DNA barcode analyses improve accuracy in fungal species distribution models

Javier Fernández-López1,2 | M. Teresa Telleria1 | Margarita Dueñas1 | Tom May3 | María P. Martín1

1Department of Mycology, Real Jardín Botánico-CSIC, Madrid, Spain
2Department of Biology, University of Massachusetts Boston, Boston, MA, USA
3Royal Botanic Gardens Victoria, Melbourne, Vic., Australia

Correspondence
Javier Fernández-López, Department of Mycology, Real Jardín Botánico-CSIC, Plaza de Murillo 2, Madrid 28014, Spain.
Email: J.FernandezLopez@umb.edu

Abstract
Species distribution models based on environmental predictors are useful to explain a species geographic range. For many groups of organisms, including fungi, the increase in occurrence data sets has generalized their use. However, fungal species are not always easy to distinguish, and taxonomy of many groups is not completely settled. This study explores the effect of taxonomic uncertainty in databases used for modeling fungal distributions. We analyze distribution models for three morphospecies from the corticioid genus Xylodon (Hymenochaetales, Basidiomycota), comparing models based on species names on vouchers specimens with models derived from species identified by DNA barcode. Differences in the contribution of predictors driving the distribution of each modeled taxon and the extent of their ranges were studied. Records under Xylodon paradoxus, X. flaviporus, and X. raduloides were obtained from fungarium collections and GenBank repository. Two grouping criteria were used: (a) specimens were grouped by their collection or sequence voucher names and (b) specimens were grouped following molecular identification using ITS sequences through barcoding gap species recognition (BGSR). Climatic, geographic, and biotic variables were used to predict the potential distribution of each taxon through MaxEnt algorithm. From the three morphospecies selected according to voucher names, up to 19 species candidates were detected using BGSR. Climatic variables were the most important predictors in distribution models made from names on vouchers specimens, but their importance decreased when BGSR was applied. In general, the extent of species distributions was more restricted for taxa under BGSR. Our results show that taxonomic uncertainty has a strong effect in Xylodon species distribution models. Misleading results can be obtained when cryptic species or identification errors mask the actual diversity of the presence records. Preserved specimens in natural history collections offer the possibility to assess whether the species name on labels matches the current species recognition criteria.

KEYWORDS
ITS, MaxEnt, natural history collections, species range, species recognition, Xylodon
1 | INTRODUCTION

In recent decades, ecological and biogeographical studies have increasingly utilized new tools for modeling species distributions from presence records (Elith et al., 2006). Modeling based on correlations between species occurrences and environmental predictors has been used to obtain maps of potential distributions for poorly studied species, to evaluate pest risks (Sutherst, 2014), or to help in design of natural reserves (Watts et al., 2009). The combination of powerful, new algorithms with the increase in environmental cartography has made it possible to apply these methodologies in a broad range of organisms, including different groups of fungi such as ectomycorrhizal (Wolfe et al., 2010) or soil biocrust (Belnap et al., 2014). Indeed, fungi have been pointed as one of the most benefited groups due to the large number of occurrence records stored in fungarium collections (Hao et al., 2020; Wollan et al., 2008).

In the modeling process, much attention has been paid to algorithm performance (Qi et al., 2015), the accuracy of predictor variables (Petitpierre et al., 2017), and the sample size and collection bias (Beck et al., 2014; Fourcade et al., 2014), but taxonomic uncertainty in presence records has attracted less interest (Elith et al., 2013). This could be due in part to the difficulty of assessing the reliability of records in reference collections (such as fungaria or herbaria) or citizen science databases (Lozier et al., 2009). On many occasions, only a list with geographic coordinates is available, and researchers must rely on the accuracy of geographic coordinates and, in particular, on the correctness of species identifications. This taxonomic uncertainty could produce misleading results with important conservation or economic consequences (Bortolus, 2008). This issue plays a major role in those groups for which taxonomy is not completely resolved, or organisms that require expertise to correctly identify the species (Smith et al., 2016).

One of the most important sources of taxonomic uncertainty in reference collections is the shift in species recognition criteria in recent decades (Bridge et al., 2003). The traditionally applied morphological species recognition, MSR (Taylor et al., 2000), has been used to identify more than 70,000 fungal species (Hawksworth et al., 1996; Taylor et al., 2000), resulting in a worldwide distribution for many of these taxa (Hallenberg, 1991). This homogeneous distribution for many fungal species has supported the Baas Becking hypothesis: “Everything is everywhere, but environment selects” (Baas Becking LGM, 1934). This idea, originally applied to microorganisms, has been extended to include fungal species due to the small size of fungal spores, the main agent of fungal dispersion (Taylor et al., 2006). The apparent unlimited dispersal ability of many fungal species has often been used to explain their cosmopolitan distributions (Davison et al., 2015). Nowadays, the development of molecular tools has allowed identification of a significant amount of hidden biodiversity with numerous cryptic or sibling species previously masked under a single species name (Fišer et al., 2018; Koufozanou et al., 1997). The shift from morphological to phylogenetic species recognition, PSR (Taylor et al., 2000), has redrawn the map of fungal distribution, and new biogeographical patterns have arisen when morphospecies were redefined following PSR criteria (May, 2018). There are already a number of situations where a single species with worldwide distribution has been redescribed as several species with regional or restricted distribution (Carlsen et al., 2011; Nilsson et al., 2003; Telleria et al., 2010). This new approach in the study of fungal diversity has promoted the idea that cosmopolitanism in fungi is just the result of the application of MSR, rather than an actual biodiversity distribution pattern (Sato et al., 2012). In this context, reference collections or DNA sequence repositories allow for a re-evaluation of the species names assigned to collections or sequence vouchers, and, therefore, the assessment of the effects of taxonomic uncertainty or misleading specimen identifications in the potential distribution inferred by species distribution models.

