
ASSESSMENT OF INVASIVE POTENTIAL OF *HOMALODISCA COAGULATA* IN WESTERN NORTH AMERICA AND SOUTH AMERICA

A. Townsend Peterson¹, Ricardo Scachetti-Pereira², and Daniel A. Kluza^{1,3}

Biota Neotropica v3 (n1) – <http://www.biotaneotropica.org.br/v3n1/pt/abstract?article+BN00703012003>

Recebido em 19/12/2003

Revisado em 14/03/2003

Publicado em 26/03/2003

¹ Natural History Museum, The University of Kansas, Lawrence, Kansas 66044 USA. E-mail town@ku.edu
² Centro de Referência em Informação Ambiental, Av. Romeu Tórtima, 388, Barão Geraldo, 13084-520 Campinas SP, Brazil. E-mail ricardo@cria.org.br

³ E-mail dakluza@ku.edu

* To whom correspondence should be addressed: town@ku.edu

Abstract

The potential of *Homalodisca coagulata* to invade South America is a question of economic importance, given its potential impact as a disease vector for several crops. We developed ecological niche models for the species on its native geographic distribution in the southeastern United States; we tested the predictivity of the models both on the native distributional area and via projections to California, where the species has long been present as an invasive species. In both cases, tests indicated high statistical significance of predictions. Projection of models to South America indicated little possibility of invasion of southeastern Brazil, where citrus diseases were of concern. However, all models agree in predicting great risk of establishment in the wine-growing regions of northern Argentina and extreme southern Brazil; great precaution is thus to be recommended when any movements of bio-materials are made from infected areas to this region.

Keywords: *ecological niche modeling, invasive species, Xylella fastidiosa, vector.*

Resumo

A possibilidade de invasão da América do Sul por *Homalodisca coagulata* é uma questão de relevância econômica, dado o potencial de seu impacto como um vetor para agentes fitopatológicos. Foram desenvolvidos modelos de nicho ecológico para a espécie na sua área de distribuição geográfica nativa, no sudeste dos Estados Unidos. A capacidade de previsão dos modelos foi testada tanto na área de distribuição nativa quanto em projeções no estado da Califórnia, onde a espécie tem estado presente por muito tempo como uma espécie invasora. Em ambos os casos, os testes indicaram alta significância estatística para as previsões. Projeções dos mesmos modelos para a América do Sul indicaram poucas chances de invasão ao sudeste do Brasil, local em que doenças em culturas de cítricos são preocupantes. Entretanto, todos os modelos são concordantes ao prever alto risco de estabelecimento da espécie na região onde se encontram as vinícolas do norte da Argentina e no extremo sul do Brasil. Recomenda-se então grande precaução no transporte de material biológico de áreas infectadas para a área em questão.

Palavras-chave: *modelagem de nicho ecológico, espécies invasoras, Xylella fastidiosa, vetor.*

Introduction

The bacterium *Xylella fastidiosa* has long been a serious concern in a number of orchard-based crops, particularly in citrus (e.g., as variegated chlorosis in citrus in Brazil) and in wine grapes (e.g., as Pierce's disease in grapes in California), although also in a much broader set of crops (e.g., as phoney peach disease in the southern United States, oleander leaf scorch in California). Its introduction into a given plant generally leads to withering and dieback, and death of the plant within a few years. A number of good summaries of the interaction between *Xylella*, its vector insect species, and its effects on agriculture are available (<http://ucce.ucdavis.edu/counties/ceventura/Agriculture977>, <http://www.ipm.ucdavis.edu/PMG/r302101211.html>, and <http://www.cnr.berkeley.edu/xylella/>).

Xylella's effects are considerably more serious when competent vector species are present. Alone, *Xylella*'s dispersal abilities are quite minimal, but a competent vector species can increase its dispersal capacity by orders of magnitude. Although native vector species are usually present in agroecosystems (e.g., *Graphocephala atropunctata*, *Draeculacephala minerva*, *Carneocephala fulgida*, in Brazilian citrus), some alien invaders can be markedly more effective at disease transmission. One of the most serious (competent) such vectors is *Homalodisca coagulata*, which is large, disperses long distances, and accumulates in large populations. Introduction of *Homalodisca coagulata* into wine-growing areas of California has already had major impacts on the industry in southern California, and threatens to affect the entire region if the species is not controlled.

