Morphometrics Within Dog Breeds Are Highly Reproducible and Dispute Rensch's Rule

Nathan B. Sutter[1,2], Dana Mosher[1], Melissa M. Gray[3] and Elaine A. Ostrander[1,4]

[1.] Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda MD 20892

[2.] Current Address: Department of Clinical Sciences, College of Veterinary Medicine, Cornell University, Ithaca NY 14853

[3.] Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA 90095

4. To Whom Correspondence May Be Addressed: Elaine A. Ostrander, Ph.D., Cancer Genetics Branch, National Institutes of Health, Building 50, 50 South Drive, Room 5351, Bethesda, MD 20892, USA; Phone: 301-594-5284; FAX 301-480-0462; eostrand@mail.nih.gov

Running Title, "Dog Breed Morphometrics"

# Supplementary Text and Figures
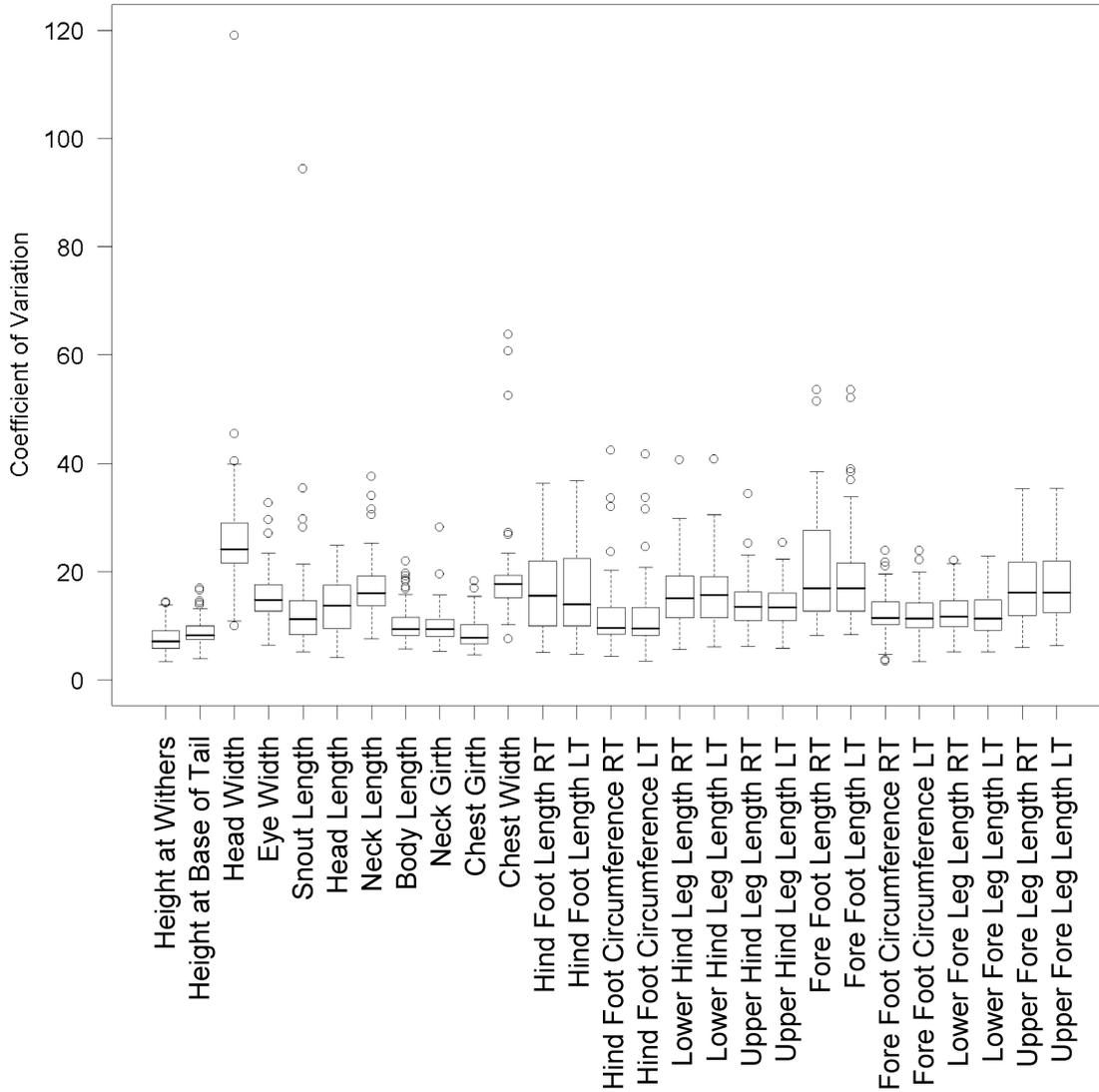
Supplementary Fig. 1. Coefficient of variation by breed for each of 27 measurements collected.
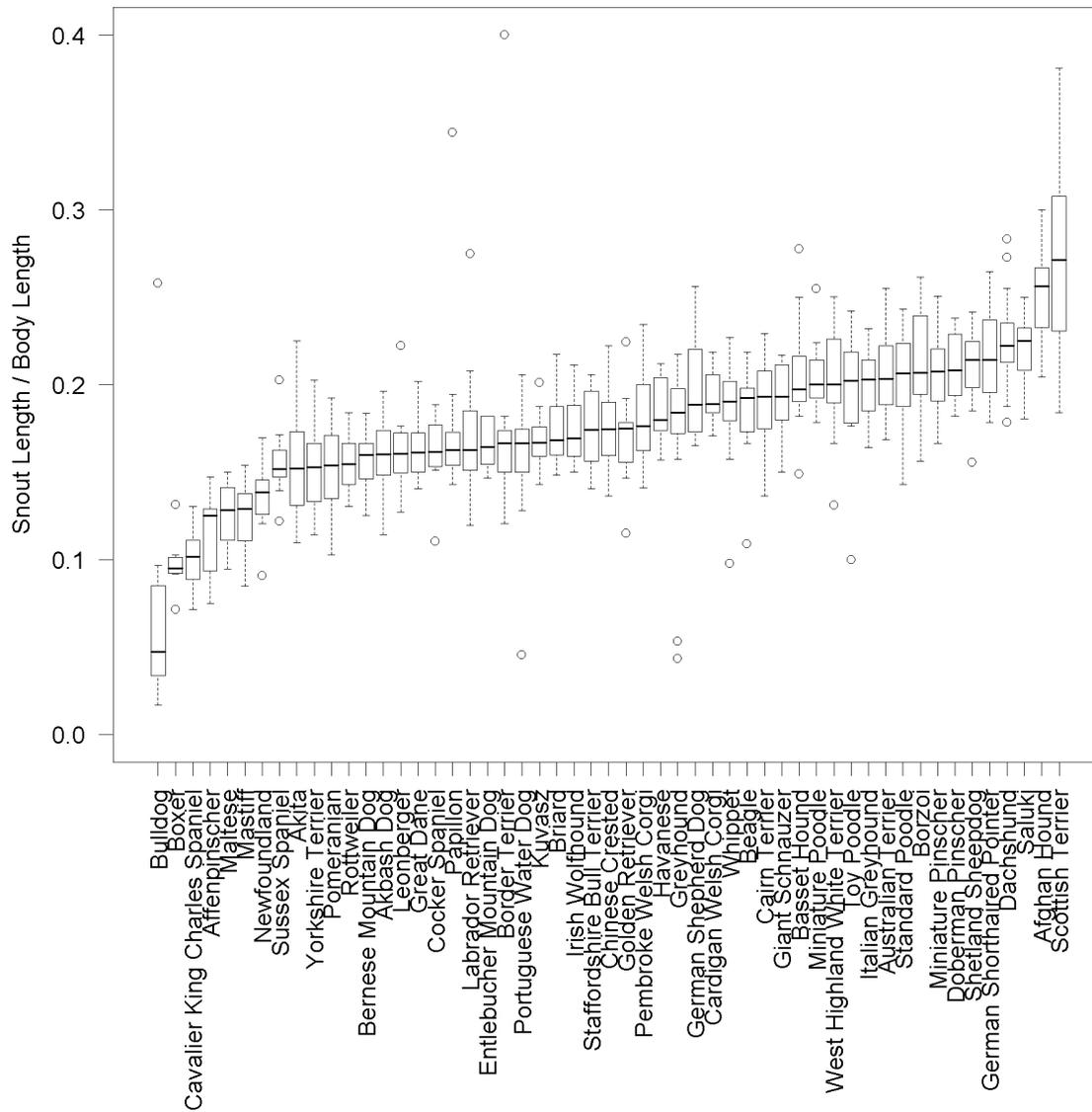
Supplementary Fig. 2. Relative snout length, defined as snout length / body length. The 53 breeds with at minimum three male and three female samples are plotted. The box central bar indicates the median, the box ends delimit the 25th and 75th percentile and the whiskers delimit the most extreme data point within 1.5 times the interquartile range. Outliers are shown as unfilled circles.

