Biogeographic Implications of the Striking Discovery of a 4,000 Kilometer Disjunct Population of the Wild Potato Solanum morelliforme in South America

Authors: Simon, Reinhard, Fuentes, Alfredo F., and Spooner, David M.

Biogeographic Implications of the Striking Discovery of a 4,000 Kilometer Disjunct Population of the Wild Potato *Solanum morelliforme* in South America

Reinhard Simon,[1] Alfredo F. Fuentes,[2] and David M. Spooner[3,4]

[1]International Potato Center, Apartado 1558, Lima 12, Peru
[2]Herbario Nacional de Bolivia and Missouri Botanical Garden, Campus Universitario, Cota Cota, Calle 27,
Casilla 10077 Correo Central, La Paz, Bolivia.
[3]USDA, Agricultural Research Service, Department of Horticulture, University of Wisconsin, 1575 Linden Drive,
Madison, Wisconsin 53706-1590 U. S. A.
[4]Author for correspondence (david.spooner@ars.usda.gov)

Communicating Editor: Fernando Zuloaga

Abstract—Solanum morelliforme is an epiphytic wild potato (*Solanum* section *Petota*) species widely distributed throughout central Mexico to Honduras. A strikingly disjunct (approximately 4,000 km) population was recently discovered in Bolivia, representing the first record of this species in South America, and the first species in the section growing in both North and Central America and in South America. Our maximum entropy analysis of 19 climatic variables matches the occurrence of the South American locality with great precision. It demonstrates the strong predictive quality of this procedure and suggests similar localities where this species may be found, especially along the eastern slopes of the Andes in the Yungas region of southern Peru and Bolivia. In addition, the presence of *S. morelliforme* in South America adds to emerging data from yet other sources to question long-held hypotheses of the origin of section *Petota* in North and Central America.

*Keywords—*Disjunction, MaxEnt, randomForest method, *Solanum morelliforme*, wild potatoes.

Plant species disjunctions have long held fascination for botanists, and have been the subject of major symposia (Solbrig 1972; Thorne 2004). While some recurrent patterns of plant disjunctions are evident, such as that between eastern North America and eastern Asia, almost every conceivable sort of disjunct range has been documented in plants (Raven 1972). Explanations of such disjunctions have been the subject of heated debate, with dispersal vs. vicariance especially so, but most recognize that both processes are likely to be the cause of disjunctions (Stace 1989). Disjunctions involve recent events of morphologically similar parent and colonizing species to those more ancient, involving taxa related only at the genus and family level; distinguished as autodisjuncts vs. allodisjuncts by Turner (1972). The focus of this study is the recent discovery of an apparent autodisjunct of an epiphytic wild potato (*Solanum* L. section *Petota* Dumort.).

Solanum section *Petota* contains about 100 wild species (Spooner 2009) distributed from the southwestern U. S. A. to Uruguay and adjacent central Argentina and Chile, and four cultivated species (Spooner et al. 2007; Ovchinnikova et al. 2011) native to South America. *Solanum morelliforme* Bitter & Muench is a diploid (2*n* = 24), self-incompatible, epiphytic member of the section, widespread throughout central Mexico (southern Jalisco to Querétato and Veracruz), south to southern Honduras, growing from 1,870–3,050 m in elevation (Spooner et al. 2004). It has small round tubers, from 0.5–1 cm in diam, and produces 5–15 seeds per fruit. The presence of tubers clearly identifies this as a member of sect. *Petota*. *Solanum morelliforme* is distinctive with its simple leaves, relatively small stature (stems 2–3 mm wide at base, 0.1–0.5 m tall), epiphytic habit, and is impossible to be confused with any other wild potato. It is most similar to *S. clarum* Correll, its sister species (Spooner et al. 2004), distributed in southern Mexico and Guatemala, from which it can be distinguished by its lanceolate to ovate-lanceolate leaves (ovate in *S. clarum*). It is the only epiphytic wild potato, growing on horizontal branches of mature *Arbutus* L., cyprus, elm, juniper, pine, or oak trees, often in moss and organic litter (Spooner et al. 2004; Fig. 1). Field studies in Mexico and Central America (Spooner

et al. 1998, 2000) showed that it is difficult to find in previously documented localities that had been logged and reforested, suggesting that its range is being reduced by deforestation.