With increasing movement and transport of biological materials, there is concern in regard to the introduction of such vector species into new areas. For instance, although *Homalodisca coagulata* has been a long-time pest in California, it has not *yet* appeared as an invader in southern South America. In South America, many potential regions and crops could be in danger. Of particular interest is citrus in southern Brazil, where *Xylella* has long been a concern, and introduction of such a competent vector might have very unfortunate consequences. The purpose of this contribution is the development and testing of ecological niche models that permit assessment of the invasive potential of *Homalodisca coagulata* in South America.

Methods

Ecological niche models were based on 116 unique occurrence points from the native range of *Homalodisca coagulata* accumulated from museum collections databases (see Acknowledgments) and from the scientific literature. An additional 22 points from the species' invaded range in California were gathered from similar sources. All occurrence points were georeferenced to the nearest 0.1' of latitude and longitude, and organized in Microsoft Excel spreadsheets for analyses.

Ecological niches were modeled using the *Genetic Algorithm for Rule-set Prediction* (GARP) (Stockwell 1999, Stockwell & Noble 1992, Stockwell & Peters 1999). In general, the procedure focuses on modeling ecological niches (the conjunction of ecological conditions within which a species is able to maintain populations without immigration) (Grinnell 1917). Specifically, GARP relates ecological characteristics of known occurrence points, such as topography, vegetation, and climate, to those of points randomly sampled from the rest of the study region, seeking to develop a series of decision rules that best summarize those factors associated with the species' presence (Peterson, et al. 2002c).

Occurrence points are divided evenly into training and test data sets. GARP works in an iterative process of rule selection, evaluation, testing, and incorporation or rejection: a method is chosen from a set of possibilities (e.g., logistic regression, bioclimatic rules), applied to the training data, and a rule is developed or evolved. Predictive accuracy is then evaluated based on 1250 points resampled with replacement from the test data and 1250 points sampled randomly from the study region as a whole. Rules may evolve by a number of means that mimic DNA evolution: point mutations, deletions, crossing over, etc. The change in predictive accuracy from one iteration to the next is used to evaluate whether a particular rule should be incorporated into the model, and the algorithm runs either 1000 iterations or until convergence.

All modeling in this study was carried out on a desktop implementation of GARP now available for download from the Internet (<http://www.lifemapper.org/desktopgarp/>). This implementation offers much-improved flexibility in choice of predictive environmental/ecological GIS data coverages. In this case, we used 35 data layers summarizing aspects of topography [elevation, slope, aspect, flow accumulation, flow direction, and topographic index (tendency to pool water) from the U.S. Geological Survey's Hydro-1K data set], aspects of climate including daily temperature range, frost days, mean annual precipitation, solar radiation, maximum, minimum,

and mean annual temperatures, vapor pressure, and wet days (annual means 1960-1990; from the [Intergovernmental Panel on Climate Change](#)); and aspects of land cover including the [University of Maryland Land Use/Land Cover classification and a coverage](#) summarizing tree cover for an area consisting of all of North America north to central Canada. For projection to South America, similar data sets were employed; however, given the confusions introduced by the opposites of winter and summer in Northern and Southern hemispheres, we also conducted analyses in which we included information for a hot month (July in north, January in south) and a cold month (January in north, July in south) along with the annual mean climate data. GARP's predictive abilities have been tested and proven under diverse circumstances (Anderson, et al. 2002a, b, Godown & Peterson 2000, Peterson 2001, Peterson, et al. 2002a, b, c, Peterson & Cohoon 1999, Peterson, et al. 2000, Peterson, et al. 2001, Peterson, et al. Submitted, Peterson, et al. 1999, Peterson & Vieglais 2001, Stockwell & Peterson 2002a, b).

Two types of errors are possible in predictive models of species' distributions: omission error (false negatives or underprediction), in which areas of actual presence are predicted absent; and commission error (false negatives or overprediction), in which areas not inhabited by the species are predicted present (Anderson, et al. 2003).

We used two manipulations based on the analysis of those two types of errors (omission and commission) to improve model performance. First, as a preliminary exploration of positive and negative effects of inclusion of particular coverages, we used a jackknife manipulation in which each coverage was omitted sequentially. Then, as a preliminary tool in exploration of effects of inclusion or exclusion of particular environmental coverages, we calculated Pearson product-moment correlation coefficients between a binary variable describing inclusion or exclusion of each coverage and omission error. Although this coefficient assumes normality of the variables, we used it as a preliminary exploratory tool, and improved approaches are under investigation. Those coverages for which these correlations were high (i.e., inclusion of the coverage in analyses associated with increased omission error), on the order of 0.05-0.1, were omitted from further analysis, and the overall jackknife procedure was repeated until all remaining coverages were either unassociated or negatively associated with omission error.

