UTTAR PRADESH JOURNAL OF ZOOLOGY

41(24): 76-83, 2020
ISSN: 0256-971X (P)

MORPHOMETRIC COMPARISON OF CHARACIDAE FISH WHEN $n < p$

DANNY VILLEGAS RIVAS[1*], MANUEL MILLA PINO[2],
TOMÁS RODRÍGUEZ BEAS[3], MIRYAM LORA LOZA[3],
MARTÍN GRADOS VASQUEZ[3], CESAR OSORIO CARRERA[4],
RICARDO SHIMABUKU YSA[5], RIVER CHÁVEZ SANTOS[6],
LUIS RAMÍREZ CALDERÓN[3], YDALIA VELÁSQUEZ CASANA[7],
WILFREDO RUIZ CAMACHO[1], HUGO EGOCHEAGA CASAS[8],
CLELIA JIMA CHAMIQUIT[9], ROBERTO PACHECO ROBLES[10],
ZADITH GARRIDO CAMPAÑA[2] AND ERICK DELGADO BAZAN[2]

[1]Faculty of Forestry and Environmental Engineering, National University of Jaén, Cajamarca, Peru.
[2]Faculty of Civil Engineering. National University of Jaén, Cajamarca, Peru.
[3]Postgraduate School, César Vallejo University, Peru.
[4]Postgraduate School, Continental University, Lima, Peru.
[5]Faculty of Mechanical and Electrical Engineering, National University of Jaén, Cajamarca, Peru.
[6]Faculty of Economic and Administrative Sciences, 'Toribio Rodríguez de Mendoza' National University of Amazonas, Perú.
[7]Postgraduate School, National University of Trujillo, Perú.
[8]Faculty of Administration and Business, Technological University of Perú, Peru.
[9]Faculty of Health Sciences, Toribio Rodríguez de Mendoza' National University of Amazonas, Perú.
[10]Faculty of Accounting and International Business and Tourism, National Autonomous University of Alto Amazonas, Perú.

AUTHORS' CONTRIBUTIONS

This work was carried out in collaboration among all authors. Authors DVR and MMP designed the study, performed the statistical analysis, wrote the protocol and wrote the first draft of the manuscript. Authors TRB, MLL, WRC, RSY and RCS managed the analyses of the study. Authors MGV, COC, YVC, LRC, HEC, RPR CJC, ZGC and EDB managed the literature searches. All authors read and approved the final manuscript.

Article Information

Editor(s):
(1) Dr. Telat Yanik, Professor, Atatürk University, Turkey.
Reviewers:
(1) Gómez Valeria, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina.
(2) Luana Pereira dos Santos Lopes, Universidade Federal de Uberlândia, Brasil.
(3) Suresh Chandra Joshi, India.

Received: 25 October 2020
Accepted: 29 December 2020
Published: 30 January 2021

Original Research Article

**Corresponding author: Email: danny_villegas1@yahoo.com;*

ABSTRACT

Currently, great importance has been given to the study of external morphology, especially in fish, when it is used as a means of identifying hybrids. This paper considers a LASSO model based on the truss protocol to compare morphological covarion patterns between specimens of *Colossoma macropomum* and the hybrid *C. macropomum* (♀) x *P. orinoquensis* (♂). In this study, 25 specimens of *C. macropomum* and 20 specimens of the hybrid *C. macropomum* (♀) x *P. orinoquensis* (♂), respectively, were analyzed. The method "Truss protocol" or "trusses" Strauss and Bookstein (1982) was used. LASSO model achieved to reduce the mean squared error. The final model obtained contains only seven covariates. LASSO model fitted on the morphological covariation patterns between specimens of *C. macropomum* and the hybrid *C. macropomum* (♀) x *P. orinoquensis* (♂) showed a good fit and allowed to correctly classify most of the specimens. Differences were observed in the area of the head and in the anterior part of the fish evidenced in covariates associated with hydrodynamic abilities and with foraging.