Recent field work in Bolivia discovered an amazingly disjunct record of *S. morelliforme*, about 4,000 km south of its previously southernmost record in Honduras: BOLIVIA. La Paz: Prov. Franz Tamayo, Parque Nacional Madidi, sector Tocoaque, between Keara y Mojos, 14°37′02.0″S, 68°57′14.2″W, 1,989 m; wet mountain yungas forest and riverside vegetation, *A. Fuentes, D. Alanes & M. Chambi 15562* (LPB) (Fig. 2). This collection was unexpected. The 4,000 km disjunction is of interest in itself, especially for a species occupying such a specialized epiphytic habitat. Prior to this discovery, there was no wild potato known to span both North and Central America and South America. This collection provides an opportunity to explore the predictive utility of probabilistic species distribution tools (Thuiller and Münkemüller 2010). In addition, in combination with recent phylogenetic data, it forces a reevaluation of long-held biogeographic hypotheses regarding the origin of section *Petota*.

Materials and Methods

Environmental Data Preparation—A georeferenced database of 76 records of *S. morelliforme* (from Spooner et al. 2004) was annotated with 19 bioclimatic variables using the Worldclim database (Hijmans et al. 2005): annual mean temperature, mean monthly temperature range, isothermality (daily temperature range as a fraction of seasonal temperature range), temperature seasonality (coefficient of variance - CV * 100), maximum temperature of warmest month, minimum temperature of coldest month, temperature annual range, mean temperature of wettest quarter, mean temperature of driest quarter, mean temperature of warmest quarter, mean temperature of coldest quarter, annual precipitation, precipitation of wettest month, precipitation of driest month, precipitation seasonality (CV), precipitation of wettest quarter, precipitation of driest quarter, precipitation of warmest quarter, and precipitation of coldest quarter. We also used latitude and altitude for a total of 21 variables.

Environmental Modeling—To test whether the new Bolivian locality falls within the environmental range of *S. morelliforme* we used the program MaxEnt (Phillips et al. 2006) with the 76 records from the northern hemisphere as a training set and the sole record from the southern hemisphere as a test record, using the 21 niche variables as predictors (model 1).

1062

FIG. 1.   Habitat of *Solanum morelliforme* (*Spooner et al. 7004*) on a horizontal branch of a large pine tree in Guatemala, Totonicipán: 6.6 km E of Totonicipán (Ronald van den Berg on David Spooner's shoulders and Victor Martínez taking notes).

As a complementary analysis we used the randomForest algorithm as implemented in the R package (Liaw and Wiener 2002) to test whether the Bolivian record could be classified as different using another method. The results of the MaxEnt algorithm were used to describe the most important niche variables determining the distribution of *S. morelliforme*. Lastly, we compared a MaxEnt model including all 77 *S. morelliforme* records (model 2) with the first model to assess whether the inclusion of the new accession would change the predicted distribution. We calculated for both predictions the areas where the probability was above the significance threshold. The logistic probability was multiplied with the area for each 'pixel' or geographic grid cell to adjust for differences of likelihood. Then the difference of both areas was calculated subtracting 'model 2' from 'model 1'. Area probability gains and losses were expressed as a percentage of the area of model 1. Violin plots were created using the R package vioplot (Adler 2005). A distribution map was created using R (R Development Core Team 2010) and the package 'maps' (Becker et al. 2010) as well as a custom script.