Second, we developed 100 replicate models of each species' ecological niche based on random 50-50 splits of available occurrence points. Unlike previous applications, which either used single models to predict species' distributions (Peterson 2001, Peterson, et al. 2002a) or summed multiple models to incorporate model-to-model variation (Peterson & Vieglais 2001), we used a new procedure (Anderson, et al. 2003) for choosing best subsets of models. The

procedure is based on the observations that (1) models vary in quality, (2) variation among models involves an inverse relationship between errors of omission (leaving out true distributional area) and commission (including areas not actually inhabited), and (3) best models (as judged by experts blind to error statistics) are clustered in a region of minimum omission of independent test points and moderate area predicted (an axis related directly to commission error). The relative position of the cloud of points relative to the two error axes provides an assessment of the relative accuracy of each model. To choose best subsets of models, we (1) eliminated all models that had non-zero omission error based on independent test points, (2) calculated the average area predicted present among these zero-omission points, and (3) identified the 10 models closest to the overall median area predicted.

Projection of the rule-sets for these models onto maps of eastern North America provided distributional predictions for the species' native range. Model quality was tested via independent sets of occurrence points set aside prior to modeling: a χ^2 test was used to compare observed success in predicting the distribution of test points with that expected under a random model (proportional area predicted present provides an estimate of occurrence points correctly predicted were the prediction to be random with respect to the distribution of the test points). For native-range predictions, we used all of eastern North America for testing model quality. For testing invaded-range predictions in California, for which relatively small areas represented full model agreement, we used an area that included the area predicted present in California by any model plus a buffer of 200 km in all directions, and thus constituted a more conservative and stringent test of model quality than a broader test region.

Results

In a first modeling exercise, annual mean climate data were used to generate predictions. Native-range occurrence points were distributed from northeastern Mexico north and east across the Gulf states to Florida and the Carolinas (Figure 1). The jackknife procedure led to the exclusion of vapor pressure, elevation, flow accumulation, and flow direction from analysis. Best-subsets ecological niche models developed from this data set predicted the species' occurrence fairly continuously throughout the region (Figure 1), including areas somewhat farther north than most known occurrences. Tests of the predictive ability of these models were uniformly highly statistically significant ($10^{-56} < P < 10^{-7}$), suggesting that these models were able to predict independent sets of occurrences with good precision.