Keywords: Morphometry; truss protocol; fishes; lambda; shrinkage regression.

1. INTRODUCTION

The family Characidae is the most diverse family of freshwater fish species in South America. The implementation of morphometric analysis in some species provides scientific knowledge that helps genetic improvement. The morphological characters are physical evidence of the expression of the genotype. Therefore, the differences between specific body characteristics can become very important to establish patterns of differentiation and inheritance [1, 2]. In continental fish, the morphometric characteristics referring to the anatomical shape have been used to evaluate the productive response in rearing both in natural environments and in captivity. Currently, there are more modern and precise morphometric analysis techniques, such as geometric morphometry [3, 4, 5, 6], which together with multivariate statistical analysis and means of direct visualization, constitute one of the most useful tools to describe the biological form and its changes.

Generally, these techniques are based on a set of measured distances between identifiable points on the organisms. In most cases, the measurements (distances between homologous points) present a high correlation, which is exploited in the models that are frequently used to compare between species. However, a variable selection model has desirable requirements: accurate predictions, interpretable models and stability, that is, small changes in the data should not cause large changes in the predictors used. Traditional methods of variable selection, such as stepwise regression, all subsets regression, or ridge regression, fail one or more of the above requirements. Modern procedures such as boosting [7], forward stagewise regression [8], and LASSO [9] generally improve stability and predictions. Although LASSO works successfully on many occasions, it has some limitations, which can be solved with the model known as Elastic Net [10]. In this sense, this paper considers LASSO models based on the truss protocol to compare morphological covarion patterns between specimens of *C. macropomum* and the hybrid *C. macropomum* (♀) x *P. orinoquensis* (♂) when $p > n$, that is, we have more variables than observations using lars and glmnet package in R.

2. METHODOLOGY

Morphological Covariation Patterns between *C. macropomum* and the hybrid *C. macropomum* (♀) x *P. orinoquensis* (♂): In this study, 25 adult specimens of *C. macropomum* and 20 adult specimens of the hybrid *C. macropomum* (♀) x *P. orinoquensis* (♂) with an average weight of 600 g, respectively, from artificial ponds of a fish farm in Portuguesa state, Venezuela, were analyzed. Within the sample of each species there are mixed male and female individuals. The method "Truss protocol" or "trusses" Strauss and Bookstein [11] was used, which achieves an exhaustive reconstruction of the shape from the distances between the homologous anatomical landmarks (see Table 1 and Fig. 1). The distances connecting these landmarks form a series of continuous quadrilaterals with their respective internal diagonals (see Fig. 1), which allows detecting differences in shape in the vertical, horizontal, and oblique directions. The limitations in this study is the number of measures necessary to achieve better efficiency in estimating parameters related to the morphology of these species.

The morphological covariation patterns between specimens of *C. macropomum* and the hybrid *C. macropomum* (♀) x *P. orinoquensis* (♂) were studied using LASSO models in R package [12].

The LASSO method (Least Absolute Shrinkage and Selection Operator): Introduced by Tibshirani [9], is a method that combines a regression model with a procedure for contracting some parameters towards zero and selecting variables, by imposing a restriction or penalty on the regression coefficients.

Below is a formulation of Lasso as an optimization problem (for details see Ramos, [10]):

Suppose we have the data $(x_i, y_i), i = 1, 2, \ldots, N$, where $x_i = (x_{i1}, \ldots, x_{ip})$ t are the predictor variables