RESULTS

The MaxEnt receiver operating curve yielded an area under the curve (AUC) of 0.997 for the training data and an AUC of 0.996 for the test; together this indicates a good model. The distribution of *S. morelliforme* as predicted by MaxEnt established a "maximum test sensitivity plus specificity" threshold of 0.443 with a significant *p*-value of 0.00428. The threshold area of predicted occurrence is shown in Fig. 3. The Bolivian accession has a threshold value within the rounding margin of 0.4426, indicating that is within the ecological niche of the

species. Also, using the randomForest algorithm as an alternative method the new locality was classified as suitable for *S. morelliforme*. The net area probability changes of model 2 (including the new accession record) were only about 0.6% (not shown); the changes due to inclusion of the new record were mainly areas around the new site becoming more probable; further confirming the robustness of the model.

The most important niche variable when used alone with MaxEnt modeling is temperature seasonality (CV of the monthly mean average temperatures expressed as a percentage of the annual mean temperature). It is an expression of the variability of the climate, the higher the more variable. Other important variables among the five variables that contribute most to the overall model include altitude, isothermality (mean diurnal range/temperature annual range, which quantifies daily temperature range in relation to seasonal temperature range), precipitation of the wettest month, and absolute latitude. In most cases the niche properties of the new site in the southern hemisphere coincide with the overall pattern with an exception of latitude (Fig. 4 for details). The results from model 2 regarding the importance of the niche variables are the same.

Other predicted sites include areas in the Dominican Republic, overlapping with two national parks. One of these, Parque Nacional Jose del Carmen Ramirez has a maximum elevation of 3,175 m, the highest point in the Antilles, and the other has areas above 1,800 m. The predicted areas in southern Brazil are in Minas Gerais around the metropolitan area of Contagem and Belo Horizonte at altitudes of 900 m and in southern Minas Gerais around the Serra das Antas at 1,500 m. One other unexpected area is close to the Peruvian coast in the Ayacucho Department around Tocyasca at 3,200–3,500 m altitude.

DISCUSSION

Utility of Probabilistic Species Distribution Analyses— The correspondence of the predicted distribution of *S. morelliforme* to its occurrence in Bolivia is striking (Fig. 3). This predicted area occurs in the Yungas Peruano-Bolivianos biogeographic province of Rivas-Martínez and Navarro (2002), along the eastern slopes of Peru and the Bolivian Andes between 1,500 and 2,500 m, occupied mainly by cloud forest. In Bolivia, the wild potato species *S. chacoense* Bitter, *S. microdontum* Biter, and *S. violaceimarmoratum* Bitter grow there (mss. of section *Petota* in southern South America in revision) and in southern Peru grow *S. chacoense*, *S. limbaniense* Ochoa, and *S. violaceimarmoratum* (Ochoa 1999). It will be interesting to see if *S. morelliforme* can be found in its predicted areas. Collecting in northwesternmost Bolivia, in the Madidi region, has only begun recently. To the southeast, the Provinces of Nor Yungas and Sud Yungas are the most intensely explored parts of all Bolivia; if *Solanum morelliforme* is found there, it may have been overlooked, but this is unlikely. To the southeast, through the Departments of Cochabamba and Santa Cruz there has not been sufficient collecting in suitable habitat to make a reasonable assessment of its presence (pers. comm., Michael Nee, New York Botanical Garden).