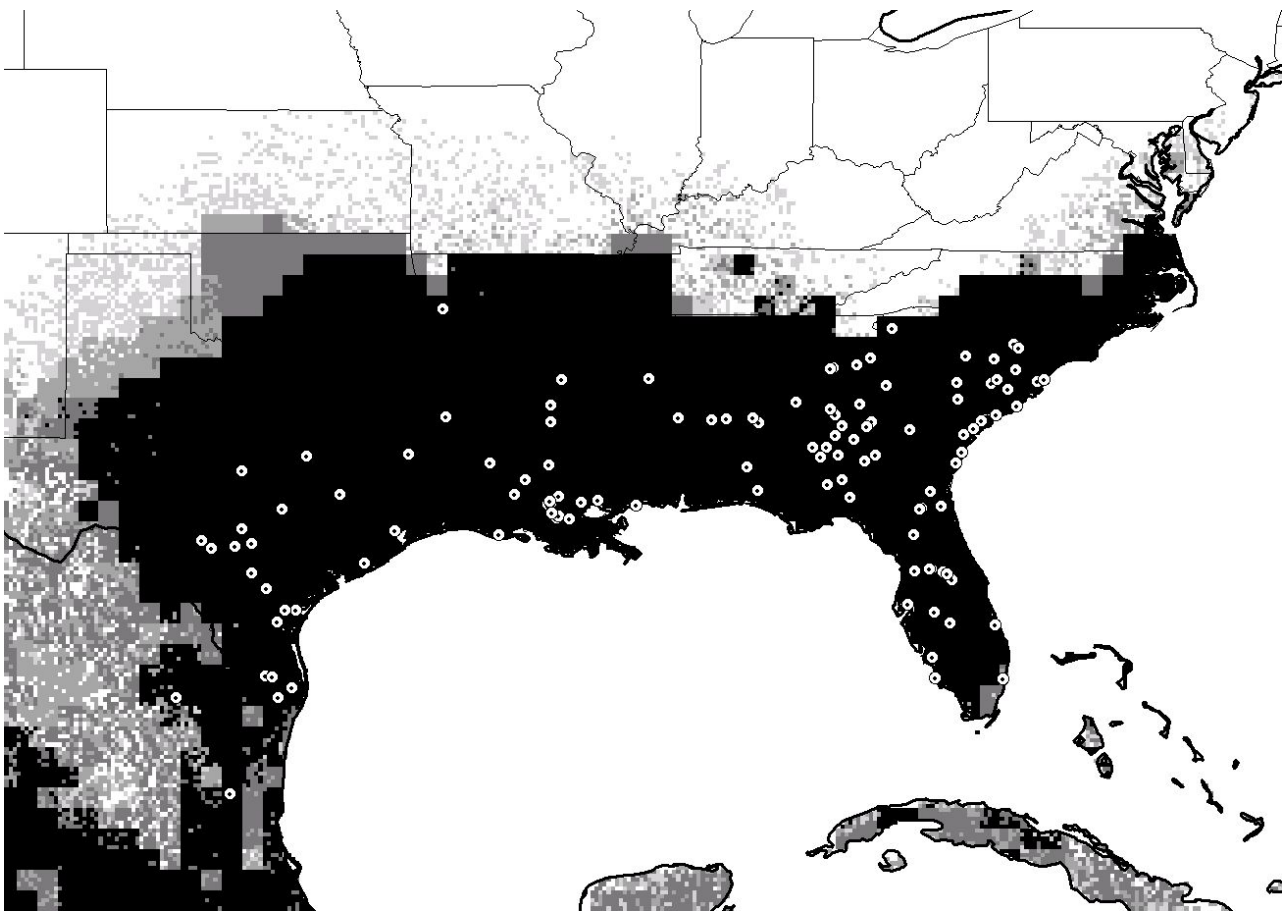


Figure 1. Native range prediction for *Homalodisca coagulata*, based on annual mean climate data and topography. Darker shading indicates areas of greater confidence in occurrence of the species. Dotted circles represent known occurrence points used in model development. Map horizontal extent is approximately 30 degrees.

Projecting these best-subsets niche models to California, where the species has been invasive episodically for decades, predicts a relatively small area as suitable for the species (Figure 2). The coincidence of 22 known occurrence points in California with these predicted areas was excellent (Figure 2). Tests of the statistical significance of the prediction (Figure 3) were highly significant at all thresholds: that is, from relatively conservative (small, all 10 models agree in predicting presence) to relatively non-conservative (large, any of 10 models predicts presence) predictions, all were considerably and significantly better than random models (all $P < 0.01$, and most considerably lower).

For predictions of potential distributional areas in South America, we used the jackknife procedure on both the annual-means-only data and the seasonal data, reducing the numbers of coverages included in analyses considerably (to flow accumulation, flow direction, slope, solar radiation, mean and maximum temperature, vapor pressure, and wet days for annual means, and to elevation, aspect, flow accumulation, topographic index, hot month solar radiation, hot month minimum temperature, annual mean minimum temperature, hot month mean temperature, and cold

month maximum temperature for seasonal data). Based on these final coverage sets, the 10 'best subsets' models were extracted. Native-range predictions were both quite similar, regardless of input coverage set, to Figure 1.

Interestingly, projections of the results of these two distinct modeling exercises to South America were qualitatively quite similar (Figures 4A, 4B). In both cases, areas of predicted presence were only marginal in southern Brazil, and more solidly in southern Paraguay and northern Argentina. Hence, the citrus-growing areas of Brazil do not appear particularly vulnerable to invasion by this potential vector for *Xylella fastidiosa*, but wine-growing areas in northern Argentina (e.g., Salta) and extreme southern Brazil appear to be much more vulnerable.