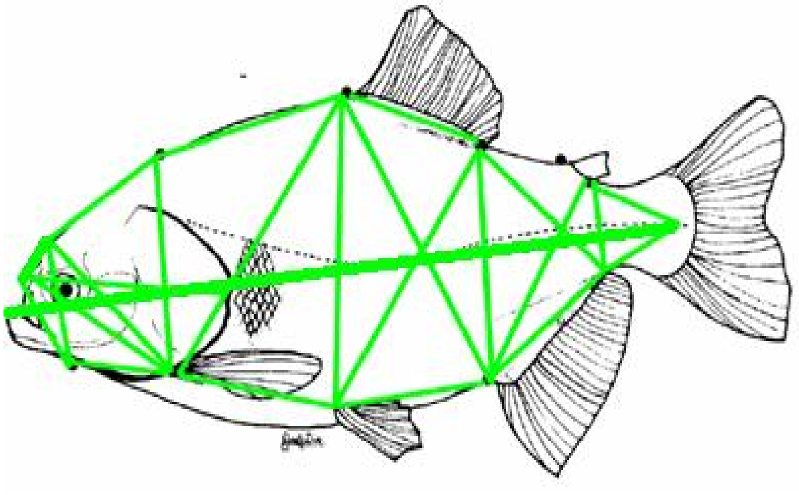

Fig. 1. Location of homologous points and distances measured on the left lateral profile of *C. macropomum* and the hybrid *C. macropomum* (♀) x *P. orinoquensis* (♂)

Table 1. Truss measurements from *C. macropomum* and the hybrid *C. macropomum* (♀) x *P. orinoquensis* (♂) specimens

Standar length (X_1)
Tip of snout to end of epiphyseal sulcus (X_2)
Tip of snout to insertion of pectoral fin (X_3)
Anterior edge of the epiphyseal sulcus to the end of the epiphyseal sulcus (X_4)
Anterior edge of the epiphyseal sulcus at the insertion of the pectoral fin (X_5)
Anterior edge of the epiphyseal sulcus when articulating (X_6)
Articulate to insertion of pectoral fin (X_7)
Posterior edge of epiphyseal sulcus to end of dorsal fin (X_8)
Posterior edge of the epiphyseal sulcus at the insertion of the pelvic fin (X_9)
Posterior edge of the epiphyseal sulcus to the insertion of the pectoral fin (X_{10})
Posterior edge of the epiphyseal groove when articulating (X_{11})
Insertion of pectoral fin to insertion of pelvic fin (X_{12})
Dorsal fin base (X_{13})
Anterior edge of dorsal fin to anterior edge of anal fin (X_{14})
Anterior edge of dorsal fin to insertion of pelvic fin (X_{15})
Anterior edge of dorsal fin to insertion of pectoral fin (X_{16})
Insertion of pelvic fin to end of anal fin (X_{17})
Posterior edge of dorsal fin to the fatty fin (X_{18})
Posterior edge of dorsal fin to posterior edge of anal fin (X_{19})
Posterior edge of dorsal fin to anterior edge of anal fin (X_{20})
Posterior edge of dorsal fin to insertion of pelvic fin (X_{21})
Anal fin base (X_{22})
Posterior edge of the fatty fin to the last scale of the lateral line (X_{23})
Posterior edge of fatty fin to posterior edge of anal fin (X_{24})
Posterior edge of the fatty fin to the anterior border of the anal fin (X_{25})
Posterior edge of the fatty fin to the anterior border of the anal fin (X_{26})
Eye diameter (X_{27})
Head length (X_{28})
Fat fin base (X_{29})

and y_i are the responses. We can consider that the x_{ij} are standardized, that is,

$$\sum_i {x_{ij}}/_{N} = 0,$$

$$\sum_i {x_{ij}^2}/_{N} = 1$$

Or in other words, they have zero mean and variance 1. If the previous condition is not verified, it is enough to classify the variables as part of the preprocessing.

If we denote $\hat{\beta} = \left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right)^{\mathrm{T}}$, the estimate of lasso $\left(\hat{\alpha}, \hat{\beta}\right)$ is defined as the optimal solution of the optimization problem:

$$\text{mín}_{\alpha,\beta} \left\{ \sum_{i=1}^{N} \left(y_i - \alpha - \sum_j \beta_j x_{ij} \right)^2 \right\}$$

subject to $\sum_j |\beta_j| \leq t$

where $t \geq 0$ is a fitting parameter.