**Xylopon** (Hymenochaetales, Basidiomycota) is a white-rot fungus considered one of the most species-rich corticioid genera (Hjortstam & Ryvarden, 2007, 2009) and plays an important role as a wood decomposer from temperate to tropical forests. It contains many species that have been traditionally cited worldwide, and its taxonomy has rapidly changed in recent years (Riebesehl & Langer, 2017). In addition, despite their macroscopic basidiocarps, the morphological traits used to distinguish among closely related species are highly homoplasic, making them prone to errors in specimen identifications.

The aim of the present study was to analyze the effect of taxonomic uncertainty and misidentifications in reference collections and sequence databases on species distribution models in Xylopon. We analyze the possible effects in two ways: First, we assessed whether a greater hidden diversity could be masked under a single species name by analyzing sequences from the ITS DNA region with barcoding gap analysis (Puillandre et al., 2012; Schoch et al., 2012); and second, we constructed species distribution models following both identification criteria (names on vouchers collections; species candidates obtained from barcoding gap analysis) and analyzed differences in the contribution of predictor variables and the distribution area sizes.

2 | MATERIALS AND METHODS

2.1 | Species studied and selection of material

A general search of preserved specimens in the Global Biodiversity Information Facility (GBIF) confirmed the worldwide distribution of presence records assigned to these three morphospecies: **Xylopon flavipurus**, **X. paradoxus**, and **X. raduloides** (Figure 1). These three **Xylopon** species traditionally known as being widely distributed were selected to discuss the effect of taxonomic uncertainty on biogeographical hypotheses supported by species distribution models. Those species have been traditionally located in **Schizopora**, but recent studies demonstrated that it is not possible to separate **Schizopora** from **Xylopon** on a morphological or molecular basis (Riebesehl & Langer, 2017), and therefore, **Schizopora** species are currently integrated into **Xylopon**.
Xylodon flaviporus (Berk. & M.A. Curtis ex Cooke) Riebesehl & E. Langer was originally described from Venezuela as Poria flavipora Berk. & M.A. Curtis. It has 24 synonyms according to Index Fungorum (Appendix S1) and has been reported from numerous hardwood substrates, such as Castanea, Eucalyptus, Fagus, and Quercus, but also on conifers, such as Picea and Pinus. Xylodon flaviporus has been considered as distributed worldwide (Figure 1), especially in warm and tropical zones. It has been reported from around the world: South America, Africa, southern Europe, and South Asia (Gilbertson & Ryvarden, 1987; Núñez & Ryvarden, 2001; Paulus et al., 2000; Ryvarden & Melo, 2014; Wu, 2000).

Xylodon paradoxus (Schrad.) Chevall. was described as Hydnum paradoxum Schrad. from Germany, and it has 52 synonyms following Index Fungorum (Appendix S1). It occurs in deciduous woodlands, mainly in Europe (Figure 1), although it has also been reported worldwide (Bernicchia, 2005; Eriksson et al., 1984; Ryvarden, 1978; Ryvarden & Johansen, 1980; Ryvarden & Melo, 2014).

Xylodon raduloides Riebesehl & E. Langer was originally described as Poria radula Pers. It was largely considered the same species as Xylodon paradoxus, but was split by Hallenberg (1983). It has 30 synonyms according to Index Fungorum (Appendix S1), and it has typically been associated with angiosperm wood. Xylodon raduloides has been reported from distant locations (Figure 1) such as Europe (Langer, 1994; Ryvarden & Gilbertson, 1994; Ryvarden & Melo, 2014), North America (Gilbertson & Ryvarden, 1987; Hallenberg, 1983), South America (Langer, 1994), temperate Asia (Hallenberg, 1983; Langer, 1994), and Australasia (Paulus et al., 2000).

All the available collections of these three morphospecies were studied from a total of five fungaria, CFMR, MA-Fungi, NY, O, and PDD (Table 1). Label information was used to assign the geographic location for each record (unprojected coordinates, WGS87 datum). When an exact location was not provided, label information such as towns or kilometer points along roads was used to obtain geographic coordinates. Only records with known coordinate uncertainty of less than 5 km (i.e., those which on average could be placed in a single 10 x 10 km cell) were considered. A basidiome fragment from fungarium specimens (less than 10 mg) was removed to perform molecular analyses.
### TABLE 1  Selected specimens and species assignation following names in fungarium and sequence vouchers (Data Set 1) and following barcoding gap species recognition (Data Set 2, Figure 2)

