1 **Appendix S1: Statistical and Analytical Details**

2 **Collyer, Heilveil, and Stockwell**

3 <u>Pool-adjusted shape values</u>

4   Univariate and multivariate analyses of variance (ANOVA and MANOVA,

5 respectively) rely on linear models for estimation of fixed and random effects. Multivariate

6 test statistics (e.g., Wilks' Λ) for MANOVA– which do not have proper distributions under

7 null hypotheses – must be approximated by F statistics to assess significance [1]. Problems

8 arise, however, in low-dimensional cases (e.g., when the number of response variables

9 exceeds the number of responses, or when the number of model parameters is large

10 compared to the number of responses) because there is no solution for conversion of

11 multivariate test statistics to F statistics in some cases [1] and type-I error rates are

12 exceedingly high for small subject-variable ratios [2]; nevertheless, effects can still be

13 estimated by linear models, and appropriate statistical tests can also still be evaluated

14 using a portion of model effects.

15   Our data suffered from low-dimensionality because the number of pools was

16 smaller than the number of fixed effects. This problem was alleviated by adjusting

17 individual fish shapes by their pool effects. The linear model for shape analyses in our

18 study had the form $\mathbf{y}_{ijkl} = \boldsymbol{\mu} + size + \mathbf{p}_i + \mathbf{s}_j + \mathbf{ps}_{ij} + \mathbf{pool}_k \mid \mathbf{ps}_{ij} + \boldsymbol{\varepsilon}_{ijkl}$, meaning that any $1 \times p$

19 vector of p shape values ($\mathbf{y}$) for lth individual fish in the ith population and jth salinity

20 treatment, assigned to the kth pool, is described by the overall mean ($\boldsymbol{\mu}$), plus

21 corresponding population ($\mathbf{p}$), salinity ($\mathbf{s}$), and interaction ($\mathbf{ps}$) fixed effects, plus the pool

22 random effect (nested within the population × salinity interaction), plus the covariate, size

23 (measured as log centroid size – see article). (Note that bold values refer to vectors.) The

24 vector, $\boldsymbol{\varepsilon}_{ijkl}$, is a 1×*p* vector of residuals. The linear model can be written in matrix form as

25 $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ to summarize that the *n* vectors of shape values in the *n*×*p* matrix, **Y**, are a

26 function of covariates and dummy variables in the *n*×*k* design matrix, **X**, for the *k*

27 parameters needed to describe the linear model above. The *k*×*p* matrix of regression

28 coefficients, $\boldsymbol{\beta}$, describes the effects of each parameter in the model, and the *n*×*p* matrix, $\boldsymbol{\varepsilon}$,

29 is a matrix of residuals. Parameter effects can be estimated as $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\left(\mathbf{X}^{\mathsf{T}}\mathbf{Y}\right)$, where

30 the superscripts, T, and $^{-1}$, refer to matrix transpose and inverse, respectively [3]. Any

31 estimated shape is thus solved as $\hat{\mathbf{y}} = \mathbf{x}\hat{\boldsymbol{\beta}}$, where **x** is a vector of appropriately coded

32 parameters.

33   An important property of $\hat{\boldsymbol{\beta}}$ is that any estimated parameter effect is calculated with

34 respect to other effects in the model. For example, population effects are estimated with

35 respect to pool effects. Shape values can thus be estimated for only effects of interest

36 without excluding extraneous sources of shape variation. Let matrix, $\hat{\boldsymbol{\beta}}'$, be a matrix of

37 coefficients found in $\hat{\boldsymbol{\beta}}$, but with pool effects removed. This procedure is the same as mean-

38 centering, if the intercept is removed [4]; therefore, the equation $\mathbf{Z} = \mathbf{X}\hat{\boldsymbol{\beta}}' + \boldsymbol{\varepsilon}$ creates

39 adjusted shape values, **Z**, which have pool effects removed, but population and salinity

40 effects are still estimated with respect to pool effects.   A linear model of the form

41 $\mathbf{z}_{ijkl} = \boldsymbol{\mu} + size + \mathbf{p}_i + \mathbf{s}_j + \mathbf{ps}_{ij} + \boldsymbol{\varepsilon}_{ijk}$ can be described such that individual fish, not pools are

42 subjects, and $\boldsymbol{\varepsilon}_{ijk}$ does not inherently contain pool effects. Analyses of variance can be

43   performed on this model without the aforementioned statistical problems; however, one

44   must be cautious that inflation of statistical power can result in incorrect inferences, [5].