Fixed β that satisfies $\sum_j |\beta_j| \leq t$, optimize in is a differentiable optimization problem in a variable, whose optimality condition is gradient equal to zero.

Prediction and estimation of the parameter *t*

We estimate the prediction error for the LASSO using a cross-validation with k-folds.

If we call

$$s = \frac{t}{\sum_{i=1}^{p} \hat{\beta}_j^{o}},$$

where $\hat{\beta}_j^{o}$ are the least squares estimators, and we vary s in a sufficiently small interval, between 0 and 1, for each value of s or respectively of t, we obtain by cross-validation an estimator $\hat{e}\,(t)$, of mean square prediction error. We thus determine t^*, value of *t* with smaller $\hat{e}\,(t)$, and this is the parameter considered.

Algorithms to find solutions: Once we have obtained an estimate of *t*, which we will call t^*, we proceed to solve the optimization problem;

$$\text{mín}_{\alpha,\beta} \sum_{i=1}^{N} (y_i - x_i^t \beta)^2$$

subject to $\sum_{i=1}^{p} |\beta_j| \leq t^*$

We observe that the previous problem has p variables, since $\beta \in \mathbb{R}^p$, and a constraint; We can transform this restriction into 2^p linear restrictions:

$$\|\beta\|_1 \leq t^*$$

$$\sum_{i=1}^{p} |\beta_j| \leq t^*$$

$$\sum_{i=1}^{p} \beta_i^+ + \beta_i^- \leq t^*$$

$$\sum_{i=1}^{p} u_i \beta_i \leq t^* \quad \forall \left(u_1, \ldots, u_p\right) \in \{-1,1\}^p$$

The previous problem is a convex quadratic optimization problem with 2^p linear constraints. It is possible to obtain an equivalent formulation with a linear number in p of constraints, expanding the number of variables. For this we make the change:

$$\beta = \beta_i^+ - \beta_i^-,$$

considering that β_i can be expressed as

$$\beta_i = \beta_i^+ - \beta_i^-,$$

With

$$\beta_i^+, \beta_i^- \geq 0.$$

from where

$$|\beta_i| = \beta_i^+ + \beta_i^-.$$

Therefore,

$$\text{mín}_{\beta^+,\beta^-} \sum_{i=1}^{N} \left(y_i - x_i^t(\beta_i^+ - \beta_i^-)\right)^2$$

subject to $\sum_{i=1}^{p} (\beta_i^+ + \beta_i^-) \leq t^*$

$$\beta^+, \beta^- \geq 0$$

This problem has 2^p variables since $\beta^+, \beta^- \in \mathbb{R}^p$, and $2p + 1$ constraints.

The lars package in R: Computes the K-fold cross-validated mean squared prediction error for LARS, LASSO, or Forward Stagewise. For details see Hastie and Efron [13].

Usage: cv. lars(x, y, K = 10, index, trace = FALSE, plot.it = TRUE, se = TRUE,type = c("lasso", "lar", "forward. stagewise", "stepwise"), mode=c("fraction", "step"), ...)

Arguments

x Input to lars
y Input to lars
K Number of folds

index Abscissa values at which CV curve should be computed. If mode="fraction" this is the fraction of the saturated |beta|. The default value in this case is index=seq(from = 0, to = 1, length =100). If mode="step", this is the number of steps in lars procedure. The default is complex in this case, and depends on whether N>p or not. In principal it is index=1:p. Users can supply their own values of index (with care).

trace Show computations?

plot. it Plot it?

se Include standard error bands?

type type of lars fit, with default "lasso"

mode This refers to the index that is used for cross-validation. The default is "fraction" for type="lasso" or type="forward.stagewise". For type="lar" or type="stepwise" the default is "step"

