Supplementary material for "Planar morphometrics using Teichmüller maps"

Gary P. T. Choi and L. Mahadevan

S1 The algorithms of the proposed Teichmüller morphometric framework

Suppose S_1, S_2 are two planar shapes with corresponding landmark points at the interior and on the boundaries. We first compute a boundary mapping $\varphi : \partial S_1 \to \partial S_2$ that matches the curvatures of the two boundary curves as well as the prescribed landmark points $\{l_k^{bdy_1}\}_{k=1}^n \leftrightarrow \{l_k^{bdy_2}\}_{k=1}^n$ on the two boundary curves. This is achieved by representing the accumulated curvature of the two boundary curves as two increasing functions and matching them using the SRVF dynamic warping method [1, 2]. With the curvature-guided boundary correspondence, we then compute a landmark-matching Teichmüller mapping $f : S_1 \to S_2$ that matches the interior landmark points $\{l_k^{int_1}\}_{k=1}^m \leftrightarrow \{l_k^{int_2}\}_{k=1}^m$ using the Quasi-conformal (QC) Iteration method by Lui et al. [3, 4].

One key algorithm in the QC Iteration method is called the Linear Beltrami Solver (LBS). Given two planar domains D_1 and D_2, a boundary correspondence $\phi : \partial D_1 \to \partial D_2$, a set of interior landmark correspondences $\{l_i^1\}_{i=1}^m \leftrightarrow \{l_i^2\}_{i=1}^m$ on D_1 and D_2 respectively, and a complex-valued function $\mu : D_1 \to \mathbb{C}$, the LBS produces a quasi-conformal mapping $h : D_1 \to D_2$ satisfying $h(l_i^1) = l_i^2$ for $i = 1, \ldots, m$ such that the Beltrami coefficient of h resembles μ as much as possible. More explicitly, suppose $h = u + iv$ and $\mu = \rho + i\tau$. h is computed by solving the following system of equations

$$\begin{cases} \nabla \cdot (A \nabla u) & = 0, \\ \nabla \cdot (A \nabla v) & = 0, \\ h(l_i^1) & = l_i^2, i = 1, \ldots, m, \\ h|_{\partial D_1} & = \phi, \end{cases} \tag{S1}$$

where $A = \begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_3 \end{pmatrix}$, $\alpha_1 = \frac{(\rho-1)^2+\tau^2}{1-\rho^2-\tau^2}$, $\alpha_2 = \frac{-2\tau}{1-\rho^2-\tau^2}$, and $\alpha_3 = \frac{(\rho+1)^2+\tau^2}{1-\rho^2-\tau^2}$. We denote the above process of applying the Linear Beltrami Solver by $h = \text{LBS}\left(\phi, \mu, \{l_i^1\}_{i=1}^m, \{l_i^2\}_{i=1}^m\right)$.

The procedure of the QC Iteration method is outlined in Fig. 1 of the paper. In our problem setting, with the curvature-guided boundary correspondence, the method starts by solving for an as-conformal-as-possible landmark-matching mapping $f_0 = \text{LBS}\left(\varphi, 0, \{l_k^{int_1}\}_{k=1}^m, \{l_k^{int_2}\}_{k=1}^m\right)$. Then, the Beltrami coefficient μ of f_0 is computed. If $|\mu|$ is not constant, the method performs a smoothing procedure on μ (denoted the result by $\tilde{\mu}$), and reconstructs an updated quasi-conformal mapping $f = \text{LBS}\left(\varphi, \tilde{\mu}, \{l_k^{int_1}\}_{k=1}^m, \{l_k^{int_2}\}_{k=1}^m\right)$. The above smoothing and reconstructing process continues until the resultant mapping f is Teichmüller, in other words, the norm of the associated Beltrami coefficient μ_f is constant over the entire domain. The convergence of the QC Iteration method has been proved [5].

With the Teichmüller mapping f and the associated Beltrami coefficient μ_f, a similarity score between S_1 and S_2 can be defined by $1 - |\mu_f|$. Our proposed method for comparing two planar shapes is summarised as Algorithm S1.

Algorithm S1 Landmark-matching curvature-guided Teichmüller map for planar shapes

Input: Two planar shapes S_1, S_2, with interior landmarks $\{l_k^{int_1}\}_{k=1}^m$, $\{l_k^{int_2}\}_{k=1}^m$, boundary landmarks $\{l_k^{bdy_1}\}_{k=1}^n$, $\{l_k^{bdy_2}\}_{k=1}^n$.
Output: A landmark-matching, curvature-guided Teichmüller mapping $f : S_1 \to S_2$, a similarity score s.

1: Compute a curvature-guided boundary mapping $\varphi : \partial S_1 \to \partial S_2$ satisfying $\varphi(l_k^{bdy_1}) = l_k^{bdy_2}$ for all $k = 1, 2, ..., n$.
2: With the boundary correspondence $\varphi : \partial S_1 \to \partial S_2$, compute a landmark-matching Teichmüller mapping $f : S_1 \to S_2$ satisfying $f|_{\partial S_1} = \varphi$ and $f(l_k^{int_1}) = l_k^{int_2}$ for all $k = 1, 2, ..., m$.
3: Compute the Beltrami coefficient μ_f of the mapping f. The score s is given by $1 - |\mu_f|$.

Now, suppose we are given a set of p planar shapes $\{S_i\}_{i=1}^p$. To cluster the shapes, we first deploy Algorithm S1 on all pairs of shapes. This gives us a similarity matrix M, where the (i, j)-entry of M is the similarity score of the landmark-matching curvature-guided Teichmüller mapping $f_{ij} : S_i \to S_j$. Then, we proceed to adaptively threshold the similarity matrix M. Our proposed adaptive thresholding algorithm is summarised in Algorithm S2.

Algorithm S2 Adaptive thresholding

Input: A $n \times n$ similarity matrix M, a thresholding parameter λ.
Output: A thresholded matrix where all entries are 0, $\frac{1}{2}$ or 1.

1: Set $M^0 = M$.
2: Set $k = 0$.
3: **repeat**
4: Update k by $k + 1$.
5: For each row i, denote $\nu_i^k = \overline{M_i^k} + \lambda\sigma_i^k$, where $\overline{M_i^k}$ and σ_i^k are respectively the mean and the standard deviation of $\{M_{it}^k\}_{t=1}^n$. Set

$$M_{ij}^k = \begin{cases} 1 & \text{if } M_{ij}^{k-1} \geq \nu_i^{k-1}, \\ 0 & \text{otherwise.} \end{cases}$$

6: Update M^k by $\frac{M^k + (M^k)^T}{2}$.
7: **until** $M^k = M^{k-1}$.

It can be proved that our proposed adaptive thresholding algorithm converges for all similarity matrices, and the rate of convergence can be estimated. The proof is as follows.

Theorem. Algorithm S2 converges for any similarity matrix M and for any thresholding parameter λ.

Proof. Note that in each iteration, we have

$$1 \to 1 \text{ and } 0 \to 0 \text{ and } \frac{1}{2} \to \begin{cases} 1, \\ \frac{1}{2}, \\ 0. \end{cases} \tag{S2}$$

Let n_k be the number of $\frac{1}{2}$ in the matrix M^k. It is easy to see that the sequence $\{n_k\}_{k=1}^\infty$ is non-increasing. Also, note that

$$0 \leq n_k \leq n_1 \leq |M|, \tag{S3}$$

where $|M|$ denotes the total number of entries in M. By Monotone Convergence Theorem [6], $\{n_k\}_{k=1}^{\infty}$ converges and $0 \leq \lim_{k\to\infty} n_k \leq n_1$.