We know of only one other similar disjunct distribution, *Solanum aligerum* Schltdl., but this is a small tree or shrub, not an epiphyte (see Solanaceae Source, http://www.nhm.ac.uk/research-curation/research/projects/solanaceaesource/ - treatment by S. Knapp).
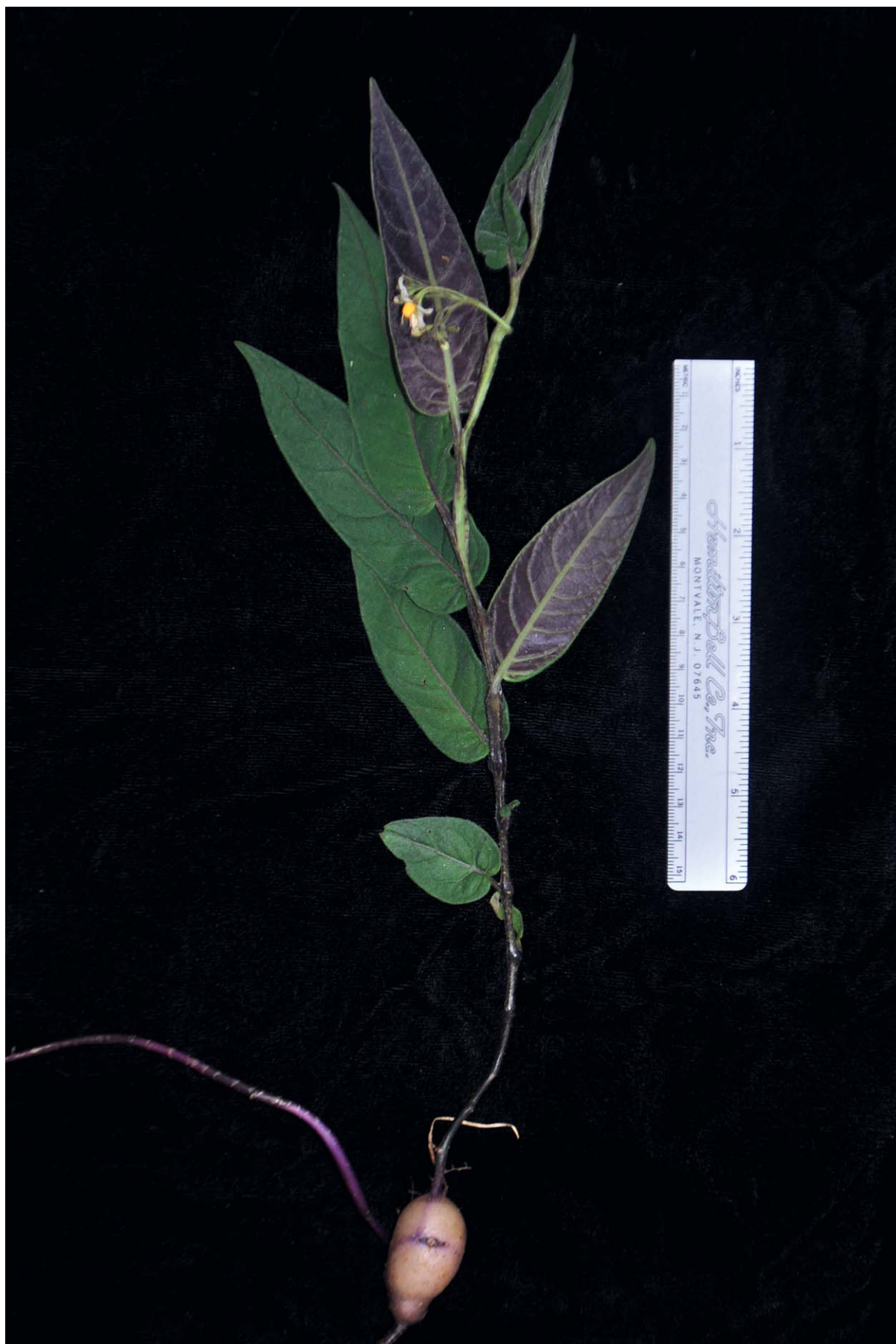
Fig. 2.   Photo of *A. Fuentes, D. Alanes & M. Chambi 15562* (LPB), the new collection of *S. morelliforme* in Bolivia.
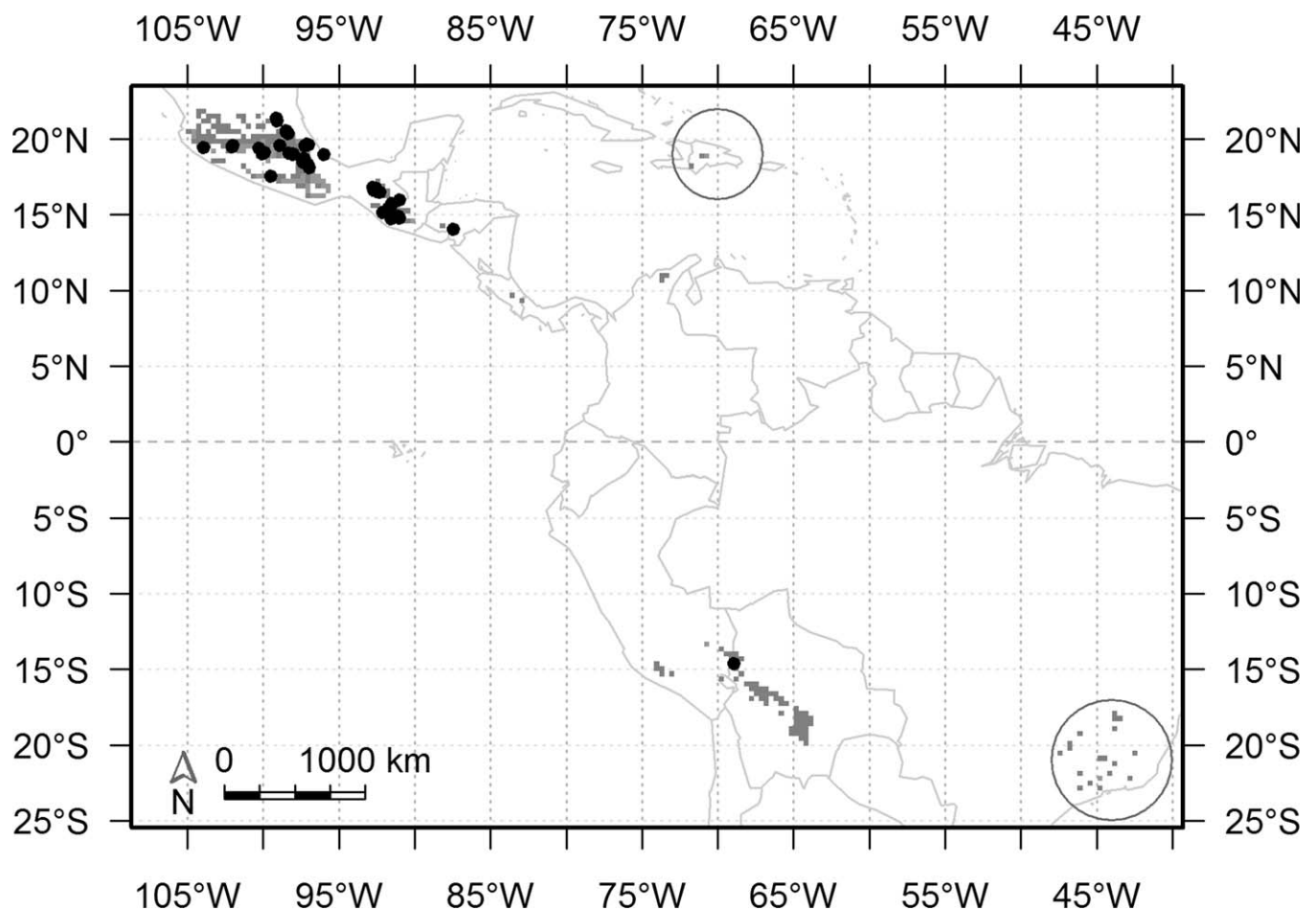
Fig. 3.   The localities of *S. morelliforme* in Mexico, Guatemala, and Honduras and the newly discovered locality 4,000 km to the southeast in northern Bolivia, north of Lake Titicaca and near the Peruvian border. Localities are set in black dots. Based on the prediction of the MaxEnt software, lighter gray squares indicate higher probabilities of the species being present, shown in southern Peru (Ayacucho Department) and in the Yungas region of Peru and Bolivia, The Dominican Republic, and southeastern Brazil.