| Label/ Voucher species name (Data Set 1) | Country | BGSR (Data Set 2) | Collection number | GenBank Accession n.* |
|----------------------------------------|---------|------------------|-------------------|----------------------|
| Xyloodon flaviporus (Berk. & M.A. Curtis ex Cooke) Riebeselh & E. Langer | Brazil  | SC-C5            | NY 1045           | MW699791             |
|                                        | Cameroon| SC-B1            | NY s.n.           | KY962843             |
|                                        | Cameroon| SC-C7            | O-F 915884        | MW699840             |
|                                        | Cameroon| SC-C3            | MA-Fungi 38220    | MW699822             |
|                                        | China   | SC-C5            | CLZhao 53         | MG231630             |
|                                        | China   | SC-C5            | CLZhao 2384       | MH114732             |
|                                        | China   | SC-C3            | CLZhao 60         | MG231631             |
|                                        | China   | SC-C5            | CLZhao 4785       | MK269030             |
|                                        | China   | SC-C5            | CLZhao 34         | MH114928             |
|                                        | China   | SC-C5            | CLZhao 116        | MG231634             |
|                                        | China   | SC-C5            | CLZhao 85         | MG231633             |
|                                        | China   | SC-C5            | CLZhao 3459       | MK269271             |
|                                        | China   | SC-C5            | CLZhao 3468       | MK269027             |
|                                        | China   | SC-C5            | CLZhao 3194       | MH114735             |
|                                        | China   | SC-C5            | CLZhao 3143       | MH114733             |
|                                        | China   | SC-C5            | CLZhao 3148       | MH114734             |
|                                        | China   | SC-C5            | CLZhao 3275       | MK269026             |
|                                        | China   | SC-C5            | CLZhao 3609       | MK269028             |
|                                        | China   | SC-C5            | CLZhao 3656       | MK269029             |
|                                        | China   | SC-C1            | SWFC 001828       | MK838854             |
|                                        | China   | SC-C1            | SWFC 004636       | MK894105             |
|                                        | China   | SC-C1            | SWFU 001902       | MK809470             |
|                                        | China   | SC-C1            | SWFC 001831       | MK838888             |
|                                        | China   | SC-C1            | SWFU 001840       | MK809478             |
|                                        | China   | SC-C1            | CLZhao 5850       | MK343690             |
|                                        | China   | SC-C1            | Wu 0211-53        | MF540763             |
|                                        | China   | SC-C1            | SWFC 001824       | MK838853             |
|                                        | China   | SC-C1            | SWFC 001817       | MK838856             |
|                                        | Costa Rica| SC-C5         | O-F 507425        | MW699833             |
|                                        | Ecuador | SC-C5            | O-F 505597        | MW699829             |
|                                        | France  | SC-C5            | MA-Fungi 70678    | MW699826             |
|                                        | France  | SC-C5            | MA-Fungi 79438    | MW699827             |
|                                        | Germany | SC-C5            | MA-Fungi 79440    | MH260071             |
|                                        | Japan   | SC-C1            | O-F 507446        | MW699836             |
|                                        | Kenya   | SC-C3            | O-F 507471        | MW699838             |
|                                        | Kenya   | SC-C3            | O-F 507406        | MW699832             |
|                                        | Lesser Antilles | SC-C5         | O-F 507388        | MW699831             |
|                                        | Malawi  | SC-C6            | O-F 507478        | MW699839             |
|                                        | Nepal   | SC-C1            | O-F 507433        | MW699834             |
|                                        | Panama  | SC-C5            | MA-Fungi 36573    | MW699819             |
| Panama  | SC-C5            | MA-Fungi 36574    | MW699820          |

(Continues)
| Label/ Voucher species name (Data Set 1) | Country | BGSR (Data Set 2) | Collection number | GenBank Accession n.º |
|-----------------------------------------|---------|------------------|-------------------|----------------------|
| Panama                                  | SC-C5   | MA-Fungi 36800   |                   | MW699821             |
| Puerto Rico                             | SC-C4   | PR 1853          |                   | MW699845             |
| Reunion                                 | SC-C2   | KAS-GEL5047      |                   | MH880203             |
| Reunion                                 | SC-C2   | FR-0249797       |                   | MH880201             |
| Romania                                 | SC-C5   | FCUG 1534        |                   | AF145573             |
| Romania                                 | SC-C5   | FCUG 1053        |                   | AF145575             |
| Rwanda                                  | SC-C6   | O-F 507449       |                   | MW699837             |
| South Korea                             | SC-C5   | KUC20130808-17   |                   | KJ668462             |
| South Korea                             | SC-C5   | KA17-0796        |                   | MK920119             |
| South Korea                             | SC-C1   | SFC20180710-24   |                   | MK992840             |
| Taiwan                                  | SC-C5   | FP 101622        |                   | MW699841             |
| Taiwan                                  | SC-C5   | ICMP 13836       |                   | AF145585             |
| Taiwan                                  | SC-C5   | KAS-GEL3462      |                   | MH880202             |
| Taiwan                                  | SC-C1   | GC 1509-71       |                   | MF540761             |
| Thailand                                | SC-C1   | O-F 507441       |                   | MW699835             |
| USA                                     | SC-C5   | DLL2011-167      |                   | KJ140665             |
| USA                                     | SC-C5   | DLL2011-134      |                   | KJ140637             |
| USA                                     | SC-C5   | DLL2011-141      |                   | KJ140642             |
| USA                                     | SC-C4   | HHB 9460         |                   | MW699843             |
| USA                                     | SC-C4   | FP 102561        |                   | MW699842             |

*Xylodon paradoxus* (Schrad.) Chevall.

| Country | SC-A4   | MA-Fungi 92327 | MW699817 |
|---------|---------|----------------|----------|
| Chile   | SC-A4   | MA-Fungi 92325 | MW699815 |
| Chile   | SC-A4   | MA-Fungi 92324 | MW699814 |
| Chile   | SC-A4   | MA-Fungi 92321 | MW699811 |
| Chile   | SC-A4   | MA-Fungi 92320 | MW699810 |
| Chile   | SC-A4   | MA-Fungi 92326 | MW699816 |
| Chile   | SC-A4   | MA-Fungi 92323 | MW699813 |
| Chile   | SC-A4   | MA-Fungi 92322 | MW699812 |
| China   | SC-A2   | CLZhao 3220    | MK269041  |
| Finland | SC-A1   | Otto Miettinen 7978 | FN907912 |
| France  | SC-A1   | MA-Fungi 79441 | MW699828  |
| France  | SC-A1   | MA-Fungi 70444 | MW699825  |
| France  | SC-A1   | MA-Fungi 81294 | MH260072  |
| Germany | SC-A1   | MA-Fungi 40866 | MW699823  |
| Germany | SC-A1   | SI59           | FJ820647  |
| Mexico  | SC-A5   | NY 8598       | MW699797  |
| Morocco | SC-A1   | MA-Fungi 5464  | MW699794  |
| Portugal| SC-B1   | MA-Fungi 26152 | MW699818  |
| Romania | SC-A1   | FCUG 1517     | AF145572  |
| Russia  | SC-A1   | FCUG 2425     | AF145571  |
| South Korea | SC-C1 | KUC 8140    | MW699844  |
| Spain   | SC-B1   | MA-Fungi 608  | KY962826  |
| Spain   | SC-B1   | MA-Fungi 22499| KY962822  |