Now, it is easy to see that if $n_{K+1} = n_K$ then $M^{K+1} = M^K$, which indicates the convergence. By the symmetry of M^k, if $n_{K+1} < n_K$, then $n_K - n_{K+1}$ must be a multiple of 2. Hence, the maximum number of iterations needed for achieving convergence is bounded above by $n_1/2$. ■

We then compute the generalised Erdős number (GEN) of the thresholded matrix using [7] and apply the GEN-based community detection method [8] to obtain the classification result. The Teichmüller-based classification scheme is summarised as Algorithm S3.

Algorithm S3 Classification of planar shapes via landmark-matching curvature-guided Teichmüller mappings

Input: A set of planar shapes $\{S_i\}_{i=1}^p$ with prescribed landmark correspondences.
Output: Community labels $\{l_i\}_{i=1}^p$.

1: Apply Algorithm S1 for all pairs of shapes (S_i, S_j), $1 \leq i, j \leq p$. Denote the similarity score between them by s_{ij}.
2: Construct a $p \times p$ similarity matrix $M = (s_{ij})$.
3: Apply Algorithm S2 on M and obtain the thresholded matrix.
4: Apply the GEN-based community detection method with the thresholded matrix and obtain the community labels $\{l_i\}_{i=1}^p$.

This completes our Teichmüller morphometric framework for planar shapes.

S2 Discretisation and implementation

In our case studies, we use a finite element discretisation of planar shapes as illustrated by Fig. S1. Every planar shape is discretised as a triangular mesh with regular triangle elements, with multiple vertices highlighted as landmarks. The landmark selection on the planar shapes is performed using the image processing software ImageJ [9]. Our proposed algorithms are implemented in MATLAB. All computations are performed on a PC with Intel i7-6700K CPU and 16 GB RAM. For triangular meshes with approximately 9000 triangle elements, the average time taken for computing each landmark-matching curvature-guided Teichmüller mapping using Algorithm S1 is less than 5 seconds. For Algorithm S3, the computations are further parallelised using the Parallel Computing Toolbox in MATLAB. The codes of our proposed algorithms are available at `http://scholar.harvard.edu/choi/files/tm.zip`.

S3 Classification of Hawaiian *Drosophila* wings

In the first case study in our paper, we apply our proposed Teichmüller morphometric framework (Algorithm S3) on *Drosophila* wings in the Hawaiian Drosophila Wing Database [10]. Table S1 lists the wing specimens used in our study. Fig. S2 shows the clustering result with specimen labels. Fig. S3, Fig. S4 and Fig. S5 show the three communities formed using our framework. Here, for every species with multiple specimens in the dataset, we put the species in one of the three figures according to the mode of the community labels of the specimens. Species with multiple modes of community labels are not included in the figures.

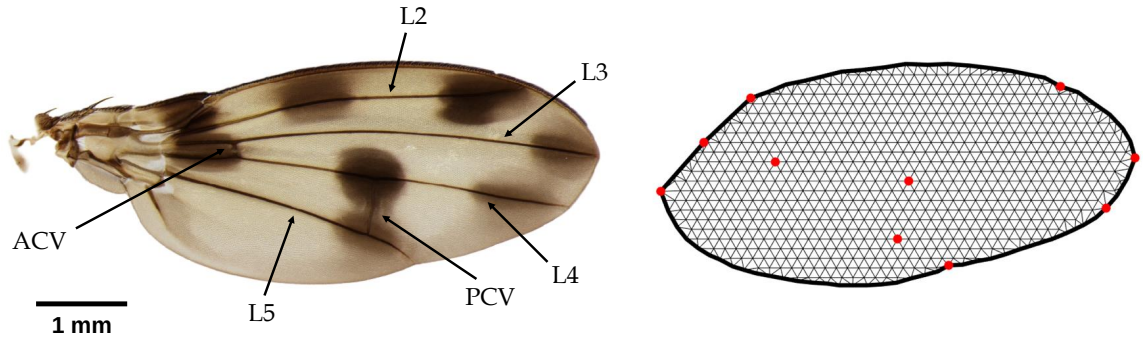

Figure S1: A Hawaiian *Drosophila* wing and the finite element discretisation of it. Red: Landmark points of the intersections between the longitudinal veins L2, L3, L4, L5, the anterior cross-vein (ACV), the posterior cross-vein (PCV) and the boundary. Wing image courtesy of the Hawaiian Drosophila Wing Database [10].

S4 Comparison between our method and the existing morphometric approaches

Fig. S6 shows a comparison between our curvature-guided, landmark-matching Teichmüller mapping approach (Algorithm S1) and four other morphometric approaches, including direct mapping, Procrustes superimposition [11], least-square conformal mapping and Thin Plate Spline [12]. Consider mapping a *D. punalua* wing onto a *D. silvestris* wing with 10 landmark correspondences. Denote the landmarks on the *D. punalua* wing and the *D. silvestris* wing by $\{\vec{p}_i\}_{i=1}^{10}$ and $\{\vec{q}_i\}_{i=1}^{10}$ respectively, where all $\vec{p}_i$ and $\vec{q}_i$ are vectors in $\mathbb{R}^2$. For the direct mapping, we directly overlay the *D. punalua* wing onto the *D. silvestris* wing and compute the intensity difference. The Procrustes and Thin Plate Spline mappings are computed using the MATLAB built-in functions `procrustes` and `tpaps` respectively, with the given landmark correspondences $\vec{p}_i \leftrightarrow \vec{q}_i, i = 1, \ldots, 10$. The Procrustes method computes an optimal linear transformation

$$f(\vec{x}) = T(b\vec{x}) + \vec{c}, \tag{S4}$$

where T is a 2×2 matrix for rotation and reflection, b is a scalar for scaling, and $\vec{c}$ is a vector for translation, such that the sum of squared errors $\sum_{i=1}^{10} \|f(\vec{p}_i) - \vec{q}_i\|^2$ is minimised. The Thin Plate Spline method computes the unique minimiser f of the weighted sum $\alpha E(f) + (1-\alpha)R(f)$, where $\alpha \in [0,1]$ is a weighting factor,

$$E(f) = \sum_{i=1}^{10} \|f(\vec{p}_i) - \vec{q}_i\|^2 \tag{S5}$$

is the landmark mismatch error measure and

$$R(f) = \iint \left[\left(\frac{\partial^2 f}{\partial x_1^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x_1 \partial x_2} \right)^2 + \left(\frac{\partial^2 f}{\partial x_2^2} \right)^2 \right] dx_1 dx_2 \tag{S6}$$

is the roughness measure. In our experiment, we set $\alpha = 1/2$. For the least-square conformal mapping, we look for a third order polynomial $f(z) = a_0 z + a_1 z + a_2 z^2 + a_3 z^3$ that deforms the *D. punalua* wing and matches the *D. silvestris* wing as much as possible, where $a_0, a_1, a_2, a_3 \in \mathbb{C}$ are parameters to be