Biogeographic Implications—Hawkes (1990) hypothesized section *Petota* to have originated in the late Cretaceous to Eocene eras in North and Central America as diploid ($2n = 2x = 24$) species possessing white stellate corollas and B genomes. He also suggested that some of these species migrated during the mid-Eocene to Pliocene eras to South America where they evolved rotate corollas and an A genome. A return migration of A genome species to North and Central America led to the formation of A × B allopolyploids. Plastid DNA phylogenetic results (Spooner et al. 1991) provided partial support for this hypothesis by showing all diploid B genome species from Mexico and Central America (all with white stellate corollas, including *S. morelliforme*) to form sister clades in section *Petota*, with A genome species as sister to these. Recent molecular clock data (Wang et al. 2008) indicated that eggplant and tomato/potato shared a common ancestor approximately 14 MYA and potato and tomato seven MYA, so it seems unlikely that the timing for the events proposed by Hawkes really took place in the Cretaceous – Eocene eras but the migration events are still possible.

There were always problems, however, with Hawkes's (1990) biogeographic hypothesis in that literally all sister groups of section *Petota* (tomato and its immediate relatives, and section *Etuberosum* Buk. & Kameraz (A. Child); Spooner et al. 1993), were confined to South America. Separately, Rodríguez et al.

(2009) showed that although a dominant phylogeny using 12 nuclear orthologs supported the North and Central American B genome species as sister to the remaining clades in section *Petota*, a minority of these orthologs supported an alternative phylogeny with some South American species (members of the "Piurana" clade) as sister. In addition, recent nuclear ortholog DNA (GBSSI) sequence data (Rodríguez et al. 2010) showed the South American species *S. circaeifolium* (possessing white stellate corollas) to form a grade with the diploid North and Central B genome species, although grouping with A genome species in the plastid data (Spooner et al. 1991), suggesting that those few South American species possessing white stellate corollas (including *S. morelliforme*) could be related to ancestral species in section *Petota*, rather than the B genome species from North and Central America.

The discovery of *S. morelliforme*, a disjunct representative of the North and of the Central American diploid clade in South America, adds yet additional data suggesting that section *Petota* could have had a South American origin. The B genome species are well defined as a clade in recent phylogenetic studies (Spooner et al. 2008; Rodriguez et al. 2009). Clearly, the South American distribution of *S. morelliforme* could have various explanations, to include its being a remnant of this clade in South America before their migration to North and Central America and the evolution of the A genome, in contrast to
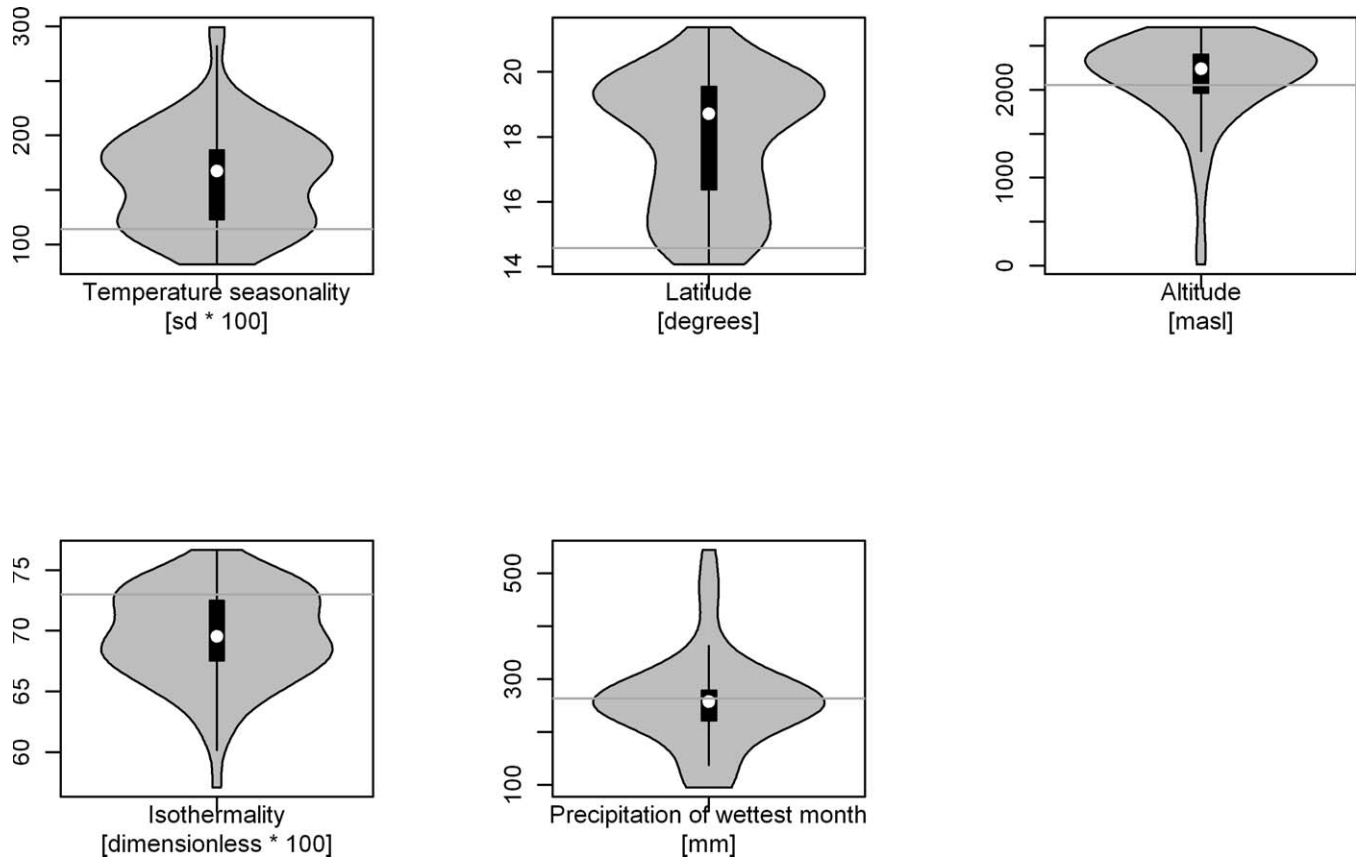
Fig. 4. The violin plots of five niche variables identified by MaxEnt as important for the geographic distribution of *S. morelliforme*. A violin plot is similar to a box plot and combines it with probability density estimates of the data. The white dot in the middle is the median value; the black bar ranges between the lower and upper quartile. The horizontal dark grey line indicates the respective values for the new locality in Bolivia; in the case of latitude the absolute value is shown.