(Continues)
### TABLE 1 (Continued)

| Label/ Voucher species name (Data Set 1) | Country | BGSR (Data Set 2) | Collection number | GenBank Accession n.º |
|----------------------------------------|---------|-------------------|-------------------|-----------------------|
| Spain SC-B1 MA-Fungi 12877             |         |                   |                   | MW699808              |
| Spain SC-B1 MA-Fungi 35643             |         |                   |                   | KY962831              |
| Spain SC-B1 MA-Fungi 75272             |         |                   |                   | KY962829              |
| Spain SC-B1 MA-Fungi 75310             |         |                   |                   | KY962825              |
| Spain SC-B1 MA-Fungi 12864             |         |                   |                   | KY962820              |
| Spain SC-A3 MA-Fungi 12880             |         |                   |                   | MW699809              |
| Spain SC-A3 MA-Fungi 12873             |         |                   |                   | MW699807              |
| Spain SC-A3 MA-Fungi 1063              |         |                   |                   | MW699792              |
| Spain SC-A3 MA-Fungi 5658              |         |                   |                   | MW699796              |
| Spain SC-A3 MA-Fungi 12772             |         |                   |                   | MW699799              |
| Spain SC-A3 MA-Fungi 12775             |         |                   |                   | MW699800              |
| Spain SC-A3 MA-Fungi 12771             |         |                   |                   | MW699798              |
| Spain SC-A1 MA-Fungi 5651              |         |                   |                   | MW699795              |
| Spain SC-A1 MA-Fungi 12794             |         |                   |                   | MW699802              |
| Spain SC-A1 MA-Fungi 46191             |         |                   |                   | MW699824              |
| Spain SC-A1 MA-Fungi 12857             |         |                   |                   | MW699805              |
| Spain SC-A1 MA-Fungi 12787             |         |                   |                   | MW699801              |
| Spain SC-A1 MA-Fungi 12844             |         |                   |                   | MW699803              |
| Spain SC-A1 MA-Fungi 12846             |         |                   |                   | MW699804              |
| Spain SC-C5 MA-Fungi 3269              |         |                   |                   | MW699793              |
| Spain SC-B1 MA-Fungi 12869             |         |                   |                   | MW699806              |
| Spain SC-B1 MA-Fungi 75244             |         |                   |                   | KY962833              |
| Spain SC-B1 MA-Fungi 12778             |         |                   |                   | KY962832              |
| Spain SC-B1 MA-Fungi 75130             |         |                   |                   | KY962824              |
| Spain SC-B1 MA-Fungi 22513             |         |                   |                   | KY962823              |
| USA SC-B2 HHB 719                      |         |                   |                   | KY962845              |
| USA SC-B3 O-F 507276                   |         |                   |                   | MW699830              |

*Xylodon raduloides* Riebesehl & E. Langer

| Country | BGSR | Collection number | GenBank Accession n.º |
|---------|------|-------------------|-----------------------|
| Argentina SC-B4 FCUG 2492 | MA-Fungi 90703 | KY962841 |
| Argentina SC-B4 FCUG 2497 | MA-Fungi 90708 | KY962839 |
| Argentina SC-B4 ICMP 13832 | MA-Fungi 90705 | KY962835 |
| Australia SC-B5 ICMP 13833 | MA-Fungi 90807 | KY962837 |
| Canada SC-B5 FCUG 678 | MA-Fungi 90402 | KY962840 |
| Chile SC-B4 MES-2446 | MA-Fungi 90702 | KY962836 |
| Chile SC-B4 MA-Fungi 90706 | MA-Fungi 90704 | KY962840 |
| Chile SC-B4 P.CH-4 | MA-Fungi 90706 | KY962838 |
| Chile SC-B4 MA-Fungi 90706 | MA-Fungi 90704 | KY962840 |
| Chile SC-B4 MA-Fungi 90457 | MA-Fungi 90704 | KY962840 |
| Denmark SC-B1 FCUG-1972 | MA-Fungi 90457 | KY962827 |
| France SC-B1 MA-Fungi 90457 | MA-Fungi 90457 | KY962827 |

(Continues)
In addition to isolation of DNA from fungarium specimens, a search at EMLB/GenBank/DDBJ and UNITE databases was performed in order to locate existing molecular information available for each studied species. Geographic locations and species identification of each sequence were obtained from GenBank/UNITE vouchers.

### 2.2 Molecular methods and candidate species assignment through molecular barcoding

DNA extractions for *Xylodon* specimens from fungaria samples were performed. For DNA isolation, DNeasy™ Plant Mini Kit (Qiagen, Valencia, California, USA) was used, following the instructions of the manufacturers. Lysis buffer incubation was done overnight at 55°C following Whiting et al. (1997). In order to detect species candidates, a barcoding gap approach was utilized, using the Internal Transcribed Spacer (ITS) because this region is a universal barcode across fungi, able to detect genetic variability at the species level (Schoch et al., 2012). The ITS5/ITS4 (White, 1990) primer combination was used to obtain DNA amplifications, of ITS1 and ITS2 regions plus 5.8S nrDNA. Amplifications were done using illustra™ PuReTaq™ Ready-To-Go™ PCR beads (GE Healthcare, Buckinghamshire, UK) as described in Winka et al. (1998), following thermal cycling conditions in Martín and Winka (2000). Negative controls lacking fungal DNA were run for each experiment to check for contamination of reagents. Results of amplifications were assayed from 5 μl aliquots by gel electrophoresis of 2% Pronadisa D-1 Agarose (Lab. Conda, Spain). Amplified DNA fragments were first separated and purified from the agarose gel using the Wizard SV Gel and PCR Clean-Up System (Promega Corporation, Madison, WI, USA) and sent to Macrogen Korea (Seoul, South Korea) for sequencing. Primers used for sequencing were those used for PCR amplifications. The ITS sequences generated were edited and assembled to obtain consensus using Geneious version 9.0.2 (http://www.geneious.com, Kearse et al., 2012). The consensus sequences were lodged in the EMLB/GenBank/DDBJ databases with the accession numbers indicated in Table 1.