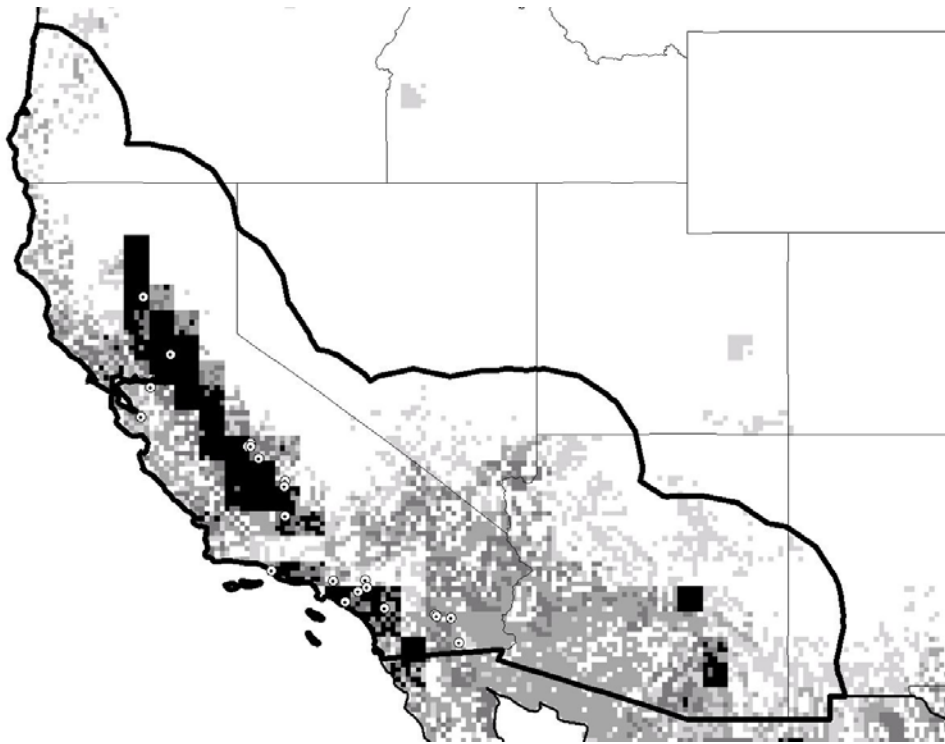


Figure 2. Projection of native-range ecological model (annual mean climate data and topography) for *Homalodisca coagulata* to California. Darker shading indicates areas of greater confidence in occurrence of the species. Dotted circles represent known occurrence points used in model testing. The dark line represents the area within which statistical significance of the prediction was assessed. Map horizontal extent is approximately 20 degrees.

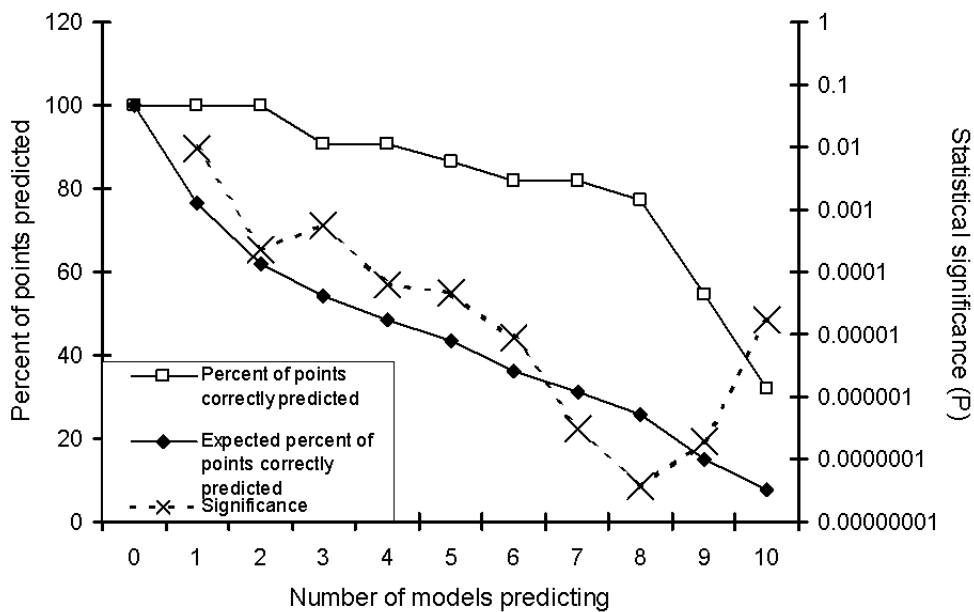


Figure 3. Multiple-threshold assessment of statistical significance of the prediction for *Homalodisca coagulata* in California. Solid diamonds show the expectations for points correctly predicted under a random model; open squares indicate actual correct prediction of points; X's indicate the statistical significance of these predictions, based on chi-square tests (right-hand axis).

