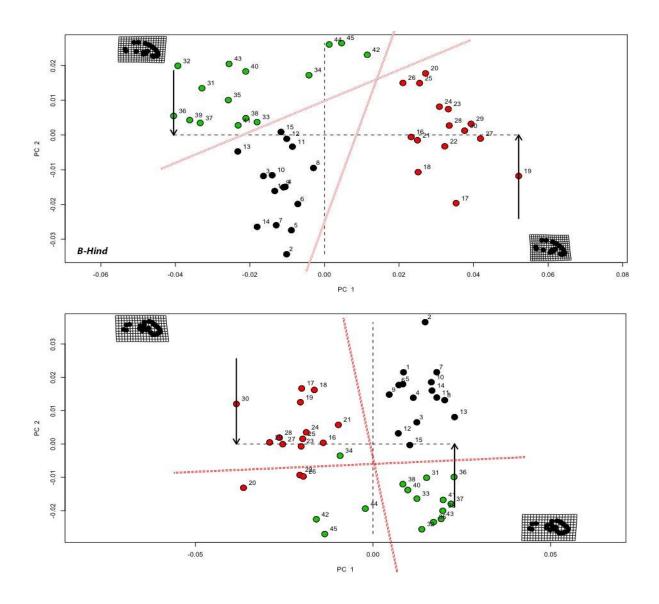


Figure 9. PCA of the full dataset sample group with clustering. Distinguishing lines separate the clusters of Species A and B, and the Unknown Females, on the (a) fore and (b) hind wings.

DISCUSSION

According to the relationship determined between the landmarks of the caddisfly individuals, intra- and interspecific differences exist within the *Hydropsychidae* family. Both the Generalized Procrustes Analysis and the Principal Component Analysis confirms variation amongst the two species, as well as between females and males. Pairwise comparisons of the points of the PCA also indicate significant variation between the three. However, females require further analysis and identification. In addition, errors from initial identification and sample preparation

will also need to be taken into consideration. Regardless, this experiment can form baseline evidence for wing morphology to be valuable for digital species identification methods such as machine learning.

Differences in Wing Size

Although Procrustes analysis revealed some morphological differences the male and female wings, as well as between species, variation overall was not significantly conclusive. The full dataset GPA showed wider variation in the wing margins in comparison to the male dataset, highlighting the wing size differences between sexes. This sexual dimorphism may be a result of differences in development to environmental differences. For the butterfly Speryeria diana, strong dimorphism exists between sexes, and for individuals at different elevation and sites, with narrower, angular forewings linked to differing flight behavior (Wells et al. 2018). For Limnephilid caddisflies, sexual dimorphism in wing size occurs depending on the species, males or females may have larger wing spans to assist in their dispersal (Müller-Peddinghaus & Hering 2013). Variation also occurred in the central crossveins of the wings, a potential indicator of variation due to environmental or evolutionary pressures. In an investigation on various Chinese dung beetle tribes, variables such as wing shape, disparity and aspect ratio were studied to determine phylogenetic differences (Bai et al. 2012). Researchers additionally used reconstructions of the beetles' ancestral forms to find evolutionary patterns. While the differences in the caddisfly wing venation could be contributed to evolutionary development, it would be impossible to confirm without more historical information on both the environment and the species in said environment.

Differences in Wing Morphology

Similarly to the GPA, the Principal Component Analysis further revealed clustering by group, and subsequently variation for each species and sex. The separation between Species A and B in the male dataset, as well as the separation between the two male clusters and the unknown females in the full dataset, suggests that enough distinction exists between the wing venation to distinguish group membership. However, the four warps showed potentially contrasting information, that the venation is not notably variable beyond the central crossveins. This variation

was also confirmed by the PCA scores, as the amount of variation explained by PCA1 and PCA2 only accounted for less than 50% of the variation, especially for the fore wings. This may be due to the higher number of morphological similarities between genus, which would make classifying variation at the species level more difficult (Müller-Peddinghaus & Hering 2013). With the PCA graphs, I was also able to categorize most of the females as either Species A or B, based on their close association. However, a couple of individual specimens needed to be investigated further in the later steps. Overall, while the variation was not strong, it provided enough evidence to confirm inter- and intraspecific differences. The pairwise and RRPP tests also confirmed significant variation between the three sample groups. According to the ANOVA results, all three comparisons had a p-value that was less than 0.05; the RRPP results also had p-values of 0.001.