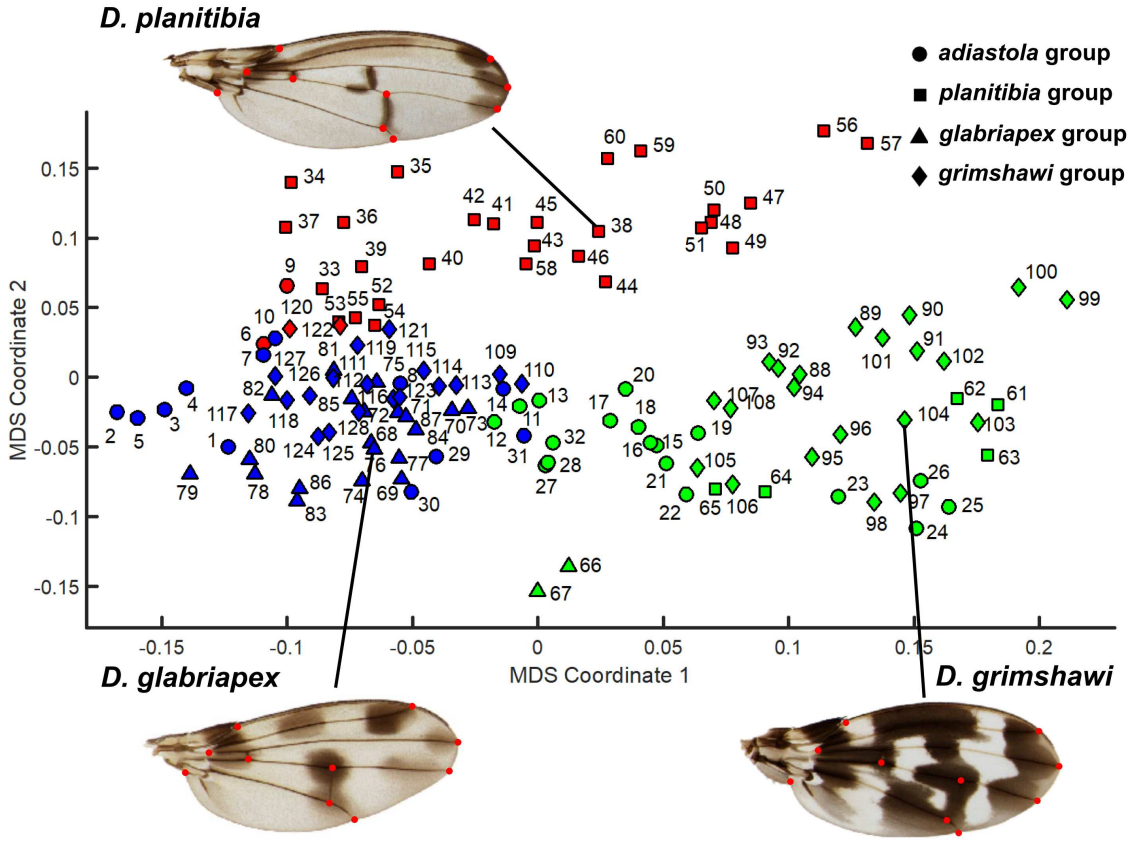

Figure S2: Visualising the community detection result obtained by our Teichmüller morphometric framework on the MDS coordinate plane. The nodes are coloured based on their communities. Blue: Community 1. Red: Community 2. Green: Community 3. The shapes of the nodes represent their phylogenetic groupings. Circle: *adiastola* group. Square: *planitibia* group. Triangle: *glabriapex* group. Diamond: *grimshawi* group. The number beside each node represents the specimen number specified in Table S1. The wing images shown are adapted from [10].

determined. The optimal f is computed by minimising the following objective function:

$$E(f) = \sum_{i=1}^{10} |f(z_i) - w_i|^2 , \tag{S7}$$

where $\{z_i\}_{i=1}^{10}, \{w_i\}_{i=1}^{10}$ are the complex representations of $\{\vec{p}_i\}_{i=1}^{10}$ and $\{\vec{q}_i\}_{i=1}^{10}$ respectively. The optimal solution is obtained using the MATLAB function `lsqnonlin`, with the Levenberg–Marquardt algorithm [13] being used.

As shown in Fig. S6, since the wings are with different size, shape and landmark locations, directly overlaying the two wings does not help analysing the difference in their geometries. Also, neither the Procrustes superimposition, the least-square conformal mapping method nor the Thin Plate Spline method can exactly match all landmarks and the wing boundaries. By contrast, our Teichmüller method guarantees exact landmark and wing boundary matching. This reflects the advantage of our proposed method for planar morphometrics.

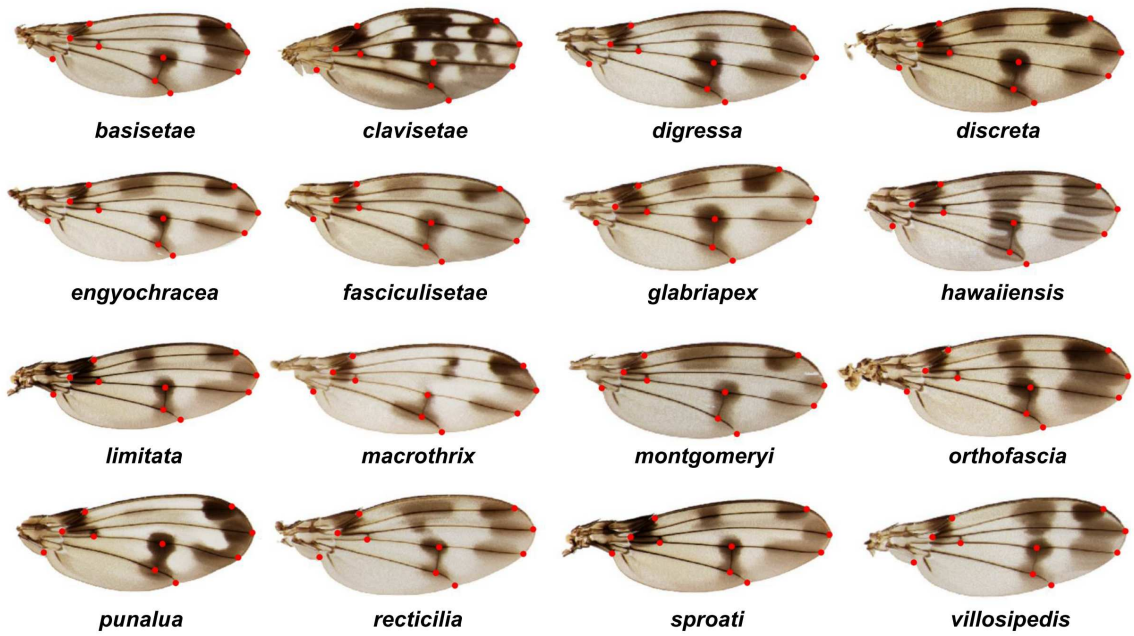

Figure S3: The species classified as Community 1 by our method. The images are adapted from the Hawaiian Drosophila Wing Database [10] (not to scale).

S5 Robustness of our method with respect to the thresholding parameter

Recall that Algorithm S2 requires an input of the thresholding parameter λ for computing the thresholded quasi-conformal similarity matrix. Fig. S7 shows the thresholding result with various λ. Using the thresholding results, we apply the GEN-based community detection algorithm [8] and obtain the clustering results as shown in Fig. S8. It can be observed that a very small value of λ may oversimplify the quasi-conformal similarity matrix and lead to the formation of large communities, while a very large value of λ may chop most entries to 0 and lead to many small communities. Nevertheless, the community detection results are consistent with various moderate values of λ between 0.5 and 1.5. This shows that our method is robust with respect to a reasonable range of λ. In practice, we take $\lambda = 1$.