Hawkes's (1990) hypothesis of a B genome origin in North and Central America. At the minimum, however, the discovery of this disjunct population in South America, the recent molecular phylogenetic data described above, and the fact that the outgroups of section *Petota* are all South American, force a reconsideration of Hawkes's (1990) hypothesis of an origin of section *Petota* in North and Central America. Analyses of additional nuclear orthologs and species in sect. *Petota* (in progress) may address this question.

Literature Cited

Adler, D. 2005. Vioplot: Violin plot. R package version 0.2. http://wsop uppenkiste.wiso.uni-goettingen.de/~dadler.

Becker, R. A., A. R. Wilks, R. Brownrigg, and T. P. Minka. 2010. Maps: Draw geographical maps. R package version 2.1–2. http://CRAN.R-project.org/package=maps.

Hawkes, J. G. 1990. *The potato: Evolution, biodiversity and genetic resources*. Oxford: Belhaven Press.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965–1978.

Liaw, A. and M. Wiener. 2002. Classification and regression by random-Forest. *R News* 2: 18–22.

Ochoa, C. M. 1999. *Las papas de sudamerica: Peru (Parte I)*. Lima, Peru.

Ovchinnikova, A., E. Krylova, T. Gavrilenko, T. Smekalova, M. Zhuk, S. Knapp, and D. M. Spooner. 2011. Taxonomy of cultivated potatoes (*Solanum* section *Petota*: Solanaceae). *Botanical Journal of the Linnean Society* 165: 107–155.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distribution. *Ecological Modelling* 190: 231–259.

R Development Core Team. 2010. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org.

Raven, P. H. 1972. Plant species disjunctions: a summary. *Annals of the Missouri Botanical Garden* 59: 235–246.