3
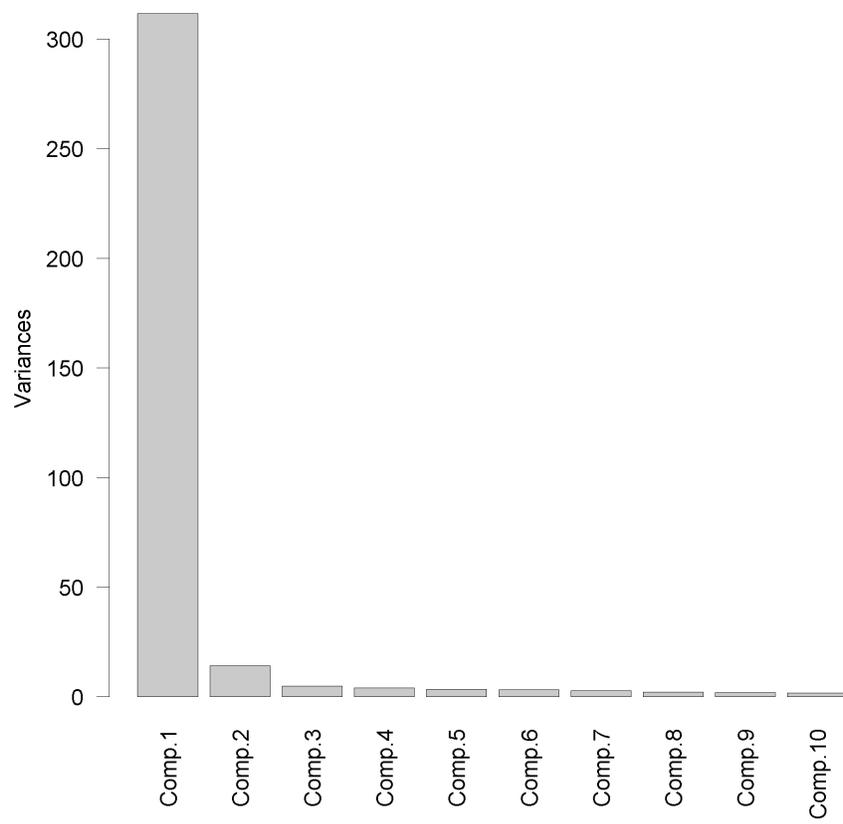
Supplementary Fig. 3. Scree plot of variance explained by first 10 principal components from principal component analysis