Additional arguments to lars

Value Invisibly returns a list with components (which can be plotted using plotCVlars)

index As above

cv The CV curve at each value of index

cv.error The standard error of the CV curve

mode As above

3. RESULTS AND DISCUSSION

Table 2 and Fig. 1 show the fit of the LASSO model on patterns of morphological covariation between *C. macropomum* and the hybrid *C. macropomum (♀) x P. orinoquensis (♂).*, where the covariates (landmarks distances): eye diameter (X_{27}), anterior edge of the epiphyseal sulcus to the end of the epiphyseal sulcus (X_4), posterior edge of the epiphyseal sulcus to the insertion of the pectoral fin (X_{10}), insertion of pectoral fin to insertion of pelvic fin (X_{12}), posterior edge of dorsal fin to the fatty fin (X_{18}), anterior edge of dorsal fin to insertion of pectoral fin (X_{16}) and posterior edge of dorsal fin to posterior edge of anal fin (X_{19}) were included in the model, suggesting there are characteristics associated with the morphological covariation patterns that allow differentiation between redundant specimens of *C. macropomum* and the hybrid *C. macropomum* (♀) x *P. orinoquensis* (♂). These covariates are associated with morphological covariation patterns that make a difference in the head area and in the anterior part of the fish. These covariates are characteristics associated with hydrodynamic abilities and to the foraging for food. Fig. 3 shows how lasso achieves, using their respective optimal values of λ, to reduce the MSE. The advantage of the final model obtained by lasso is that it is much simpler since it contains only seven covariates. These results coincide with those reported by Pineda et al., [14] who used principal component analysis for the morphometric comparison

Table 2. LASSO model fitted on morphological covariation patterns between *C. macropomum* and the hybrid *C. macropomum (♀) x P. orinoquensis (♂)*

Land marks distance	LASSO model coefficients
Intercept	3.7965292496
Anterior edge of the epiphyseal sulcus to the end of the epiphyseal sulcus (X_4)	-0.0116715619
Posterior edge of the epiphyseal sulcus to the insertion of the pectoral fin (X_{10})	0.0001085698
Insertion of pectoral fin to insertion of pelvic fin (X_{12})	0.0159858221
Anterior edge of dorsal fin to insertion of pectoral fin (X_{16})	0.0015759937
Posterior edge of dorsal fin to posterior edge of anal fin (X_{19})	0.0011496300
Eye diameter (X_{27})	-0.1496373453
Fat fin base (X_{29})	-0.0200950546
% Deviance	87.13
Optimum Lambda (λ)	0.02401

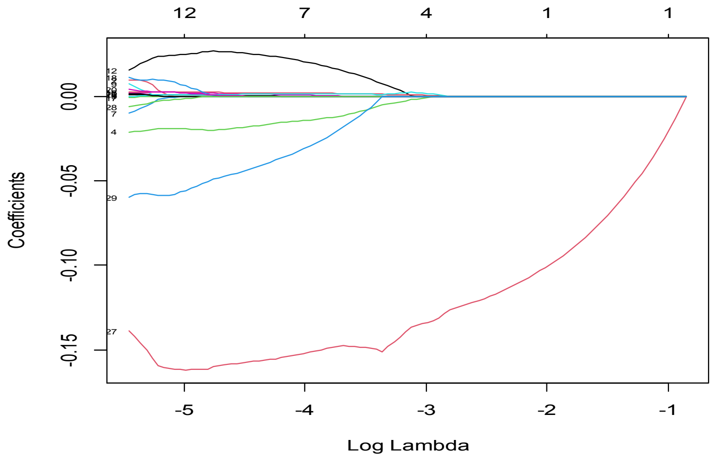

Fig. 2. LASSO adjustment on morphological covariation patterns between *C. macropomum* and the hybrid *C. macropomum* (♀) x *P. orinoquensis* (♂)