#### Table 1 (Continued)

| Label/ Voucher species name (Data Set 1) | Country | BGSR (Data Set 2) | Collection number | GenBank Accession n.º |
|----------------------------------------|---------|------------------|-------------------|-----------------------|
| France SC-B1 | MA-Fungi 79442 | KY962834 |
| France SC-B1 | MA-Fungi 79314 | KY962830 |
| France SC-B5 | MA-Fungi 78658 | KY962828 |
| France SC-B5 | MA-Fungi 74919 | KY962842 |
| New Zealand SC-B5 | NZFS:4546 | MH409968 |
| New Zealand SC-B5 | ICMP 13841 | AF145579 |
| New Zealand SC-B5 | ICMP 13838 | AF145578 |
| New Zealand SC-B5 | ICMP 13829 | AF145577 |
| New Zealand SC-B5 | ICMP 13840 | AF145576 |
| New Zealand SC-B6 | PDD 91616 | GQ411525 |
| Romania SC-B1 | FCUG 1055 | AF145569 |
| Russia SC-B1 | FCUG 2433 | AF145570 |
| Spain SC-B1 | MA-Fungi 90709 | KY962844 |
| Spain SC-B1 | MA-Fungi 75310 | KY962825 |
| Spain SC-B1 | FCUG 2136 | AF145565 |
| Turkey SC-B1 | FCUG 2239 | AF141613 |
| USA SC-B2 | S.D. Russell MycoMap 8118 | MK575271 |
| USA SC-B2 | DLL2011 142 | KJ140643 |
| USA SC-B2 | DLL2009 049 | JQ673187 |
| USA SC-B2 | DLL2009 087 | JQ673189 |
| USA SC-B2 | DLL2009 082 | JQ673188 |
| USA SC-B7 | UC2022947 | KP814552 |

Note: New sequences in bold.
A first general analysis was conducted with the whole ITS DNA alignment in order to delimit major clades using pairwise distances under JC69 model and Neighbor-joining algorithm (Figure 2). After that, the Automatic Barcode Gap Discovery algorithm (ABGD, Puillandre et al., 2012) was applied to each major clade in order to obtain species candidates (barcoding gap species recognition, BGSR). ABGD algorithm uses the gap in the distribution of pairwise distances (i.e., the first statistically significant peak in the slope of ranked pairwise genetic distance values) of DNA barcode regions, assuming the divergence among organisms belonging to the same species is smaller than divergence among organisms from different species (Schoch et al., 2012). Consequently, specimens are grouped into species hypotheses or candidates that can be assessed later through the inclusion of molecular data from other DNA regions or other sources of evidence (morphology, ecological preferences, biogeography, etc.).

The ABGD analysis was performed at https://bioinfo.mnhn.fr/abi/public/abgd/. The relative gap width was set to 0.5 to allow for the detection of closely related taxa. The remaining parameters were set to default (Jukes–Cantor distance (JC69), Pmin = 0.001, Pmax = 0.100, Steps = 10, Number of bins = 20).

### 2.3 Species distribution modeling

Three kinds of predictor variables were included in distribution models, representing different factors that usually affect species distributions: abiotic, biotic, and geographic variables (Soberón, 2007). We used 19 environmental climatic (abiotic) layers from the Worldclim 2 database (https://www.worldclim.org/data/index.html, Fick & Hijmans, 2017). Variance Inflation Factor (VIF) scores were used to evaluate multicollinearity among abiotic predictors; the final set of climatic predictors used for modeling had VIF values lower than 2 (Zuur et al., 2010). After that, a total of 19 environmental climatic variables (Table 2) were included taking into account the importance of these factors reported by other studies (Wollan et al., 2008). Percentage of tree cover (biotic) from MODIS project (https://modis.gsfc.nasa.gov/data/dataprod/mod44.php) was also included as a predictor since Xylodon species are wood-decay fungi and depend on the existence of wood to grow and maintain their populations. Finally, to include pure geographic constraints that could affect species distributions limiting their dispersal or colonization capacity, latitude and longitude were included as predictor variables (Acevedo et al., 2012). Due to the circular character of longitude, the sine and cosine components were used instead (Pewsey et al., 2013). All predictor layers were used at 10 × 10 km resolution grid.

Two data sets of presence records were built, depending on the species recognition criterion used to create them from the specimens analyzed: in Data Set 1, modeling groups were made based on taxonomic information of each specimen as recorded in fungaria or sequence databases (Table 1); in Data Set 2, barcoding gap analyses results were used to re-group specimens following candidate species proposed by molecular barcoding analyses (BGSR). When the number of presences reported for a species candidate by BGSR was too low (less than 6), distribution model was not performed for such species candidates due to the small sample size (Pearson et al., 2006; van Proosdij et al., 2016). The modeling approach was exactly the same for all arrangements in both data sets. We used the MaxEnt algorithm to conduct distribution models (Phillips et al., 2006, 2017). MaxEnt has been reported to perform well when only presence data (i.e., museum and herbarium/fungarium data) are used, as in our case. Moreover, this algorithm has demonstrated acceptable accuracy for small-sized samples (Pearson et al., 2006). For each candidate species, we ran 10 replicates with internal AUC (Area Under the Curve Receiver Operating Characteristic, Fielding & Bell, 1997) validation (80% of presences for model calibration versus. 20% for model evaluation). Only linear, quadratic, and product features were used to promote model interpretability (Merow et al., 2013) and regularization multiplier = 1 and cloglog output were selected (Phillips et al., 2017). In order to control the possible sample bias from our presence data sets, a layer of human footprint index was included as bias grid in MaxEnt, to represent those areas with more human accessibility as more probably sampled (Elith et al., 2011; Phillips et al., 2009). This index was obtained from “Last of the Wild Project”, version 2. It consists of an overlay of a number of global data layers