S6 Robustness of our method with respect to the community detection method

We are also interested in the robustness of our proposed method with respect to the community detection method used in Algorithm S3. Fixing $\lambda = 1$, we apply various clustering algorithms, including the GEN-based community detection algorithm [8], hierarchical clustering, density-based clustering and spectral clustering, on the thresholded similarity matrix and obtain the community detection results. For hierarchical clustering, the MATLAB function `cluster` in the Statistics and Machine Learning Toolbox is used, with the required input maximum number of clusters set to be 4. For density-based clustering, the DBSCAN algorithm [14] is used. For the spectral clustering, the algorithm by Shi and

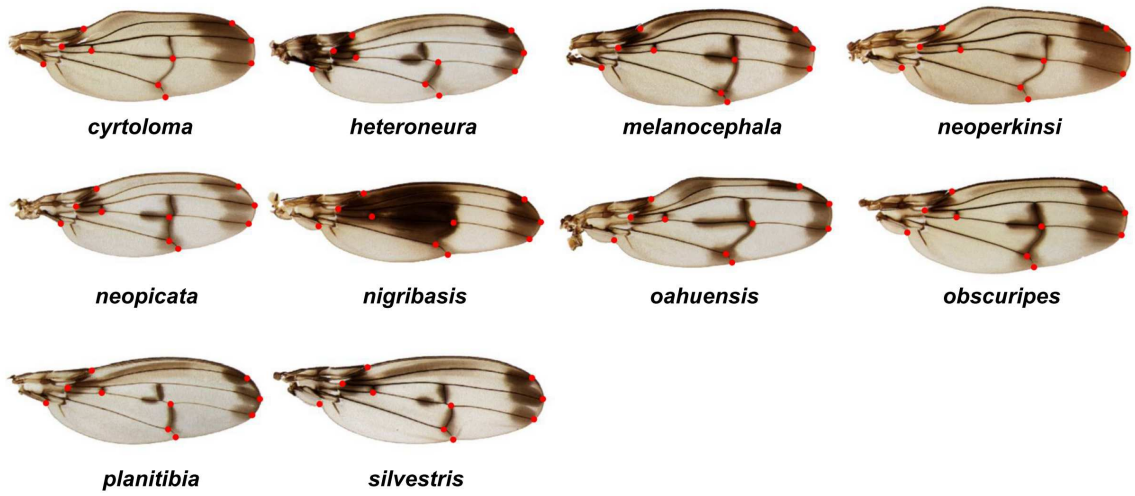

Figure S4: The species classified as Community 2 by our method. The images are adapted from the Hawaiian Drosophila Wing Database [10] (not to scale).

Malik [15] in the MATLAB Community Detection Toolbox is used, with the required input number of clusters set to be $n = 3$ and $n = 4$.

Fig. S9 shows the community detection results. It can be observed that the algorithms produce consistent clustering results, which suggests that our proposed method is robust with respect to the community detection method. It is noteworthy that the GEN-based community detection method is advantageous as it does not require any prescribed information of the number of clusters.

S7 Detecting subtle shape dissimilarities using our method

Our Teichmüller-map morphometric method is capable of detecting subtle shape dissimilarities such as microevolutionary patterns and bilateral asymmetry. In Fig. S10, we compare the wings of a specimen of *D. heteroneura* by computing the landmark-matching Teichmüller mapping from the right wing to the left wing. The resulting quasi-conformal similarity score $1 - |\mu_f|$ is approximately 0.944, indicating a high overall similarity between the two wings.

To study the subtle dissimilarity between them, we consider the intensity difference between the mapping result and the left wing and observe a subtle difference between the longitudinal veins on the two wings, especially L2 and L5. It can also be observed that the pigment patterns near the distal tips of L2, L3 and L4 on the two wings are almost identical, while those near the cross-veins are slightly different.

Furthermore, analogous to the study of the temporal development of Lepidoptera wings discussed in Section 4(b) of the paper, we can study the bilateral asymmetry by considering the deformation of small circles to small ellipses under the Teichmüller mapping. This time, it is natural to quantify the local area difference by the quantity $\delta_A = \dfrac{\text{Area of ellipse}}{\text{Area of circle}} - 1$ for each small circle. For each local region on the right wing, a positive δ_A indicates that the region is smaller than the corresponding local region on the left wing, a negative δ_A indicates that it is larger, and a perfect symmetry between the two regions yields $\delta_A = 0$. From the heat map of δ_A in Fig. S10, we observe that the most significant local area asymmetry occurs near the distal tip of L2, as well as the intersections between L4, L5 and PCV.

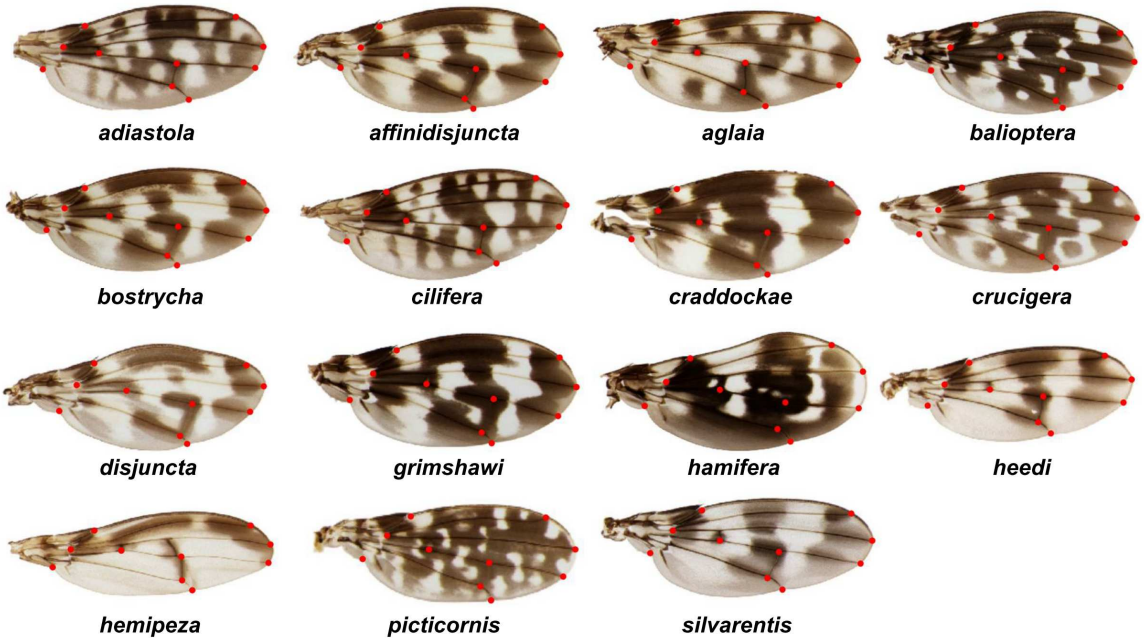

Figure S5: The species classified as Community 3 by our method. The images are adapted from the Hawaiian Drosophila Wing Database [10] (not to scale).

S8 Assessing the vein difference under Teichmüller mappings

Suppose S_1 and S_2 are two insect wings with prescribed landmark correspondence. Under the landmark-matching Teichmüller mapping $f : S_1 \to S_2$, the boundary arcs and all of the landmarks of them are matched, while the interior veins, in general, are not. In this section, we outline an approach for assessing the vein difference under Teichmüller mappings.

