

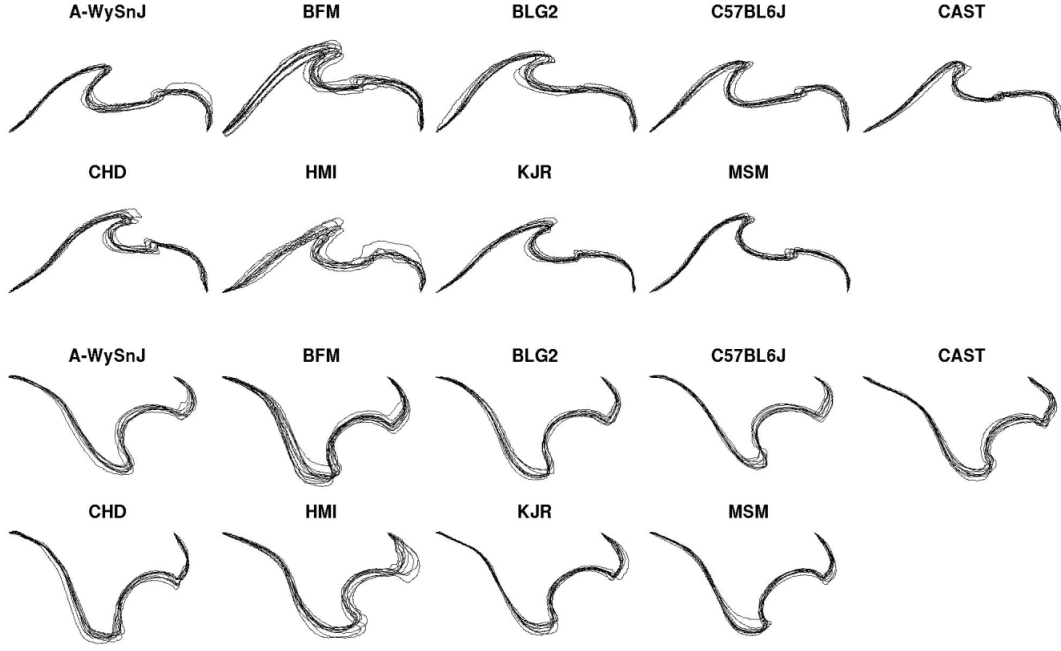


Fig. 5. 93 open curves extracted from mouse mandibular outlines of nine inbred strains with each strain consisting of 10–13 shapes. Curves of same strain are superposed (above: upside curves, below: downside curves).

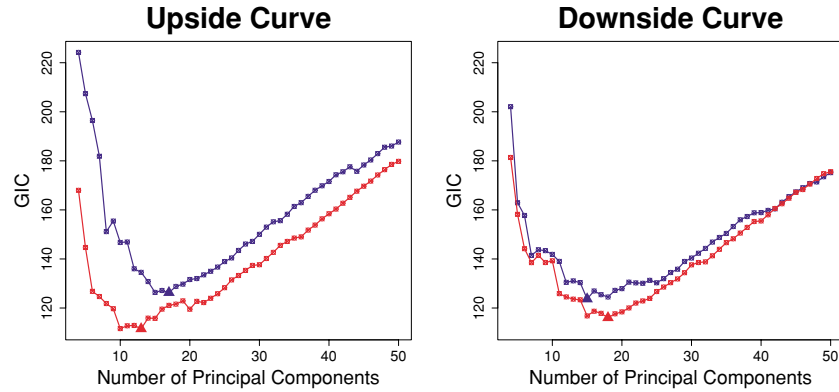


Fig. 6. GIC of the logistic regression model vs. the number of dimensions of the model, with the data represented by TFDs (blue graph) and PFDs (red graph). Triangles indicate the smallest peak values.

and downside curves) extracted from nine inbred strains of mouse mandibular shapes, represent all open curves as 200-dimensional feature vectors using the TFD and the PFD, (2) using principal component analysis, divide the whole variation in data into independent components, (3) fit the multinomial logistic regression model of arbitrary dimension K to the data using the first K principal component scores (assuming that the essential difference among distinct inbred strains is contained in the first K principal component scores), (4) information-theoretically evaluate the goodness of the model using the GIC, and compare the GIC values in the cases of the TFD and the PFD. Using the descriptor which captures the inter-strain variability of shapes more clearly, the GIC value is supposed to be smaller compared with the one obtained with the other descriptor.

4.2 Experimental results

In the multidimensional shape space, we do not know which direction the essential difference among distinct in-

bred strains is represented most clearly. So, assuming that the essential feature is contained in the first K principal component scores, we fitted the model of arbitrary dimension K to the K dimensional data, and calculated GIC value of the model. And the dimension where the GIC value has the smallest peak value is the optimal dimension of the model. The graphs of GIC value vs. K (the dimension of the model) are shown in Fig. 6. Left and right figures show the results in the cases of upside and downside curves respectively. Red and blue graphs represent GIC values of the models with the data represented by PFDs and TFDs respectively. In either result for upside or downside curve, and for each number of dimensions of the data, GIC values of the models are smaller using PFDs than when using the TFDs, and the difference between the smallest peak values of GIC with the data represented by PFDs and TFDs appears more clearly in the case of upside curves than it is in the case of downside curves.