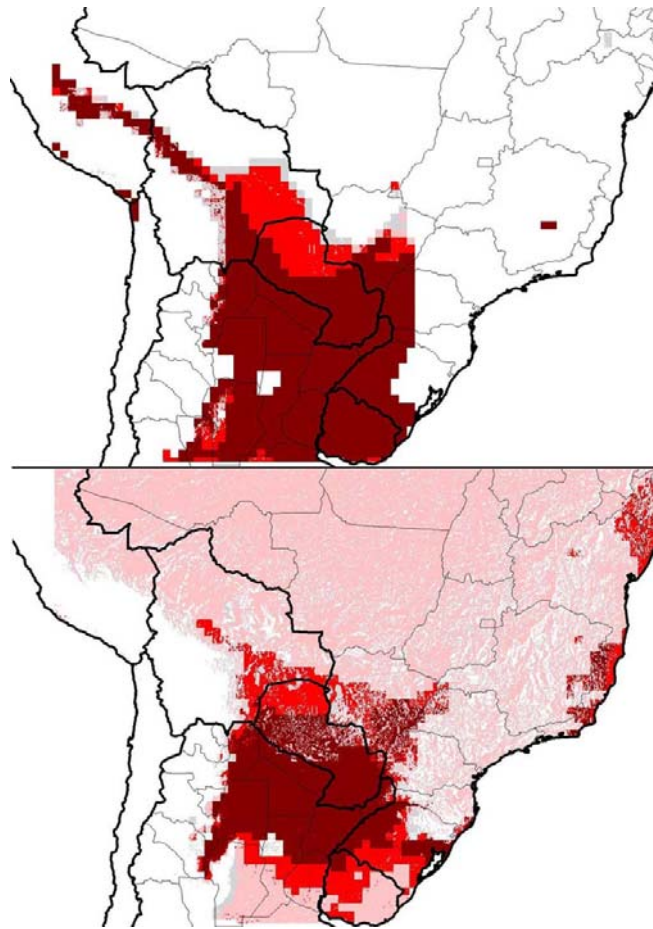


Figure 4. Projection of native-range ecological model for *Homalodisca coagulata* to South America. Top: annual mean climate data and topography. Bottom: Annual and monthly mean climate data and topography. Darker shading indicates areas of greater confidence in potential occurrence of the species. Horizontal extent of both maps is approximately 40 degrees.

Discussion

The development of predictive technologies for species' invasions (Peterson & Vieglais 2001) provided a first proactive view into this major economic problem that is a manifestation of biological processes. In the present case, models developed on the native range, and tested both on the native range and on another invaded range (California) provided predictions of the potential geographic distribution of *Homalodisca coagulata* in South America. Results indicate that *Homalodisca* may be of minor concern as a pest in citrus in South America, but may be of considerable concern for wine-growing regions further south. Great care is certainly recommended for any movements of biological movements between infected areas (southeastern United States, California) and these vulnerable areas.

Of course, such predictions are only as good as the models on which they are based. In the present case, the models employed were highly predictive of the species' already-in-process invasion of California wine-growing areas, indicative of the probably predictive nature of the models for the species' invasive potential in South America. Such predictivity has been encountered in numerous addi-

tional tests with similar methodologies (Peterson, et al. Submitted, Peterson & Vieglais 2001), but a much broader spectrum of tests is desirable to ascertain the frequency with which invasive species do *not* obey the ecological 'rules' that appear to govern their distributions on the native distributional area.

Acknowledgments

We thank David A. Vieglais and David Stockwell for their important contributions to the development of these predictive technologies. Occurrence data were provided by the University of Kansas Natural History Museum, University of Georgia Museum of Natural History, Enns Entomology Museum of the University of Missouri-Columbia, Florida State Collection of Arthropods, Florida Department of Agriculture and Consumer Services, Louisiana State Arthropod Museum, and Clemson University Arthropod Collection. This research was supported by a grant from the U.S. Environmental Protection Agency; ATP's work in Brazil was supported by a grant from the Fundação de Amparo à Pesquisa do Estado de São Paulo.

