Appendix
Relocation Error

The averaged errors in terms of percentage associated with relocating landmarks (two replicates) are minimal for both *Lanistes* and *Melanoides*, but errors varied somewhat along the utilized landmarks (Table A1-1). The largest errors in the *Lanistes* data set are associated with landmark L2 *x*, L3 *xy*, L5 *y*, L6 *y*, and L7 *y*, of which four (L2, L5, L6, and L7) are considered Bookstein type 3 landmarks (Bookstein 1991). Although L3 is of type 2, it is problematic to relocate it with high accuracy. In *Melanoides*, the largest errors are obtained for L2 *x* (type 3) and L5 *x* and *y* (type 2).

TABLE A1-1. Mean error relative to centroid size.

	Lanistes		*Melanoides*	
Landmark	Mean	SD	Mean	SD
L1 *x*	0.299	0.275	0.193	0.158
L1 *y*	0.088	0.265	0.166	0.161
L2 *x*	0.465	0.467	0.441	0.402
L2 *y*	0.057	0.067	0.172	0.289
L3 *x*	0.980	0.746	0.045	0.082
L3 *y*	0.607	0.505	0.294	0.225
L4 *x*	0.121	0.143	0.106	0.108
L4 *y*	0.243	0.228	0.135	0.142
L5 *x*	0.136	0.232	0.339	0.340
L5 *y*	0.691	0.686	0.697	0.678
L6 *x*	0.056	0.082	0.124	0.148
L6 *y*	0.741	0.714	0.191	0.274
L7 *x*	0.032	0.049	0.105	0.096
L7 *y*	0.620	0.549	0.114	0.101
L8 *x*	0.148	0.371	0.103	0.095
L8 *y*	0.105	0.176	0.124	0.119
L9 *x*	0.143	0.184	0.122	0.126
L9 *y*	0.117	0.130	0.165	0.139
L10 *x*	0.226	0.342	0.090	0.099
L10 *y*	0.103	0.256	0.140	0.150
L11 *x*	0.241	0.328	0.107	0.109
L11 *y*	0.112	0.184	0.123	0.151
L12 *x*	-	-	0.061	0.061
L12 *y*	-	-	0.278	0.278
L13 *x*	-	-	0.096	0.091
L13 *y*	-	-	0.240	0.249
Total	0.288	0.210	0.183	0.133

Mean error in terms of percentage associated with relocating landmarks and its standard deviation (SD); each for the *x* and *y* coordinate of every landmark.

Reference

Bookstein, F.L. 1991. Morphometric tools for landmark data: geometry and biology. Cambridge University Press, Cambridge.

Appendix 2

Principal Components Analysis

In addition to nonmetric multidimensional scaling (nmMDS), we applied principal components analysis (PCA) to our data sets. The results of both techniques are overall very similar, but there are some difficulties with PCA. In nmMDS, the number of dimensions (ordination axes) is chosen a priori and the reliability of the data representation in these dimensions (goodness-of-fit) is determined subsequently by reconstructing the amount of stress. For PCA, however, such tests do not exist, making it difficult to evaluate the number of principal components (PCs) that are of significant importance in explaining the variation in the data sets (Hammer and Harper 2006). Moreover, some of the assumptions of PCA, including multivariate normality, which was not obtained in our data sets, are still under debate (Hammer et al. 2001).

We indicate the similarity between nmMDS and PCA by presenting the PCA plots for the data sets of the elliptic Fourier analysis (EFA) of both genera. For *Lanistes*, the first two PCs account for 82.50 % of the variance in the data set. Although overall very similar, the PCA plot (Fig. A2-1) shows the morphospace occupation to be somewhat arched. This arch results from nonlinear relationships between the variables. Because the plot was obtained via linear scaling (PCA), the two axes should not be interpreted as being independent. The EFA nmMDS plot appears to be affected to a much lesser extent, whereas a similar pattern is absent from the nmMDS plots of the landmark and semi-landmark data sets. The PC1 versus PC2 plot (Fig. A2-2) of *Melanoides*

accounts for 87.42% of the variance in the data set. The plot is nearly identical to the nmMDS plot. This suggests that both nmMDS and PCA results are reliable.

