# Testing the variation of landmark regions within and between species

To understand the [WHATHEVER THE PAPER IS ABOUT] we tested whether some landmark regions (e.g. the landmarks on the zygomatic arch) varied significantly more or less than the rest of the skull within and between each species. To test these variation, we applied a “landmark partition test” that tests whether the maximum variation between landmarks position in a set of landmarks is significantly different than any random same size set of landmarks. This test is a kind of permutation test testing with a two-sided alternative . If can be rejected it means that the statistic measured from the set of landmarks (see below) is different than the overall statistic from the population of landmarks (in other words: the landmark set is different than the all the landmarks combined).

We measured two statistic, the overall difference in landmark size (area difference - see below) and the probability of size overlap between two landmarks sets (Bhattacharyya Coefficient - see below). These two statistics were measured between the two most extreme specimens selected on the range of variation between the specimens landmarks coordinates (see below). We tested the differences for three cranium partitions ([WHAT ARE THEY?]) and three mandible ones ([WHAT ARE THEY?]) for each species (*Wombat ursinus*, *Lasiorhinus krefftii* and *Lasiorhinus latifrons*) and for the three species combined. To account for type I error due to multiple testing and for the number of replicates involved in the landmark partition tests, we adjusted the *p*-values using a Bonferonni-Holm [[CITE HOLM](https://scholar.googleusercontent.com/scholar.enw?q=info:5ZyBqM-YIwUJ:scholar.google.com/&output=citation&scisig=AAGBfm0AAAAAWsWbTUBK3jllx3lcdeQmf7gkwWclv4OR&scisf=3&ct=citation&cd=-1&hl=en)] correction and lowered our *p-value* rejection threshold to 0.01 (see below). The whole procedure, including the landmark variation is described in details below and implemented in R: <https://github.com/TGuillerme/landmark-test>.

## Selecting the range of landmark variation

One approach for selecting the range of landmark variation is based on a specific axis of an ordination matrix (e.g. PCA [CITE SOME PAPER THAT DOES IT]). This approach has several advantages namely (1) its intuitiveness and (2) the fact that the selected range is based on a specific axis of variation (e.g. the range on the first ordination axis is based on first axis of variance/covariance in the dataset). However, this approach suffers also from several statistical drawbacks namely (1) that not all the variance/covariance is taken into account (even if based on an ordination axis that contain a great amount of the datasets’ variance) and that (2) the resulting range of landmark variation is based on the ordination of the variance and not directly on the range of the actual variation *per se* (i.e. in an Euclidean space).

Here we propose a an approach based directly on the Procrustes superimposed landmarks’ differences rather than their ordination (i.e. we compare the two skulls or mandibles with the most different landmark’s position). We calculate this difference as the length of the radius in spherical coordinates between a pair of homologuous landmarks. Spherical coordinates are a coordinate system measured in a sphere and are expressed by three parameters: (1) the radius (, the euclidean distance between the two landmarks); (2) the azimuth angle (, the angle on the equatorial plan); and (3) the polar angle (, the angle measured on the polar plan). Since we are only interested in the *magnitude* of difference () regardless of the direction ( and ) we will use only the radius below.

To calculate the maximum range of landmark variation (the two individuals with the most different radii) we use the following algorithm:

1. Calculate the radii for all landmarks between the mean Procrustes superimposition and each specimen’s superimposition.
2. Rank each set of radii and measure the area under the resulting curve (see eq. 1 below).
3. Select the specimen with the highest radii area. This is now the “maximum” Procrustes (i.e. the specimen with the most different shape compared to the mean).
4. Calculate the radii for all landmarks between the “maximum” Procrustes and the remaining individual ones.
5. Repeat step 2 and 3. The resulting specimen with the highest radii area is now the “minimum” Procrustes (i.e. the specimen with the most different shape compared to the “maximum” Procrustes).

These two “maximum” and “minimum” Procrustes superimpositions are not the biggest/smallest, widest/narowest, etc. skulls or mandibles *per se* but rather the two extremes of the overall distribution of landmark variation (). Because of the multidimensionality aspect of the problem, it is probably impossible to even determine which one of the two selected “maximum” and “minimum” craniums/mandibles has actually the most variation. However, since we are not interest in *direction* of difference, these two selected craniums/mandibles are sufficient to give us the magnitude of differences between specimens (while taking into account *all* the landmarks the three dimensions).

For reducing the effect of any eventual outlier, we also performed this step based on the 95% confidence interval only. To realize this interval we performed the same algorithm as above but before selecting the specimen with the highest *n* radii, we removed the specimen with *n* radii superior to the highest quantile of the distribution. For example, with a 95% confidence interval, we remove any specimen about the 97.5th quantile (i.e. ).