Rivas–Martínez, S. and G. Navarro. 2002. Mapa biogeográfico de América del Sur. Pp. 42–43 in *Geografía ecológica de Bolivia: Vegetación y ambientes acuáticos*, eds. G. Navarro, and M. Maldonado. Cochabamba, Bolivia: Centro de Ecología Simón I, Patiño–Departamento de Difusión.

Rodríguez, F., M. Ghislain, A. M. Clausen, S. H. Jansky, and D. M. Spooner. 2010. Hybrid origins of cultivated potatoes. *Theoretical and Applied Genetics* 121: 1187–1198.

Rodríguez, F., F. Wu, C. Ané, S. Tanksley, and D. M. Spooner. 2009. Do potatoes and tomatoes have a single evolutionary history, and what proportion of the genome supports this history? *BMC Evolutionary Biology* 9: 191, doi: 10.1186/1471-2148-9-191.

Solbrig, O. T. 1972. Disjunctions in plants: a symposium. *Annals of the Missouri Botanical Garden* 59: 105–106.

Spooner, D. M. 2009. DNA barcoding will frequently fail in complicated groups: an example in wild potatoes. *American Journal of Botany* 96: 1177–1189.

Spooner, D. M., G. J. Anderson, and R. K. Jansen. 1993. Chloroplast DNA evidence for the interrelationships of tomatoes, potatoes, and pepinos (Solanaceae). *American Journal of Botany* 80: 676–688.

Spooner, D. M., R. Hoekstra, R. G. van den Berg, and V. Martínez. 1998. *Solanum* sect. *Petota* in Guatemala: taxonomy and genetic resources. *American Journal of Potato Research* 75: 3–17.

Spooner, D. M., J. Núñez, G. Trujillo, M. del Rosario Herrera, F. Guzmán, and M. Ghislain. 2007. Extensive simple sequence repeat genotyping of potato landraces supports a major reevaluation of their gene pool structure and classification. *Proceedings of the National Academy of Sciences USA* 104: 19398–19403.

Spooner, D. M., A. Rivera-Peña, R. G. van den Berg, and K. Schüler. 2000. Potato germplasm collecting expedition to Mexico in 1997: taxonomy and new germplasm resources. *American Journal of Potato Research* 77: 261–270.

Spooner, D. M., F. Rodríguez, Z. Polgár, H. E. Ballard Jr., and S. H. Jansky. 2008. Genomic origins of potato polyploids: GBSSI gene sequencing data. *The Plant Genome, a supplement to Crop Science* 48(S1): S27–S36.

Spooner, D. M., K. J. Sytsma, and E. Conti. 1991. Chloroplast DNA evidence for genome differentiation in wild potatoes (*Solanum* sect. *Petota*: Solanaceae). *American Journal of Botany* 78: 1354–1366.

Spooner, D. M., R. G. van den Berg, A. Rodríguez, J. Bamberg, R. J. Hijmans, and S. I. Lara-Cabrera. 2004. Wild potatoes (*Solanum* section *Petota*) of North and Central America. *Systematic Botany Monographs* 68: 1–209 + 9 plates.

Stace, C. A. 1989. Dispersal versus vicariance: no contest. *Journal of Biogeography* 16: 200–201.

Thorne, R. 2004. Tropical plant disjunctions: a personal reflection. *International Journal of Plant Sciences* S137–S138.

Thuiller, W. and T. Münkemüller. 2010. Habitat suitability modeling. Pp. 77–85 in *Effects of climate change on birds*, eds. A. P. Moller, W. Fiedler, and P. Berthold. Oxford: Oxford University Press.

Turner, B. L. 1972. Chemosystematic data: their use in the study of disjunctions. *Annals of the Missouri Botanical Garden* 59: 152–164.

Wang, Y., A. Diehl, F. Wu, J. Vrebalov, J. Giovannoni, A. Siepel, and S. D. Tanksley. 2008. Sequencing and comparative analysis of a conserved syntenic segment in the Solanaceae. *Genetics* 180: 391–408.