Mathematically, every interior vein of a wing can be regarded as a parameterised curve $\gamma : [0, 1] \to \mathbb{R}^2$. A classical result in differential geometry is that the shape of a plane curve is uniquely determined by its signed curvature k_γ, up to translation and rotation.

Suppose $\gamma_1 : [0, 1] \to \mathbb{R}^2$ and $\gamma_2 : [0, 1] \to \mathbb{R}^2$ are two corresponding veins on S_1 and S_2 respectively, and their endpoints are corresponding landmark pairs. Under the landmark-matching Teichmüller mapping $f : S_1 \to S_2$, we have $f(\gamma_1(0)) = \gamma_2(0)$ and $f(\gamma_1(1)) = \gamma_2(1)$. These boundary conditions determine the translation and rotation. Therefore, the difference between $f(\gamma_1)$ and γ_2 is captured by the difference in their signed curvatures $k_{f(\gamma_1)}$ and k_{γ_2}. It is natural to define the distance between γ_1 and γ_2 by

$$d_f(\gamma_1, \gamma_2) = \|k_{f(\gamma_1)} - k_{\gamma_2}\|_2 = \left(\int_0^1 (k_{f(\gamma_1)}(t) - k_{\gamma_2}(t))^2 dt \right)^{1/2}. \tag{S8}$$

In practice, the above distance can be easily evaluated by considering the image of uniformly sampled points on γ_1 under f and uniformly sampled points on γ_2.

We remark that it is possible to obtain a more explicit expression of $d_f(\gamma_1, \gamma_2)$. Recall that the first order approximation of f around z_0 is given by

$$f(z) \approx f(z_0) + f_z(z_0)(z - z_0) + f_{\overline{z}}(z_0)\overline{z - z_0}. \tag{S9}$$

This suggests that the infinitesimal orientation change under f is $\frac{1}{2}\arg\mu_f(z_0)$. As curvature is related to the rate of change of angle between neighbouring tangents, it is expected that $k_{f(\gamma_1)}$ can be described in terms of $\arg\mu_f$ and k_{γ_1}.

S9 Clustering the Hawaiian *Drosophila* wings by the traditional geometric morphometrics approach

From the clustering result obtained by our Teichmüller morphometric framework, it can be observed that the *D. adiastola* wings are split into two communities. We compare the result with that obtained using traditional geometric morphometrics [16]. Instead of performing non-rigid mappings between the wing shapes, we apply the Procrustes superimposition to align all wings based on the 10 landmarks on each of them. Then, we perform principal component analysis on the 20 coordinates of the 10 landmarks of all wings. Figure S11 shows the principal component analysis result. The k-means clustering algorithm is further applied to form communities.

Since this approach only considers the 10 landmarks on each wing but not the shape of the entire wing boundary, the community detection result is different from that obtained by our Teichmüller morphometric framework. However, it can be observed that the *D. adiastola* wings are again split into two communities under this approach. This suggests that the split of the *D. adiastola* wings obtained by our proposed approach is more likely to be reflecting the multimodal wing pattern of *D. adiastola*, rather than an error made by our method.

References

[1] Srivastava A, Wu W, Kurtek S, Klassen E, Marron JS (2011) Registration of functional data using Fisher-Rao metric. Preprint, arXiv:1103.3817v2.

[2] Kurtek SA, Srivastava A, Wu W (2011) Signal estimation under random time-warpings and nonlinear signal alignment. Adv. Neural. Inf. Process. Syst. 675-683.

[3] Lui LM, Lam KC, Gu X, Yau ST (2014) Teichmüller mapping (T-map) and its applications to landmark matching registration. SIAM J. Imaging Sci. 7(1):391-426.

[4] Meng TW, Choi GPT, Lui LM (2016) TEMPO: feature-endowed Teichmüller extremal mappings of point clouds. SIAM J. Imaging Sci. 9(4):1922-1962.

[5] Lui LM, Gu X, Yau ST (2015) Convergence of an iterative algorithm for Teichmüller maps via harmonic energy optimization. Math. Comp. 84(296):2823-2842.

[6] Rudin W (1964) Principles of mathematical analysis (Vol. 3). New York: McGraw-hill.

[7] Morrison G, Mahadevan L (2011) Asymmetric network connectivity using weighted harmonic averages. EPL 93:40002.

[8] Morrison G, Mahadevan L (2012) Discovering communities through friendship. PLOS ONE 7(7):e38704.

[9] Abràmoff MD, Magalhães PJ, Ram SJ (2004) Image processing with ImageJ. Biophotonics International 11(7):36-42.

[10] Edwards KA, Doescher LT, Kaneshiro KY, Yamamoto D (2007) A database of wing diversity in the Hawaiian Drosophila. PLoS One 2(5):e487.

[11] Gower JC (1975) Generalized Procrustes analysis. Psychometrika 40(1):33-51.

[12] Bookstein FL (1989) Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Trans. Pattern Anal. Mach. Intell. 11(6):567-585.

[13] Marquardt D (1963) An algorithm for least-squares estimation of nonlinear parameters" SIAM J. Applied Math. 11(2): 431-441.

[14] Ester M, Kriegel HP, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. Kdd. 96(34):226-231.

[15] Shi J, Malik J (2000) Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 22(8):888-905.

[16] Blackith RE, Reyment RA (1971) Multivariate morphometrics. Academic Press.

Species	Phylogenetic group	Specimen number
clavisetae	*adiastola*	1–5
ornata	*adiastola*	6–7
setosimentum	*adiastola*	8–10
adiastola	*adiastola*	11–14
cilifera	*adiastola*	15–20
hamifera	*adiastola*	21–26
spectabilis	*adiastola*	27–32
heteroneura	*planitibia*	33–37
planitibia	*planitibia*	38
silvestris	*planitibia*	39–40
nigribasis	*planitibia*	41–42
cyrtoloma	*planitibia*	43–46
melanocephala	*planitibia*	47–48
neoperkinsi	*planitibia*	49–51
neopicta	*planitibia*	52–55
oahuensis	*planitibia*	56–57
obscuripes	*planitibia*	58–60
hemipeza	*planitibia*	61–63
picticornis	*planitibia*	64–65
aglaia	*glabriapex*	66–67
basisetae	*glabriapex*	68–70
digressa	*glabriapex*	71–72
discreta	*glabriapex*	73–74
fasciculisetae	*glabriapex*	75
glabriapex	*glabriapex*	76–77
macrothrix	*glabriapex*	78–80
montgomeryi	*glabriapex*	81–83
punalua	*glabriapex*	84–87
affinidisjuncta	*grimshawi*	88–89
balioptera	*grimshawi*	90–91
bostrycha	*grimshawi*	92–93
craddockae	*grimshawi*	94
crucigera	*grimshawi*	95–98
disjuncta	*grimshawi*	99–103
grimshawi	*grimshawi*	104
heedi	*grimshawi*	105–106
silvarentis	*grimshawi*	107–108
limitata	*grimshawi*	109–112
engyochracea	*grimshawi*	113–115
hawaiiensis	*grimshawi*	116–118
murphyi	*grimshawi*	119–120
orphnopeza	*grimshawi*	121–122
orthofascia	*grimshawi*	123
recticilia	*grimshawi*	124–125
sproati	*grimshawi*	126–127
villosipedis	*grimshawi*	128

Table S1: The list of wing specimens adapted from [10] in our study of Hawaiian *Drosophila* wings. The specimen numbers represent the row/column numbers corresponding to the specimens in the similarity matrix.