| TABLE 2 Description of environmental predictors used in species distribution models |
|----------------------------------|----------------------------------|
| **Climatic**                     | **Geographic**                   |
| Isothermality—BIO3               | Latitude                          |
| Temperature Seasonality—BIO4    | Sin (Longitude)                   |
| Mean Temperature of Wettest Quarter—BIO8 | Cos (Longitude)                 |
| Mean Temperature of Driest Quarter—BIO9 | Biotic                          |
| Mean Temperature of Warmest Quarter—BIO10 | Tree cover                     |
| Mean Temperature of Coldest Quarter—BIO11 |                             |
| Precipitation of Wettest Quarter—BIO16 |                             |
| Precipitation of Driest Quarter—BIO17 |                             |
| Precipitation of Warmest Quarter—BIO18 |                             |
| Precipitation of Coldest Quarter—BIO19 |                             |
that represent the location of human population distribution, urban areas, roads, navigable rivers, and various agricultural land uses (http://sedac.ciesin.columbia.edu/data/collection/wildareas-v2).

In order to evaluate the importance of each kind of predictor variable, percent contributions in model predictions were analyzed after checking for model comparability-based AUC scores. To compare distributional range size between each species recognition criterion, distribution probability maps were transformed to presence/absence maps using equal test sensitivity and specificity threshold and the area occupied was calculated.

3 | RESULTS

A total of 150 collections were considered in this study, of which 83 were newly sequenced (Table 1). Following genetic distance tree results (Figure 2), sequences were separated into three major clades, each corresponding to one morphospecies: Clade A—*Xylodon paradoxus*, Clade B—*Xylodon raduloides*, and Clade C—*Xylodon flaviporus*. Up to 19 species candidates were detected following BGSR: Data Set 2; Recursive Partition, prior intraspecific divergence \( p = .002 \) (Table 1). The number of collections assigned to each species candidate under BGSR varies from one to 37 (Table 1). In general, no clear correspondences were found among voucher collection names and species candidates detected under BGSR, that is, no nested patterns were found (Table 1). For the specimens named *X. flaviporus*, eight different species candidates were detected under BGSR, while under the names *X. paradoxus* and *X. raduloides*, ten and six species candidates were detected through BGSR, respectively (Table 1).

Three distribution models were performed from Data Set 1 obtained following names on vouchers collections, one for each morphospecies. From the 19 species candidates detected under BGSR, the distribution models of only 9 species candidates were performed, those groups for which sample size was greater than 6 specimens (see Table 3, Figure 2).

The importance of each predictor in distribution models is shown in Table 3. Climatic variables obtained the highest percent contribution in models from names in fungarium and sequences vouchers (71.37% on average), followed by geographic predictors (21.64% on average) and finally by biotic variables (tree cover, 6.98% on average). However, for the models performed from species candidates detected by BGSR, the contribution of climatic variables was generally lower (41.68% on average; but see SC- A4, SC- B5, and SC- C5). Geographic predictors were most important for five of the nine species candidates under BGSR (53.74% of contribution on average), and tree cover had the least predictive value (6.98 percent contribution on average).

Distribution models built from names on voucher collections showed worldwide distributions and lacked biogeographic patterns (Figure 3). The extent of those distributions ranged from 14% to 19% of total worldwide emerged lands (Table 3).

### TABLE 3  Modeling results using both data sets: (1) Label and sequence voucher names and (2) Barcoding Gap Species Recognition (BGSR)

| Data Set 1: Label and sequence voucher names | AUC | % Occupied area | Predictors % contribution |
|--------------------------------------------|-----|-----------------|--------------------------|
|                                            |     |                 | Climate | Geography | Tree cover |
| *X. flaviporus* (n = 62)                   | 0.90| 19%             | 76.47 | 10.27 | 13.26 |
| *X. paradoxus* (n = 50)                    | 0.93| 16%             | 68.22 | 29.12 | 2.66 |
| *X. raduloides* (n = 38)                   | 0.91| 14%             | 69.43 | 25.54 | 5.03 |
| **Average**                                | **71.37** | **21.64** | **6.98** |

| Data Set 2: Following Barcoding Gap Species Recognition (BGSR) | Species Candidates | AUC | % Occupied area | Predictors % contribution |
|---------------------------------------------------------------|---------------------|-----|-----------------|--------------------------|
|                                                              |                     |     |                 | Climate | Geography | Tree cover |
| SC-A1 (n = 16)                                                | 0.95                | 9%  | 26.33           | 72.95 | 0.72 |
| SC-A3 (n = 7)                                                 | 0.99                | 3%  | 23.70           | 70.51 | 5.78 |
| SC-A4 (n = 8)                                                 | 0.99                | <1% | 60.28           | 39.47 | 0.24 |
| SC-B1 (n = 25)                                                | 0.95                | 10% | 22.04           | 76.32 | 1.64 |
| SC-B2 (n = 6)                                                 | 0.99                | <1% | 30.52           | 64.61 | 4.86 |
| SC-B4 (n = 12)                                                | 0.99                | <1% | 48.02           | 49.09 | 2.89 |
| SC-B5 (n = 8)                                                 | 0.92                | 14% | 56.08           | 32.60 | 11.32 |
| SC-C1 (n = 15)                                                | 0.99                | <1% | 42.50           | 55.32 | 2.18 |
| SC-C5 (n = 37)                                                | 0.90                | 18% | 65.63           | 22.81 | 11.55 |
| **Average**                                                   | **41.68**           | **53.74** | **4.57** |

Abbreviation: AUC, Area Under the receiver operating characteristics Curve.
FIGURE 3 Presence records and distribution models for specimens arranged following labels and vouchers species names

X. flaviporus

X. paradoxus

X. raduloides
FIGURE 4 Presence records and distribution models for specimens arranged following ITS barcoding gap analyses (BGSR)
In contrast, distribution models obtained for species candidates detected by BGSR showed in most cases local or restricted distributions (Figure 4). The distributions predicted from these models were in general smaller, with the exception of the species candidates SC-B5 and SC-C5. AUC values were always high, independent of the arrangement criterion used, with minimum and maximum between 0.90 and 0.99 (Table 3).