45   In order to perform multivariate analysis of variance (MANOVA) without

46   introducing problems of statistical power, we used a permutation procedure, which

47   randomized individual shape values and recalculated model parameter estimates with each

48   permutation. Sums of squares and cross products (SSCP) matrices were calculated for each

49   effect [3] and the traces of these matrices – which are the total sums of squares for each

50   effect – were used as test statistics. P-values were determined as percentiles of the

51   observed test statistics in the empirical distributions of randomly generated statistics. This

52   procedure has three key advantages. First, the test statistics are not plagued by improper

53   degrees of freedom. Second, the test works equally well for univariate data, as the trace of

54   a SSCP matrix is the sums of squares of the effect, itself. Third, R^2 values can be calculated

55   for each effect, and can be compared within and between different analyses of univariate or

56   multivariate shape data. We performed ANOVA and MANOVA for both male and female

57   data sets using this method with 10,000 random permutations (the observed values

58   constituting one permutation).

59

60   Visualization of multivariate reaction norms

61   Compared to univariate reaction norms in Figs 2 & 3, males and females had some

62   differences in comparison of multivariate reaction norms, shown as vectors in the principal

63   component plots of Figs S1 & S2. First, reaction norms were more aligned with the first PC

64   for males, indicating that phenotypic plasticity in body shape was rather consistent with

65   differences due to population source (Fig S1). Reaction norms for females indicated that

66   phenotypic plasticity was associated more so with the second PC, suggesting that body

67   slendering in response to salinity was somewhat independent of evolved differences in

68   body depth. For females more so than males, phenotypic plasticity in body shape for SS

69   fish (principally along PC 2) was strikingly more independent of evolved differences in

70   shape (principally along PC1). This visual result confirms why the *population source ×*

71   *salinity* interaction was significant in the MANOVA (Table 2) for females.

72   Although not obvious in the PC plots, differences between MM and SS body shapes were

73   more pronounced in males. Procrustes distance, *d*, between SS high salinity fish and MM

74   low salinity fish – the greatest inter-group difference – was 0.044 for males, compared to

75   0.027 for females. Females and males had similar ranges of phenotypic plasticities: Males:

76    $0.013 \leq d \leq 0.024$ and Females: $0.012 \leq d \leq 0.021$. These results confirm the differences

77   between males and females in terms of the ANOVA and MANOVA results. Males tend to

78   have greater differences in shape based on population source. These results also suggest

79   that the response to salinity changes is similar in scale for the two native strains; however,

80   the differences in direction of shape change in the morphospace (Figs S1 and S2) indicate

81   that the two populations might have different genetic covariances in the traits that make up

82   body shape. Comparison of shape changes between pure Salt Creek fish (SS) and the other

83   mesocosm types, between low and high salinity, suggests that SS fish deepen the dorsal

84   aspect of the bodies in low salinity, whereas body deepening is more general for Malpais

85   Spring (MM) fish and hybrids (either MS or SM).

86   Finally, maternal effects are more apparent in females, at least for hybrids from Malpais

87   Spring female parents. MS hybrids were more similar to MM fish than SM fish. For both

88   males and females, hybrid reaction norms were closer in the morphospace to the MM

89 reaction norm than the SS reaction norm, although this is more readily apparent for

90 females. These results suggest that shape differences between MM and SS fish might be

91 exacerbated by maternal effects. Based on the close proximity of MS and MM reaction

92 norms, and the distinction of SM and SS reaction norms, females from less saline

93 environments are prone to produce deep-bodied offspring but females from saline

94 environments are not prone to produce slender-bodied offspring.