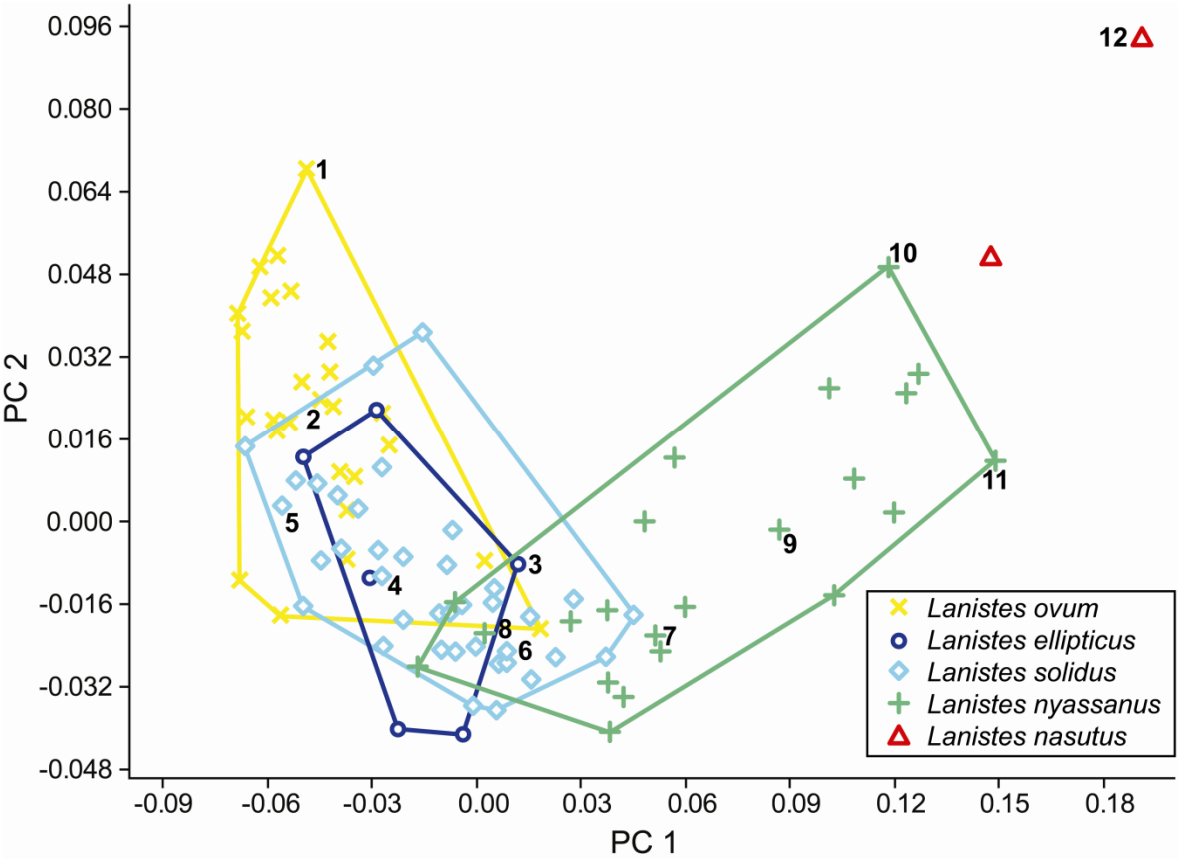

FIGURE A2-1. PCA plot for the *Lanistes* EFA data set. The first two PCs in principal components analysis account for 82.50 % of the variance in the data set of the *Lanistes* elliptic Fourier coefficients. Although overall very similar to the plot of the nonmetric multidimensional scaling for the same treatment, the configuration in the PCA plot is curved.