4 | DISCUSSION

The development of new statistical tools to predict species distributions has promoted the use of presence-only databases such as natural history collections (Elith & Leathwick, 2007). Herbaria/fungaria or museums have been an important source of vouchers. Nowadays, those techniques are commonly applied for a broad range of purposes, from assessing pest invasion risks to conservation management (Franklin, 2013). They have also been used to evaluate the environmental factors that drive fungal species distributions (Wollan et al., 2008; Yuan et al., 2015) or to predict the potential distribution of ectomycorrhizal fungi under different climate change scenarios (Guo et al., 2017). However, the effects of taxonomic uncertainty have rarely been assessed in fungal distribution models (Elith et al., 2013). Xylodon is an appropriate case study to understand those effects due to its high diversity and the lack of macroscopic diagnostic characters in many of its species (Riebesehl & Langer, 2017).

Our results distinguished up to 19 species candidates under only three species names using molecular tools (Figure 2, Table 3). Although these species candidates are not all confirmed because a deeper study is needed, it draws a more realistic picture about the actual diversity in our presence records. In fungi, it is becoming commonplace to detect many phylogenetic species when molecular data are analyzed for within a single morphospecies (Cai et al., 2014; Fernández-López et al., 2020). Taxonomic issues are not fully solved in our analyses since only one DNA region was used, and multiple sources of evidences in an integrative framework are recommended to correctly define species boundaries (Dayrat, 2005). However, it has been demonstrated that the ITS barcoding region generally performs well in fungal species delimitation (Schoch et al., 2012) and barcoding region analyses are broadly used in fungal environmental studies (Tedersoo et al., 2014). Therefore, the species candidates delimited in this study are an appropriate first step to understand the complexity in the available Xylodon data in different reference collections. A deeper study of those candidates could be useful to detect new morphological or ecological traits to distinguish among species in Xylodon.

Genetic analyses pointed toward two sources of misleading information in the studied material: first taxonomic uncertainty through cryptic speciation processes inside each morphospecies, since several subclades could be distinguished in the three major clades delimited (Figure 2), and second, a significant amount of incorrect identifications even for the broadly defined morphospecies, especially between Xylodon paradoxus and X. raduloides. These results could be expected due to the morphological similarities of these two species. In addition, X. raduloides was split from X. paradoxus only in the late twentieth century (Hallenberg, 1983), and therefore, it is probable that several X. raduloides collections were still labeled under its old name.

Despite the relatively small sample size used in this study, our presence records described well the scope of the general distribution of the material available in reference collections (Figures 1 and 3). Predicted areas from models using label/voucher information described cosmopolitan distributions for the three morphospecies. Predicted areas occupied up to 19% of the world’s emerged lands, and the three morphospecies can be found in Africa, America, Asia, Europe, and Oceania. However, models derived from the molecular analysis showed local or restricted distributions in most cases (except SC-B5 and SC-C5), with a biogeographic pattern (Figure 4). These reduced distributions support a more realistic picture of fungal diversity, since it has been demonstrated that genetic lineages remained at least partially isolated from each other in many fungi (Peay et al., 2010; Sato et al., 2012). It should be noted that the number of presence records for most of the candidate species is too small (Data Set 2) to affirm that predicted distributions reflect the actual species range, that is, species candidates SC-A3 and SC-A4 (Table 3). Thus, the lack of occurrences scattered over the actual species range could produce overfitted predictions, and therefore, distribution ranges can be underestimated. Moreover, the addition of new presence records could affect the distribution pattern described for each species candidates, especially for those with a smaller sample size. However, differences in distributions obtained between Data Set 1 and Data Set 2 are in accordance with similar patterns that have been reported in many Basidiomycota, for which there has been a transition from a few cosmopolitan species to numerous species with a regional distribution (Petersen & Hughes, 1999). The distinct geographic distributions of each lineage in the molecular analysis is in itself support for recognition of the lineages as distinct taxonomic entities, although distribution on its own would not be sufficient for recognition of segregate species.

Among the species candidates delimited by the BGSR approach, SC-B5 and SC-C5 maintained a worldwide distribution, with no biogeographic pattern supporting a genetic structure (Figure 4). In the case of SC-B5, this is due to two specific samples (one from France and another from Canada) and could be explained by human-mediated translocation, commonly reported for wood-decay fungi, for example, timber trade (Fernández-López et al., 2019; Paulus et al., 2000), since the rest of the samples are located in Australia–New Zealand. On the other hand, SC-C5 presents a much more complex pattern, with closely related genetic samples distributed around the world. This pattern could be due to the inability of the barcoding approach to distinguish between these close-related species and therefore other sources of evidence or more DNA regions should be used to confirm this result (Balasundaram et al., 2015; Martín et al., 2018). Nevertheless, the hypothesis that the specimens
arranged in this group conform a species with a worldwide distribution cannot be discarded. New methodologies that explicitly account for long-distance dispersal should be performed to resolve this issue.

The distribution predicted for species arranged following fungarium and sequence vouchers was mainly driven by climatic predictors rather than geographic or tree cover predictors (Table 3). It has been demonstrated that variables such as temperature or precipitation play a central role in fungal distributions (Hao et al., 2020). However, for distribution modes derived from the genetic barcoding gap approach, although climatic factors remained important, they generally lost part of their predictive power in favor of geographic variables (Table 3). In our analyses, tree cover had less contribution than climatic or geographic factors. However, its contribution could be masked by climatic factors due to collinearity among predictors, and therefore, it should not be evaluated. Moreover, the resolution of cartographic layers could be too low to reflect the actual wood availability in small patches, where corticioid fungi can be present in isolated trees (Abrego et al., 2017).