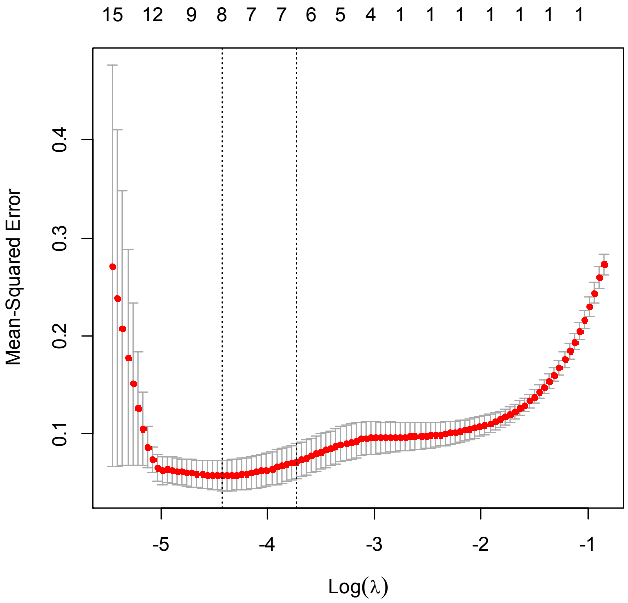

Fig. 3. Mean squared error for lambda (λ) in a LASSO model on morphological covariation patterns between *C. macropomum* and the hybrid *C. macropomum* (♀) x *P. orinoquensis* (♂)

between males and females of *C. macropomum* maintained in ponds, and those reported by Villegas et al., [15] in a multivariate analysis that allowed a morphometric comparison of a hybrid originated from *C. macropomum and P. orinoquensis*, and those reported by Villegas et al., [16] when studying the redundancy in morphological covariation patterns between *C. macropomum and P. orinoquensis*. However, the results differ from those indicated by Villegas et al [17] when using a multiple logistic model to study the morphological covariation patterns between the mentioned species. The foregoing reveals what was indicated by Porras-Rivera and Rodríguez-Pulido [18] and Conte-Grand et al., [19], who point out that external morphology is not always reliable when used as the only means of identification, particularly for hybrid individuals beyond the first generation.

4. CONCLUSIONS

LASSO model achieved, using their respective optimal values of λ, to reduce the mean squared error. The final model obtained by LASSO it was much simpler since it contains only seven covariates. LASSO model fitted on the morphological covariation patterns between specimens of *C. macropomum* and the hybrid *C. macropomum (♀) x P. orinoquensis (♂)* showed a good fit and allowed to correctly classify most of the specimens. Differences were observed in the area of the head and in the anterior part of the fish between the hybrid and its parent. The morphological differences between these two species were evidenced in covariates associated with hydrodynamic abilities and with foraging. Finally, the results of this research suggest the use of the LASSO model to compare morphological covariation patterns between the hybrid *C. macropomum* (♀) x *P. orinoquensis* (♂) and *P. orinoquensis* when the sample size is less than the number of landmarks ($n < p$).