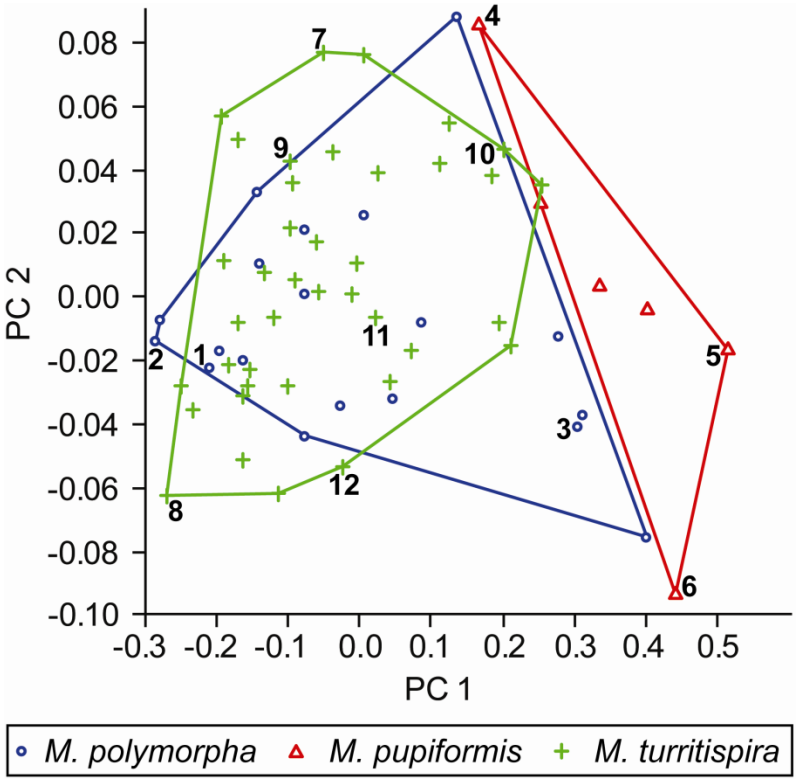

FIGURE A2-2. PCA plot for the *Melanoides* EFA data set. The first two PCs in principal components analysis account for 87.42 % of the variance in the data set of the *Melanoides* elliptic Fourier coefficients. The plot is almost identical to the nmMDS plot.

References

Hammer, Ø., and D. A. T. Harper. 2006. Paleontological data analysis. Blackwell, Oxford.

Hammer Ø., D. A. T. Harper, and P. D. Ryan. 2001. PAST: paleontological statistics software package for education and data analysis. Palaeontologia Electronica 4:1-9.

Appendix 3

Classification Comparison

We have compared the performance of support vector machines (SVMs) to that of linear canonical variates analysis (CVA) for the semi-landmark analysis of both *Lanistes* and *Melanoides*. Canonical variates analysis requires reduction of data dimensionality utilizing principal component analysis, prior to classification and cross-validation. The CVA method and data reduction with a variable number of PC's utilized are described in Sheets et al. (2006). The program utilized is CVAGen6o (Sheets 2008).

Results are shown in Table A3-1, which suggests a comparable performance of CVA and SVM for the *Lanistes* data set, but a lower performance of CVA for *Melanoides*. Classification with CVA seems thus to experience significant problems for more complex objects. This is likely because the methods of dimensionality reduction adopted here may blur between-group differences and within-group variation, affecting the subsequent classification. In order to formulate robust conclusions concerning preferences in classification methods a more detailed comparison of SVMs and CVA is required. This falls beyond the scope of the present paper, because it would demand documenting the performance for multiple SVM training algorithms, several of which require accurate and time-consuming parameterization.

TABLE A3-1. Comparison of CVA and SVM performances.

Taxon	Treatment	PC's	mean	SD	95% confidence interva	
Lanistes	SLM-CVA	15	0.739	0.047	0.616	0.826
Lanistes	SLM-CVA	18	0.741	0.053	0.605	0.826
Lanistes	SLM-SVM-2CV	n.a.	0.745	0.042	0.663	0.826
Lanistes	SLM-SVM-10CV	n.a.	0.740	0.040	0.661	0.819

Melanoides	SLM-CVA	12	0.528	0.081	0.375	0.703
Melanoides	SLM-CVA	15	0.523	0.081	0.344	0.703
Melanoides	SLM-SVM-2CV	n.a.	0.647	0.103	0.445	0.849
Melanoides	SLM-SVM-10CV	n.a.	0.800	0.050	0.702	0.897

Comparison suggests that the classification rate, obtained by support vector machines (SVMs) is similar to or higher than that obtained by linear canonical variates analysis (CVA). Although CVA appears to be a reliable classification method for *Lanistes*, at least if the number of PCs utilized during the dimensionality reduction is optimized, SVM performs significantly better for the more complex *Melanoides* shells. The 95% confidence interval is derived from bootstrapping.

References

Bookstein, F. L. 1991. Morphometric tools for landmark data: geometry and biology. Cambridge University Press, Cambridge.

Hammer, Ø., and D. A. T. Harper. 2006. Paleontological data analysis. Blackwell, Oxford.

Hammer, Ø., D. A. T. Harper, and P. D. Ryan. 2001. PAST: paleontological statistics software package for education and data analysis. Palaeontologia Electronica 4:1-9.

Sheets, H. D. 2008. IMP-Integrated Morphometrics Package. Department of Physics, Canisius College, Buffalo, N.Y.

Sheets, H. D., K. M. Corvino, J. M. Panasiewicz, and S. R. Morris. 2006. Comparison of geometric morphometric outline methods in the discrimination of age-related differences in feather shape. Frontiers in Zoology 3:15.