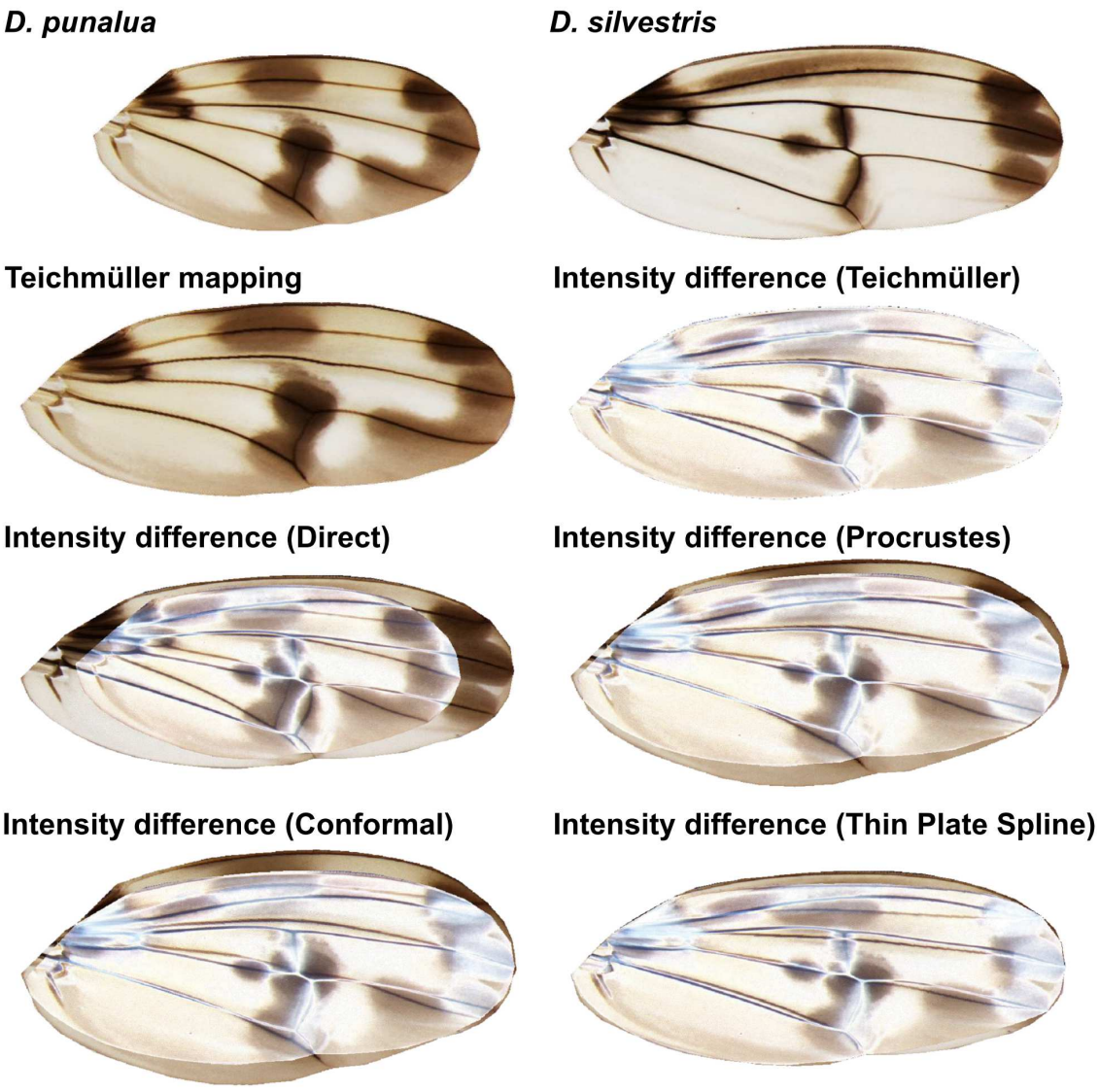

Figure S6: A comparison between our method and four existing morphometric approaches. The first row shows the *D. punalua* and *D. silvestris* wings. Row 2 shows the Teichmüller mapping of *D. punalua* onto *D. silvestris*, and the intensity difference between the Teichmüller mapping result and the *D. silvestris* wing. Row 3-4 show the intensity differences computed using direct mapping, Procrustes superimposition [11], least-square conformal mapping and Thin Plate Spline [12]. The *D. punalua* and *D. silvestris* wing images are adapted from [10].

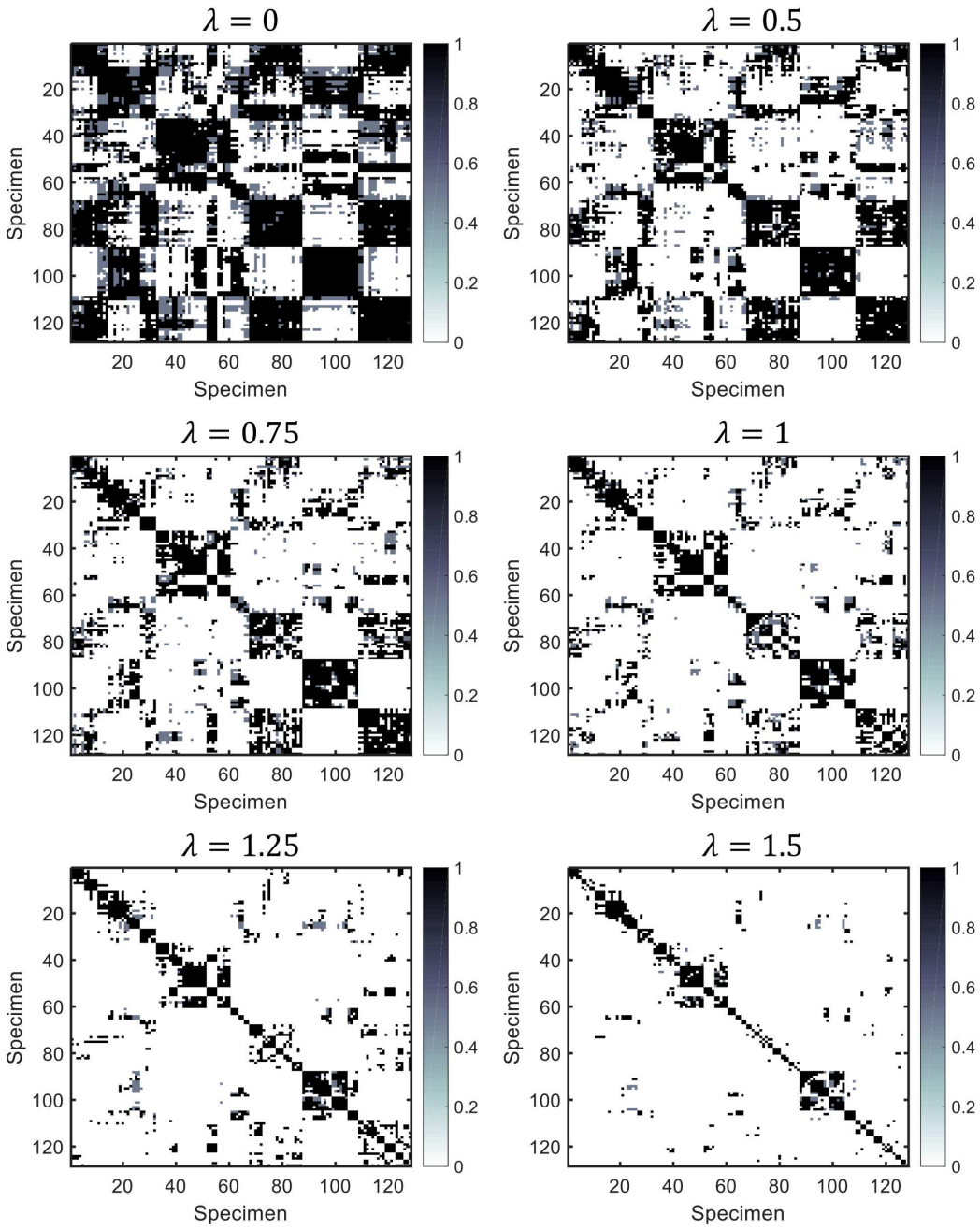

Figure S7: The thresholded similarity matrices produced by our adaptive thresholding algorithm (Algorithm S2) with different choices of the thresholding parameter λ.