ETHICAL APPROVAL

As per international standard or university standard ethical approval has been collected and preserved by the authors.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Mateo JF, Delgado N, López H. Caracterización morfométrica del híbrido yaque pintado (Pseudoplatystoma fasciatum x Leiarius marmoratus) y sus progenitores (Siluriformes: Pimelodidae). Rev. Fac. Cienc. Vet. U. C. V. 2008;49(1):47-60.
2. Salas D, Véliz D, Scott S. Diferenciación morfológica en especies del género Cheirodon (Ostariophysi: Characidae) mediante morfometría tradicional y geométrica. Gayana. 2012;76(2):142-52.
3. Bookstein FL, Chernoff R, Elder J, Humpries G, Smith Y, Strauss R. Morphometrics in evolutionary biology. The Academy of Natural Sciences of Philadelphia, Michigan; 1985.
4. Trapani J. Geometric morphometric analysis of body – form variability in Cichlasoma minckleyi, the Cuatro Cienagas cichlid. Environmental Biology of Fishes. 2003;68:357–369.
5. Adams DC, Rohlf FJ, Slice DE. Geometric morphometrics: ten years of progress following the "Revolution". Italian Journal of Zoology. 2004;71:5–16.
6. Shipunov AB, Bateman RM. Geometric morphometrics as a tool for understanding Dactylorhiza (Orchidaceae) diversity in European Russia. Biological Journal of the Linnean Society. 2005;85(1):1-12.
7. Freund Y, Schapire R. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences. 1997;55(1):119–139.
8. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning; Data mining, Inference and Prediction, Springer Verlag, New York; 2001.
9. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological). 1996;58(1):61–93.
10. Ramos L. LASSO regression. Facultad de Matemáticas. Universidad de Sevilla. España. 2018;61.
11. Strauss R, Bookstein F. The truss: body form reconstructions in morphometrics. Syst. Zool. 1982;31(2):113–135.
12. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2020.
13. Hastie T, Efron B. Paquete lars en r. urlhttps://CRAN.Rproject.org/package=Lars; 2013.
14. Pineda H, Restrepo L, Ángel M. Comparación morfométrica entre machos y hembras de Cachama Negra (Colossoma macropomum, Cuvier 1818) mantenidos en estanque. Revista Colombiana de Ciencias Pecuarias. 2004;17:24-29.
15. Villegas D, Milla M, Castillo O, Durant K. Análisis multivariado en la comparación morfométrica del híbrido Colossoma macropomum X Piaractus brachypomus) y sus parentales. Revista INDES. 2013;1(1):29-36.
16. Villegas D, M. Milla Y, Pérez S. Villegas Z. Garrido, Delgado E, Ruiz W, Velasquez Y, De Souza B. Redundancy in morphological covariation patterns between Colossoma macropomum and Piaractus orinoquensis. Uttar Pradesh Journal of Zoology. 2020a;41(14):37-46.
17. Villegas D, Milla M, Garrido Z, Grados M, Osorio C, Delgado E, Velasquez Y, Ruiz W, Shimabuku R, Paredes J, Lizarzaburu M. On a logistic model for morphological covariation patterns between colossoma macropomum and the hybrid colossoma macropomum (♀) x piaractus orinoquensis (♂). Uttar Pradesh Journal of Zoology. 2020b;41(18):28-34.

18. Porras-Rivera G, Rodríguez-Pulido JA. Morphometric comparison and characterization of the hybrid (Pseudoplatystoma metaense x Leiarius marmoratus) and its parental lines (Siluriformes: Pimelodidae). Int. J. Morphol. 2019;37(4):1409-1415.

19. Conte-Grand C, Sommer J, Orti G, Cussac V. Populations of Odontesthes (Teleostei: Atheriniformes) in the Andean region of Southern South America: Body shape and hybrid individuals. Neotropical Ichthyology. 2015;13(1):137–150.

APENDIX 1

R code for LASSO adjustment on morphological covariation patterns between *C. macropomum* and the hybrid *C. macropomum (♀) x P. orinoquensis (♂).*

```
> d<-datos
> d
> library(caret)
> library(glmnet)
> library(doParallel)
> library(doParallel)
> library(corrplot)
> set.seed(666)
> inTrainingSet <- createDataPartition(d$Especie, p = 0.5, list = FALSE)
> train <- d[inTrainingSet, ]
> test <- d[-inTrainingSet, ]
> predictors <- names(d)[!names(d) %in% "Especie"]
> x = train[,predictors]
> y = train$Especie
> x = as.matrix(x)
> set.seed(666)
> modelos_lasso <- glmnet(x = x, y = y, alpha = 1)
> plot(modelos_lasso, xvar = "lambda", label = TRUE)
> set.seed(1)
> cv_error_lasso <- cv.glmnet(x = x, y = y, alpha = 1, nfolds = 7)
> plot(cv_error_lasso)
> cv_error_lasso$lambda.min
> cv_error_lasso$lambda.1se
> # Se reajusta el modelo con todas las observaciones empleando el valor de
> # lambda óptimo
> modelo_final_lasso <- glmnet(x = x, y = y, alpha = 1, lambda = cv_error_lasso$lambda.1se)
> coef(modelo_final_lasso)
> plot(cv_error_lasso,ylab = "Mean Square Error lasso")
```