## Difference statistic between partitions

As mentioned above, we will use two different statistics between to compare the partitions to the rest of the cranium/mandible: (1) the overall difference in landmark size between the “maximum” and “minimum” cranium/mandible (this statistic is a proxy for length difference between landmarks ranges); (2) the probability of overlap between the size differences (between the “maximum” and “minimum”) in the partition and the rest of the cranium/mandible (this statistic is a proxy for measuring whether both partitions comes from the same distribution).

### Overall difference in landmark size (area difference)

To measure the overall differences in the length of the radii, we can calculate the area difference as follows:

(eq. 1)

Where is minimum number of comparable landmarks and and are ranked functions (i.e. ) for the landmarks in the partition and all the landmarks respectively. If one of the functions or have elements (with ) , a rarefied estimated of the function with elements is used instead.

(eq. 2)

Where is the number of rarefaction replicates. is chosen based on the Silverman’s rule of thumb for choosing the bandwidth of a Gaussian kernel density estimator multiplied by 1000 with a result forced to be 100 1000 [[CITE SILVERNMAN](https://scholar.googleusercontent.com/scholar.enw?q=info:UGz7W84HloAJ:scholar.google.com/&output=citation&scisig=AAGBfm0AAAAAWsWcW-x_EpNItS_u-3x_T_-uVXdptYdA&scisf=3&ct=citation&cd=-1&hl=en)].

(eq. 3)

With being the minimum of the standard deviation and the interquartile range of the distribution. This allows the rarefaction algorithm to be fast but “fair” (i.e. reducing the number of re-sampling the when the distribution is more homogeneous [CITE SILVERMAN](https://scholar.googleusercontent.com/scholar.enw?q=info:UGz7W84HloAJ:scholar.google.com/&output=citation&scisig=AAGBfm0AAAAAWsWcW-x_EpNItS_u-3x_T_-uVXdptYdA&scisf=3&ct=citation&cd=-1&hl=en)). If the area difference is positive, the landmark’s variation in the partition is bigger than the overall landmark’s variation, if the difference is negative, the variation is smaller.

### Probability of overlap between size differences (Bhattacharyya Coefficient)

The Bhattacharyya Coefficient calculates the probability of overlap of two distributions [[CITE BHATTA](https://scholar.googleusercontent.com/scholar.enw?q=info:K1BkX8mmZF8J:scholar.google.com/&output=citation&scisig=AAGBfm0AAAAAWsWdBttRNb84L1_PaId7yTjwjqdFKUwT&scisf=3&ct=citation&cd=-1&hl=en); [CITE GUILLERME](https://scholar.googleusercontent.com/scholar.enw?q=info:xINpsBGGa2YJ:scholar.google.com/&output=citation&scisig=AAGBfm0AAAAAWsWdPqo9fA2pYyANT1k-HZLIG_LeMBET&scisf=3&ct=citation&cd=-1&hl=en)]. When it is equal to zero, the probability of overlap of the distributions is also zero, and when it is equal to one, the two distributions are entirely overlapping. It forms an elegant and easy to compute continuous measurement of the probability of similarity between two distributions. The coefficient is calculated as the sum of the square root of the relative counts shared in bins among two distributions.

(eq. 4)

Where and are the number of counts in bin for the distributions and respectively divided by the total number of counts in the distribution and respectively.

was determined using the Silverman’s rule of thumb (see equation 3). We consider two distributions to be significantly similar when their Bhattacharyya Coefficient is . Conversely, they are significantly different when their Bhattacharyya Coefficient is . Values in between these two threshold just show the probability of overlap between the distributions but are not conclusive to assess the similarity or differences between the distributions.

## Random permutation test

To test whether the landmarks’ variation (i.e. the radii between the “maximum” and “minimum” cranium/mandible) one partition is different than the rest of the landmark’s variation, we used a kind of permutation test (based on a modification from ade4::as.randtest [[CITE THIOULOUS](https://scholar.googleusercontent.com/scholar.enw?q=info:D7YFXN9dhNoJ:scholar.google.com/&output=citation&scisig=AAGBfm0AAAAAWsWfFOTSWlZRE1b2yJ_7REIccOHOla7d&scisf=3&ct=citation&cd=-1&hl=en)]). This is a pretty intuitive yet powerful test aiming to calculate whether the observed difference in statistic (either the area difference or the probability of overlap) is different than the same statistic drawn randomly from the same population (here, the distribution of all the landmark’s variation). First we measured the statistic between the landmark partition of interest and all the other landmarks (including the ones from the partition). Second, we generated 1000 statistics by randomly sampling the same number of landmarks as in the partition in the whole distributions and compared them again to the full distribution. This resulted in 1000 null comparisons (i.e. assuming the null hypothesis that the statistic in the partition is the same as the statistic in the total distribution). We then calculated the *p* value based on:

(eq. 5)

Where is the number of random draws (1000 bootstrap pseudo-replicates), is the statistic from the comparison of the randomly drawn partition and the distribution and is the observed statistic between the partition and the distribution. An increased number of bootstrap pseudo-replications increases the type I error. We therefore lowered our threshold for accepting to 0.01 instead of the commonly used 0.05.