Species Association

By analyzing the genitalia of the individual specimens, I was able to confirm each caddisfly species. The males were more easily identified via a caddisfly identification manual that revealed the distinctions between genitalic structures. The females, however, could not be identified as easily. With no complex genitalic structure, identification was more limited. The fact that these designations were limited may be a strong indicator of lacking information. In future experiments, barcoding may provide more efficient and accurate results for species classification. Overall, geometric morphometrics via wing venation may not be as particularly helpful at identifying species, due to less variation; however, this method could be more applicable on the family or order level.

Limitations and Future Directions

Most of the limitations resulted from lack of information on the specimens, as well as potential data errors. Compared to identified natural history museum specimens, the original sample population from Curry Canyon was only preliminarily identified to the family level; potential errors may have arisen from identifying to the species level, as there is no formal species list for the stream or county. In addition, genitalia could not be properly cleared, as no official key for identification exists for Trichoptera in California. Instead, an unpublished manual (Burdick

2010) was utilized in identification for this study, some parts of which did not have sufficient information on the species.

In the future, I would confirm the species identification before the experiment begins, so I could see whether the GPA and PCA were different to to species classification, rather than analysis error. The study would be enhanced by including individuals from a different genus, or different areas of the streams. There is a good likelihood that wing development may have differed extremely due to flight and dispersal patterns. Prior research has already confirmed a strong correlation for morphology between biological and environmental variability (Landerio et al. 2012). Using GIS, I could also confirm a relationship between these factors for populations in this region, and how other variables such as stream intermittency contribute.

Broader Implications

This experiment explored the relationship between wing morphology and species identification, but also serves as the baseline information for efficient identification tools and resources. Machine learning is a newly emerging identification approach in the field of biology, using a database of near perfect images to "teach" a computer program to identify novel images. For leaf venation, heat maps developed for leaf venation accurately identified specimens to the order and family level (Wilf et al. 2016). Unfortunately, this approach required over 7000 images from over 2000 genera, a sample size not feasible for caddisflies. Future applications for aquatic insect identification can explore machine learning, using the dataset started with caddisfly wings. Machine learning could be the future in species identification for insect species, especially when paired with traditional approaches relying on specialist knowledge of male genitalia. This method could make identification accessible to researchers, more encompassing of sexual dimorphism, and more widely applicable to other insect studies.

ACKNOWLEDGEMENTS

I want to thank my mentor Patina Mendez for her time, care and patience in helping me with my project, as well as the rest of Team 175: Kurt Spreyer, Leslie McGinnis, and Ellen Plane. I also

want to thank my ES peers for their editing and suggestions, and my family and friends for their love and support.