performed on 1155 measured dogs.
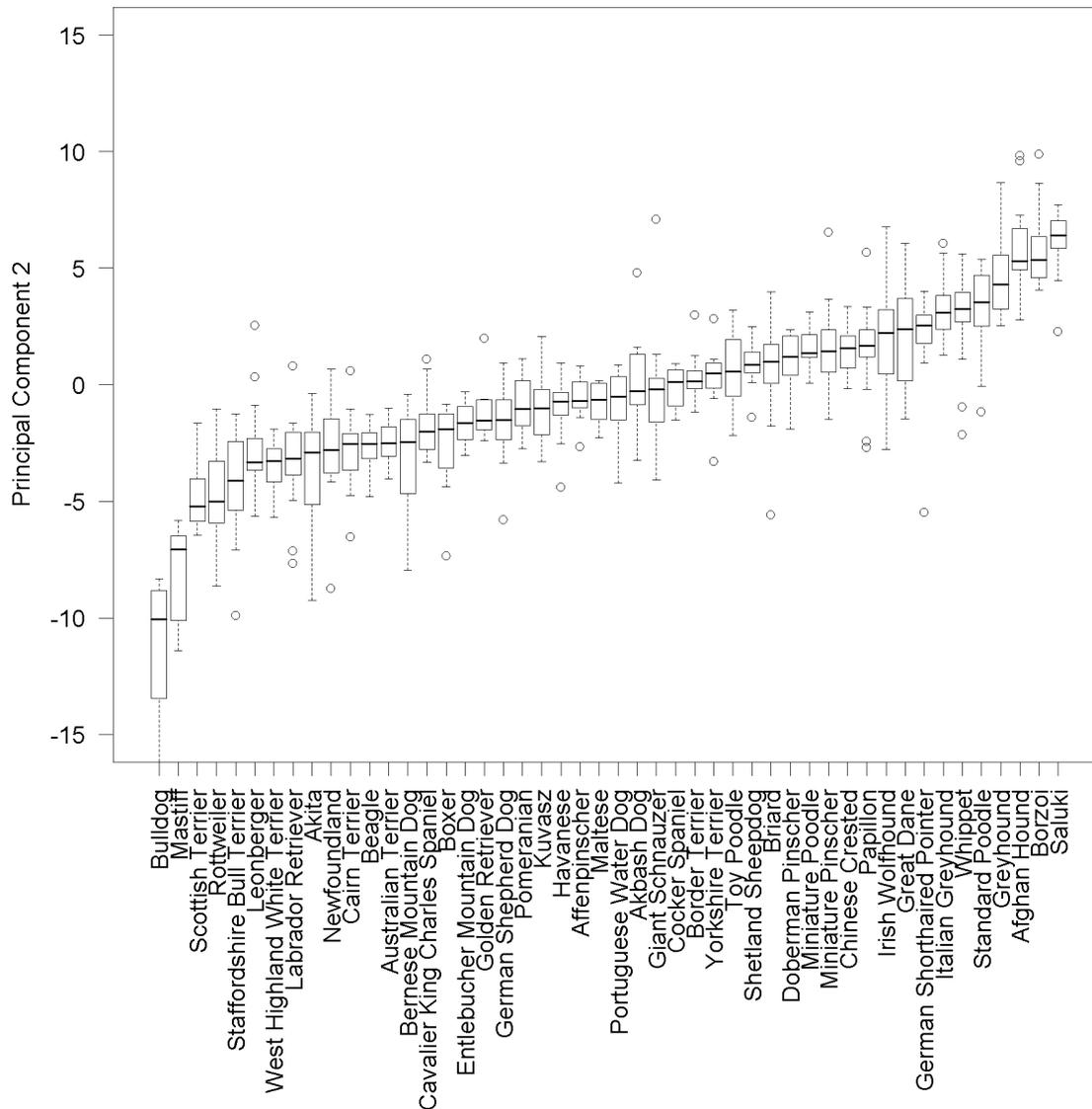
Supplementary Fig. 4. Boxplots of each breed's distribution of the second principal component