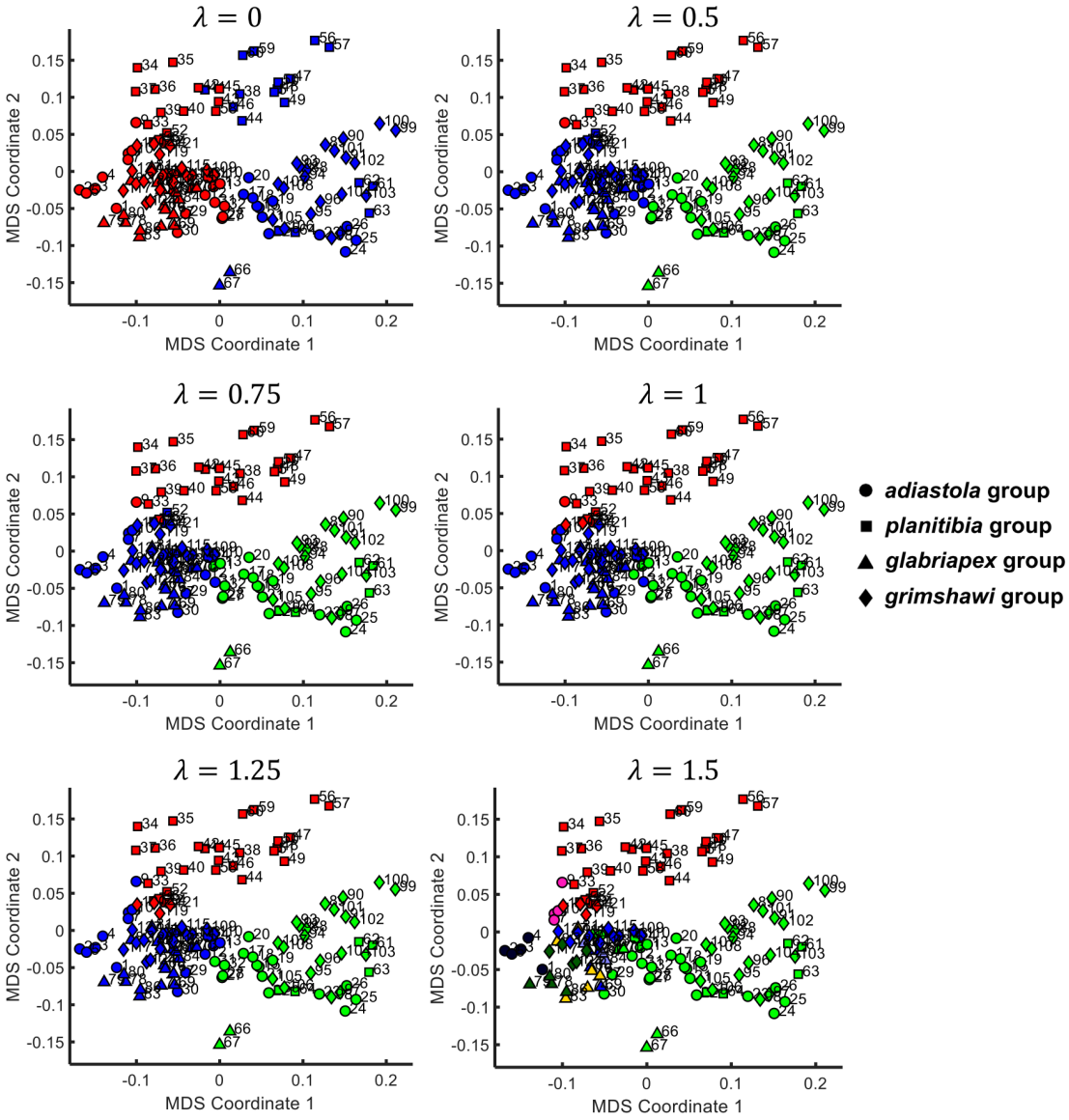

Figure S8: The community detection results with different choices of λ. Every specimen is represented as a node on the plane constructed by multidimensional scaling. The node colours represent the communities detected by the GEN-based community detection algorithm [8]. The four shapes of the nodes represent the four phylogenetic groupings.