References

1. ANDERSON, R. P., GOMEZ, M. & PETERSON, A. T. 2002a. Geographical distributions of spiny pocket mice in South America: Insights from predictive models. *Global Ecol. and Biogeogr.*, v. 11, p. 131-141.
2. ANDERSON, R. P., LAVERDE, M. & PETERSON, A. T. 2002b. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos*, v. 93, p. 3-16.
3. ANDERSON, R. P., LEW, D. & PETERSON, A. T. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecol. Model.*, v. 162, p. 211-232.
4. GODOWN, M. E. & PETERSON, A. T. 2000. Preliminary distributional analysis of U.S. endangered bird species. *Biodivers. Conserv.*, v. 9, p. 1313-1322.
5. GRINNELL, J. 1917. Field tests of theories concerning distributional control. *Am. Natur.*, v. 51, p. 115-128.
6. PETERSON, A. T. 2001. Predicting species' geographic distributions based on ecological niche modeling. *Condor*, v. 103, p. 599-605.
7. PETERSON, A. T., BALL, L. G. & COHOON, K. C. 2002a. Predicting distributions of tropical birds. *Ibis*, v. 144, p. e27-e32.
8. PETERSON, A. T. & COHOON, K. C. 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. *Ecol. Model.*, v. 117, p. 159-164.
9. PETERSON, A. T., EGBERT, S. L., SANCHEZ-CORDERO, V. & PRICE, K. P. 2000. Geographic analysis of conservation priorities using distributional modelling and complementarity: Endemic birds and mammals in Veracruz, Mexico. *Biol. Conserv.*, v. 93, p. 85-94.
10. PETERSON, A. T., ORTEGA-HUERTA, M. A., BARTLEY, J., SANCHEZ-CORDERO, V., SOBERON, J., BUDDEMEIER, R. H. & STOCKWELL, D. R. B. 2002b. Future projections for Mexican faunas under global climate change scenarios. *Nature*, v. 416, p. 626-629.
11. PETERSON, A. T., SANCHEZ-CORDERO, V., SOBERON, J., BARTLEY, J., BUDDEMEIER, R. H. & NAVARRO-SIGUENZA, A. G. 2001. Effects of global climate change on geographic distributions of Mexican Cracidae. *Ecol. Model.*, v. 144, p. 21-30.
12. PETERSON, A. T., SCACHETTI-PEREIRA, R. & HARGROVE, W. W. Potential distribution of Asian longhorned beetles (*Anoplophora glabripennis*) in North America. *J. Econ. Entomol.*, v. p. Submitted.
13. PETERSON, A. T., SOBERON, J. & SANCHEZ-CORDERO, V. 1999. Conservatism of ecological niches in evolutionary time. *Science*, v. 285, p. 1265-1267.
14. PETERSON, A. T., STOCKWELL, D. R. B. & KLUZA, D. A. 2002c. Distributional prediction based on ecological niche modeling of primary occurrence data. In: J. M. Scott (Ed.) *Predicting species occurrences: Issues of scale and accuracy*. Washington, D.C.: Island Press.
15. PETERSON, A. T. & VIEGLAIS, D. A. 2001. Predicting species invasions using ecological niche modeling. *BioScience*, v. 51, p. 363-371.
16. STOCKWELL, D. R. B. 1999. Genetic algorithms II. In: A. H. Fielding (Ed.) *Machine learning methods for ecological applications*. Boston: Kluwer Academic Publishers. p. 123-144.
17. STOCKWELL, D. R. B. & NOBLE, I. R. 1992. Induction of sets of rules from animal distribution data: A robust and informative method of analysis. *Math. Comput. Simulat.*, v. 33, p. 385-390.
18. STOCKWELL, D. R. B. & PETERS, D. P. 1999. The GARP modelling system: Problems and solutions to automated spatial prediction. *Int. J. Geo. Inf. Syst.*, v. 13, p. 143-158.
19. STOCKWELL, D. R. B. & PETERSON, A. T. 2002a. Controlling bias in biodiversity data. In: J. M. Scott (Ed.) *Predicting species occurrences: Issues of scale and accuracy*. Washington, D.C.: Island Press.
20. STOCKWELL, D. R. B. & PETERSON, A. T. 2002b. Effects of sample size on accuracy of species distribution models. *Ecol. Model.*, v. 148, p. 1-13.

Title: Assessment of invasive potential of *Homalodisca coagulata* in western North America and South America

Authors: A. Townsend Peterson, Ricardo Scachetti-Pereira and Daniel A. Kluza

Biota Neotropica, Vol. 3 (number 1): 2003
<http://www.biotaneotropica.org.br/v3n1/pt/abstract?article+BN00703012003>

Date Received 19/12/2003 Revised 14/03/2003
 Accepted 26/03/2003

ISSN 1676-0611