95

96 Multivariate AIC

97 The formula for multivariate AIC is given by Bedrick and Tsai [6] as

98 $\text{AIC} = n\left[\ln\left(\frac{|\mathbf{E}|}{n^p}\right) + p\right] + 2\left[pk + 0.5p(p+1)\right]$, where n is the number of subjects, k is the number

99 of model parameters, and p is the number of response variables. This equation has two

100 parts: the first part express the log likelihood of a model, where $|\mathbf{E}|$ is the determinant of

101 the sums of square and cross-products matrix of the model error; the second part is the

102 parameter penalty. For univariate data the parameter penalty simplifies to $2(k + 1)$, which

103 is often called $2K$ for convention [7].

104 Burnham and Anderson [7] recommended that a Δ AIC value of less than 2 suggests

105 that neither of two competing models is sufficiently better than the other. This "rule of

106 thumb", however, only applies to univariate response data. This value suggests that based

107 on AIC scores, one would not expect two models with the same log-likelihoods to differ by

108 more than one parameter. For example, if two models had the same log-likelihood but

109 differed by one parameter (i.e., one model had $k + 1$ parameters compared to the k

110 parameters of the other) for univariate response data, then Δ AIC =

111     $2\big[(k+1)+1\big]-2\big[k+1\big]=2$. Thus, a difference in AIC values of 2 more or less suggests that

112   these models would not be expected to differ by more than one parameter if their log

113   likelihoods were similar.

114   Using the multivariate development of this logic [6] for the p possible shape

115   variables in a morphmetric study, two models with the same likelihood and which differ by

116   one parameter would also differ by Δ AIC = $2p$, because:

117
118

$$\text{AIC}_1-\text{AIC}_2 = \left| n_1\left[\ln\left(\frac{|\mathbf{E}_1|}{n_1^p}\right)+p\right]+2\big[p(k+1)+0.5p(p+1)\big] \right| - \left| n_2\left[\ln\left(\frac{|\mathbf{E}_2|}{n_2^p}\right)+p\right]+2\big[p(k)+0.5p(p+1)\big] \right|$$

119         $\approx 2\big[p(k+1)+0.5p(p+1)\big]-2\big[p(k)+0.5p(p+1)\big]$

120         $\approx 2\big[p(k+1)-pk\big]$

121         $\approx 2\big[pk+p-pk\big]$

122         $\approx 2p$

123   Thus, it is difficult to either use the "rule of thumb" frequently used by many in model

124   comparisons or to compare results of model comparisons between different numbers of

125   principal components used to describe shape variables. The "rule of thumb" scales

126   proportionally to the number of shape variables used; therefore, multiplying AIC values by

127     $1/p$ nullifies this scaling, i.e., $\text{AIC}^* = n\left[\dfrac{\ln\left(\dfrac{|\mathbf{E}|}{n^p}\right)}{p}+1\right]+2k+p+1.$

128   Applying the same logic (of same log likelihoods) yields $\text{AIC}_1-\text{AIC}_2 \approx 2(k+1)-2k=2$. We

129   used this adjustment of AIC so that univariate and multivariate models would be more

130   comparable.

131   **Literature Cited for Appendix S1**

132

133   1. Rencher, AC (2002) Methods of multivariate analysis. New York: Wiley.

134   2. Elliot, RS, Barcikowski, RS (1993) Multivariate test statistics and their

135   approximations: some problems. Education Resources Information Center, report

136   #ED359231.

137   3. Rencer, AC, Schaalje, GB (2008) Linear models in statistics. New York: Wiley.

138   4. Sokal, RR, Rohlf, FJ (1995) Biometry, 3rd ed. San Francisco: Freeman.

139   5. Hurlbert, SH (1984) Pseudoreplication and the design of ecological field

140   experiments. Ecological Monographs 54:187-211.

141   6. Bedrick, EJ, Tsai, C (1994) Model selection for multivariate regression in small

142   samples. Biometrics 50:226-231.

143   7. Burnham, KP, Anderson, DR (2002) Model selection and multimodel inference: a

144   practical information-theoretic approach, 2nd Edition. New York: Springer-Verlag.

145