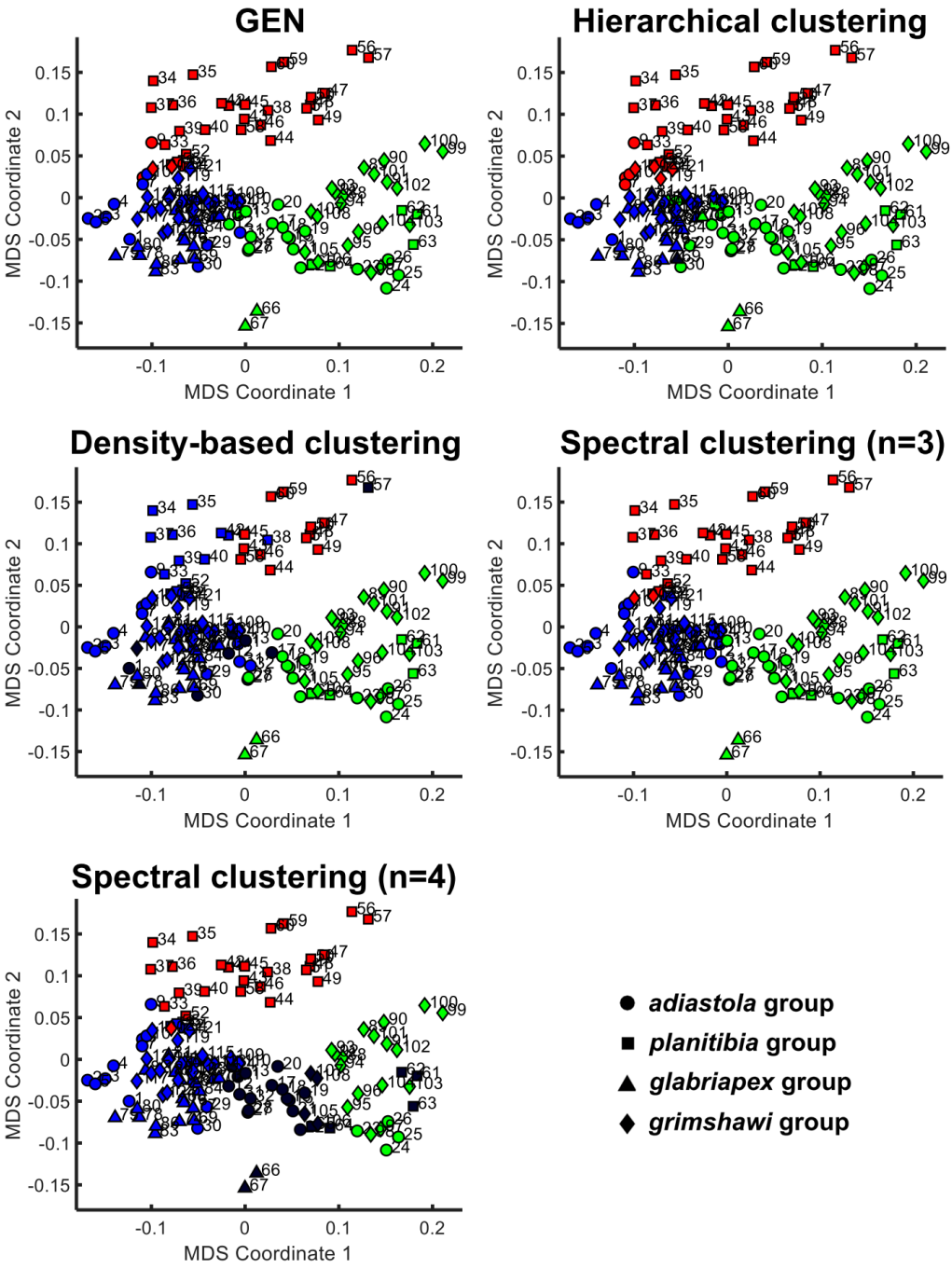

Figure S9: The community detection results on the thresholded similarity matrix (with the thresholding parameter $\lambda = 1$) by various clustering algorithms.

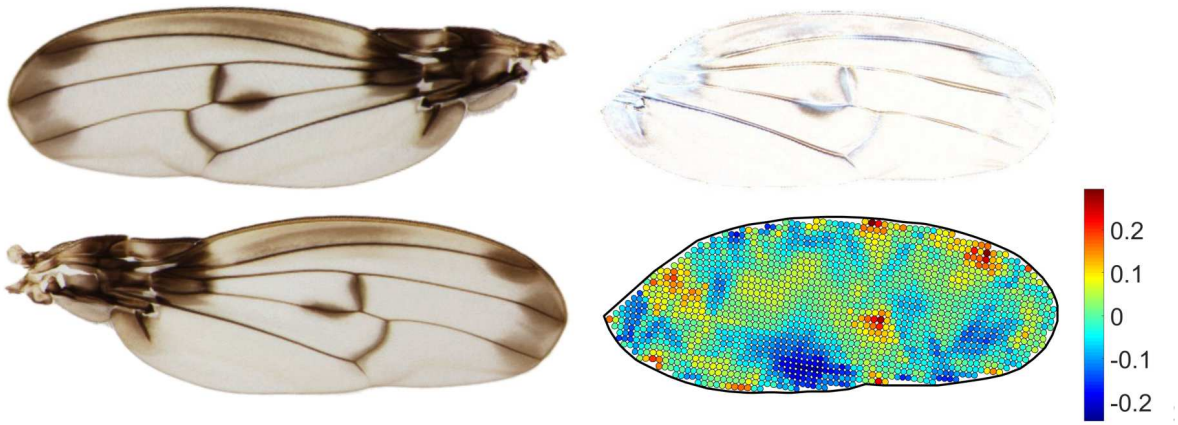

Figure S10: Analysing the bilateral asymmetry of a specimen of *D. heteroneura*. The first column shows the right wing and the left wing. We compute a landmark-matching Teichmüller map from the right wing to the left wing. The second column shows the intensity difference between the Teichmüller mapping result and the left wing, together with the heat map of the local area difference δ_A. The wing images are adapted from [10].

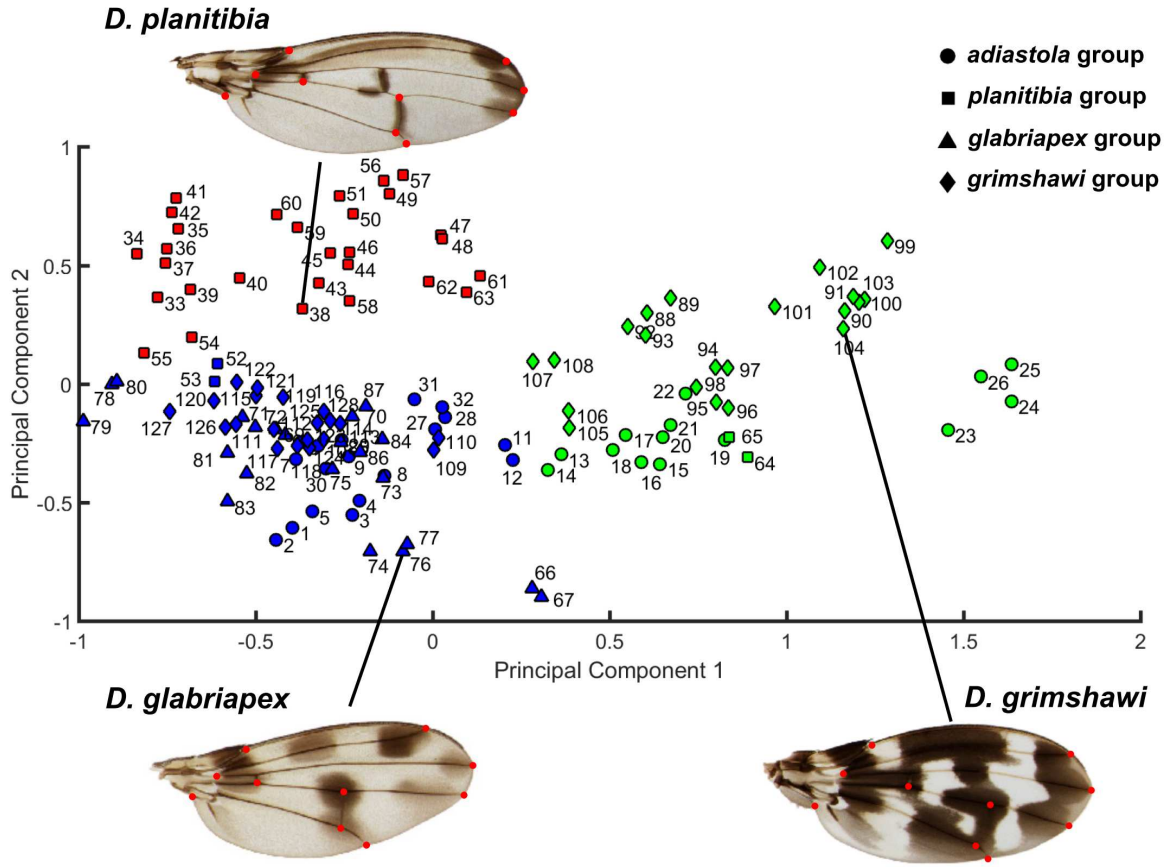

Figure S11: The principal components analysis result of the Hawaiian *Drosophila* wings obtained by traditional geometric morphometrics. The nodes are coloured based on the k-means clustering result. Blue: Community 1. Red: Community 2. Green: Community 3. The shapes of the nodes represent their phylogenetic groupings. Circle: *adiastola* group. Square: *planitibia* group. Triangle: *glabriapex* group. Diamond: *grimshawi* group. The number beside each node represents the specimen number specified in Table S1. The wing images shown are adapted from [10].