REFERENCES

- iNaturalist.org. (n.d.). . https://www.inaturalist.org
- Kumar, N., P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I. C. Lopez, and J. V. B. Soares. 2012. Leafsnap: A Computer Vision System for Automatic Plant Species Identification. Pages 502–516 in A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, editors. Computer Vision ECCV 2012. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Zhou, X., P. B. Frandsen, R. W. Holzenthal, C. R. Beet, K. R. Bennett, R. J. Blahnik, N. Bonada, D. Cartwright, S. Chuluunbat, G. V. Cocks, G. E. Collins, J. deWaard, J. Dean, O. S. Flint, A. Hausmann, L. Hendrich, M. Hess, I. D. Hogg, B. C. Kondratieff, H. Malicky, M. A. Milton, J. Morinière, J. C. Morse, F. N. Mwangi, S. U. Pauls, M. R. Gonzalez, A. Rinne, J. L. Robinson, J. Salokannel, M. Shackleton, B. Smith, A. Stamatakis, R. StClair, J. A. Thomas, C. Zamora-Muñoz, T. Ziesmann, and K. M. Kjer. 2016. The Trichoptera barcode initiative: a strategy for generating a species-level Tree of Life. Philosophical Transactions of the Royal Society B: Biological Sciences 371:20160025.
- Armstrong, K. F., and S. L. Ball. 2005. DNA barcodes for biosecurity: invasive species identification. Philosophical Transactions of the Royal Society B: Biological Sciences 360:1813–1823.
- Will, K. W., and D. Rubinoff. 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics 20:47–55.
- Su, J., K. Guan, J. Wang, and Y. Yang. 2015. Significance of hind wing morphology in distinguishing genera and species of cantharid beetles with a geometric morphometric analysis. ZooKeys 502:11–25.
- Gladkov, G., C. Hubble, K. Will, and P. Mendez. 2018. Trichoptera Emergence Patterns in Lower Curry Canyon Near Mount Diablo. UC Berkeley CNR.
- Blahnik, R. J., and R. Holzenthal. 2004. Collection and curation of Trichoptera, with an emphasis on pinned material. Neotropical Trichoptera Newsletter 1:21.
- Mendez, P. K., S. Lee, and C. E. Venter. 2018. Imaging natural history museum collections from the bottom up: 3D print technology facilitates imaging of fluid-stored arthropods with flatbed scanners. ZooKeys 795: 49-65.
- Schefter, P. 2005. Re-evaluation of genera in the subfamily Hydropsychinae (Trichoptera:

- Hydropsychidae). Aquatic Insects 27:2. 133-154.
- Geometric Morphometrics. Department of Earth and Atmospheric Sciences. Indiana University Bloomington. (n.d.). . 1-5. https://www.indiana.edu/~g562/Handouts/Collecting%20Landmarks.pdf
- Sherratt, E. 2014. Quick Guide to Geomorph v.2.0. 1-71.
- Adams, D. C., M. L. Collyer, and A.Kaliontzopoulou. 2019. Geomorph: Software for geometric morphometric analyses. R package version 3.1.0. https://cran.r-project.org/package=geomorph
- Collyer, M. L. and D. C. Adams. 2019. RRPP: Linear Model Evaluation with Randomized Residuals in a Permutation Procedure. https://cran.r-project.org/web/packages/RRPP
- Collyer, M. L. and D. C. Adams. 2018. RRPP: RRPP: An R package for fitting linear models to high-dimensional data using residual randomization. Methods in Ecology and Evolution. 9(2): 1772-1779.
- Wells, C., A. Munn, and C. Woodworth. 2018. Geomorphic Morphometric Differences between Populations of Speyeria diana (Lepidoptera: Nymphalidae). Florida Entomologist 101:195–202.
- Bai, M., R. G. Beutel, K.-Q. Song, W.-G. Liu, H. Malqin, S. Li, X.-Y. Hu, and X.-K. Yang. 2012. Evolutionary patterns of hind wing morphology in dung beetles (Coleoptera: Scarabaeinae). Arthropod Structure & Development 41:505–513.
- Müller-Peddinghaus, E., and D. Hering. 2013. The wing morphology of limnephilid caddisflies in relation to their habitat preferences. Freshwater Biology 58:1138–1148.
- Landeiro, V. L., L. M. Bini, A. S. Melo, A. M. O. Pes, and W. E. Magnusson. 2012. The roles of dispersal limitation and environmental conditions in controlling caddisfly (Trichoptera) assemblages: Environmental and spatial factors controlling caddisfly distributions. Freshwater Biology 57:1554–1564.
- Wilf, P., S. Zhang, S. Chikkerur, S. A. Little, S. L. Wing, and T. Serre. 2016. Computer vision cracks the leaf code. Proceedings of the National Academy of Sciences 113:3305–3310.