obtained from PCA with the following six achondroplastic breeds removed from the dataset:

basset hound, dachshund, Pembroke and Cardigan Welsh corgi, Sussex spaniel, and Dandie

Dinmont terrier. Plotting is as for Fig 7.
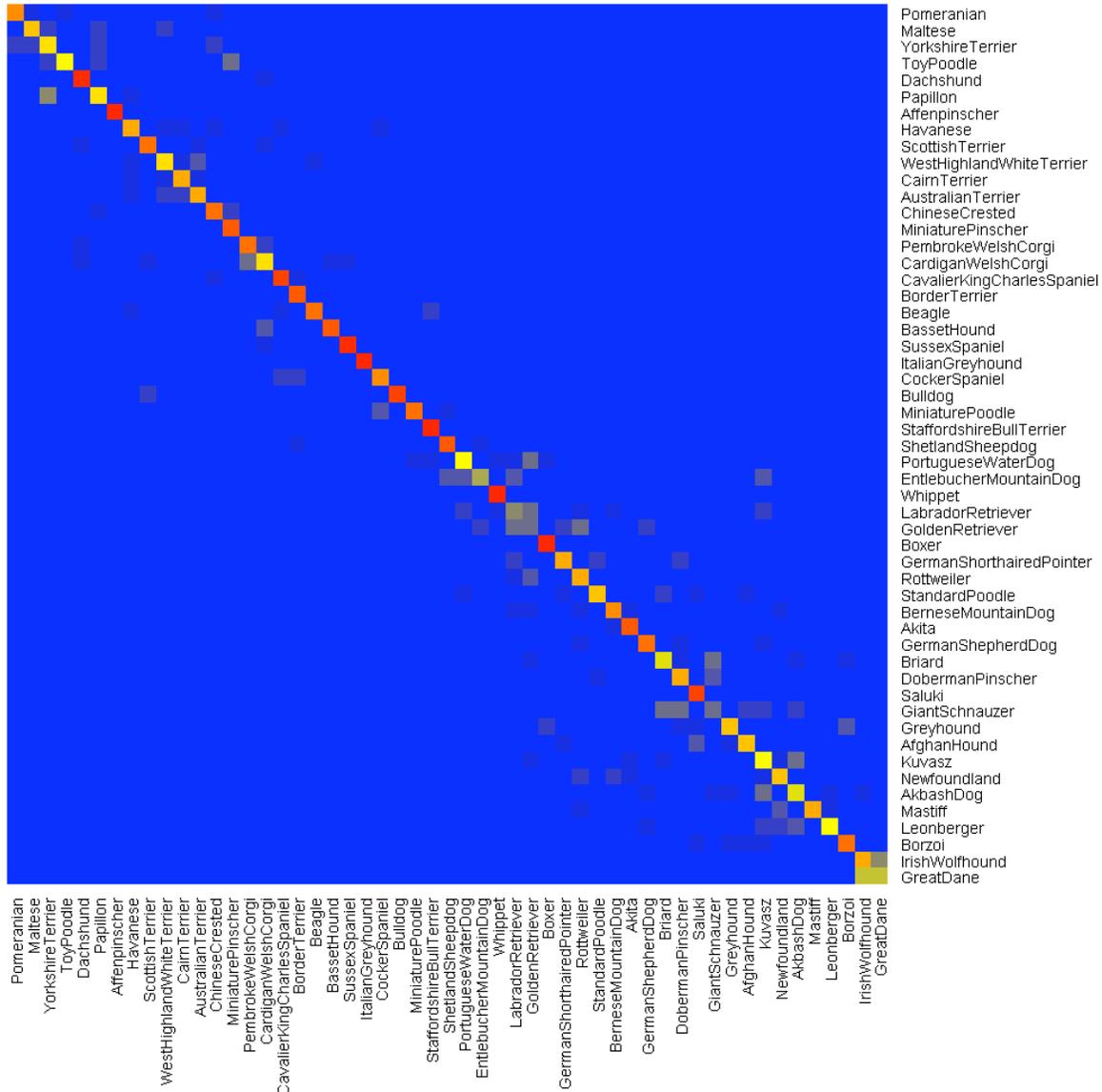
Supplementary Fig. 5. Heatmap showing the percentage of dogs classified for each breed with a minimum of three male and three females measured for 19 traits. Actual breed identity is shown on each row while column values indicate the predicted breed membership; correct classification is along the diagonal and mis-classification is off diagonal. The color values range from blue to yellow to red in 5% increments with blue blocks indicating 0% and red blocks indicating 100% classification. The matrix is sorted on the first discriminate function of the breed centroids value which is strongly correlated to height at withers.