It is important to highlight the inability of internal model validation to detect taxonomic uncertainty. Internal cross-validations are made by partitioning presence sample in training and test sets. Since test data have the same origin as training data, MaxEnt internal AUC is unable to detect wrong identifications in the occurrences. For this reason, AUC values for all models were always high (>0.90) independent of the Data Set used. In addition, it is known that the geographic coverage of a model influences AUC scores (Lobo et al., 2008). Since our study area is worldwide, AUC scores for our models could be a misleading measure of model performance.

5 | CONCLUSION

Our results demonstrate the important role that taxonomic uncertainty plays in the inferences obtained from species distribution models (Elith et al., 2013). Distribution patterns obtained from models based on names on fungarium collections and sequence vouchers appear to support the Baas Becking hypothesis “Everything is everywhere, but environment selects” in Xylo
don. These unrealistic and overestimated distributions could similarly be assumed for other species that are involved in conservation programs or pest management plans, resulting in biological and economic losses (Bortolus, 2008; Fernández-López et al., 2018). In this context, preserved specimens in natural history collections offer the possibility to reevaluate occurrence data sets by sequencing when taxonomic uncertainty may compromise the results obtained from species distribution models (Elith & Leathwick, 2007).

ACKNOWLEDGMENTS

Thanks to the curators of CFMR, NY, O, and PDD for their invaluable assistance arranging specimens and culture loans. Thanks to Marian Glenn (Seton Hall University), Jane Elith and Tianxiao Hao (University of Melbourne), and three anonymous reviewers for comments to the manuscript. The funding for this research was obtained from Consejo Superior Investigaciones Científicas, CSIC (Proyecto Intramural Especial del CSIC nº202030E), CSIC/Fundación Endesa/Fundación San Ignacio de Huinay (2011HUN10, 2013CL0012), and the Spanish Ministry of Economy and Competitiveness (BES-2013-066429, CGL2012-35559, CGL-2015-67459-P).

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

Javier Fernández-López: Conceptualization (lead); data curation (lead); formal analysis (lead); methodology (lead); writing—original draft (lead); writing—review and editing (equal). M. Teresa Telleria: Conceptualization (lead); funding acquisition (lead); project administration (lead); writing—original draft (equal); writing—review and editing (equal). Margarita Dueñas: Conceptualization (lead); data curation (lead); resources (lead); writing—original draft (equal); writing—review and editing (equal). Tom May: Conceptualization (lead); writing—original draft (equal); writing—review and editing (equal). María P. Martín: Conceptualization (lead); formal analysis (lead); methodology (lead); writing—original draft (equal); writing—review and editing (equal).

DATA AVAILABILITY STATEMENT

DNA sequences and voucher information: Genbank accession numbers in Table 1. Presence records for MaxEnt input files: Dryad https://doi.org/10.5061/dryad.z8w9ghxbv.

ORCID

Javier Fernández-López https://orcid.org/0000-0003-4352-0252

REFERENCES

Abrego, N., Christensen, M., Bässler, C., Ainsworth, A. M., & Heilmann-Clausen, J. (2017). Understanding the distribution of wood-inhabiting fungi in European beech reserves from species-specific habitat models. Fungal Ecology, 27, 168-174. https://doi.org/10.1016/j.functec.2016.07.006
Acevedo, P., Jiménez-Valverde, A., Melo-Ferreira, J., Real, R., & Alves, P. C. (2012). Parapatric species and the implications for climate change studies: A case study on hares in Europe. Global Change Biology, 18(5), 1509-1519. https://doi.org/10.1111/j.1365-2486.2012.02655.x
Baas Becking LGM (1934). In W. P. Van Stockum, & N. V. Zoon (Eds.). Geobiologie of inleiding tot de milieukunde. Serie 18/19. The Netherlands.
Balasundaram, S. V., Engh, I. B., Skrede, I., & Kauserud, H. (2015). How many DNA markers are needed to reveal cryptic fungal species? Fungal Ecology, 119, 940-945. https://doi.org/10.1016/j.funeco.2015.07.006
Beck, J., Böller, M., Erhardt, A., & Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species’ geographic distributions. Ecological Informatics, 19, 10-15. https://doi.org/10.1016/j.ecoinf.2013.11.002
Belnap, J., Miller, D. M., Bedford, D. R., & Phillips, S. L. (2014). Pedological and geological relationships with soil lichen and moss distribution in the eastern Mojave Desert, CA, USA. Journal of Arid Environments, 106, 45-57. https://doi.org/10.1016/j.jaridenv.2014.02.007
Wollan, A. K., Bakkestuen, V., Kauserud, H., Gulden, G., & Halvorsen, R. (2008). Modelling and predicting fungal distribution patterns using herbarium data. *Journal of Biogeography*, 35(12), 2298–2310. https://doi.org/10.1111/j.1365-2699.2008.01965.x

Wu, S. H. (2000). Studies on *Schizopora flavipora* sl, with special emphasis on specimens from Taiwan. *Mycotaxon*, 76, 51–66.

Yuan, H. S., Wei, Y. L., & Wang, X. G. (2015). MaxEnt modeling for predicting the potential distribution of Sanghuang, an important group of medicinal fungi in China. *Fungal Ecology*, 17, 140–145. https://doi.org/10.1016/j.funeco.2015.06.001

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1), 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Fernández-López J, Telleria MT, Dueñas M, May T, Martín MP. DNA barcode analyses improve accuracy in fungal species distribution models. *Ecol Evol*. 2021:11:8993–9009. https://doi.org/10.1002/ece3.7737