Supplemental Material S1. Methods.

Modularity Analysis

Broadly speaking, a modular architecture refers to the presence of groups of traits (modules) which exhibit relatively strong inter-connections (internal integration) but weaker connections to traits in other modules. Modularity may reflect shared developmental origins or functional roles and can lead, via correlational selection, to genetic integration. Hypotheses of modularity can be tested against a null hypothesis of no modularity which occurs when traits within and between modules are of similar magnitude. If that magnitude is high, then the trait is highly integrated. If that magnitude is low, then there is very little integration among traits of that structure.

For the current study, we treat each laryngeal cartilage as an a priori module. Thus, the coordinates which define the semilandmarks and landmarks of that cartilage are the traits assigned to that module. The steps involved in acquiring landmark and semilandmark data from the surface renderings of each cartilage from the CT data were described in the main text, as was the process of superimposing all cartilages of the same type (e.g., cricoid) via generalized Procrustes analysis. We will now present the workflow for the analysis of modularity hypotheses for the laryngeal cartilages as these methods are new to the field of laryngeal morphometrics. The following draws heavily from the work of Adams (2016) and Adams and Collyer (2019) who developed these statistics and the appropriates tests of statistical significance (see also Collyer et al., 2015; Adams and Collyer, 2016).

To test hypotheses of modularity, we must first operationalize our definition of modularity. The covariance ratio (CR) represents an appropriate statistic for this purpose. To calculate the CR, we first generate a single matrix, **C**, where rows contain all variables for a single observation and the columns contain all variables for a pair of cartilages – the cricoid and arytenoid in the following example. A partitioned covariance matrix, **X**, is generated from **C** as follows:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X_{CC}} & \mathbf{X_{CA}} \\ \mathbf{X_{AC}} & \mathbf{X_{AA}} \end{bmatrix}$$

$\mathbf{X_{CC}}$ and $\mathbf{X_{AA}}$ are covariance matrices within the cricoid and arytenoid modules, respectively. The $\mathbf{X_{CA}}$, and $\mathbf{X_{AC}}$ modules are covariance matrices between the two modules ($\mathbf{X_{AC}} = \mathbf{X_{CA}}^t$). The CR statistic is calculated as follows:

$$\mathrm{CR_{CA}} = \sqrt{\frac{tr(\mathbf{X_{CA}}\mathbf{X_{AC}})}{\sqrt{tr(\mathbf{X^*_{CC}}\mathbf{X^*_{CC}})tr(\mathbf{X^*_{AA}}\mathbf{X^*_{AA}})}}}$$

Where *tr* indicates the trace of a matrix (sum of the diagonal elements) and $\mathbf{X^*_{CC}}$ and $\mathbf{X^*_{AA}}$ are the within-module covariance matrices with zeros on the diagonal. The ratio represents the total between module covariance scaled by the within module covariance (excluding individual trait variance). The value of CR is between 0 and positive values. When modules have low between- and high within- module covariance (a modular organization), this value approaches zero. When within- and between-module covariance is nearly equal (no modular signal), the value

approaches 1. The CR statistic cannot discern between high or low overall integration when values are near 1. Statistical significance of the CR statistic is assessed via a permutation procedure where the landmarks are randomly assigned to the two modules and the CR value is recalculated. This procedure produces an empirical distribution of CR values representing the null hypothesis of no modularity. The expected value of this distribution is 1. The percent of these resampled values which are less than the observed CR (CR_{CA}) is an estimate of statistical significance (p). In our example, CR_{CA} = 0.65 (p = 0.001). This implies that the cricoid and arytenoid cartilage shapes are more modular than by chance.

To produce a measure which is more broadly comparable across studies, we next calculate a z-score (a standardized effect size) which assess the magnitude of the effect. The z-score for CR is:

$$Z_{CR_{CA}} = \frac{CR_{CA} - \hat{\mu}_R}{\hat{\sigma}_R}$$

Where $\hat{\mu}_R$ is the average CR value calculated from the resampled distribution and $\hat{\sigma}_R$ is the standard deviation of that distribution. By standardizing the difference between CR_{CA} and $\hat{\mu}_R$ by variation around $\hat{\mu}_R$, it is possible to assess the magnitude of the modularity signal measured between the cricoid and arytenoid cartilages. As CR values less than 1 indicate modularity, and the expectation for $\hat{\mu}_R$ is 1, more negative values indicate a larger effect size. As reported in Table 3 of the main text, $Z_{CR_{CA}}$ = -20.9.

Finally, we can compare the effect sizes (Z_{CR}) for each pair of modules to ascertain whether a pair of modules (cartilages) exhibit a stronger signal of modularity than other pairs using the pairwise $\hat{Z}$ statistics. For example, we can compare the $Z_{CR_{CA}}$ value of -20.9 to the $Z_{CR_{TA}}$ value of -24.2 calculated for the thyroid and arytenoid cartilages. Pairwise $\hat{Z}$ is calculated as follows:

$$\hat{Z}_{CA-TA} = \frac{\left|\left(CR_{CA} - \hat{\mu}_{R_{CA}}\right) - \left(CR_{TA} - \hat{\mu}_{R_{TA}}\right)\right|}{\sqrt{\hat{\sigma}^2_{R_{CA}} + \hat{\sigma}^2_{R_{TA}}}}$$

The values in the above equation are the same quantities used in the calculation of Z_{CR} above. The numerator is the absolute value of the difference in effect sizes for each pair of cartilages, and the denominator is the pooled within-sample standard deviation. Statistical significance of pairwise $\hat{Z}$ is assessed by reference to a standard normal distribution. To complete our example, the effect size of the thyroid-arytenoid pair is significantly larger than that of the cricoid-arytenoid pair. Thus, the thyroid and arytenoid cartilages have a stronger signal of modularity than the cricoid and arytenoid cartilages.

References

Adams, D. C. (2016). Evaluating modularity in morphometric data: Challenges with the RV coefficient and a new test measure. *Methods in Ecology and Evolution, 7,* 565–572.

Adams, D. C., & Collyer, M. L. (2016). On the comparison of the strength of morphological integration across morphometric datasets. *Evolution, 70,* 2623–2631.

Adams, D. C., & Collyer, M. L. (2019). Comparing the strength of modular signal, and evaluating alternative modular hypotheses, using covariance ratio effect sizes with morphometric data. *Evolution, 73,* 2352–2367.

Collyer, M. L., Sekora, D. J., & Adams, D. C. (2015). A method for analysis of phenotypic change for phenotypes described by high-dimensional data. *Heredity, 115,* 357–365.