6
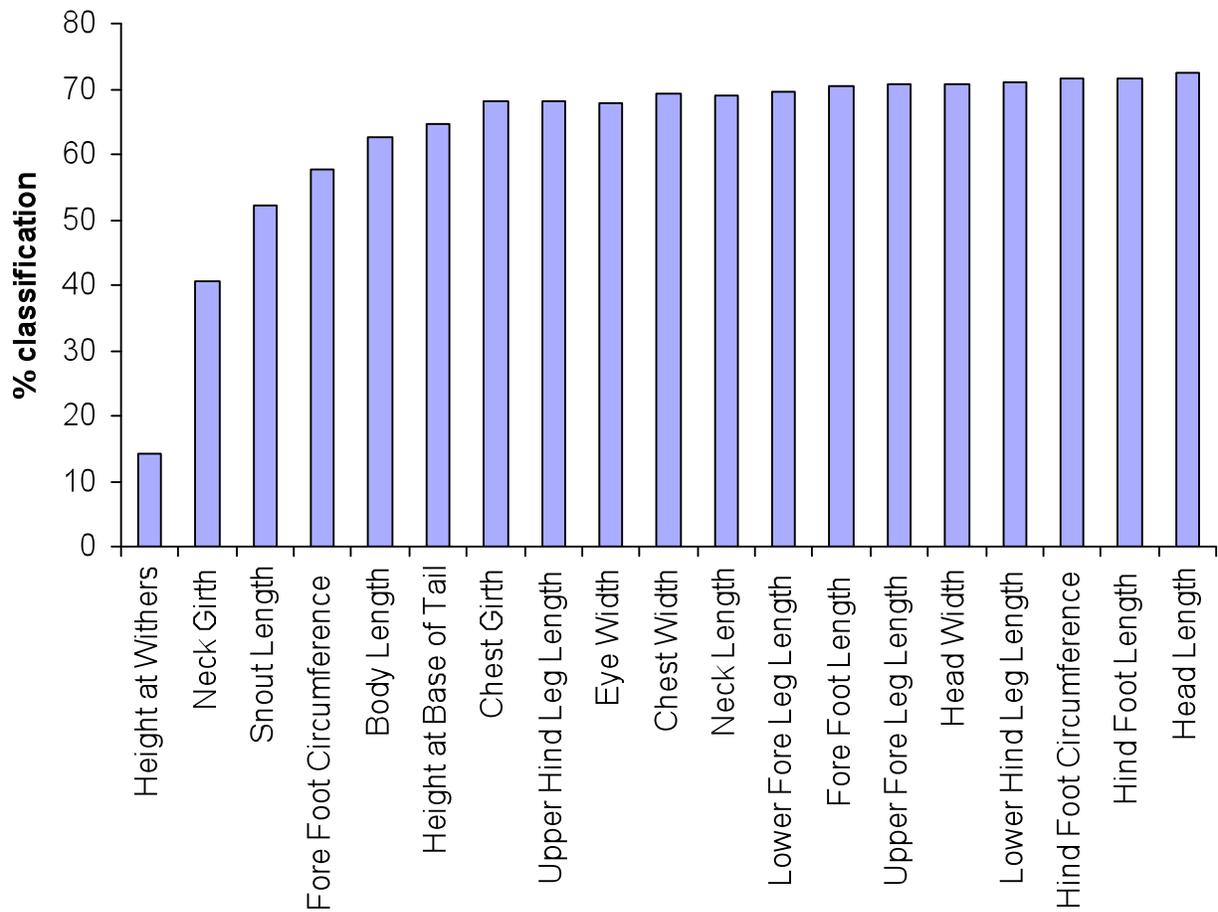
Supplementary Fig. 6. The percentage of successful classifications to breed with stepwise introduction of variables into the DFA. The x axis label indicates the variable introduced at that step. For example the second bar from the left is based on the two variables height at withers and neck girth.